Keiichi Shirase
Seiji Aoyagi
Editors

Service Robotics and Mechatronics

Selected Papers of the International Conference on Machine Automation ICMA2008

Springer

Service Robotics and Mechatronics

Keiichi Shirase · Seiji Aoyagi

Service Robotics and Mechatronics

Selected Papers of the International Conference on Machine Automation ICMA2008

Springer

Editors
Dr. Keiichi Shirase
Kobe University
1-1 Rokko-dai
Nada
657-8501 Kobe
Japan
shirase@mech.kobe-u.ac.jp

Dr. Seiji Aoyagi
Kansai University
3-3-35 Yamate-cho
Suita
564-8680 Osaka
Japan
aoyagi@iecs.kansai-u.ac.jp

ISBN 978-1-84882-693-9 e-ISBN 978-1-84882-694-6
DOI 10.1007/978-1-978-1-84882-694-6
Springer London Dordrecht Heidelberg New York

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2009938199

Cover design: eStudioCalamar, Figueres/Berlin

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

It is our great pleasure to hold the 2008 International Conference on Machine Automation at Awaji Yumebutai International Conference Center. The conference is held every 2 years in Finland and Japan by turns, and this is the 7th conference sponsored by Kansai University and the Japanese Council of International Federation for Theory of Machines and Mechanism, in cooperation with 6 academic societies and 2 institutes and organizations.

In a society with declining birth rate and rapidly aging population, even the senior citizen need to take up active roles in daily life and social activities. In order to provide a safe living environment for them, and for those who are handicapped, new technologies that can support their activities adequately are required. Furthermore, in order to facilitate coexistence between man and machine, the design and application of systems that take into account human psychology and movement need to be considered. The shape and movement of the support robots that assist human beings in daily life must be designed with judgment on the effects of their applications on human psychology. While more detailed surveillance are required for monitoring system that monitor the daily life of human beings and the safety of the society, it is also important that such systems do no overburden those being monitored. In addition to those, mechatronics is also highly applicable in the area of reconstruction works after disasters, and restoration works on polluted or abandoned environment due to chemical substances, land mines or other factors. In production plant, it is expected the workers' burden can be reduced by implementing autonomous distributed or unmanned factory. Simply put, in order to ensure peace and safety to human society, the next generation robotics and mechatronics systems are required to systematically analyze human psychology and movement, and based on these results new system and machines will be developed.

This conference focuses on the topics and applications of such service robotics and mechatronics, and 79 papers are accepted after careful screening for the presentation in 18 technical sessions including rehabilitation, medical applications, robot, manufacturing, sensor, control, simulation, etc. Finally, two key note papers and 66 papers are selected to publish the book titled "Service Robotics and Mechatoronics" after the conference.

We would express our sincere appreciation and thanks to all the members of the Advisory Committee, the Organizing Committee, the Program Committee, the reviewers, the chairs, and the staffs as well as the participants for their good contribution and supports to hold this exciting conference.

Keiichi Shirase and Seiji Aoyagi
General Co-Chairs

Conference Committees

General Co-Chairs:

Aoyagi, S. (Kansai Univ., Japan)
Shirase, K. (Kobe Univ., Japan)

Advisory Committee:

Arai, E. (Osaka Univ., Japan)
Arai, T. (Univ. of Tokyo, Japan)
Arai, T. (Osaka Univ., Japan)
Asama, H. (Univ. of Tokyo, Japan)
Dohi, T. (Univ. of Tokyo, Japan)
Fujie, M. (Waseda Univ., Japan)
Fujimoto, H. (Nagoya Inst. of Tech., Japan)
Fukuda, T. (Nagoya Univ., Japan)
Higuchi, T. (Univ. of Tokyo, Japan)
Hirose, S. (Tokyo Inst. of Tech., Japan)
Hosaka, H. (Univ. of Tokyo, Japan)
Inaba, M. (Univ. of Tokyo, Japan)
Kamiya, Y. (Kanazawa Univ., Japan)
Kaneko, M. (Osaka Univ., Japan)
Kawamura, S. (Ritsumeikan Univ., Japan)
Kivikoski, M. (Tampere Univ., Finland)
Kosuge, K. (Tohoku Univ., Japan)
Mae, Y. (Osaka Univ., Japan)
Mitsuishi, M. (Univ. of Tokyo, Japan)
Morimoto, Y. (Wakayama Univ., Japan)
Sato, T. (Univ. of Tokyo, Japan)
Sugano, S. (Waseda Univ., Japan)

Organizing Committee:

Aoyama, H. (Univ. of Electro-Communications, Japan)
Hirai, S. (Ritsumeikan Univ., Japan)
Kaihara, T. (Kobe Univ., Japan)
Kume, Y. (Kinki Univ., Japan)
Kurata, J. (Kansai Univ., Japan)
Nakamoto, K. (Kobe Univ., Japan)
Osumi, H. (Chuo Univ., Japan)
Sakaguchi, T. (Kobe Univ., Japan)
Sasaki, K. (Univ. of Tokyo, Japan)
Seki, H. (Kanazawa Univ., Japan)
Sugimura, N. (Osaka Pref. Univ., Japan)
Taura, T. (Kobe Univ., Japan)
Tejima, N. (Ritsumeikan Univ., Japan)
Tsumaya, A. (Kobe Univ., Japan)
Umeda, K. (Chuo Univ., Japan)
Wakamatsu, E. (Osaka Univ., Japan)

Program Committee:

Airila, M. (Helsinki Univ. of Tech., Finland)
Conrad, F. (Technical Univ. of Denmark, Denmark)
Ellman, A. (Tampere Univ. of Tech.
/ Seinäjoki Univ. Consortium, Finland)
Fukushima, F. (Tokyo Inst. of Tech., Japan)
Halme, A. (Helsinki Univ. of Tech., Finland)
Handroos, H. (Lappeenranta Univ. of Tech., Finland)
Hata, S. (Kagawa Univ., Japan)
Hikitsu, M. (Kanazawa Univ., Japan)
Hirata, Y. (Waseda Univ., Japan)
Hosoda, K. (Osaka Univ., Japan)
Inoue, K. (Yamagata Univ., Japan)
Ivantysynova, M. (Purdue Univ., USA)
Iwamura, K. (Osaka Pref. Univ., Japan)
King, T. (Univ. of Leeds, U.K.)
Komeda, T. (Shiraura Inst. of Tech., Japan)
Koivo, H. (Helsinki Univ. of Tech., Finland)
Krus, P. (Linköping Univ., Sweden)
Matsubara, A. (Kyoto Univ., Japan)
Mitsuyuki, K. (DENSO, Japan)
Miwa, M. (Tokushima Univ., Japan)
Mizugaki, Y. (Kyushu Inst. of Tech., Japan)
Morita, T. (Univ. of Tokyo, Japan)
Narita, H. (Nagoya Inst. of Tech., Japan)
Peussa, P. (VTT Industrial systems, Finland)
Ravina, E. (Univ. of Genoa, Italy)
Röning, J. (Univ. of Oulu, Finland)
Salminen, V. (Lappeenranta Univ. of Tech., Finland)
Sashio, K. (Kobe Univ., Japan)
Shiota, Y. (Polytechnic Univ., Japan)
Sugita, N. (Univ. of Tokyo, Japan)
Takubo, T. (Osaka Univ., Japan)
Tani, K. (Gifu Univ., Japan)
Tanimizu, Y. (Osaka Pref. Univ., Japan)
Vähä, P. (VTT Electronics, Finland)
Virvalo, T. (Tampere Univ. of Tech., Finland)
Wikander, J. (The Royal Inst. of Tech., Sweden)
Yamamoto, M. (Kyushu Univ., Japan)

Contents

Legged robots

Control and Simulation

Position measurement

Actuators

Manufacturing systems

Applicaion of vision

Driving support and Wheel chair

Optimization and Reasoning

Robots

Sensors

Designing Mobile Work Machines in Cyber Space

Asko Ellman

Dept. of Mechanics and Design, Tampere University of Technology, Tampere, Finland

Abstract

Rapid development in ICT has made possible to accomplish detailed simulations of complete machine system in meaningful time and effort. Due to this, mechanical structures, power transmission, actuation and control systems can be designed more precisely. Furthermore, real-time simulation and hardware-in-the-loop simulation has enabled including of a human operator and real control system on the simulation. VR-technology together with user-centred design methods enables constructing of machines with good user experience.

Keywords:

Virtual Design; Mobile Work Machine, User Experience, Virtual Reality

1 INTRODUCTION

In most products of Mechanical Engineering the Time-to-Market and an understanding customer needs are today major factors in product success [1]. The speed of design process is very much gained during past 10 years due to development in 3D CAD, modelling and simulation methods. Functionality of rather complex machine can be simulated off-line which improves to optimize the design and reduce uncertainty. Increasing of computation power has made it possible to real-time simulation of simplified machine systems. This has enables hardware-in-the-loop simulations where part of the real machine can be simulated against simulation models, Figure 1.

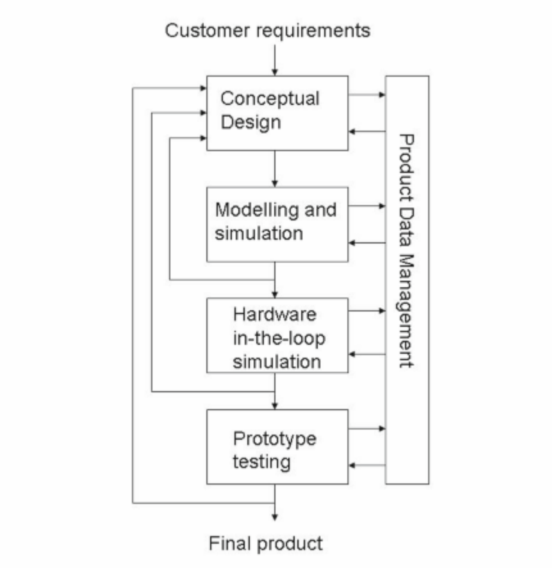


Figure 1. Phases in simulation based product development.

Cyberspace is virtual dimension created by computers and their interconnections such as Internet. In engineering this can be used for evaluation of prototypes by mathematical product models and Visualization. This will become more useful due to technical development:

- In 10 years time computational cost will be reduced with a factor of 1000.
- Internet backbone speed will be increased with a factor of 100.

The design will also be done distributed globally in collaborative context.

2 MOBILE MACHINES IN FINLAND

Typical mobile machines made in Finland, as shown in Figure 2, are highly specialized and they produced in small series. Often concurrent engineering is needed because there are number of different engineering teams attending design of such machines. Designing the control cabin of a mobile machine itself is challenging because it necessitates numerous trade-offs that need to be done between the various design quantities such as visibility, functionality, ergonomics, safety and industrial design.

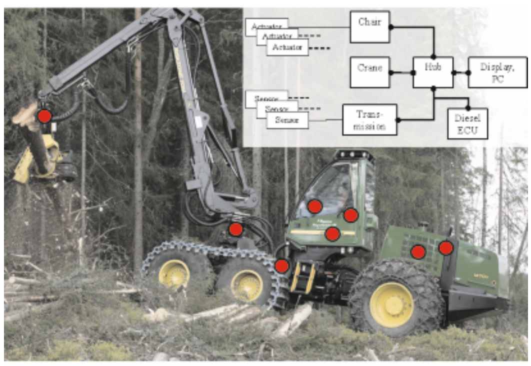


Figure 2. Mobile Machine.

Mobile Machines are challenging objects for Virtual Design due to their small series and rather tight time schedule which then leaves not much time and effort for Virtual Design. Therefore the machine design process need to be rethink so that there are more resources and time reserved at begin of the project for finding the customer needs and checking the plan. This time will be saved when revisions and corrections need not do be done at the end of the process.

At the moment Mobile machines are mainly manually driven machines. However, there is already number of teleoperated semi-autonomous machines which do not need a human driver but which need remote control by a human operator. Examples of this can be found from mining and harbour equipments. Next step of the development are fully autonomous machines which do not need human supervision. Due to complexity of their work environment there are no examples of this in Mobile Work Machines but in housekeeping there are already a hovering machine and lawn mower. As a next step in development is a autonomous machine group which is capable to operate together without human control, Figure 3. Complexity of even manually driven machines will increase as a function of time because more work supporting functions based on ICT will be added. As a result, complexity of Mobile Work Machines will increase significantly.

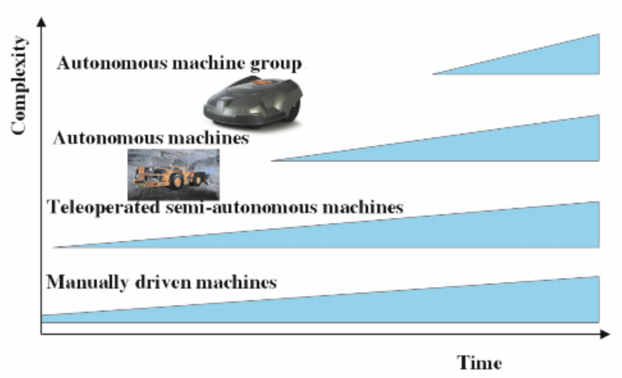


Figure 3. Increase of complexity in development of Mobile Machines.

3 TESTING OF CONTROL SYSTEM

Complexity of control system makes their testing very challenging. First, in distributed control system also the software is distributed in number of Electronic Control Units (ECU). Second, the amount of software has increased significantly due to number functions that are included in the system. In the future when machines will develop to be more automated and even autonomous the amount and complexity of the control system can be expected to increase furthermore. This is acknowledged to be a problem already at car industry.

One option for testing such system is using a Hardware-in-the-loop simulation where real control system of the machine including all the software is connected on real-time simulation model of the machine, Figure 4. In this approach test computer is feeding set of different input sequences on control system which is connected to real-time simulation computer via I/O –cards. In this manner functionality of the control system can broadly be tested and also some of the safety test can be accomplished.

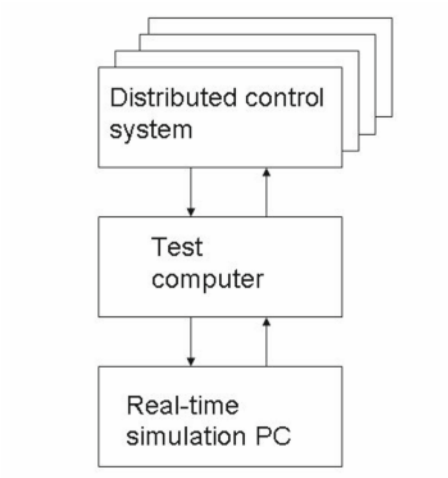


Figure 4. Automatic testing of control system using real-time simulation.

Depending on content and quality of real-time simulation model of the machine, different features of the control system can be tested

- Harness
- Functionality of the software
- Tuning of control system parameters
- Behavior of automated machines

Testing of harness is not trivial due to distributed nature of the control system. However, it can be done by I/O-mapping and some auxiliary functions that are necessary for supporting communication between ECU's.

Testing of functionality of software can be done in similar manner as harness testing if functionality of the software is described in state machine model. This testing enables finding of most logical errors which exists in control software code. Also accomplishing some safety testing may become possible in the future if safety testing standards and regulations will allow it.

If simulation model of the machine is accurate enough, it can be used for tuning of control system parameters. Generally a dynamic model is required which at the moment is challenging for hydraulically driven mechanisms.

In the case of testing of a single autonomous machine or even machine group requirements for testing becomes even more challenging due to complexity of work process and work environment. This will be especially highlighted is humans are allowed to enter to work space of a machine group.

4 DESIGN OF COCPIT WITH VIRTUAL REALITY TECHNOLOGY

Virtual Reality (VR) technology is very promising technology for acquiring customer requirements and conceptual design because it allows exploring the design in natural size with number of people [2]. VR technology has been used in large manufacturing companies, especially relating car and aerospace industry [3], [4] despite of its high cost. However, there has been significant development in software and hardware available on market. As a result exploring 3D CAD models in VE can be accomplished with a low cost PC devices and also conversion to VR platforms is no longer problem. Applying this technology has already been subject of interest in many companies.

Virtual prototype can be studied in several displays in VE allowing its exploring with number of people, Figure 5. This enables evaluation of design alternatives at conceptual design phase between customers, engineers, industrial designers and other stakeholders. VE offers also 3D vision of the design and allows the user to move in respect to picture. This gives a very realistic feeling on moving inside a 3D picture. This realism can further be increased by haptics [5] which gives force feedback to the user and makes virtual prototype tangible.

Figure 5. Evaluation of Virtual Prototype in VE.

In spite of its realism, VE provides also inconsistent sensations which easily disturbs the user and can cause simulator sickness in more serious case. This is emphasized due to fact that in many cases such test user are at first time in VE and they have not learn to tolerate these impacts as VR professionals have. This issue has been discussed in [6] and [7]. This can at least be partially be solved by proper test arrangements. A test situation can also be improved by introducing advanced interaction techniques [8], [9], [10] such as haptic glove shown in Figure 6.

Because there are always humans involved in VR as well in work process, some input from humanistic sciences is needed. Activity theory presented in work sociology [11],[12] user (subject), tool and outcome are the main components of an activity (a work task) as presented in Figure 7. Outcome of the work is it's result such as excavation or harvesting.

Figure 6. Evaluation of Human-Machine interface using a haptic glove.

There are interesting interactions between the actors in the triangle; Tool is loading the operator. This is generally linked with ergonomic studies. Also the operator is loading the Tool. As a result user's different behaviour is affecting fuel consumption, machine's productivity and lifetime of the machine. Another issue relating on this is that sometimes user's may misuse the machine and cause unexpected loading. This is generally due to two reasons: First, users do not know how designer has taught some work to be done. Second, users are using the machine in different task or way than designer has planned. This can be studied using a virtual prototype.

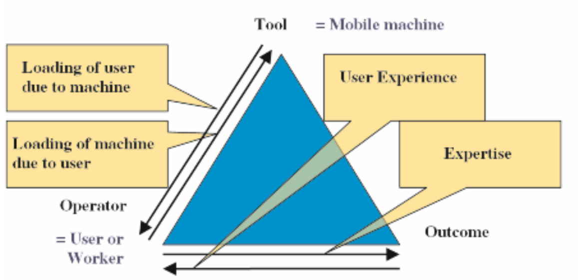


Figure 7. User experience in the case of workers' use mobile machine

Interesting interaction lies between the Operator and Outcome; Users are keen to put all their expertise on achieving a good outcome of their work. It is something that they are proud and which gives satisfaction. This important feedback is called user experience (UX) and it is considered to be very important part of products competence. Therefore its understanding is vitally important. The term UX is commonly connected to user interface design focusing on the joy to use a product, which connects the products to leisure time use [13].

This kind of experiment can be arranged in VE but for analysing the results knowledge on qualitative methods used in ethnography and anthropology is needed [14], [15], [16], [17].

In user studies of our laboratory we have made test arrangement shown in Figure 8. In the test Operator is driving the Virtual prototype and he/she is assisted by designer or operator of the VE system. An external Observer is needed for observing how the operator is managing in different test situations. This is necessary because the designer of the system may not see the issues that are in the system that he/she has designed. Usually test group of 6-8 persons is adequate for the test.

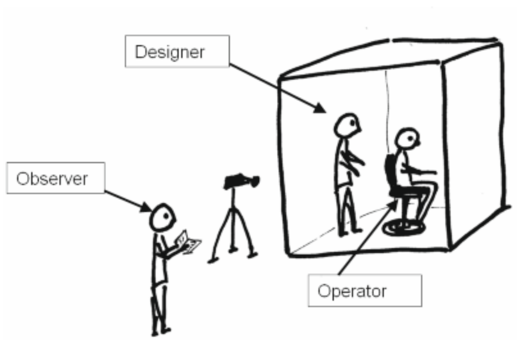


Figure 8. Test setup inspired by ethnography and anthropology.

For the understanding the emotions related driving to the machine a motion platform is connected to real-time simulation model and visualization as shown in Figure 9.

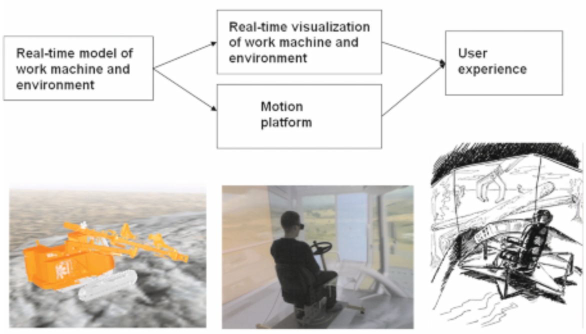


Figure 9. Test setup inspired by ethnography and anthropology.

5 SUMMARY

Development and testing of Mobile machines will increasing be made with product models, real-time-simulation and Hardware-in-the-loop simulation. This due to both increased complexity of machines as well as shortened design time. Especially testing of control systems of autonomous machines and machine groups will be complex.

Importance of Conceptual design phase can be extended with VR-tools and User testing. This supports construction of mobile work machines with good user experience.

Design of Mobile Work Machines will increasingly take place in Cyber Space.

6 REFERENCES

[1] Ulrich, K., Eppinger, S., 2003, Product Design and Development, McGraw-Hill, Boston, 3rd. Ed.

[2] Seron, F.J., Gutierrez, D., Magallon, J.A., Sobreviela. E.A., and Gutierrez, J.A., 2004, A CAVE-like environment as a tool for full-size train design. Virtual Reality, 7(2): 82-93.

[3] Dmitriev, K., Annen, T., Krawczyk, G., Myszkowski, K., Seidel., H.P., 2004, A CAVE System for Interactive Modeling of Global Illumination in Car Interior. MPI Informatik, Saarbrücken, Germany. VRST'04, November 10-12. pp 137-145.

[4] Schabacker, M., Sabeur, M., Vajna, S., 2006, INSIDES – A New Design and Simulation Platform for Virtual Prototyping in Automotive and Aerospace Industry. Proceedings of ASME 2006 IDETC/CIE Conference. Philadelphia. DETC2006-99228. 8p.

[5] Jayaram S., Joshi, H., Jayaram, U., Kim, Y., Kate, H., 2006, Embedding Haptics-Enabled Virtual Tools in CAD for Training Applications. Proceedings of ASME 2006 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, Philadelphia, DETC2006-99656. 11 p.

[6] Tiainen, T, Ellman, A, Katajamäki, T, and Kaapu, T., 2006, Occasional Users' Experience on Visiting in a Virtual Environment. In E. Alba, S.J. Turner, D. Roberts, and S.J.E. Taylor (Eds.) Proceedings of Tenth IEEE / ACM International Symposium on Distributed Simulation and Real-Time Applications (2-4.10.2006, Torremolinos, Malaga, Spain), IEEE Computer Society, Los Alamitos, California, USA (ISBN 0-7695-2697-7). pp. 63-69.

[7] Tiainen, T., Ellman, A., Kaapu, T., Davis, R., 2007. Effect of Navigation Task on Recalling Content: The Case of Occasional Users in Restricted, Cave-like Virtual Environment. To be published in E. Alba, S.J. Turner, D. Roberts, and S.J.E. Taylor (Eds.) Proceedings of 11th IEEE / ACM International Symposium on Distributed Simulation and Real-Time Applications, 22-34.10.2007, Chania, Greece, 8 p.

[8] Ellman, A., Laitinen, J., Tiainen, T., 2007. Combination of Virtual and Physical Objects in User-centered Design of Mobile Work Machine Cabin. Proceedings of IMECE2007, 2007 ASME International Congress and Exposition, November 11-15, 2007, Seattle, Washington, USA, 6 p.

[9] Iltanen, M., Ellman, A., Laitinen, J., 2007. Wearable haptic device for an IPT system based on pneumatic muscles. Proceedings of 2007 ASME International Design Engineering Technical Conferences IDECT/CIE 4-7.9.2007, Las Vegas, USA, 7 p.

[10] Kuusisto, J., Ellman, A., 2006. Speech control interfaces in virtual environments. ICMA 2006 International conference on mechatronics 7.-8.6.2006. Seinäjoki, Finland. 7 p

[11] Vygotskij, L.S., 1978, Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA, Harward University Press.

[12] Engeström, Y., 1987, Learning by expanding, Orienta-konsultit, Helsinki.

[13] Nielsen, 2009, Nielsen Norman Group, Strategies to enhance the user experience. http://www.nngroup.com /about/userexperience.html obtained 30.1.2009

[14] Eriksen, T. H., 2001. Small places, big issues: An introduction to social and cultural anthropology. London: Pluto. (First edition 1995).

[15] Marton, F., 1981, Phenomenography – describing conceptions of the world around us, Instructional science, (10):177-200.

[16] Marton, F., 1982, Towards phenomenography of learning, Integratial experiments aspects, University of Gothenburg Dept. Education, Gothenburg.

[17] Meyer, C., and Schwager, A., 2007. Understanding Customer Experience. Harvard Business Review, February, 117-126.

Proposal of 3D Micro Prototyping Using Synchrotron Radiation and Its Application to Bio-Microsystems

Yuichi Utsumi

Laboratory of Advanced Science and Technology for Industry , University of Hyogo, Kamigori, Ako, Hyogo, Japan

Abstract

A new X-ray microfabrication system and succeeding molding process as the 3D micro prototyping process have been developed using synchrotron radiation (SR) lithography and nano-imprinting technique. We adapted the process to the achievement of 3D microfluidic platforms for bio chemical applications which will be used point-of-care diagnostics and drug compound screenings. An enzyme linked immunosorbent assay method has been applied to the analysis of the endocrine disrupter using proposed fluidic platforms. Drastic improvement of the analysis sensitivity and decreasing of the total analysis effort and required time have confirmed

Keywords:

synchrotron radiation; microfabrication; microfluidics; Lab on chip; assay ; ELISA ;DNA

1 INTRODUCTION

There have been rapid developments in the application of microsystems in advanced industries such as intelligent information systems, energy and environment conservations, and medical and biochemical applications. Microsystems typically consist of different types of precise parts for various micro structures. The realization of 3D microstructures integrating multiple functions, such as electrical, optical, mechanical, and chemical sequential operations, in a restricted space will bring many advantages to the industry development. Systems of this type have been fabricated using micro-electro-mechanical system (MEMS) processes. Recently, however, fabrication techniques with a higher precision and a higher aspect ratio than those conventionally achievable have become increasingly important. This is due to the fact that natural phenomena on which these device functions are based, such as electrostatic fields, surface tension, and surface chemical reactions, tend to become more pronounced as specific surface area increases. To realize these requirement, we developed a new “3D micro prototyping process” based on X-ray microfabrication system equipped at the “NewSUBARU” SR facility [1,2]. and succeeding molding process using synchrotron radiation (SR) lithography and nano-imprinting technique. The stacking process as the device packaging of obtained micro structures is more essential where it needs various surface treatment and succeeding bonding process with different materials [3,4].

Meanwhile, the progress of life science has been increasing rapidly and the development of the platform technologies that supports it become more significant. Miniaturization and integration technology that have been successfully developed at microelectronics fields are nowadays adapting also to automated chemical analysis and synthesis, based on miniaturized total chemical system, so-called “μTAS “ or “Lab on Chip” made from microfluidic components. We adapted proposed “3D micro prototyping process” to the achievement of the integrated microfluidic platform and have confirmed advantages of 3D micro-integration of chemical functions in one platform for some biochemical applications. A high sensitive and rapid enzyme linked immunosorbent assay (ELISA) method are demonstrated using micro 3D-structured system consists of 3D microfluidic channel network and vertical chemical operation chamber with fluid control filter and mixer. Drastic improvement of the analysis sensitivity and decreasing the total analysis effort and required time have found in the applications to the environmental analysis and DNA analysis.

2 3D MICRO PROTOTYPING USING SYNCHROTRON RADIATION

2.1 3D X-ray Microfabrication System

The “lithographite, galvanoformung, abformung” (LIGA) process, which consists of deep-X-ray lithography, electroforming, and molding, is a promising candidate for 3D microfabrication[5]. The LIGA process starts from the fabrication of high-aspect-ratio polymer microstructures with heights greater than a few hundred microns using deep-X-ray lithography of a photosensitive polymer (resist) [6]. In the next step, a metal replica structure is formed by electroforming using the fabricated polymer master. The obtained metal replica structure can be used as a component in a

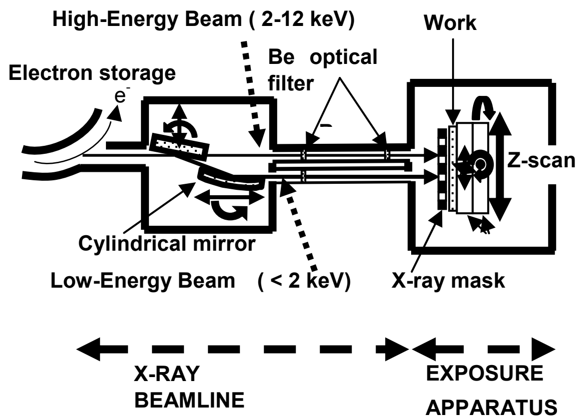


Figure 1: Schematic diagram of newly developed lithography system using SR.

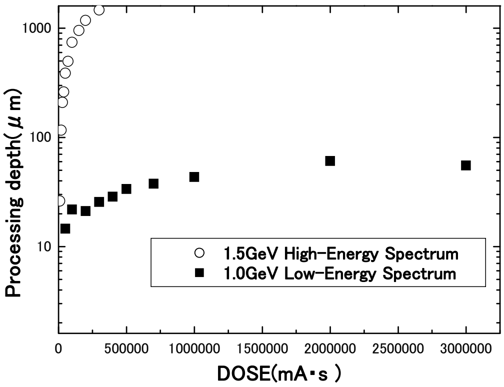


Figure 2: Exposure-dose dependence of processing depth. Circles show the data using the high-energy spectrum for 1.5 GeV operation. Squares show the data using the low-energy spectrum for 1.0 GeV operation.

microsystem or as a mold insert for final molding processes, such as hot embossing and injection molding. It is important to develop a next-generation process that can achieve both a large process area and more fine fabrication property simultaneously. No such process satisfying the demands has yet been realized. We established a new X-ray lithography system, “BL2”, with a large exposure area of up to A4 size, developed at the “NewSUBARU” synchrotron radiation facility [2]. Figure 1 shows a schematic diagram of the newly developed apparatus for lithography. It consists of a beamline with X-ray optics for energy selection and an exposure apparatus with multi motion stages. The main feature of our lithography system is that it can continuously select X-rays from 1 to 12 keV using mirrors and the filters. Each energy region can be selected in accordance with the desired size and shape of the fabricated microstructures. Figure 2 shows the relationship between the processing depth of a PMMA resist after 3 h of development and the X-ray exposure dose at operation energies of 1.0 and 1.5 GeV. A processing depth of more than 1600 μm was obtained with exposure to the high-energy X-ray beam during 1.5 GeV operation. It is clearly shown that exposure photon energy changes dose dependence on processing depth of the work (PMMA), since the penetration depth of the X-rays into the resist varies according to their energy. By scanning the 210-mm-wide SR beam along the longitudinal axis with a 800 mm span, full A4-size exposure can be achieved. Figure 3 shows that

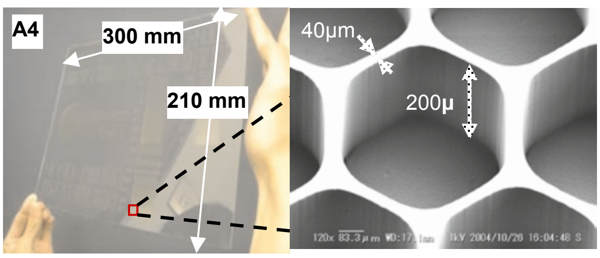


Figure 3: PMMA resist patterns fabricated using large-area deep x-ray lithography system. A4-size deep x-ray lithography, and right magnified image shows patterns with high accuracy.

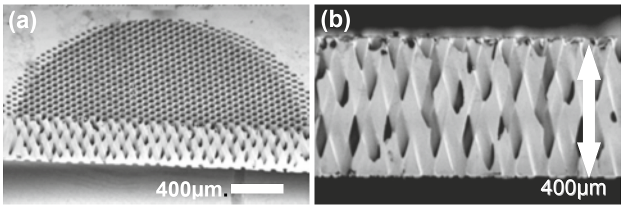


Figure 4: (a);Three-dimensional microstructures patterned on a 400 micron PMMA sheet using SR with multi step exposure. (b); magnified image of the cross-linked micro capillaries.

large-area patterning of an area of 220 mm × 300 mm (A4-size) with a highly uniform pattern thickness is successfully demonstrated using this system. The uniformity of the pattern thickness along the horizontal direction was estimated to be less than 5%. Patterning with an aspect ratio greater than 20 was confirmed. Using this large-area X-ray lithography system, high-aspect-ratio microstructures with a wide range of sizes, from submicron to millimeter, can be achieved by using different type of X-ray masks in the same exposure chamber.

In order to apply this system to the fabrication of advanced microsystems with more integrated functions for automated operations, it is necessary to develop a high-precision 3D microfabrication technique capable of sizes less than several dozen of microns. Several 3D microfabrication studies have been reported so far [7-8]. In this study, three-dimensional microstructures can be obtained by exposing the X-ray beam intermittently over work substrate (PMMA) with different incident SR angles set by multi axis stages [9] as shown in Figure 4 (a), (b). This structure shows the cross-linked micro capillaries for mixing of micro fluids with small Reynolds number, where high mixing efficiency was confirmed from the reaction speed measurement of the enzyme and substrate solution [10]. This 3D structure of cross-linked micro capillaries can be fabricated by tilting the exposure stage. The 400μm thick PMMA sheet and X-ray mask are mounted on an X-ray exposure stage, and the stage is tiled from the vertical axis, and PMMA sheet is exposed to X-ray. Then, the stage is reversely tilted by same angle and exposured again. The distance between crossing axies of the capillaries can be controlled by rotating the X-ray mask a few degrees. The mixing efficiency was significantly affected by the distance between center axies of each capillary. Figure 5 (a) shows the results of the computational fluid dynamics (CFD) simulation of water mixed after 0.5 ms, which suggests the great improvement of the mixing efficiency by using this 3D structured capillaries. The color bar indicates mass fraction of the colored water. Figure 5 (b) shows time shift of standard deviation of the mass fraction of the ideal separated waters inpoured from both side of the two cross-linked capillaries and mixed in capillaries as the mixing progress. The capillary structural parameter is 40 μm and 400μm in diameter and length. The result shows that mixing efficiency in cross-linked capillary is much increased at the structure of which distance between center axies of each capillary is half in diameter (half cross-linked structure). Moreover, the mixing will be finished within 1 ms by the use of the mixer with half cross-linked micro capillaries.

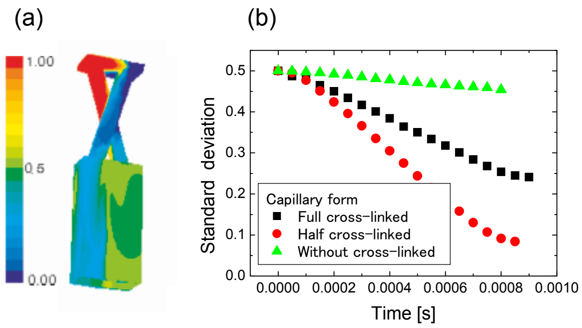


Figure 5: (a); Simulated mixing behavior. (b); time shift of standard deviation of the mass fraction of the ideal separated waters inpoured from both side of the two cross-linked capillaries.

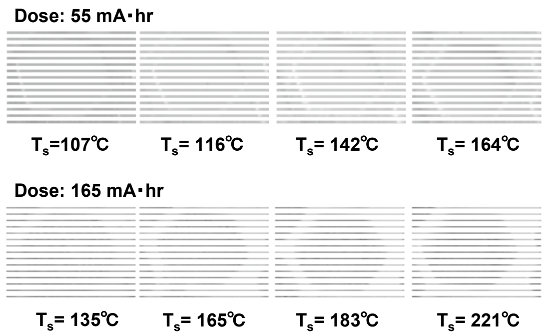


Figure 6: SEM images of each etched surface irradiated by SR with different substrate temperature and x-ray doses.

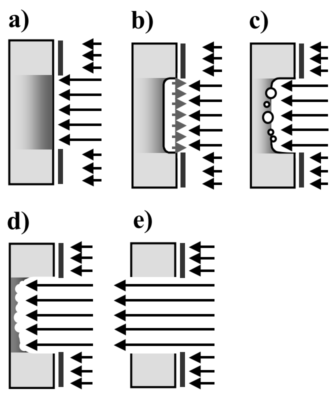


Figure 7: Speculated mechanism of PTFE etching; (a) PTFE decomposition, fragment desorption and decrease of melting point, (b) formation of melting block at the SR-irradiated area, (c) growth of the bubbles in melting block, (d) continuous etching of melting block by SR-induced reaction and thermal desorption, (e) finish of PTFE etching at SR-irradiated area.

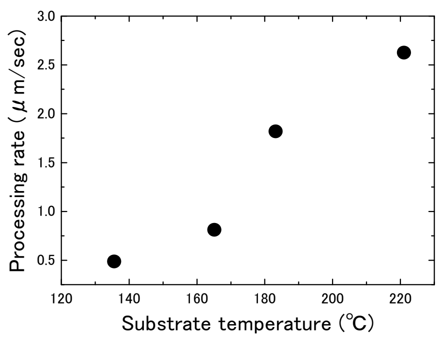


Figure 8: Etching rate dependence on substrate temperature.

2.2 Microfabrication of PTFE by SR-induced Direct Etching

The X-ray microfabrication system can also apply the SR-induced photo-chemical reaction to direct etching of fluorinated polymer molecules such as polytetrafluoroethylene (PTFE) and polytetrafluoroethylene-co-perfluoroalkoxy vinyl ether [11-12]. PTFE has big potential for various applications, such as chemical, bioscience, electronics and environmental applications, for its electrical, chemical, mechanical and thermal properties. Due to its property microfabrication of PTFE with high precision has been difficult so far by the use of conventional machining, molding, and semiconductor processes. In order to accomplish wider practical applications of PTFE, microfabrication with large work-area is inevitable. We have demonstrated the PTFE mocrofabrication using the developed X-ray Microfabrication System and investigated the processing characteristics of PTFE direct etching.

At the first step, processing characteristics of the PTFE direct etching using SR had been studied. In the experiments, relatively high-energy region of 2 to 6 keV and 2 to over 10 keV were used at the condition of storage electron-beam energy of 1.0GeV and 1.5GeV, respectively. To investigate heating effects during X-ray exposure, large-area hotplate was mounted onto the stage and PTFE substrates, which cleaned with methanol, were attached on the hotplate to face the X-ray beam.

Figure 6 shows For the lower dose of 55 mA·hr the irradiated surface shows very smooth and fine morphology. On the contrary, there were found bubble-like structures at the higher dose of 165 mA·hr. The morphology is closely related to the etching mechanism as speculated below. As shown in Figure 7, it can be speculated that melting of PTFE less than original melting point occurs by SR irradiation, because high energy soft-x-ray photons induce continuous decomposition of PTFE. The volatile decomposed fragments of PTFE desorb from the SR-irradiated surfaces(Fig. 7(a)). As x-rays penetrate into PTFE substrate deeply, the bond-brake progresses in deeper part of the substrate. Furthermore, as elevated heat

contributes to rapid diffusion of decomposed PTFE molecules, it leads to the formation of large melting block expanding in entire exposed aria(Fig. 7(b)). This melting of PTFE was supposed o result in the abrupt rate increase around 165 degrees as shown in Figure 8. Decomposed fragments will agglutinated and results in the growth of the bubbles in melting block (Fig. 7(c)). Etching of melting block by SR-induced reaction and thermal desorption continuously progresses during these steps and finally the irradiated part will be selectively etched (Fig. 7(d,e)). By these processes, PTFE will be precisely etched by the use of developed lithography system.
Next, the PTFE micro capillariy filters with functions as microvalve, micromixer, and microreactor, were fabricated.
Figure 9 is SEM images of fabricated PTFE filter with about two thousand arrayed micro capillaries. The thicknesses and capillaty diameter are rangeing from 130 to 2000 µm and 25 to 40 µm, respectively so the maximum aspect ratio of capillary was 80. Patterning to A5 size PTFE sheet using wet-etched stencil mask was also demonstrated. It is useable for large-area microstructure applications such as lab-on-a-chip devices for medical and environmental high throughput analysis.

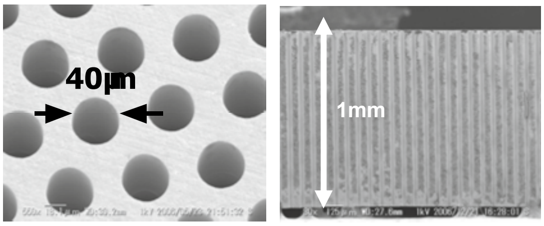


Figure 9: SEM images of fabricated PTFE filter with about two thousand arrayed micro capillaries.

2.3 Stacking of Microstructure using Direct Bonding and Nanoimprint Technique

The next process following after deep x-ray lithography, molding process was investigated using nanoimprint technique. Figure 10 shows the SEM images of the microchannel structure of micro capillary electrophoresis (MCE) chip for DNA analysis formed after molding of PMMA sheet by nanoimprint apparatus. The accuracy of the line dimension was less than 4% from the nickel mold inserts. And the surface roughness of the micro channel wall was less than few nm, which show the enough precision for assembly the whole DNA chip structure with multiple functions. In order to achieve the microsystem with desired functions, the assembly of the individual micro parts is inevitable. We have attained the assembly of the plastic micro parts by the

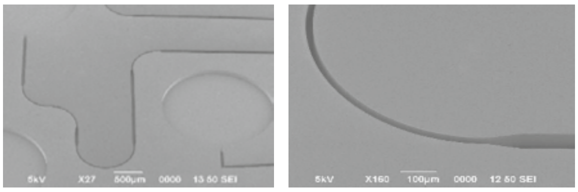


Figure 10: SEM images of the micro channel structure of MCE.

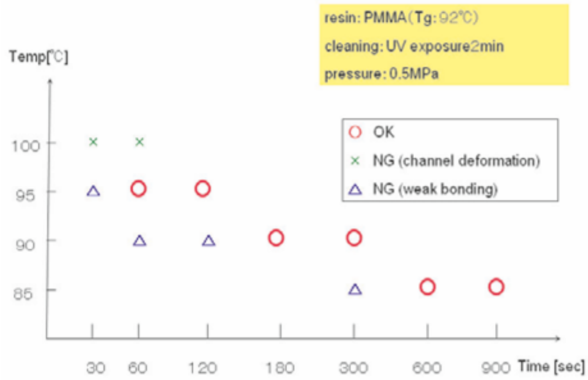


Figure 11: The result of the trials for fusion bonding.

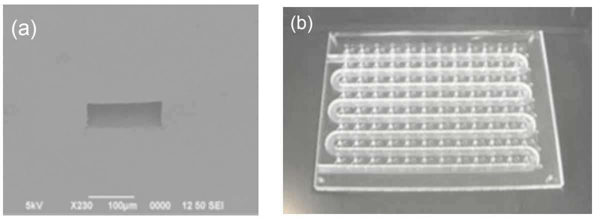


Figure 12: Outside view of the stacked structure of chip for DNA analysis.

use of the direct fusion bonding using nanoimprint apparatus to press the stacked molded structures in head loads. The fusion bonding without binding agent is necessary to avoid the inclusion of the binding agent into the micro fluid channel and to attain the dimension accuracy, which is degraded by the insertion of binding layer with low thickness stability.The fusion bonding of each three PMMA layer; top layer of a sample injection, second electrodes layer for DNA amplification and separation with 96 MCE units, and bottom heat exchanger layer, was demonstrated after some treatments of the bonding surface. Figure 11 shows the result of the trials for fusion bonding of two PMMA layers at different temperature and bonding time. The cross sectional SEM images as shown in Figure 12(a) shows the sealed microchannel with no defects and air bubble at the sealed interface and Figure 12 (b) shows the whole outside view of the stacked structure of chip with 96 MCE units for DNA analysis, which size is 88mm x 130 mm[13].

3 APPLICATION TO BIO MICORSYSTEMS

3.1 Proposal of Bio Microreactor with Vertical Chemical Operation

The automation of bio-operations has became a key subject to realize point-of-care (POC) diagnostics and high throughput screenings (HTS) of various matters, such as medicine, DNA, proteins in post-genome analysis, products from immunoassays. From this point of view, “Lab on Chip” system, on which multiple chemical operations are sequentially performed through microfluidic components have been attracted great attentions, because of their leading features such as compactness, small sample volume, fast and precisely controlled reaction, high analysis sensitivity, shortage of required time, reproducibility of output data, and low cost reliance. Such properties originated from down sizing of fluidic channels and components integrated monolithically which results in a large surface-area-to-volume ratio, a rapid thermal diffusion, a high pressure gradient, a high reagent

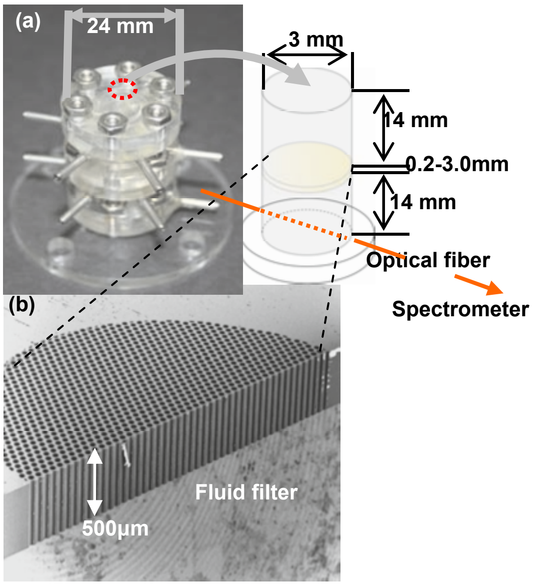


Figure 13: Out view of the vertical reactor with capillary bundle structure.

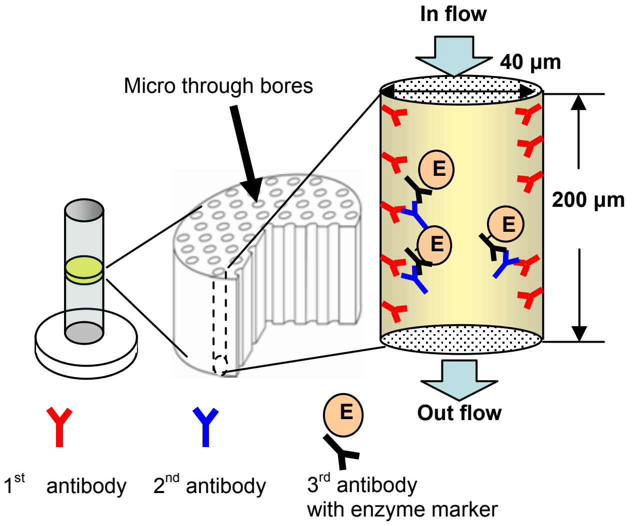


Figure 14: Scheme of the high sensitive ELISA analysis using immunoreaction arisen with the antibody stacks immobilized inside the micro capillary wall.

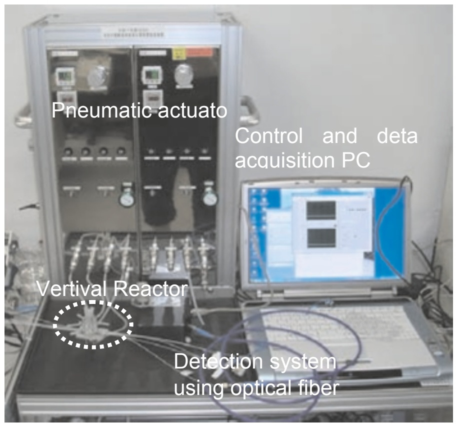


Figure 15 : Out view of the automatic ELISA system using proposed vertical immuno reactor fabricated by SR lithography.

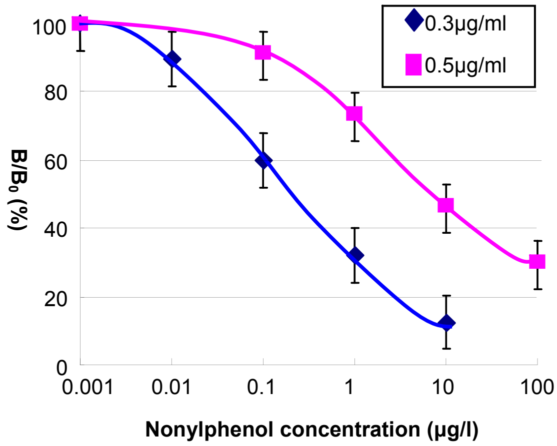


Figure 16: Calibration curve of nonylphenol concentration in the competitive ELISA.

solution concentration, and a mechanical toughness. However, typical structure of the conventional “Lab on Chip” systems assembled on a two-dimensional planar substrate has restrictions of integration of every fluidic components necessary for attaining whole sequential chemical operations specified in an analysis and synthesis.

Achieving 3D integration of multiple chemical modules and these interconnections within a finite small space will results in the integration of multiple functions which composes total sequential chemical operation, and will develop several major application fields, such as POC diagnositics, HTS of the matters, fine chemical synthesis, combinatorial synthesis.

As the first step to achieve 3D integration of microfluidics, we proposed and fabricated a new chemical reactor with a vertical fluid flow operation. As shown in Figure 13, the reactor utilizes fluid filter in which thousands of micro-through-bores are bundled, which sustains the liquids loaded from the first upper unit reservoir. For the step of the transportation of the sustained liquids towards the lower unit reservoir, pneumatic pressure will be loaded on the surface of the liquids. In the first step of loading of the reagents from the upper unit reservoir, different reagents are not mixed sufficiently due to the lowReynolds number of the reservoir. By the use of this fluid filter, it is assumed from the reaction rate measurement and computational fluidic dynamics simulations that different reagents are mixed entirely in a few milliseconds during their transportation from upper to lower reservoir and rapid chemical reactions can also be achieved; whereas, this filter can stock the reagent solutions with antibodies immobilized on the surface of the micro-through-bores, at which immuno reaction for ELISA occurs as shown in Figure 14 (sandwich method). We investigated the possibility of the analysis using the proposed vertical micro reactor for competitive enzyme-linked immunosorbent assay (ELISA), which enable highsensitive detection of an endocrine disrupter (nonylphenol) by a series of vertical fluidic operation. In this analysis, competition between nonylphenol and nonylphenol-Horseradish-peroxidase conjugate occurs at the binding of these molecules to anti-nonylphenol monoclonal antibody (anti-NP-antibody) immobilized on surface of the micro through bores as shown in Figure 9. In the assay using filters pretreated with 0.3 µg/ml of anti-NP-antibody, even at 0.01ng/ml of free NP, we still observed the inhibition of the binding (B/B_0=90%). Whole analysis protocol have been performed using automatically operating system shown in Figure 15. Under this condition, calibration curve of NP was obtained at the range of 0.01-10ng/ml as shown in Figure 16. This sensitivity was two orders higher than the sensitivity (5-100 ng/ml) obtained by ordinary ELISA using 96-wells-micro-titer plate and the same anti-NP-antibody. These assays gave reproducible results, reactor to set at the centers of the filters[14-15].

3.2 3D Micro Fluidic Network

The miniaturized analysis chip such as microchip electrophoresis have been well developed, however the micro chemical systems with whole bio-chemical processes, including sample pre-treatment, have not been developed, while it requires much efforts and time over separation and detection. The key-technology for the solution is assembly and interconnection technologies because seamless functional connection of various devices is necessary to acueve totally automated microchip systems. The various chemical operations must be realized on one platform to realize totally automated bio-process. We adapted proposed "3D micro prototyping process" to the achievement of such advanced systems and have confirmed the advantages of 3D micro-integration of chemical functions in one chip for DNA analysis and some immunoassay applications. We proposed 3D fluidic platform system in which total chemical unit operation necessary to achieve without human operations can perform automatically, which brings high analysis sensitivity and less total analysis effort and required time. In this concept whole procedure for immunoassay can be achievable by stacking multiple CD fluids with individual functions corresponds to each sequential step for bio-chemical operation to cover whole procedure for assay. This concept is shown as the schematic diagram of a new CD-like microfluidic platform with 3D fluid networks in Figure 17. In this system only necessary manual operation is loading of sample and reagent solutions into inlets. The analysis processes typically start with this sample purification followed by some sample preparation dilution and amplification. The purified samples are then mixed with buffer solutions containing another regents such as enzyme-labelled haptens by increasing the rotation speed of CD. The regents are pre-loaded in the other chamber formed on the bottom disk of the stack structure. The specific volumes of regent solutions and sample liquids are automatically injected into mixing chamber located in the middle disk of the stacked structure consists of three-dimensionally crossing channel network, and dispensed to desired volume ratios for designed protocol. For complete mixing of sample and regents the high-efficiency mixing device is necessary. For this purpose a 3D mixer with opposed-capillary structure is feasible at the mixing chamber. The mixture will next injected into biochemical reaction chamber with antibody immobilized 3D microstructure. Specific area of immobilized antibody become to be several decades times larger than conventional planar reactor microchannels, which results in the enhancement of the reaction rate and the analysis sensitivity, reproducibility of the immuno systems. The solution of product of bio-chemical reaction will then injected into detection chamber located after the reaction chamber and amount of the product will measured by using optical detection. Normally, optical path of micro channel on a conventional planer fluidic platform, which corresponds to the depth of the fluidic channel, is less than 100 micron and it has restrictions for detection such as difficulty of alignment and small signal due to the lack of optical path for absorption and fluorescence detection. On the other hand stacked CD structure can increase optical path up to several millimetres by forming detection channel across the stacked CDs.

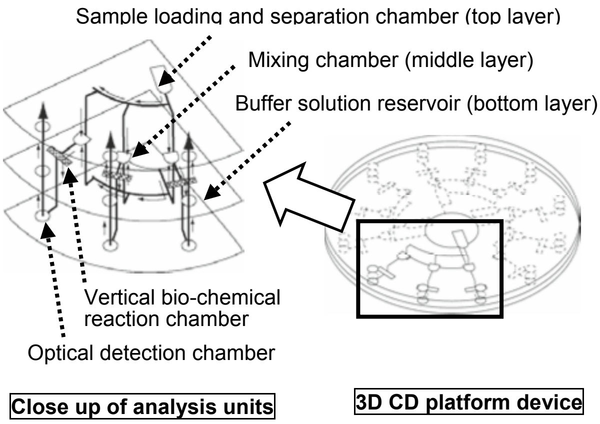


Figure 17: Schematic diagram of a new CD-like microfluidic platform with 3D fluid networks.

3.3 Application of 3D Fluidic System to High Sensitive Elisa Immunoassay

The other benefits of the automation are realization of highly precise and reliable operations by reducing human errors or mistakes. One promising way to realize automation for immuassay is to utilize CD-like microfluidic platforms as mentioned above. In this section we will describe the detailed structure and protocol for ELISA with high-sensitive and high-speed, and high-reliable property of 3D CD platform devices. For automatic ELISA, the liquids, sample, wash, substrate, and reaction aborting solution are pre-loaded into reservoirs and sequentially injected into 3D bio-reaction chamber by spinning CD. For the first sequencing, the mixture of sample and enzyme conjugated hapten will be injected and competitively conjugate with 1st antibody immobilized on the

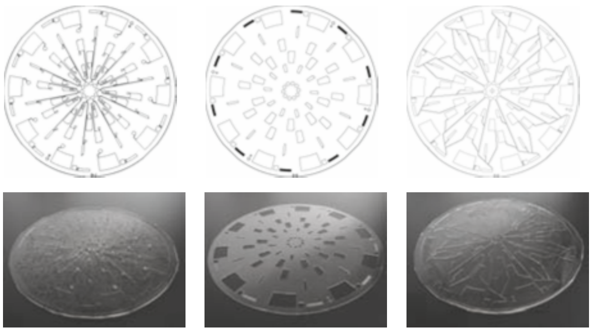


Figure:18 Illustration and photographs of designed and fabricated CD platfrom consist of three layers

surface of 3D microstructure. After incubation, the washing buffer will injected from second reservoirs by accelerating hundreds rpm, and excess samples will be washed away. The next, substrate reagent to quantify the extent of competitive reaction of sample and enzyme labelled haptens will be injected from third reservoir into reaction chamber and incubated to amplify the signal of products according to designed protocol. Finally the enzyme reaction will stopped by injecting reaction stopping solution released from fourth reservoir, and amount of products will measured in optical chamber and calibration curve will obtained. Figure 18 shows illustration and photographs of designed and fabricated devices. The devices consist of three-stacked layers with ten assay units for parallel assay of 10 samples. Each unit consists of four liquid loading reservoirs and biochemical chamber with filter structure as mentioned at chapter 3.1 and optical detection chamber. All CDs are aligned and stacked together by using self-adhesion of poly-dimethylsiloxane (PDMS). The thickness of the reservoirs and chambers set to adjust the suitable sample, reagent volumes and optical path length of the reaction products for UV absorption and fluorescence. The high aspect ratio structure of liquid loading reservoir is effective to integrate assay units by reducing planar area of reservoirs. To demonstrate the automatic sequencing in 3D CD platform, we loaded protein solution into the reservoirs (0.1% bovine serum albumin phosphate buffer solution: 0.1%BSA) and CD platform is spun for the sequencing. The snapshot images are obtained by strobe scope system to observe and measure the burst frequency of each chamber. The schematic illustration of strobe scope system is shown in Figure 19. The optical sensor will generate

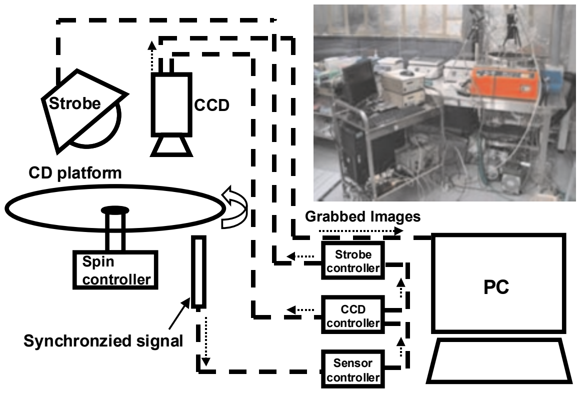


Figure :19 Schematic illustration and photograph of strobe scope system

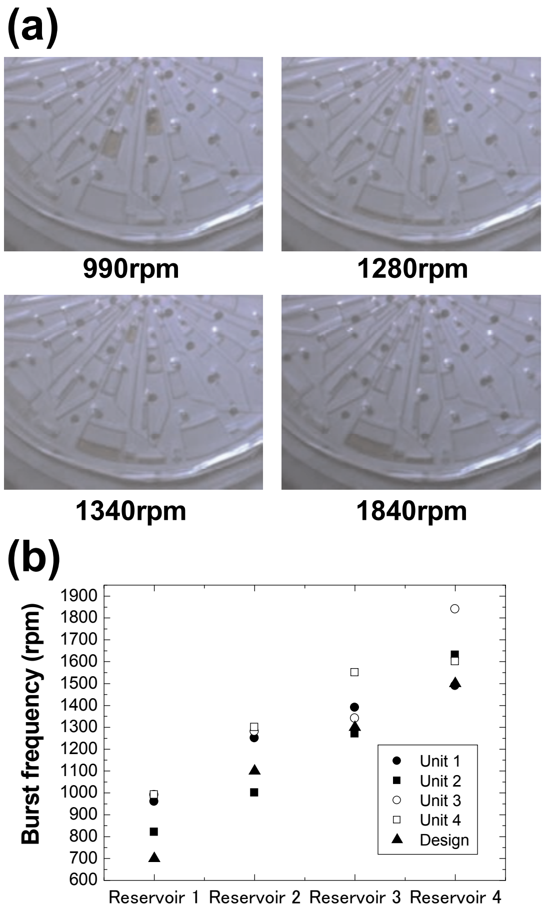


Figure: 20 (a) Images of flow sequencing of BSA solution and (b) measured burst frequency the trigger signals for the synchronized control of CCD and strobe. The taken images are automatically grabbed with time information into PC and burst frequency was calculated by checking holding state and the time.Figure15 (a) show the obtained images of automatic flow sequencing in 3D CD platform and measured burst frequencies are shown in Figure 20 (b). As shown in Figure20 (b) we succeeded in sequential transportation of BSA solution in four individual reactor units in 3D CD-like platforms. The result suggests the proposed 3D microfluidcs platform is available to automated ELISA analysis.

4 SUMMARY

3D micro prototyping process have been developed using synchrotron radiation (SR) lithography and nano-imprinting technique. Large-area patterning up to A4-size area was also successfully performed with a highly uniform pattern thickness. The X-ray microfabrication system can also apply the SR-induced photo-chemical reaction to direct etching of fluorinated polymer molecules. The stacking process as the device assembly and interconnections for obtained micro structures is also demonstrated using some types of surface treatment and succeeding nano-imprinting techniques for assembling 3D functional fluidic structure. We adapted proposed “3D micro prototyping process” to the achievement of 3D micro fluidic platforms and have confirmed the novel properties of high analysis sensitivity, speed, and low reagent consumption. for some assay applications. An enzyme linked immunosorbent assay method has been investigated using proposed micro 3D fluidic platforms in which thousands of micro capillary are integrated as a fluid control filter and bio-chemical reaction space in 3D fluid networks. Drastic improvement of the analysis sensitivity and decreasing the total analysis effort and required time have found in the applications for the environmental analysis. We also succeeded in sequential transportation of BSA solution in four individual reactor units in 3D CD-like platforms.

5 ACKNOWLEDGMENTS

This work has performed with the aides of New Energy Development Organization, Japan Society for the Promotion of Science, and COE program of Hyogo Prefecture, Japan.

6 REFERENCES

[1] A. Ando, T. Hattori, K. Hosono, K. Kanda, H. Kinoshita, S. Matsui, H. Mekaru, M. Niibe, Y. Utsumi and T. Watanabe : Journal of The Japanese Society for Synchrotron Radiation Research, **15**, (2002) p. 336.

[2] Y.Utsumi, and T. Kishimoto, Journal of. Vacuum. Science and Technologies. B, **23**, 6, 2903-2909 (2005).

[3] D.Fukuoka, and Y.Utsumi, p387., International Conference on Electronics Packaging 2008, JAPAN, June 2008

[4] Y. Utsumi, T. Ikeda, M. Minamitani, .K Suwa, p.125., 7th International Workshop on High-Aspect-Ratio Micro-Structure Technology 2007, FRANCE, June 2007

[5] F. J. Pantenburg, S. Achenbach and S. Mohr, Microsystems Tech. **4**, (1998) 89.

[6] E. W. Becker, W. Ehrfeld, P. Hagmann, A. Maner and D. Munchmeyer, Microelectronic Eng. **4**, (1986) 35.

[7] W. Ehrfeld : Proc. High Aspect Ratio Micro Structure Technology, 2001,(2001) p. 05.

[8] J. Goettert, G. Aigeldinger, Y. Desta, Z.L.Ling, and L. Rupp, Proc. Synchrotron Radiation Instrumentation, 2001, (2001) p. 102.

[9] Y. Utsumi, M.Minamitani and T. Hattori : Electrical Engineering in japan , **165,1** (2008) p.52.

[10] K. Fujiwara, Y. Ukita, M. Takeo, S. Negoro, T. Kanie, M. Katayama, and Y. Utsumi, p.297., 7th International Workshop on High-Aspect-Ratio Micro-Structure Technology 2007, FRANCE, June 2007

[11] Y. Ukita, M. Kishihara, Y.Haruyama, K. Kanda, S.Matsui, K. Michiji, and Y.Utsumi, Japanese Journal of Applied Physics, 47, pp. 337-341 (2008).

[12] M. Kishihara, Y. Ukita, Y. Utsumi, and I. Ohta, p.301., 7th International Workshop on High-Aspect-Ratio Micro-Structure Technology 2007, FRANCE, June 2007

[13] Y.Utsumi, M.Minamitani and T. Hattori : Electrical Engineering in japan , **165,1** (2008) p.52.

[14] Y.Utsumi, Y.Ukita, and T. sano, 51st International Conference on Electron Ion and Photon Beam Technology and Nanofabrication, USA, May 2007

[15] K. Matsui, I. Kawaji, Y. Utsumi, Y. Ukita, T. Asano, M. Takeo, D. Kato, and S. Negoro, Journal of Bioscience and Bioengineering,104, 4, 347-350 (2007).

[16] M. J. Madou and G. J. Kellogg, Proc. SPIE, **3259**,1998, p80-93

[17] S. Lai, S. Wang, J. Luo, L.J. Lee, S-T. Yang, and M. J. Madou, Anal. Chem., 76, 1832-1837, 2004

[18] Y-K. Cho, J-G. Lee, J-M Park, B-S. Lee,Y. Lee, and C. Ko, Lab. Chip, **7**, 565-573, 2007.

Surgical Tool Based Preoperative Planning System for Total Hip Arthroplasty

Makoto Yagihashi[1], Hirohisa Narita[1] and Hideo Fujimoto[1]

[1] Nagoya Institute of Technology, Nagoya, Japan

Abstract

In preoperative planning of total hip arthroplasty, a type, a size and a place of a femoral stem is planned on plane radiographs or computed tomography images. However, the shape of the natural medullary cavity of the femur on these images differs from a prepared one expanded with reamers and rasps to insert the stem. This difference makes it hard to select an appropriate stem and its position without imagining behavior of surgical tools. In this paper, we propose a surgical tool based preoperative planning system for total hip arthroplasty to obtain a minimally invasive position of the stem.

Keywords:

Total hip arthroplasty; Preoperative planning; Minimally invasive surgery

1 INTRODUCTION

Appropriate selection and placement of a femoral stem lead to a good surgical result. The femoral stem which does not suit patient's femur causes leg length discrepancy and shortens the longevity of the artificial hip joint. Even if a surgeon found that the selected femoral stem is not fit the femur during the operation, reselecting another femoral stem prolongs operation time to increase the burden for patients and risk of infectious disease.

In the preoperative planning of total hip arthroplasty (THA), the surgeon selects the femoral stem subjectively. A type, a size and a place of the femoral stem is planned on plane radiographs or computed tomography (CT) images. A template of the femoral stem is placed on these images to find an appropriate position of the femoral stem referring to a shape of the medullary cavity. There are some studies to find the optimal position of the femoral stem automatically [1][2], but there is still no gold standard except the surgeon's experiences.

The shape of the medullary cavity of femur is also determined by the surgeon subjectively. The shape of the natural medullary cavity is given as a contoured one with a certain threshold of Hounsfield Unit (HU). The surgeon imagines the contour of the medullary cavity on plane radiographs or sets proper HU to shape one on a preoperative planning system. This procedure is called as segmentation, and there are some attempts to make segmented bone model objectively, but automatic segmentation is not still free from the surgeon's experiences [3].

In this paper, we present a novel method to determine the position of the femoral stem objectively by predicting the shaped medullary cavity of the femur. The femoral stem is inserted into a medullary cavity of femur after the surgeon expands the medullary cavity with surgical tools such as reamer and rasps. This prepared shape of the medullary cavity differs from the natural shape of one. The surgical tools are designed to shape the medullary cavity into a little smaller shape of the femoral stem. Predicting the prepared shape of the medullary cavity, the position of the femoral stem could be determined properly. Thus, our goal is simulating behavior of the surgical tools as a preoperative planning system of THA to obtain the appropriate position of the stem.

The expert surgeon pays attention not to hurt cortical bone with the tools while expanding the medullary cavity [4]. This prepared medullary cavity could be regarded as a minimally invasive shape for inserting the femoral stem. In order to simulate this surgical technique, we introduce an evaluation function, Invasive Parameter (IP) to estimate damage to femur. IP has been given as a sum of squares of the apparent densities of bone's elements which will be destroyed by the surgical tools. By minimizing IP, we could get an optimal position of the femoral stem. This proposed algorithm needs no segmentation process. Efficiency of proposed method is evaluated on the real CT images.

2 PREPARING THE MEDULLARY CAVITY OF FEMUR

In a surgical operation, the medullary cavity of the femur is expanded with reamers and rasps as shown in Figure 1(a). The reamers dig in the longitudinal direction of femur to form the axis of the femur. Then, the rasps shave the medullary cavity along with the femoral axis for preparing the stem insertion. Because the shapes of the rasps are rather smaller ones of the femoral stem, the shape of prepared medullary cavity would be ready for insertion of the femoral stem.

In the proposed preoperative planning system, the surgical operation which shapes the medullary cavity of the femur is simulated on real CT images as shown in Figure 1(b). In the simulator, CT images are used as are, and would not be segmented. Pixels would be removed by segmentation are used for calculating IP. A procedure of the simulator is following. (1) The femoral axis is determined by simulating behavior of the reamers minimizing IP. (2) Inserting the rasp along with the femoral axis rotating around the axis to find the position with minimal IP. Finally, the position of the optimal position of the femoral stem is given as one of the final rasp.

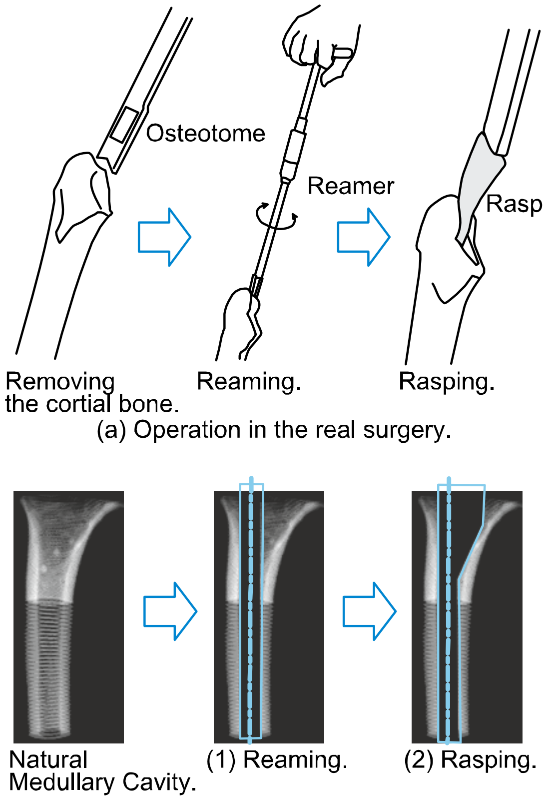


Figure 1: Preparing a shape of a medullary cavity of femur.

3 ESTIMATING INVASIVENESS

3.1 A Model of the femur

Before explaining about IP, the model of femur for the proposed simulation is to be described. The femur is modelled as several slices of CT images as shown in Figure 2. Each CT slice has zero thick and placed referring to its DICOM header. Ordinarily, voxels are used in the field of visualization and analysis of medical data, however handling CT images as voxels increases calculating time for collision detection and errors cased by similar reason to the partial volume effect. A collision between the bone and the tool is detected by simplly finding the pixels inside of the models of the surgical tools.

3.2 Invasive Parameter (IP)

The reamers and rasps are operated according to not only the preoperative plan, but also the reaction force through the tools in the operation. The tool working as a probe allows the expert surgeon to decrease invasiveness while expanding the medullary cavity. Invasive Parameter: IP is introduced to estimate this reaction force through the tools. IP has been given as following,

$$\mathrm{IP} = k \sum_{i=1}^{n_{CT}} \sum_{j=1}^{n_{pixel}} (HUij)^2, \tag{1}$$

where,

k is constant can be 1 here,

n_{CT} is number of CT images,

n_{pixel} is number of the pixels interfering the tool

HU_{ij} is Hounsfield Units (HU) of the pixels

($HU_{ij} = 0$ if $HU_{ij} < 0$ here)

The reaction force is regarded as being proportionate to a sum of compressive strength of each element of the bone is to be removed by the tool. The compressive strength can be assumed to be proportionate to the square of the bone density [6]. Thus, IP is proportionate to a sum of the squared HU of the each pixel in the CT images. A minimally invasive position of the tool is estimated by minimizing IP.

4 REAMING SIMULATION

4.1 A model of the reamer

In THA, tapered reamers are used to expand the medullary cavity of femur. In the reaming simulator, the reamer is modelled as a cylinder to ignore the depths of insertion of the reamer to simplify the model. A schematic diagram of the reaming simulation is shown in Figure 3. A diameter of the reamer starts from 1mm and increased by 1mm to 7mm in this evaluation.

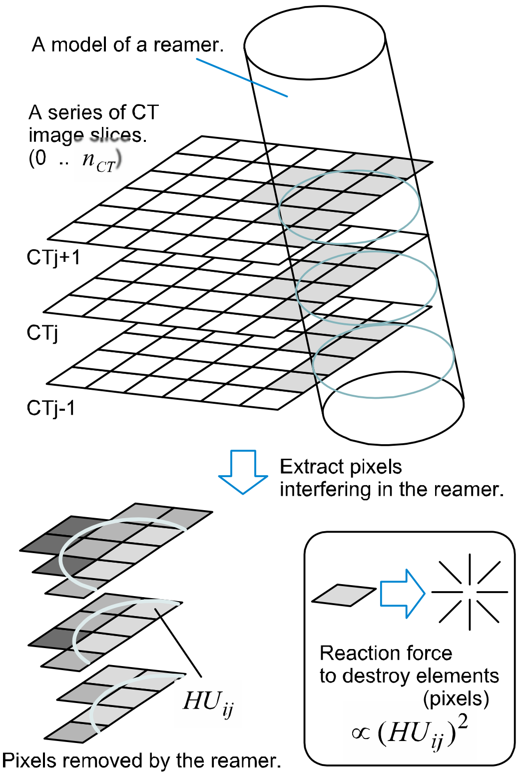


Figure 2: A model for calculating invasiveness.

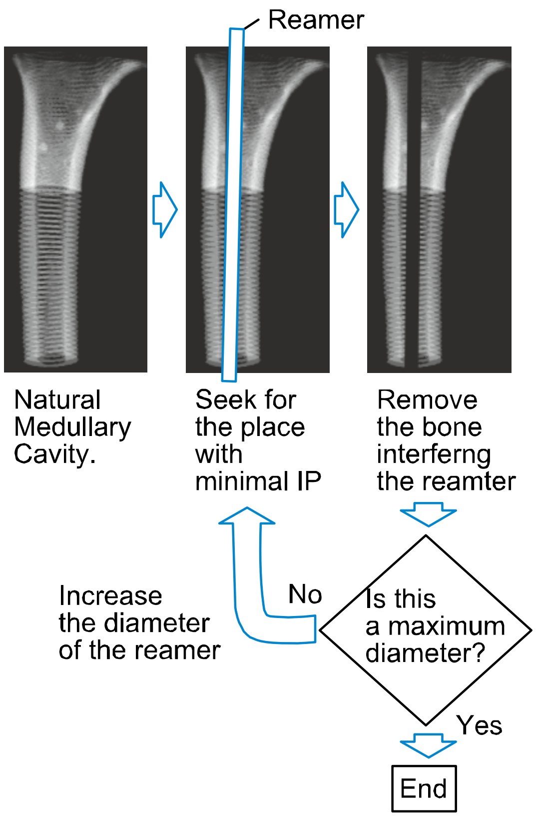


Figure 3: A schematic diagram of the reaming simulation.

IP is calculated and compared with ones of the neighbours to find the position with minimal IP. When the reamer reaches the position with minimal IP, elements of femur interfering in the reamer is removed, in other words, all HU of the pixels being interfered with the reamer are set to zero. Then, the diameter of the reamer is increased and the reamer finds the minimal IP position to the maximum size of the reamer. We have used real CT images provided by VAKHUM project [7] for the evaluation. Table 1 shows the conditions of an evaluation of the simulation.

4.2 Result of the Reaming Simulation

The reamer was initially placed inside of the medullary cavity arbitrary. In all cases, the reamer had been converged on an unique position finally.

Figure 4 shows the result of the reaming simulation. The validity of proposed method has been confirmed visually. The result suggests the possibility of automated determination of the femoral axis. A calculation time for a convergence was within 12 seconds with Intel Core™ 2 Duo E6600 (2.4GHz, with one thread).

In order to confirm the convergence of IP with the reaming simulation, IP around the minimally invasive position was calculated as shown in Figure 5. The reamer translated in the radial direction to the femoral axis, and IP was obtained at each position of the reamer. Figure 5 shows that IP changes rapidly according to the translation of the reamer, which indicates strong convergence of IP, and suggests the robustness of the proposed algorithm.

Table 1: Conditions of the real CT images.

Number of CT images	108
Pixel spacing of CT image [mm]	0.84 x 0.84
Space between each slices [mm]	1 (proximal part) 3 (distal part)
Size of CT image [in pixels]	512 x 512

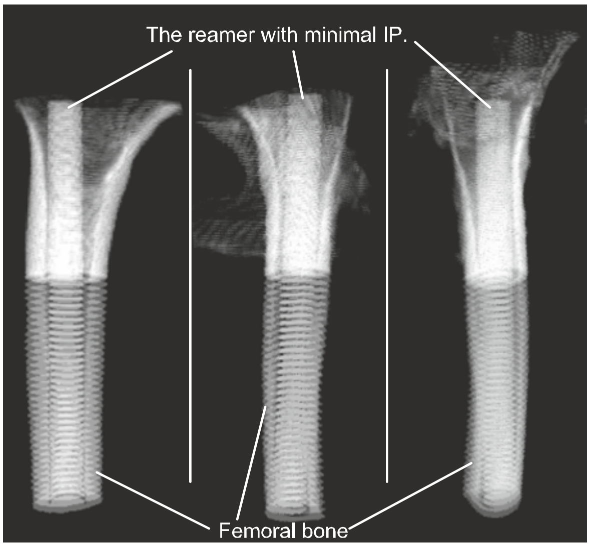


Figure 4: Result of the reaming simulation seen from several directions.

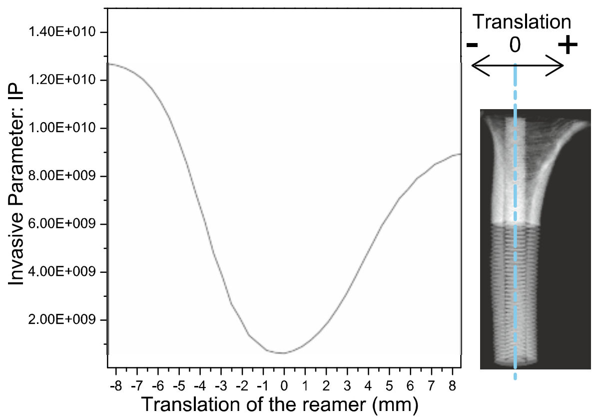


Figure 5: IP's change according to the position of the reamer.

5 RASPING SIMULATION

5.1 A model of the rasp

The shape of the rasp for THA has similar shape of the femoral stem with rather small size. The rasps are designed for each femoral stem to prepare the medullary cavity of femur for inserting the stem. The rasp expands the reamed medullary cavity of femur to fit the shape of the femoral stem. In the rasping simulation, the shape of the raps is modelled as a parametric shape as shown in Figure 6. Using parametric shape make it possible to reduce calculation time for detecting collisions. The rasp model moves in two degrees of freedom. One is translation along with the femoral axis determined by the rasping simulation, and the other is rotation around the femoral axis. These two movements simulate the insertion of the rasp and the behavior to seek an appropriate angle of the rasp with feeling the reaction force through it in the real surgical operation.

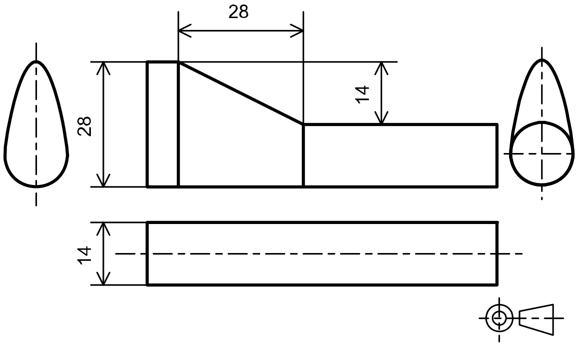


Figure 6: A model of the rasp.

5.2 Result of the rasping simulation

A rasping simulation was performed after the reaming simulation. Appropriate rotation angle of the rasp with minimal IP was determined by setting the depth of insertion to 55mm. Here, this depth is determined by an operator subjectively referring CT image and the shape of the rasp. As mentioned later in discussion; section 6.3, proper depth of insertion could not be determined by just comparing IPs. Then, the rasp rotated from 0 degree and increased by 1 degree to 360 degree. The result of the rasping simulation is shown in Figure 7. The final position of the rasp seems to fit the medullary cavity well to the eye. A calculation time for a convergence was within 70 seconds with Intel Core™ 2 Duo E6600 (2.4GHz, with one thread).

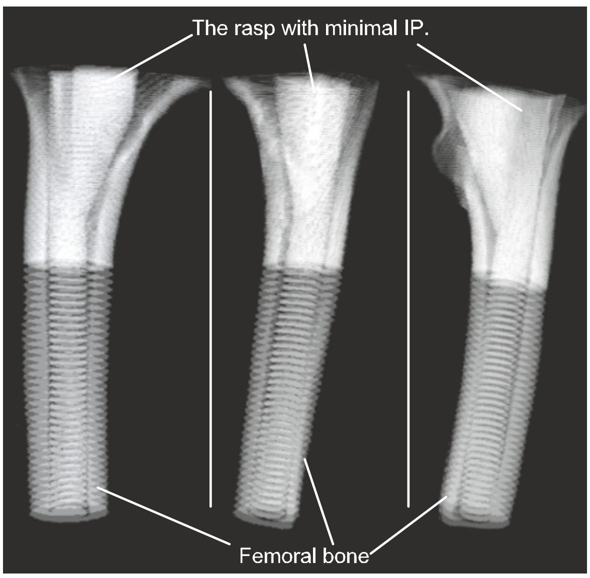


Figure 7: Result of the rasping simulation seen from several directions.

In order to confirm the result objectively, the convergence of IP has been considered numerically. IP was calculated according to the changes of the depth of insertion and the rotation around the femoral axis as shown in Figure 8. IP increases according to the depth of insertion. There are a few local minimums of IP observed when the rasp was not inserted deeply. However by inserting the rasp deeply, an unique minimal IP indicated by an arrow sign in the Figure 8 can be obtained. The position of the rasp on Figure 7 is the position with this minimal IP. Because the shapes of the medullary cavity and the rasp have the same morphological characteristics such as swelling in one side, it's natural that there is a point with such minimal IP. The result suggests the possibility of automated placement of the rasp, and the femoral stem.

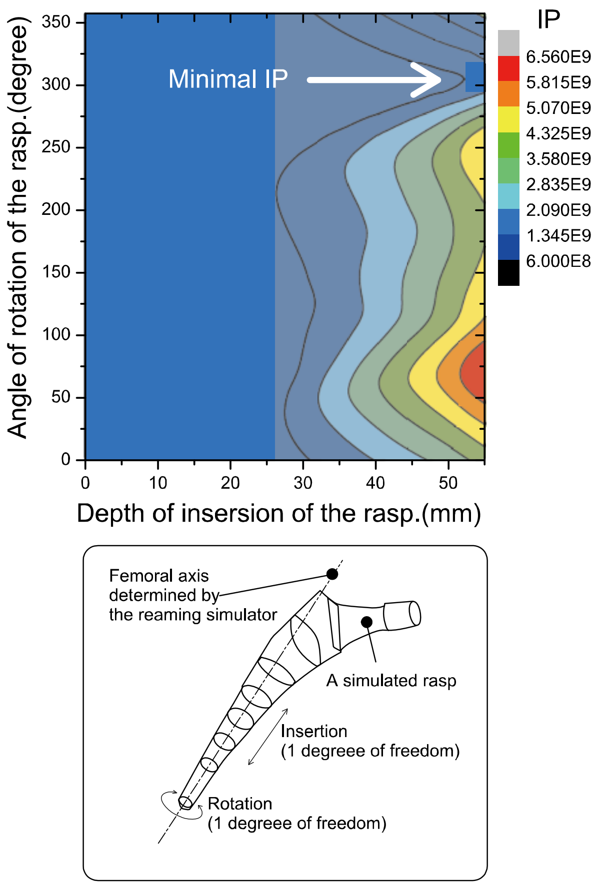


Figure 8: IP's change according to the insertion and the rotation of the rasp.

6 DISCUSSION

6.1 Segmentation or surgical tool based shaping

Conventional preoperative planning systems for THA need segmentation to operate a medullary cavity of femur and a femoral stem. These systems handle a task to place the femoral stem as one of contact problem. To solve a contact problem, two or more different surfaces, their physical properties are needed. In the case of THA, two surfaces are the shapes of a medullary cavity of femur and a femoral stem. The shape of the femoral stem is known, but the shape of the medullary cavity is ambiguous. Here is example as shown Figure 9, the shape of the medullary cavity changes depending on a threshold value of HU. There are three cases of thresholds and contours created by converting pixels of CT images into binary value at each threshold. Contours on Figure 9 explain difficulty of making shape of the medullary cavity of femur properly. So, conventional approaches are not free from current gold standard of determining threshold, which is made by the expert surgeon. There are yet another complicated problems with physical properties, we never mention about them here.

Our proposed algorithm does not need segmentation, so that this approach is free from a contact problem. This algorithm handles the medullary cavity and the surgical tools as solid rigid bodies. Only relative strength of each element of the bone is needed for simulating behavior of surgical tools. Searching a position with minimal IP leads the surgical tool model to the optimal position. Here, 'optimal' means having least invasive to the bone.

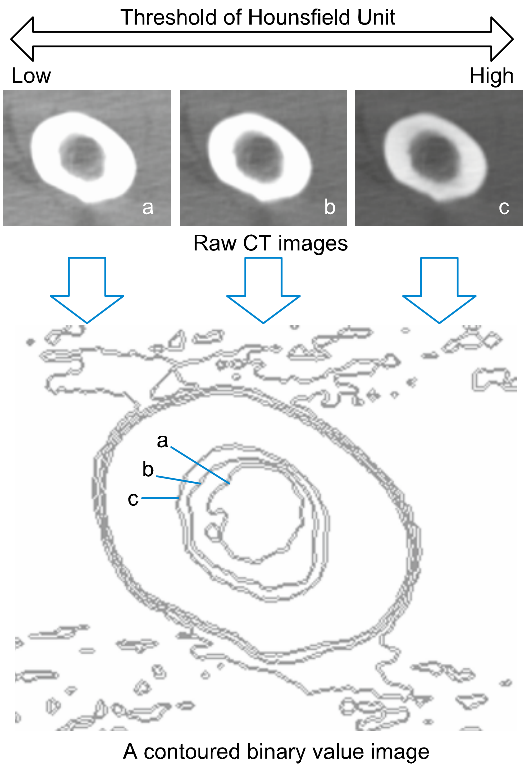


Figure 9: Contoured shape changes depending on the threshold of Hounsfield Unit.

6.2 Calculating time with a proposed algorithm

This proposed algorithm brings rapid results. Medical data tend to be handled as voxels or meshes for the reason of visualization or the restriction of the application software. Detecting collision between voxels or meshes requires long time to calculate. In the proposed algorithm, CT image is handled as groups of pixels, and the surgical tool is described in a parametric manner to reduce calculating time. Of course, the surgical tools can be described in polygons such as a STL format, which would not increase calculation time too much.

6.3 Using this algorithm in preoperative planning

Predicted prepared shape of the medullary cavity of femur is almost the same shape of the inserted femoral stem, so that the proposed algorithm could be used for preoperative planning of THA. To use this algorithm clinically, the limit of IP should be determined to stop the insertion of the rasp. The rotation of the rasp can be determined minimizing IP as shown in Figure 8, while there is no mean to determine the proper depth of the insertion of the rasp. The limit of insertion could not be determined by merely comparing IPs. The limit of the insertion would be determined comparing the results of the simulations with clinical results.

7 CONCLUSIONS

Conclusions are summarized by the followings,

1) A novel algorithm for preoperative planning for THA has been proposed, which could place the femoral stem in the medullary cavity of femur at an unique position automatically.
2) A prepared shape of the medullary cavity shaped by surgical tools has been predicted uniquely with strong convergence of IP, which suggests the robustness of the proposed algorithm.
3) A prepared shape of the medullary cavity composed of 108 of CT images formed by the reaming and rasping simulator within 80 seconds with the proposed algorithm with Intel Core™ 2 Duo E6600 (2.4GHz, with one thread).

8 ACKNOWLEDGMENTS

The authors thank Prof. Hirotaka Iguchi (Nagoya City University, Japan) for his advices. Part of this study was supported by Okuma Corporation.

9 REFERENCES

[1] Sakai, T., Nakamoto, M., Kagiyama, Y., Sato, Y., Sugano, N., Nishii, T., Akazawa, K., Tada, Y., Tamura, S., Yoshikawa, H., 2003, Automated Preoperative Planning of Femoral Stem Placement for Total Hip Replacement Based on Femoral Canal Shape (in Japanese), Japanese Society for Medical and Biological Engineering, 41/4:237-244.

[2] Adachi, T., Tomita, Y., Kuunimoto, H., Tsubota, K., 2002, Shape Design Simulation of Artificial Hip Joint Stem Based on Stress Uniformity at Bone-Stem Interface (in Japanese), Japanese Society for Clinical Biomechanics, 23:405-410.

[3] Yokota, F., Shimada, R., Okada, T., Kagiyama, Y., Nakamoto, M., Takao, M., Sugano, N., Sato, Y., Yoshikawa, H., Tada, Y., 2007, Semi-automated Segmentation of the Femur from 3D CT Images Using Multi-Level Statistical Shape Model (in Japanese with English abstract), The Institute of Electronics, Information and Communication Engineers(IEICE) technical report, 107/58:47-52.

[4] Hirosaki, K., Shintani, K., Kaneuji, A., 2008, Development of Custom-made Design Method of Artificial Hip Joint Stem (in Japanese), Journal of the Japan Society for Precision Engineering, 74/3:313-318.

[5] Kubo, T., 2005, Orthopaedic Surgery Knack and Pitfalls essential and blind points for hip surgery (in Japanese), bunkodo.,:253

[6] Carter, D.R., Hayes, W.C, 1977, The Compressive Behavior of Bone as a Two-Phase Porous Structure, The Journal of Bone and Joint Surgery Am., 59/7:954-962.

[7] Virtual Animation of the Kinematics of the Human for Industrial, Educational and Research Purposes (online), available from <http://www.ulb.ac.be/project /vakhum/>, (accessed 2007-02-19).

A Trial of Numerical Calculation of Wear for Artificial Hip Joints

Hirohisa Narita [1], Kakeru Harusawa[1], Makoto Yagihashi [1], Hideo Fujimoto [1]
[1] Nagoya Institute of Technology, Aichi, Japan

Abstract
A numerical analysis method for the wear of the artificial hip joints due to the walking is proposed and a wear prediction system for them is developed in this research. In order to develop the system, EHL, "elasto-hydrodynamic lubrication theory" and a model for asperity load sharing are applied. The feasibility of the developed wear prediction system for artificial hip joints is shown and the relationship between the wear and various parameters of the artificial hip joint is described in this paper.

Keywords:
Artificial hip joint ; Elasto-hydrodynamic lubrication ; Wear ; Lubricating film thickness

1 INTRODUCTION

Total hip replacement is often used to treat joint failure caused by osteoarthritis. The life span of artificial hip joints is about 20 years; hence some attempts related to stress analyses, material developments, manufacturing methods, surface treatment methods and wear analyses have been conducted in order to extend the their life time. Especially, the wear deteriorates functions of the hip joint, and the wear-particle induces osteolysis mediated aseptic loosing [1].

Regarding the wear and the lubricant analysis models, some researches has been carried out for artificial joints. Wearindex is applied to predict the wear for the artificial knee joint [2]. This method is very simple to use, but this doesn't include all parameters of the artificial joint. Hence, the analyzed results can't be used for some artificial joints design. The lubricant model for the artificial hip joints has been proposed [3, 4] based on Elasto-hydrodynamic lubrication (hereinafter EHL). This method is suitable for the analysis for the distribution of the pressure and the lubricant status with considering the parameters related to the artificial hip joint. This isn't, however, suitable for the wear prediction, because the asperity contact can't be considered.

In this research, a numerical analysis method for the wear of the artificial hip joints due to the walking is proposed and a wear prediction system for them is developed based on EHL theory. The novel aspect of this research is to realize the wear calculation with consideration of the detailed lubrication status of the artificial hip joint due to the dynamic change during one walking cycle. By using the wear prediction system, the artificial hip joint of metal-on-metal type is analyzed in this paper.

2 2 WEAR PREDICTION SYSTEM FOR ARTIFICIAL HIP JOINTS

2.1 System Overview

There are three lubrication status : hydrodynamic, boundary and mixed lubrication [5] as shown in Figure 1. The lubricating film supports the load in the hydrodynamic lubrication, hence the wear hardly occurs. The bodies come into closer contact at their asperities in the boundary lubrication, hence The plenty of wear occurs. The mixed lubrication means the intermediate condition between the boundary and the hydrodynamic lubrications

In order to estimate the wear, the lubrication status should be distinguished propery. Hence, the we construct a prediction system with considering the aforementioned problem.

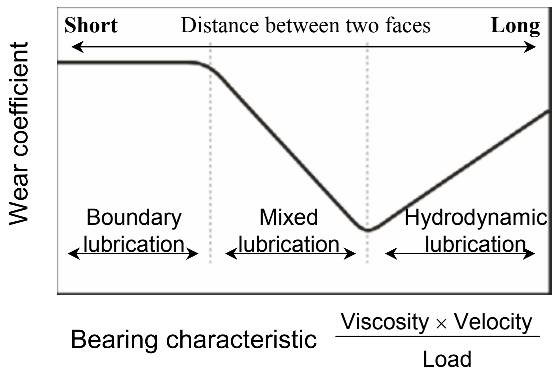


Figure 1: Stribeck curve and lubricantion status .

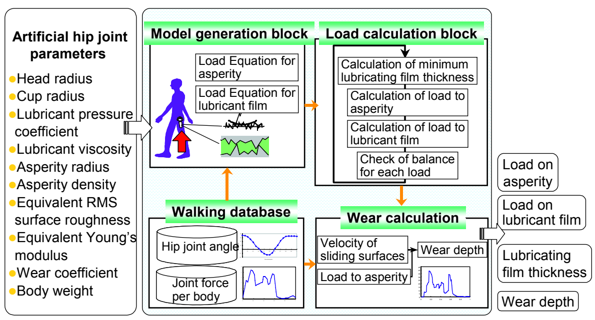


Figure 2: System overview of wear prediciton system.

The wear prediction system develped in this research consists of four blocks: a model generation, a load calculation, a wear calculation and a walking database blocks as shown in Figure 2.

The model generation block generate an analysis model for the micro-geometry of the contact, when a cup and a head radii of a femoral component, an equivalent RMS (Root Mean Square) surface roughness, an asperity density, an equivalent Young's modulus, an asperity radius, a lubricant pressure coefficient, a lubricant viscosity, a wear coefficient and a body weight are input to the system.

The load calculation block calculates each load on asperities and the lubricating film. In order to obtain the each load, a minimum lubricating film thickness is estimated based on "elasto-hydrodynamic lubrication theory", which considers the elastic deformation of the lubricating surface due to the pressure. The each load is determined from the obtained minimum lubricating film thickness.

The wear calculation block calculates the wear of the artificial hip joint due to the walking with consideration of the load on each asperity and the velocity of the sliding surfaces.

The walking database block stores a change of the hip joint angle during one walking cycle and a joint force per body. The change of the hip joint angle during one walking cycle is used to calculate the velocity of the sliding surfaces between the cup and the head of the femoral component, and the joint force per body weight is used to calculate the two kind of loads on each asperity and the lubricating film.

2.2 Model generation block

It is assumed that a mixed lubrication status is occurred between a cup and a femoral head shown in Figure 3. This means that two surfaces are separated by a gap distance. Then, there is a lubricant film and are some asperities to share the total load.

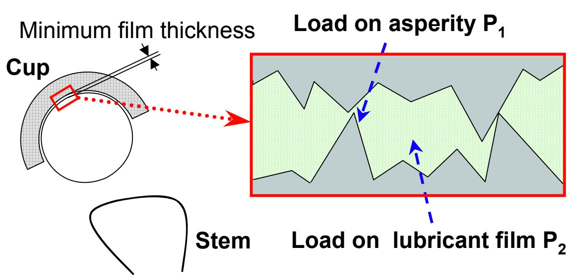


Figure 3: Load in boundary lubrication.

In order to calculate two kinds of load on each asperity and the lubricant film theoretically, a contact between two surfaces are regarded as the one of a cylinder and a flat surfaces shown in Figure 4. Then, an equivalent radius, B, and an equivalent RMS surface roughness [6] of the flat surface are obtained by the following equations.

$$\frac{1}{B} = \frac{1}{R_1} \pm \frac{1}{R_2} \tag{1}$$

$$\sigma_1^2 + \sigma_2^2 = \sigma^2 \tag{2}$$

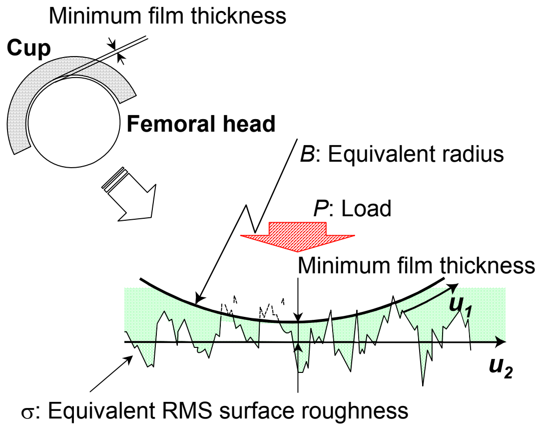


Figure 4: Model transformation for two surfaces contact.

In equation (1), R_1 and R_2 mean the radii of a cup and a femoral head, respectively. Plus and Minus signs also correspond to Figure 5.

In equation (2), σ_1 and σ_2 mean the RMS surface roughness of each surface, respectively. In the Figure 4, u_1 and u_2 also mean the sliding velocities of each surface, respectively. Here, the asperity height distribution is regarded as a normal distribution, in which a standard deviation is σ.

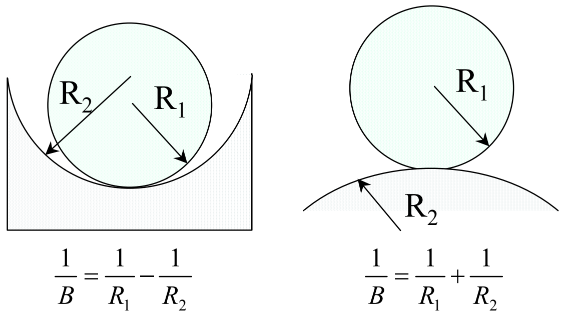


Figure 5: Definision of eqivalent radius.

2.3 Load calculation block

Total load, P, is obtained by summing the load on the asperity, P_1, and the load on the lubricant film, P_2, as follows.

$$P_1 + P_2 = P \tag{3}$$

In order to calculate the load, we must obtain the minimum lubricating film thickness, which influence each load (P_1 and P_2).

The minimum lubricating film thickness, d_0, is calculated with consideration of EHL as follows [7].

$$d_0 = \frac{2.65\alpha^{0.54}(\mu_0 u)^{0.7}(E')^{-0.03}B^{0.43}}{P_2^{\,0.13}} \tag{4}$$

In equation (4), α, μ_0, u and E' mean a pressure-viscosity coefficient, a lubricant fluid viscosity at conditions of entry to contact, an average velocity between two surfaces and an equivalent Young's modulus, respectively. This equation is often used for analyses of petroleum-derived oil, but this equation is applied for genral purpose. In equation (4), Equivalent Young's modulus E' is calculated as follows.

$$\frac{1}{E'} = \frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2} \tag{5}$$

In equation (5), ν is Poisson's ratio. If the materials are the same, E' is half this. Then, d_0 is replaced by a nondimentional separation, h_0, which is defined as follows.

$$d_0 = \sigma\, h_0 \tag{6}$$

Based on equations (4) and (5), P_2 is calculated as follows.

$$P_2 = \left[2.65\alpha^{0.54}(\mu_0 u)^{0.7}(E')^{-0.03}B^{0.43}(\sigma h_0)^{-1}\right]^{7.7} \tag{7}$$

On the other hand, P_1 is obtained by the surface contact theory [8, 9]. That is to say, P_1 is obtained by assuming that a reactive force is proportional to an elastic deformation of the asperity as follows.

$$P_1 = 0.8\eta L E' \beta^{1/2}\sigma^{3/2}\left[\frac{1}{\sqrt{2\pi}}e^{-h_0^2/2} - \frac{h_0}{2}\, erfc(h_0/\sqrt{2})\right] \tag{8}$$

In equation (8), *erfc* means complementary error function defined as follows.

$$erfc(x) = \frac{2}{\sqrt{\pi}}\int_x^{\infty}\exp(-t^2)\, dt \tag{9}$$

h_0 is obtained by solving equations (3), (7) and (8) with iterative calculation based on Secant method [10] in this research. The secant method is identical to the univariate Newton method, except that it replaces the derivative of a function with an approximation constructed from the function values at the two previous iterates.

2.4 Wear calculation block

Generally, there are four wear types: adhesive, abrasive, fatigue and corrosive wears. The adhesive wear is expected to be a dominant factor due to the walking. The adhesive wear proportional to load on the contact surface and the velocity of the sliding surface. Hence, the wear depth, W, is obtained by the following equation [5]

$$W = K\int P_1 V\, dt \tag{10}$$

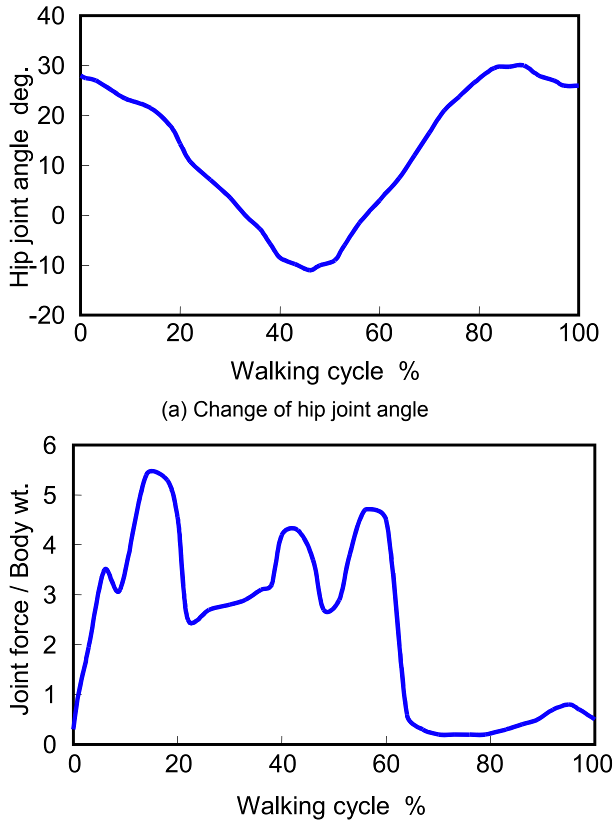
In equation (10), V, t and K mean the velocity of the sliding surface, time and the wear coefficient.

In one walking cycle, the velocity of the sliding surface and the load to the hip joint change drastically. Generally, conventional EHL theory is suitable for constant velocity and the load and isn't suitable drastic change of the velocity and the load. Hence, the walking cycle is segmentalized in time domain and EHL theory is applied by treating the wear of the artificial hip joint as microscopical phenomena. In other words, an instantaneous velocity and an instantaneous load are applied to calculate an instantaneous wear and the total wear is calculated by summing all instantaneous wear in one walking cycle.

2.5 Walking database block

The walking database block stores a change of the hip joint angle during one walking cycle and a joint force per body as shown in Figure 6 [11, 12].

The walking cycle consists of the stance phase and the swing phase. The horizontal line in the figures, the stance phase corresponds to about 60% in a first part and the swing phase corresponds to about 40% in a second part. These changes are considered to calculate the wear in one walking cycle.



(a) Change of hip joint angle

(b) Change of joint force per body weight

Figure 6: Change of hip joint angle and load in walking cycle

3 CASE STUDY

3.1 Parameter setting

The artificial hip joint of Metal-on Metal type is analyzed. Material is Co-Cr alloy for medical purpose. The lubricant fluid for the artificial hip joint is generally the body's natural synovial fluid.

The lubrication is occurred by the body's natural synovial fluid, but the lubricant viscosity and the lubricant pressure coefficient of it is unknown. Hence, we use water data [13], because water is main factor. The lubricant pressure coefficient is generally calculated as follows [14].

$$\eta_p = \eta_0 \exp(\alpha p) \tag{11}$$

η_p and η_0 are viscosity at pressure, p, and one at atmospheric pressure. Parameters for artificial hip joint analysis are summarized in Table 1. The walk count in one day is also regarded as 7378 [15] from the average walk count of Japanese.

Table 1: Parameters of an artificial hip joint and a lubricant fluid

Radius of artificial femoral head R_1 [m]	2.0×10^{-2}
Radius of cup R_2 [m]	2.015×10^{-2}
Lubricant pressure coefficient α [Pa^{-1}]	2.46×10^{-10}
Lubricant viscosity μ_0 [Pa·s]	8.0×10^{-3}
Average asperity of radius β [m]	5.4×10^{-5}
Asperity density η [m^{-2}]	6.0×10^{9}
Equivalent surface roughness σ [m]	3.4×10^{-8}
Equivalent modulus of elasticity E' [Pa]	2.53×10^{11}
Coefficient of wear K [m/N]	1.6×10^{-15}
Body weight bw [kg]	60.0

3.2 Lubrication and wear analysis

Figure 7 shows the change of the minimum film thickness due to the joint force in one walking cycle. The minimum film thickness is from 3.0×10^{-8} to 10.0×10^{-8} m and is changed according to the variation of the joint force. Generally, the film thickness doesn't depend on the load in the hydrodynamic lubrication. Hence, the mixed lubrication or the boundary lubrication is occurred.

Figure 8 shows the ratio of asperity load and lubricant film load in one walking cycle. Obviously, asperity load is dominant, hence the boundary lubrication is obviously occured for the artifical hip joint.

Figure 9 shows the wear change in one walking cycle. Almost wear is occurred in stance phase and little wear is occurred in swing phase. Using this data, the wear depth per year is calculated. It becomes 5μm and coincides with the real data [16]. A feasibility of the developed wear prediction system is shown through this case study.

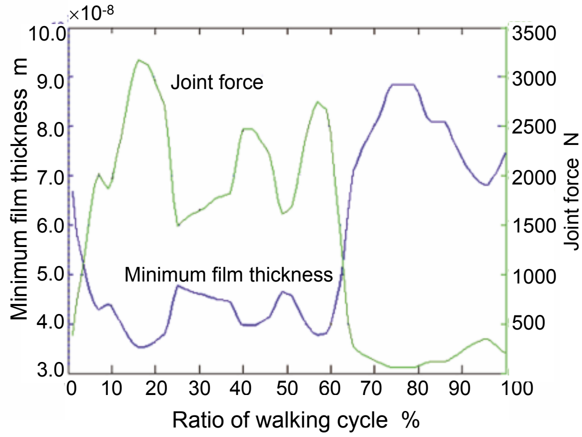


Figure 7: Change of Minimum film thickness due to the joint force in one walking cycle

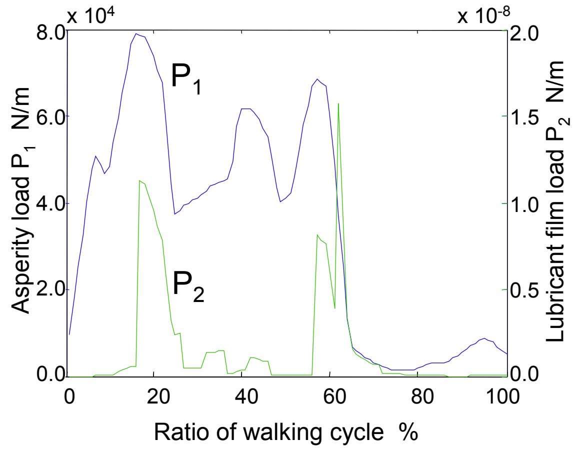


Figure 8: Ratio of asperity load and lubricant film load in one walking cycle

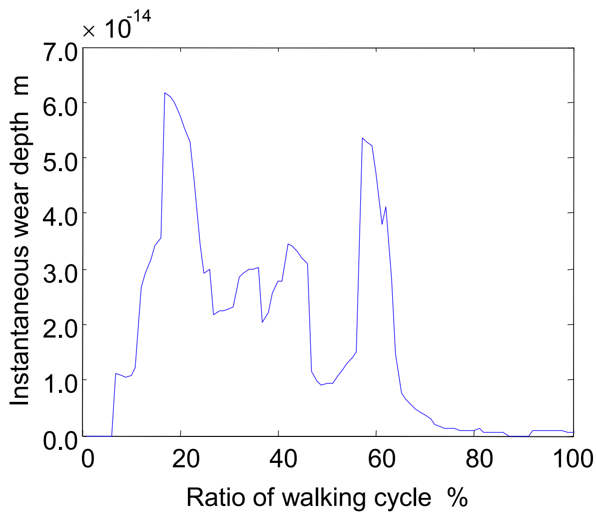


Figure 9 : Wear change during one walking cycle

Next, some parameters of an artificial hip joint is discussed for realizing the hydrodynamic lubrication. There are some parameters such as the radii of femoral head and the cup, the surface roughness, the asperity density, the Young's modulus, the asperity radius, the lubricant pressure coefficient, a lubricant viscosity for the design conditions. The radius of the femoral head in this paper, because average pressure on the surface decrease and the lubricant film might be generated easily.

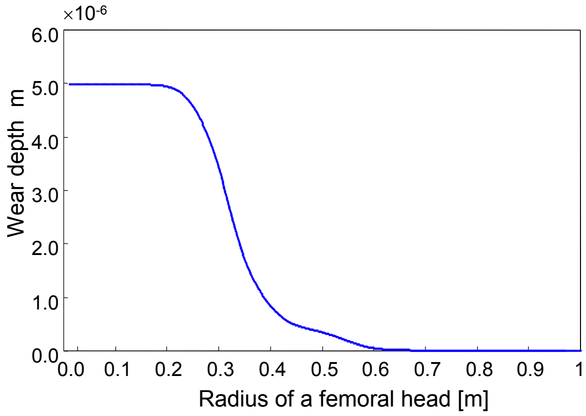


Figure 10 : Relationship between total wear and various radii of a femoral head

Figure 10 shows the calcurated results between total wear and various radii of the femoral head. As shown in Figure 10, the wear reduction strats about 180mm and the hydrodynamic lubrication starts about 620 mm. It means that the boundary lubrication is dominant under about 180mm and the mixed lubrication is occured from about 180mm to about 620mm. These kinds of large femoral head can't be insert in a human body, hence the hydrodynamic lubrication very difficult for the artificial hip joints.

Figure 11 shows the calcurated results between total wear and equivalent Young Modulus. The wear is decreased from one point near equivalent Young Modulus of lead (Pb). However, there are no implant materials which have biocompatibility and low Young modulus simultaneously, hence the hydrodynamic lubrication very difficult currently.

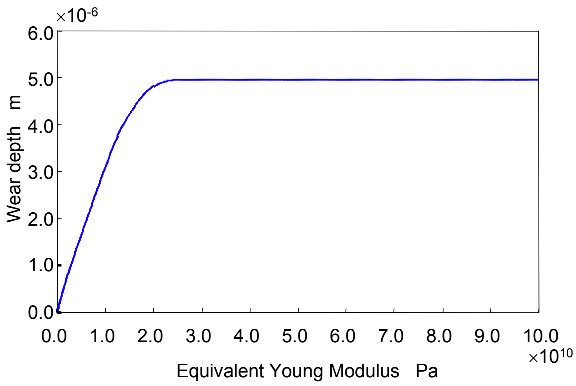


Figure 11 : Relationship between total wear and equivalent Young Modulus

Figure 12 shows the calcurated results between total wear and equivalent surface roughness. The wear is decreased from one point. The atomic diameter is about 1.0×10^{-9} and then high accuracy is required very much to manufacture the femoral head.

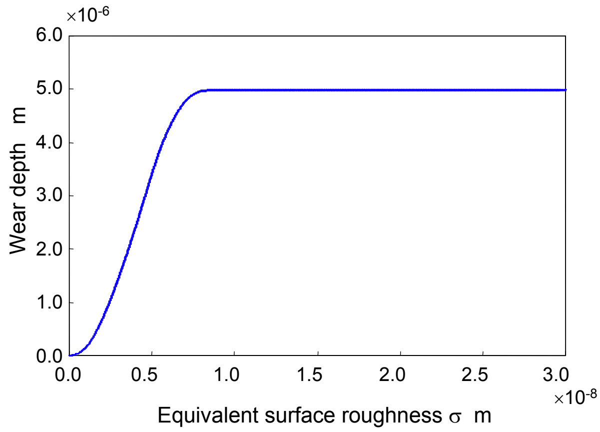


Figure 12 : Relationship between total wear and equivalent surface roughness

4 CONCLUSIONS

Conclusions are summarized by the followings

1) A wear prediction model for artificial hip joints has been proposed based on elasto-hydrodynamic lubrication theory and surface contact theory.
2) A predicted wear coincides with a real one, hence a feasibility of the developed system has been demonstrated.
3) It has been found that the boundary lubrication is dominant and the hydrodynamic lubrication is very difficult for the artificial hip joints by the analyzed result.
4) A relationship between the wear and some parameters of an artificial hip joint is described

Future work is that the lubrication model is modified for biological phenomena and design formula to analyze the wear is constructed by using the developed system.

5 REFERENCES

[1] Eyerer, P., Kurth, M., Mckellup, HA, and Mittlemeier, T. 1987, Characterization of UHMWPE Hip Cups Run on Joint Simulators, J. Biomed. Master. Res., 21: S.275-291.

[2] Yasutake, S., Higaki, H., Shimoto, T., Ogata, T., Shiraishi, Y., Yoshimura, D., Nakanishi, Y. , 2005, Evaluation of wear prediction of UHMWPE using Wearindex parameters, Proceedings of the 16th JSME Conference on Frontiers in Bioengineering : 93-94. (in Japanese)

[3] Sasada, T., Tsukamoto, Y. and Mabuchi, K., 2000, Biotribology –Wear and Lubrication in Joints-, Sangyo Tosyo: 88-93. (in Japanese)

[4] Wang, F.C., Jin, Z.M., 2005, Elasto-hydrodynamic Lubrication Modeling of Artificial Hip Joints Under Steady-State Conditions, Trans. of ASME, J. of Trib., 127:729-739.

[5] Muraki. M., 2007, Grahpic Explanation, Tribology, Science-of-friction and lubrication technique, THE NIKKAN KOGYO SHIMBUN, Ltd. (in Japanese)

[6] Tallian, T. E., Chiu, Y. P., Huttenlocher, D. F., Kamenshine, J. A., Sibley, L. B., Sindlinger N. E., 1964, Lubricant Films in Rolling Contact of Rough Surface, ASLE Transactions, 7, 2, 109-126.

[7] Dowson, D. and Higginson, G. R., 1966, Elasto-Hydrodynamic Lubrication, Pergamon Press.

[8] Greenwood, J.A., Williamson, J.B.P, 1966, Contact of Nominally Flat Surfaces, Proc. of Royal Society, A.295 : 300-319.

[9] Thompson, R. A. and Bocchi, W., 1972, A Model for Asperity Load Sharing in Lubricated Contacts, ASLE Transactions, 15, 1, 67-79.

[10] Suzuki, S., Iida, Y. And Ishizuka, Y., Numerical Computing with C Language, 2000. (in Japanese)

[11] Patmurray, M., Drought, A. B. and. Kory, R. C, 1964, Walking Patterns of Normal Men, JBJS, 46-A :335-360.

[12] Seireg, A. and Arvikar, R.J., 1975, The Prediction of Muscular Load Sharing and Joint Forces in The Lower Extremities During Walking, J. of Biomechanics, 8: 89-102.

[13] Bett, K. E. and Cappi, J. B., 1965, Nature, 207: 620-621.

[14] Barus C., 1893, Isotherms, Isopiestics, and Isometrics Relative to Viscosity, The American Journal of Science, 45:87-96.

[15] Ministry of Welfare, 1995, National Nutrition Survey report. (in Japanese)

[16] Kubo, T., 2005, orthop edic surgery Knack & Pitfalls essential and blind points for hip surgery, bunkodo. (in Japanese)

A Non-invasive Method to Measure Joint Range of Motion for Hip Joints

Noriko Masuta[1], Makoto Yagihashi[1], Hirohisa Narita[1] and Hideo Fujimoto[1]
[1] Nagoya Institute of Technology

Abstract
In order to reduce the risk of dislocation after total hip arthroplasty, a hip joint range of motion should be evaluated properly. The non-invasive measuring method which obtains the joint range of motion without using computed tomography or X-ray is proposed in this research. A hip joint center inside a body is estimated by two positional sensors which are attached on a body surface. The joint range of motion is calculated by the estimated hip joint center and positions and orientations of the pelvis and the femur. The effectiveness of the proposed method is discussed.

Keywords:
Joint range of motion; Hip joint center; Total hip arthroplasty; Non-invasive measurement

1 INTRODUCTION

The purpose of this paper is to measure the hip joint range of motion in order to reduce the risk of dislocation after total hip arthroplasty (THA). This paper proposes a non-invasive measuring method which does not use computed tomography (CT) or X-ray.

The hip pain is relieved by THA, but hip joints after THA can be dislocated [1, 2]. In order to evaluate safe range of motion which realize no dislocation, a system of four-dimensional motion analysis after THA has been developed [3, 4]. The dislocation after THA is caused by excessive motion of the joint of a hip prosthesis. Thus, the joint range of motion needs to be measured properly.

Not only measuring motions of the femur and the pelvis but also the hip joint center (HJC) should be detected to measure the joint range of motion. The HJC is measured by using CT or X-ray conventionally because the HJC is inside a body.

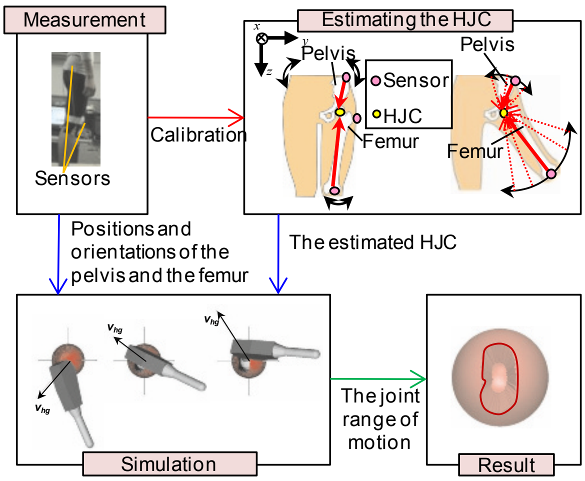


Figure 1: The overview of the proposed measuring method of the joint range of motion.

As patients cannot move actively, the HJC of healthy bodies need to be measured to obtain the joint range of motion in daily activities. However, the method using CT or X-ray is limited to patients only because this method exposes healthy bodies to radiation. Davis has proposed an estimation method of the HJC based on X-ray images of frontal plane and sagittal plane for 25th hip joint [5]. Vaughan has proposed a method to estimate the HJC by using X-ray and referring anthropometric dimension measured by Chandler [6]. But in these methods, only a static HJC could be measured.

In this paper, a non-invasive method to measure the HJC is proposed. This method is an estimation method of the HJC with two points measured by two three-dimensional magnetic positional sensors (POLHEMUS). Generally, the hip joint which consists of the pelvis and the femur is regarded as a ball joint between two rigid bodies [7]. The HJC is estimated by the measured positions and orientations of the pelvis and the femur.

2 MEASURING METHOD

2.1 The overview of measuring joint range of motion

Figure 1 shows the overview of the proposed measuring method of the joint range of motion. First, the movement of healthy bodies in daily activities is measured by POLHEMUS attached on the body surface. Table 1 shows a summary of the POLHEMUS specification. Next, the HJC is estimated by the positions and orientations of the pelvis and the femur. The joint range of motion is calculated by the estimated HJC and the positions and orientations of the pelvis and the femur. Finally, the joint range of motion described.

2.2 The method to estimate the HJC

The relation between the HJC and hip joint motion

The hip joint consists of a femur and a pelvis. The femur and the pelvis are regarded as rigid body as described in Figure 2 (a). HJC is decided by measuring positions and orientations of the femur and the pelvis if hip joint could be regarded as a ball joint as described in Figure 2 (b).

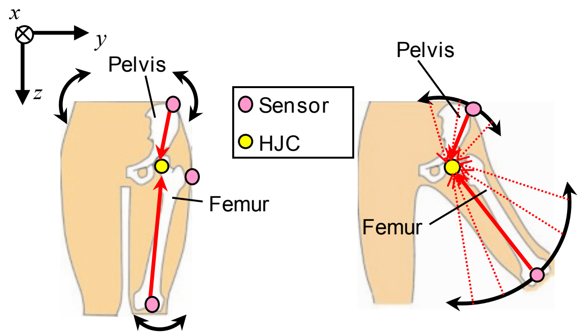
Procedure to estimate the HJC

The following is the procedure to estimate the HJC:

1. The temporary HJC is decided.
2. Two initial vectors; one is from the pelvis to the temporary HJC ($\boldsymbol{v_p}$) and the other is from the femur to the temporary HJC ($\boldsymbol{v_f}$), are calculated by the initial positions of the pelvis, the femur and the temporary HJC while standing as described in Figure 3 (a).
3. The gap between the end point of $\boldsymbol{v_p}$ and $\boldsymbol{v_f}$ occurs due to the difference between the true HJC and the temporary one as described in Figure 3 (b).
4. Finally, the HJC is defined so as to minimize the gap because there would be no gap if the hip joint is a ball joint.

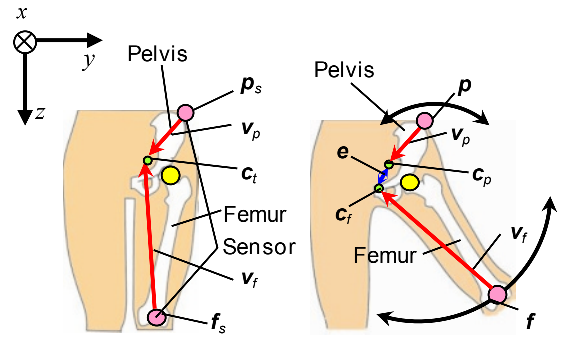
Table 1: POLHEMUS specification.

	Position	Angular
Coverage	76cm	All-attitude
Static accuracy	0.08cm	0.15°
Output	Cartesian coordinates	Euler orientation
Update rate	Two receivers 60updates/second	Angular



(a) (b)

Figure 2: The relation between the HJC and hip joint motion. (a) Standing. (b) In motion.



(a) (b)

Figure 3: The calculation method of HJC. (a) The temporary HJC while standing, (b) The HJC in motion.

The method to calculate the gap

The method to calculate the gap is explained. The vector representing the gap ($\boldsymbol{e}$) is expressed by two located vectors; one is the end point of $\boldsymbol{v_p}$ ($\boldsymbol{c_p}$) and the other is one of $\boldsymbol{v_f}$ ($\boldsymbol{c_f}$) as described in Figure 3 (b).

$$\boldsymbol{e} = \left| \boldsymbol{c}_f - \boldsymbol{c}_p \right| \tag{1}$$

The homogeneous transform matrix is expressed by POLHEMUS.

$$\mathbf{T} = \begin{bmatrix} \mathrm{CA*CE} & \mathrm{CA*SE*SR - SA*CR} & \mathrm{CA*SE*CR + SA*SR} & x \\ \mathrm{SA*CE} & \mathrm{CA*CR + SA*SE*SR} & \mathrm{SA*SE*CR - CA*SR} & y \\ -\mathrm{SE} & \mathrm{CE*SR} & \mathrm{CE*CR} & z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2}$$

Where :

$$\mathrm{CA} = \cos(\mathrm{azimuth})$$
$$\mathrm{CE} = \cos(\mathrm{elevation})$$
$$\mathrm{CR} = \cos(\mathrm{roll})$$
$$\mathrm{SA} = \sin(\mathrm{azimuth})$$
$$\mathrm{SE} = \sin(\mathrm{elevation})$$
$$\mathrm{SR} = \sin(\mathrm{roll})$$

Two homogeneous transform matrices; one is the attitude matrix of the pelvis ($\boldsymbol{T_p}$) and the other is the attitude matrix of the femur ($\boldsymbol{T_f}$) are calculated by substituting the measurement data which are the angle and position of the pelvis ($\boldsymbol{p}$) and the femur ($\boldsymbol{f}$) in equation (2) as described in Figure 3 (b). $\boldsymbol{c_p}$ and $\boldsymbol{c_f}$ are calculated by $\boldsymbol{T_p}$, $\boldsymbol{T_f}$, $\boldsymbol{v_p}$ and $\boldsymbol{v_f}$.

$$\boldsymbol{c}_p = \boldsymbol{T}_p \boldsymbol{v}_p \tag{3}$$

$$\boldsymbol{c}_f = \boldsymbol{T}_f \boldsymbol{v}_f \tag{4}$$

Two homogeneous transform matrices, one is the initial attitude matrix of the pelvis ($\boldsymbol{T_{ps}}$) and the other is the initial attitude matrix of the femur ($\boldsymbol{T_{fs}}$) are calculated by substituting the measurement data which are the initial position and angle of the pelvis ($\boldsymbol{p_s}$) and the femur ($\boldsymbol{f_s}$) for the equation (2). $\boldsymbol{v_p}$ and $\boldsymbol{v_f}$ are calculated by the temporary located vector ($\boldsymbol{c_t}$), $\boldsymbol{T_{ps}}$ and $\boldsymbol{T_{fs}}$ as described in Figure 3 (a).

$$\boldsymbol{v}_p = \boldsymbol{T}_{ps}^{-1} \boldsymbol{c}_t \tag{5}$$

$$\boldsymbol{v}_f = \boldsymbol{T}_{fs}^{-1} \boldsymbol{c}_t \tag{6}$$

$\boldsymbol{e}$ is calculated by substituting equations (3) to (6) in equation (1).

$$\boldsymbol{e} = \left| \left(\boldsymbol{T}_f \boldsymbol{T}_{fs}^{-1} - \boldsymbol{T}_p \boldsymbol{T}_{ps}^{-1} \right) \boldsymbol{c}_t \right| \tag{7}$$

The initial located vector of HJC ($\boldsymbol{c_s}$) is defined so as to minimize the magnitude of $\boldsymbol{e}$ as described in Figure 4 (a).

2.3 Measurement method of the joint range of motion

Procedure to measure the joint range of motion

The following is the procedure to measure the joint range of motion:

1. Two vectors; one is from the pelvis to the HJC ($\boldsymbol{v_{ph}}$) and the other is from the femur to the great trochanter of the femur ($\boldsymbol{v_{fg}}$), are calculated by their initial positions of the

pelvis, the femur and the estimated HJC as described in Figure 4 (a).

2. The joint range of motion is calculated by the motion of the vector from the HJC to the great trochanter of the femur position ($\boldsymbol{v}_{hg}$) as described in Figure 4 (b).
3. The joint ranges of motions for sitting on a chair, getting into the bathtub, walk, stoop, going up and down at stairs, sitting cross-legged on a floor, kneeling on one's knees, squat and sitting on a floor with one's arms holding together the bent knees are recorded.

The method to calculate the joint range of motion

The joint range of motion is expressed by motion of $\boldsymbol{v}_{hg}$ as described in Figure 4 (c). $\boldsymbol{v}_{hg}$ is expressed by the located vectors of the great trochanter of the femur ($\boldsymbol{g}$) and the HJC ($\boldsymbol{c}$) as described in Figure 4 (b).

$$\boldsymbol{v}_{hg} = \boldsymbol{g} - \boldsymbol{c} \tag{8}$$

$\boldsymbol{g}$ and $\boldsymbol{c}$ are calculated by $\boldsymbol{v}_{ph}$, $\boldsymbol{T}_p$, $\boldsymbol{v}_{fg}$ and $\boldsymbol{T}_f$ as described Figure 4 (b).

$$\boldsymbol{c} = \boldsymbol{T}_p \boldsymbol{v}_{ph} \tag{9}$$

$$\boldsymbol{g} = \boldsymbol{T}_f \boldsymbol{v}_{fg} \tag{10}$$

$\boldsymbol{v}_{ph}$ and $\boldsymbol{v}_{fg}$ are calculated by the initial located vectors of the HJC ($\boldsymbol{c}_s$) and the great trochanter of the femur ($\boldsymbol{g}_s$), $\boldsymbol{T}_{ps}$ and $\boldsymbol{T}_{fs}$ as described in Figure 4 (a).

$$\boldsymbol{v}_{ph} = \boldsymbol{T}_{ps}^{-1} \boldsymbol{c}_s \tag{11}$$

$$\boldsymbol{v}_{fg} = \boldsymbol{T}_{fs}^{-1} \boldsymbol{g}_s \tag{12}$$

$\boldsymbol{v}_{hg}$ is calculated by substituting equations (9) to (12) in the equation (8).

$$\boldsymbol{v}_{hg} = \boldsymbol{T}_f \boldsymbol{T}_{fs}^{-1} \boldsymbol{g}_s - \boldsymbol{T}_p \boldsymbol{T}_{ps}^{-1} \boldsymbol{c}_s \tag{13}$$

The axis of rotation ($\boldsymbol{n}$) is expressed by the vector before rotation ($\boldsymbol{v}_{hg}$) and that after rotation ($\boldsymbol{v'}_{hg}$). The joint range of motion (θ) is calculated by $\boldsymbol{v}_{hg}$ and $\boldsymbol{v'}_{hg}$.

$$\left.\begin{aligned} n_x &= x_{hg} z'_{hg} - z_{hg} y'_{hg} \\ n_y &= -x_{hg} z'_{hg} + z_{hg} x'_{hg} \\ n_z &= x_{hg} y'_{hg} - y_{hg} x'_{hg} \end{aligned}\right\} \tag{14}$$

$$\theta = \cos^{-1}\left(\frac{x_{hg} x'_{hg} + y_{hg} y'_{hg} + z_{hg} z'_{hg}}{\sqrt{x_{hg}^2 + y_{hg}^2 + z_{hg}^2}\sqrt{x'^2_{hg} + y'^2_{hg} + z'^2_{hg}}} \right) \tag{15}$$

Where;

$$\boldsymbol{n} = (n_x, n_y, n_z)$$

$$\boldsymbol{v}_{hg} = (x_{hg}, y_{hg}, z_{hg})$$

$$\boldsymbol{v}'_{hg} = (x'_{hg}, y'_{hg}, z'_{hg})$$

2.4 Procedure to display the joint range of motion

The display style of the joint range of motion is explained. A sphere centering on the HJC is shown as described in Figure 5 (a). The diameter of the sphere which is regarded as the cup is 50 mm. The sphere is trimmed by the movement of $\boldsymbol{v}_{hg}$ as described in Figure 5 (b). The trimmed edge is the maximum joint range of motion as described in Figure 5 (c).

2.5 The procedure to measure the position and the angle by POLHEMUS

POLHEMUS is used to measure positions and orientations. The initial positions of the great trochanter of the femur, the pelvis and the femur are measured as described in Figure 6.

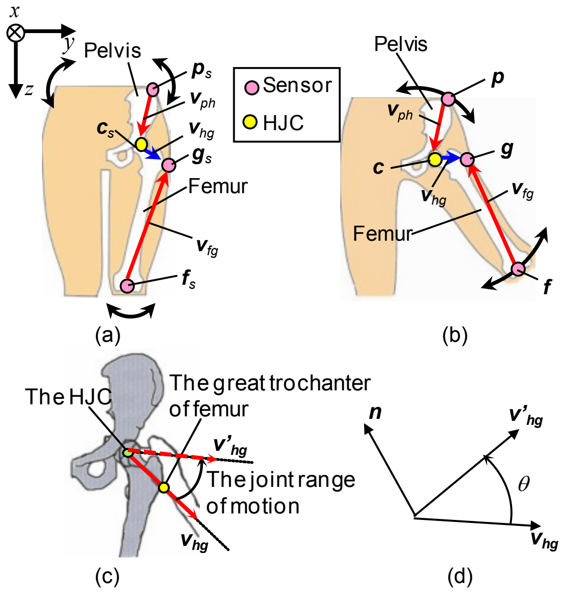


Figure 4: The calculation method for the joint range of motion. (a) All vectors with standing. (b) Vectors in motion. (c) The joint range of motion due to moving vectors. (d) The calculation method for the unit normal vector and the angle.

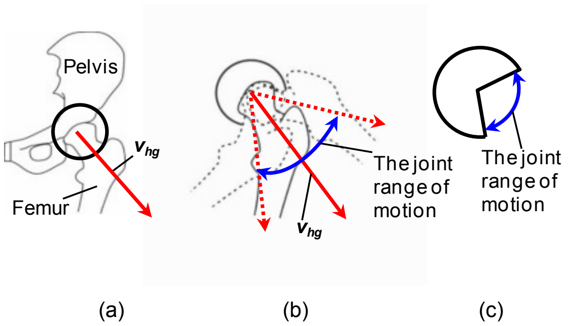


Figure 5: Procedure to display the joint range of motion. (a) A sphere centering on the HJC. (b) Trimming due to the moving vector. (c) Joint range of motion obtained.

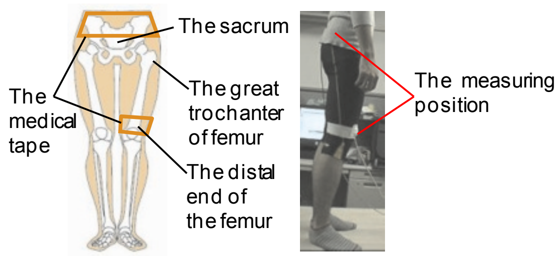


Figure 6: Sensor positions.

The sensors attached to the sacrum and the distal end of the femur regions, where the displacement caused by the movement of muscle and skin can be reduced, are fixed tenaciously by a medical tape and movements are measured. The motions for sitting on a chair, getting into the bathtub, walk, stoop, going up and down at stairs, sitting cross-legged on a floor, kneeling on one's knees, squat and sitting on a floor with one's arms holding together the bent knees are recorded.

3 SIMULATUON

The model of hip prosthesis

Figure 7 shows the model of hip prosthesis. The diameter of femoral head is 26 mm and the length of hip prosthesis is 165 mm. The radius of femoral head and the length of hip prosthesis are of variable length. The angle of femoral anteversion and neck are calculated by the measured positions and orientations of pelvis and femur.

Simulation

Figure 8 shows the simulation result of sitting on a chair. The joint range of motion and the HJC are calculated by the measured position and orientation of pelvis and femur and the hip prosthesis move in real time. The shaft trims the sphere by the radius of the shaft when the model of hip prosthesis moves. The trimmed sphere shows the joint range of motion.

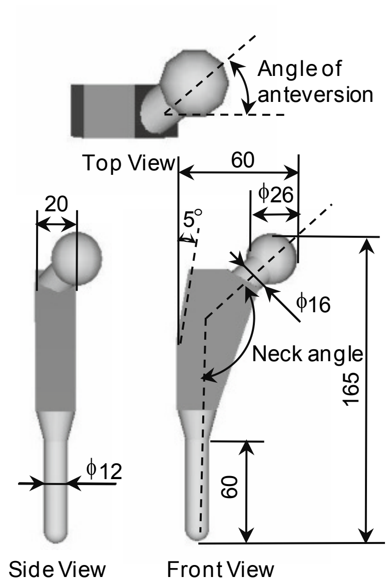


Figure 7: The model of hip prosthesis.

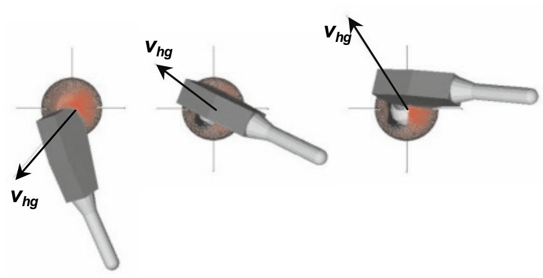


Figure 8: The simulation result for squat.

4 RESULT

The joint ranges of motions in daily activities are calculated by the positions and orientations of the pelvis and the femur and the estimated HJC. Figure 9 (a) to (i) shows the joint range of motion for sitting on a chair, getting into the bathtub, walk, stoop, going up and down at stairs, sitting cross-legged on a floor, kneeling on one's knees, squat and sitting on a floor with one's arms holding together the bent knees.

5 DISCUSSION

The gap of hip joint center

In this section, the possibility of proposed method is discussed and the gap of the HJC is examined. Table 2 shows summaries of the gap between the end points of $\boldsymbol{v_p}$ and $\boldsymbol{v_f}$.

The minimum gap for stoop motion is 2.4 cm. The maximum gap for the all recorded motions is 8.7cm when sitting on a floor with one's arms holding together the bent knees. The gap value of a shoulder joint center which was measured by Takafumi has been 10 cm [8].

The gap of the HJC is caused by: the error of measurement by POLHEMUS; regarding hip joint as a ball joint; and moving sensor by skin or muscle. The measurement error by POLHEMUS has few influences on the gap of the HJC. The gap of the HJC is caused by regarding the hip joint as a ball joint because the hip joint is closes to a roller bearing. The HJC cannot be estimated accurately because the positions and orientations of sensors are shifted by the skin and muscle movement.

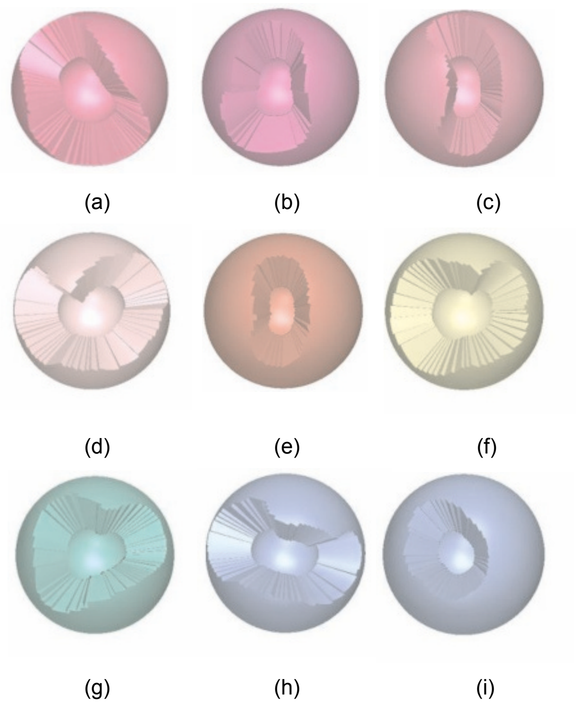
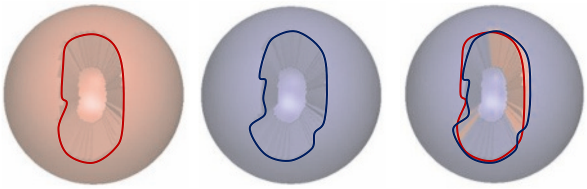


Figure 9: The joint range of motion. (a) Sitting cross-legged on a floor. (b) Kneeling on one's knees. (c) Squat. (d) Sitting on a floor with one's arms holding together the bent knees. (e) Sitting on a chair. (f) Getting into the bathtub. (g)Walk. (h)Stoop. (i) Going up and down at stairs.

The error of the joint range of motion

The joint ranges of motion are calculated by the HJCs estimated from femur and pelvis. The errors of some joint ranges of motion are described in Figure 10 to Figure 13. Parts (a) and (b) of Figure10 to Figure 13 show the joint range of motion calculated by the HJC estimated from pelvis and femur respectively.

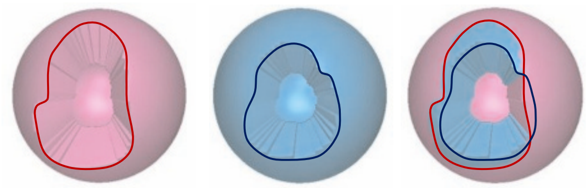


(a) (b) (c)

Figure 10: The of joint range of motion for sitting on a chair. (a) Calculated by the HJC estimated from plvis. (b) Calculated by the HJC estimated from femur. (c) Calculated by two HJCs estimated from pelvis and femur.

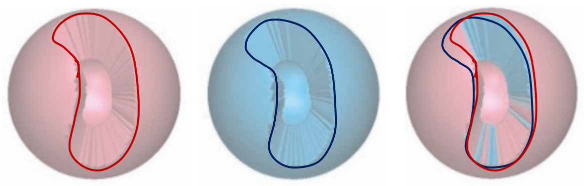
(a) (b) (c)

Figure 11: The of joint range of motion for sitting cross-legged on a floor. (a) Calculated by the HJC estimated from plvis. (b) Calculated by the HJC estimated from femur. (c) Calculated by two HJCs estimated from pelvis and femur.



(a) (b) (c)

Figure 12: The joint range of motion for kneeling on one's knees. (a) Calculated by the HJC estimated from plvis. (b) Calculated by the HJC estimated from femur. (c) Calculated by two HJCs estimated from pelvis and femur.



(a) (b) (c)

Figure 13: The joint range of motion for squat. (a) Calculated by the HJC estimated from plvis. (b) Calculated by the HJC estimated from femur. (c) Calculated by two HJCs estimated from pelvis and femur.

Part (c) of Figure 10 to Figure 13 shows the joint range of motion calculated by two HJCs estimated from pelvis and femur. Table 3 shows the percentage of the joint range of motion and the error of the joint range of motion per the area of a sphere. The joint ranges of motion for sitting on a chair and squat which are the only flexural motion calculated by two HJCs estimated from pelvis and femur overlap considerably as described in Figure 10 (c) and Figure 13 (c). However, the areas of no overlaps for sitting cross-legged on a floor and kneeling on one's knees which includes adduction or inner rotation are large as described in Figure 11 (c) and Figure 12 (c). The percentage of the error of joint range of motion for sitting on a chair, sitting cross-legged on a floor and kneeling on one's knees are 0.7%, 0.3% and 3.1% as shown by Table 3. The percentage of the error of joint range of motion for sitting cross-legged on a floor and kneeling on one's knees are small but the areas of no overlaps are large. Therefore the error of joint range of motion would be caused by not only the gap of the HJC but also displacement in direction of the gap of the HJC.

The conventional cup is designed as hemispherical shape and the persentage of the joint range of motion for the conventional cup per the area of a sphere is more than 50%. In this experiment, the persentages of the joint range of motion per the area of a sphere are less than 50% as shown by Table 3. The conventional cup would have a risk of dislocation because the joint range of motion is too large.

Design the cup to prevent dislocation

In order to reduce a risk of dislocation after total hip arthroplasty, the cup should be designed as the shape which results from trimming the actually used the joint range of motion from a sphere. Figure 14 (a) to (c) shows the cup whose shape is obtained by trimming the joint range of motion. The dislocation could be avoided by trimming the joint range of motion because the area covering the femoral head becomes large and the impingement between the cup and the stem does not occur. Furthermore the cups suitable for variety of lifestyles could be designed. If the cup designed as described above, it has to be placed accurately. However, the decision of the direction of placing the cup is difficult because the shape of the cup is quite complicated.

Table 2: The gaps between the end points of $\boldsymbol{v}_p$ and $\boldsymbol{v}_f$.

	The gap of the HJC [cm]
Sitting cross-legged on a floor	5.4
Kneeling on one's knees	6.3
Squat	6.8
Sitting holding one's knees with arms	8.7
Sitting on a chair	6.7
Getting into the bathtub	6.7
Walk	4.0
Stoop	2.4
Going up and down at stairs	6.7

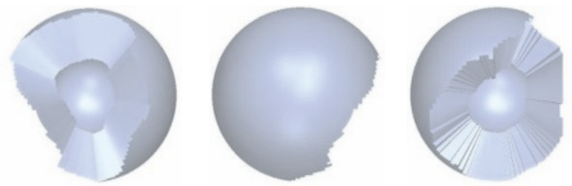
6 CONCLUSIONS AND FUTURE WORK

The non-invasive measuring method which obtains the joint range of motion without using CT or X-ray was proposed. The invisible HJC has been estimated by two three-dimensional magnetic positional sensors which were attached on a body surface. The joint range of motion was calculated by the estimated HJC and the positions and orientations of the pelvis and the femur.

The maximum gap between two HJCs estimated by pelvis and femur is 8.7 cm. The percentage of the maximum error of the joint range of motion per the area of a sphere is 12.0%. The error of joint range of motion would be caused by the gap of the HJC and displacement in direction of the gap of the HJC.

In the future works, the method of fixing sensors shall be improved so as not to move sensors. The gap shall be corrected by analyzing tendency of the gap and performing additional experiments.

To reduce a risk of dislocation, the cup should be designed for the joint range of motion.



(a) (b) (c)

Figure 14 : The cup whose shape is obtained by trimming the joint range of motion. (a) From frontal view of the body. (b) From rear view of the body. (c) From right side view of the body.

7 ACKNOWLEDGMENTS

We would like to express to Dr. Blagovest Vladimirov for soulful support.

8 REFERENCES

[1] Yasuhiro, H. and Satoru, Y., 2005, The survey of dislocation after total hip arthroplasty – the analysis of the case and the cause for thirteen years, the Journal If Japanese Physical Therapy Association, 32/2: 470.

[2] Yasuhiro, N. et al, 2005, The investigation of dislocation example after total hip arthroplasty- the investigation of the dislocation cause concentrating on the installation angle of component, The Central Japan Association of Orthopaedic Surgery & Traumatology, 48: 303-304.

[3] Otake, Y. et al, 2003, 4-dimensional computer-based motion simulation after total hip arthroplasty, Stud Health Technol Inform 94: 251-257.

[4] Hagio, K. et al, 2004, A novel system of four-dimensional motion analysis after total hip arthroplasty, J Orthop Res 22/3: 665-670.

[5] Davis, R.B., Ounpuu, S., Tyburski, D. and Gage, J.R., 1991, A gait analysis data collection and reduction technique, Human Movement Sciences, 10/5: 575-587.

[6] Vaughan, C.L., Davis, B.L., and O'Conners, J.C., 1992, dynamics of human gait, 15-43, Human Kinetics Publisher.

[7] Seiji, N., 2006, Tom Atlas of Anatomy, BUNKODO CO, 244-247.

[8] Takafumi, T. and Shinichi, H., 1999, Building Linkage Model of Human Arm through Shoulder Joint Motion Measurements, 20 :70, society of biomechanisms.

Table 3: The percentage of the joint range of motion and the error of the joint range of motion per the area of a sphere

	The joint range of motion[%] (Calclated by the HJC estimated from pelvis)	The joint range of motion[%] (Calclated by the HJC estimated from femur)	The error of the joint range of motion[%]
Sitting cross-legged on a floor	40.5	40.2	0.3
Kneeling on one's knees	29.3	26.2	3.1
Squat	24.8	24.0	0.8
Sitting holding one's knees with arms	23.6	21.1	2.5
Sitting on a chair	20.6	19.9	0.7
Getting into the bathtub	33.2	45.2	12.0
Walk	41.7	43.7	2.0
Stoop	15.8	15.5	0.3
Going up and down at stairs	18.2	22.4	4.2

Robot Task of Pin Insertion to a Hole without Chamfering and with Small Clearance Using Fuzzy Control

Masaya Sanji[1], Tomonori Nakamura[1], Masato Suzuki[1], Seiji Aoyagi[1]

Contact person: Seiji Aoyagi , Kansai University, aoyagi@iecs.kansai-u.ac.jp, Fax 06-6330-3154

[1] Dept. of Mechanical Engineering, Faculty of Engineering Science, Kansai University, Osaka, Japan

Abstract

Almost the robot tasks are performed by the teaching playback method. This method is based on the considerably high repeatability of the robot arm, thus the environment is required to be strictly constant. However, a human can achieve the task such as assembly with adapting to the environmental change. In this research, the pin insertion task to a hole without chamfering and with small clearance is focused on. And this pin insertion task is aimed to be achieved by imitating a human motion, extracting and defining the rule using membership functions, and executing the task based on fuzzy control.

Keywords:

Fuzzy control; Pin insertion; Hole without chamfering

1 INTRODUCTION[5][6]

While the automation of machining parts has been advanced from early time in the industrial history, the automation of assembly has been delayed, because the assembling process needs complex operations accompanied critical judgment. However, there are strong needs for assembly automation in the viewpoints of saving workers, lowering production cost, etc. In the course of achieving successful assembly automation, there should be advances in element technologies, computer technologies, robot technologies, etc.

It is difficult for robots to perform assembling tasks automatically, because the robot technologies at present stage cannot cope with the uncertainties in the production process. For example, if there is an error in a hole clearance based on a stochastic machining error, it is rather difficult for an industrial robot to accomplish a pin insertion task to this hole by only employing the teaching playback method. In a teaching playback method, the transitions of joint angles are memorized while performing a given task in a teaching stage, then these joint angles are positioned to the memorized angles for many times in a playback stage. This teaching playback method relies on considerably high robot's repeatability, thus the environment is required to be strictly constant for all the trials. So, a robot cannot cope with even a slight change of shape and position of parts owing to uncertainty. In this method, sensor feedback is usually not used.

To cope with above-mentioned uncertainties, sensor feedback is inevitable for a robot. For example, in the pin insertion task, force or impedance control using a force sensor is necessary for reducing the interference force between pin and hole. A human also can skillfully achieve the task such as assembly with adapting to the environmental change. In this process, he uses the sensor feedback, i.e., he adjusts the motion based on the sensor information.

In addition, it is a problem how to process information from the sensor. For example, a human can perform a complex task by appropriately judging it, even in the situation that only rough information is given about the relative position and orientation between objects.

In this research, fuzzy control, which is able to deal with the uncertainty, is applied. And, a pin insertion task to a hole with considerably small clearance is aimed to be achieved by imitating a human motion.

2 SYSTEM COMPOSITION

Figure 1 shows the system composition of this study. As a robot arm, 7-DOF articulated robot (Mitsubishi Heavy Industries, Ltd., type PA10) was employed, and the six axis force sensor (BL AUTOTEC, Ltd., type F/T-10/100) and CCD camera (SONY, Ltd., type XC-ST50) are installed at the tip of the arm. The diameter of a pin is 11.97 mm, and that of a hole without chamfering is 12.01 mm, so the clearance is ±20 μm.

3 DETECTION OF HOLE POSITION BY IMAGE PROCESSING

Human recognizes the position of the object by his sight. Similarly, a rough position of a hole is detected by image data processing in this study. A CCD camera takes a picture of the

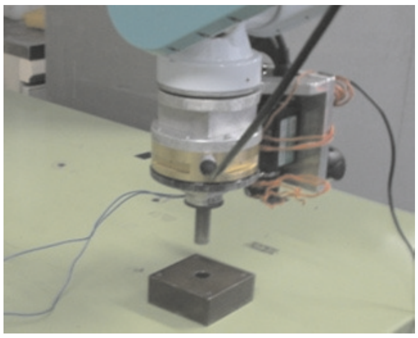


Figure 1: System composition.

hole from overhead. Surrounding the hole, four marks are set, the adjacent distance between which is precisely known in advance. Thus, a real distance corresponding to adjacent pixels of the CCD image can be calculated by using the ratio of the known distance to the existing pixel number between two marks' images.

By the above-mentioned way, two dimensional position of the hole relative to the camera centre is obtained by image processing. Practically, there is an offset between the pin and the camera (see Figure 1). So the detected relative position between the hole and the camera becomes same as that between the hole and the pin, after shifting the robot arm's tip by this offset.

4 PIN INSERTION IMITATING HUMAN MOTION

One of purposes of this study is to increase the performance of a robotic pin insertion so that it becomes human equivalent. For this purpose, it is effective to imitate the human motion. Therefore, at first in this study, a human pin insertion task was observed in detail for 10 testees.

As the result, it was proven that (s)he first explores the contact state between the pin and the hole, by inclining, shifting, and rotating the pin. Finally, (s)he can recognize the relative position of the pin to the hole.

In this study, imitating this observed motion, the relative position of a hole to a pin is aimed to be recognized by the exploring motion as follows: the pin is moved downward till it becomes contact with the surface, on which the hole exists. Then, the pin is inclined while keeping the tip position of the pin. Following that, the pin is translationally shifted on the surface. If the direction of shifting is wrong, the pin is rotated like wobbling motion, and the correct direction of shifting is searched. After the pin is shifted, it is set up to the vertical orientation from inclination. Then, a virtual compliance control is carried out while inserting a pin downward.

5 FUZZY CONTROL

5.1 Outline of fuzzy control [3][4]

In fuzzy control, the control variable is decided based on the fuzzy rule, which is composed of antecedent part ("if" part) and consequent one ("then" part). Each variables used in this rule bears a membership function, which can deal with the uncertainty.

5.2 Pin insertion procedure using fuzzy control

The pin is roughly positioned by a robot arm over the hole by visual feedback, then, it is moved downward vertically till the force F_Z in Z direction, i.e., vertical direction, becomes 4.9 N (0.5 kgf). Here, the pin is inclined by 10 deg, and shifted horizontally by 1 mm. If the shifting direction is correct to the center of the hole, the moment around X axis (M_X) does not change so much. Compared to this case, if the shifting direction is wrong, the moment changes much.

In the former case, the shifting distance is decided by using the components of the detected force/moment according to membership functions and a fuzzy rule, which are acquired by using many test data in advance. Hereinafter in this section, the former case is focused on. As for the latter case, it is focused on in the next Section 6. Geometric relationship between the pin and the hole is shown in Figure2. The extent how the pin should be shifted is decided mainly based on the change in M_X before and after the pin shifting of 1 mm, which is expressed as ΔM_X , Example of the experimental transition of *Mx* is shown in Figure 3. The change in case of 1 mm shifting in F_X , F_Y , and M_Y are expressed as ΔF_X , ΔF_Y , and ΔM_Y , respectively. Each of force/moment information has its corresponding membership function. Consequent part also has a membership function. Here, membership function outputs the fitness value, which shows the extent how the antecedent or consequent reasoning is true.

As for ΔM_X , the membership function is decided on the basis of experimental data shown in Fiugure 3. The resultant membership function for ΔM_X (or ΔM_Y) is shown in Figure 4. Other membership functions for ΔF_X (or ΔF_Y) and F_Z are shown in Figures 5 and 6, respectively. The membership function for the consequent part is shown in Figure 7.

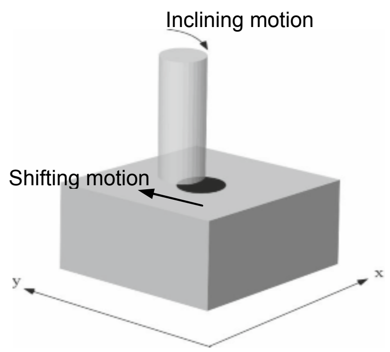


Figure 2: pin and hole.

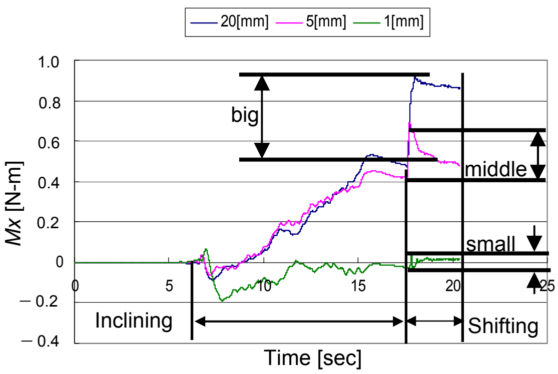


Figure 3: Change in *Mx*.

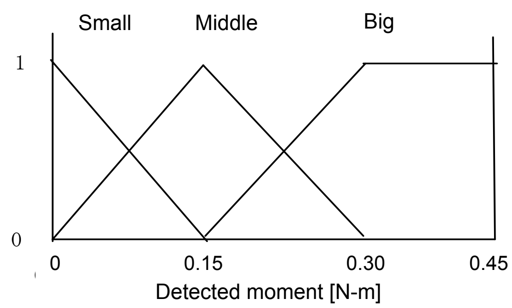


Figure 4: Membership function for Δ *Mx* (or Δ *My*).

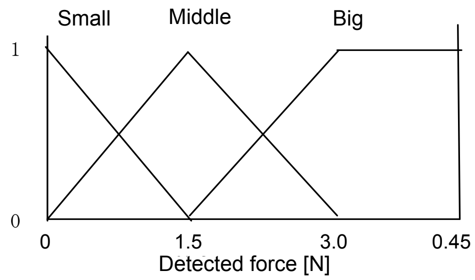


Figure 5: Membership function for ΔFx (or ΔFy).

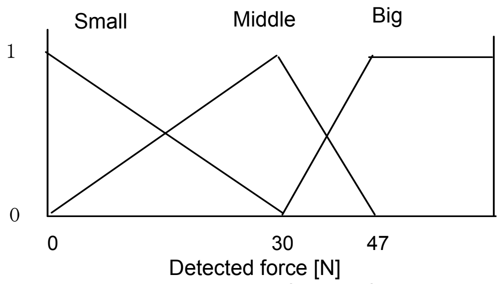
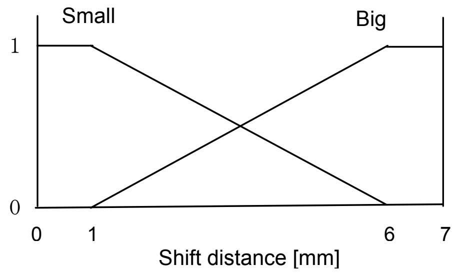


Figure 6: Membership function for *Fz*.



Firgure 7: Member ship function of the consequent part.

R_A: IF *Fz*, ΔF_Y, ΔM_x are small,
and ΔFx, ΔM_Y are middle,
THEN Move to negative in *y* direction a little.

R_B: IF Fz, ΔF_Y, ΔM_x are middle,
and ΔFx, ΔM_Y are middle,
THEN Move to negative in *y* direction greatly.

Figure 8: Fuzzy rule.

The extent how the pin should be shifted in *Y* direction is decided based on fuzzy rules (see Figure 8) and above-mentioned membership functions.

6 DETECTION OF CORRECT SHIFTING DIRECTION TO THE HOLE BY ROTATING A PIN

When the pin is inclined and shifted by 1 mm, there is a possibility that the antecedent part of the fuzzy rule might not be satisfied. This means that the hole does not exist in the inclined direction. When such a case is encountered, the correct shifting direction to the hole should be detected in some method.

To address this problem, the pin is rotated like wobbling motion keeping the inclined angle and keeping the contact position of the pin's tip, which is performed by controlling the pitch and the yaw angles of the robot arm's tip. The movement of pin rotation around the hole is schematically shown in Figure 9.

It was empirically confirmed for many trials that F_Z becomes small and both F_X and F_Y become approximately zero when the direction of the centerline of the pin intersects with that of the hole, i.e., when the hole exists in the inclined direction of the pin. The reason for this fact is supposedly as follows: contact state changes from one-point contact state to two-points one, when the inclined direction of the pin is identical to the direction in which the hole exists. Thus, the contact force F_Z changes rapidly at this moment.

For example, Figure 10 shows the experimental transition of detected force during the pin is rotated. At the state marked by circles in this figure, the hole exists in the inclined direction of the pin. So, in case of this example, the above-mentioned criterion for detecting the hole existing direction, i.e., F_Z becomes small and both F_X and F_Y become approximately zero, surely holds true.

According to this criterion, the correct shifting direction to the hole centre is searched, followed by shifting the pin based on the fuzzy control which is mentioned the previous Section 5.

7 VIRTUAL COMPLIANCE CONTROL

7.1 Virtual compliance control [2]

After the pin is shifted, the pin is set up to the vertical orientation from inclination. Here, a virtual compliance control, which is usually employed in robotic compliant assembly motion, is executed while inserting a pin downward in *Z* direction.

Since mechanical compliance of a robot is usually stiff without applying any control contrivance, even the small positioning error causes the great increase of contact force, finally making the insertion task to fail. In the pin insertion task, this

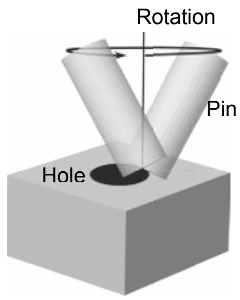


Figure 9: Schmatic of pin rotation.

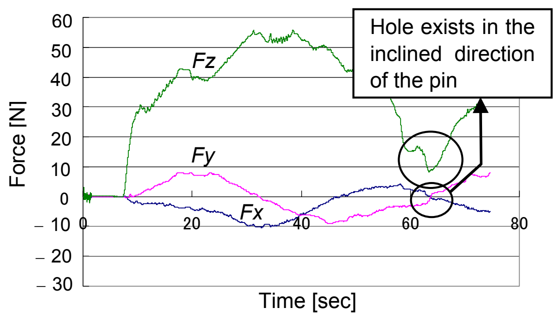


Figure 10: Changes in force.

situation is possible.

To address this problem, in this study, virtual compliance is given by control software, in order to reduce the mechanical compliance. In this method, stiffness (reciprocal of compliance), viscosity, and inertia are adjusted to desired values by controlling the velocity of robot arm's tip based on the force feedback.

7.2 Control rule [1]

The rule of virtual compliance control is shown by the following expression:

$$\dot{\boldsymbol{x}}_{ref} = M^{-1}\int(\boldsymbol{f}_{ext} - \boldsymbol{K}\boldsymbol{x} - D\dot{x})dt \cdot \quad (1)$$

This equation is discretized as follows:

$$\boldsymbol{v}_k = TM^{-1}(\boldsymbol{f}_{ext} - \boldsymbol{K}\boldsymbol{x}) + (I - TM^{-1}D)\boldsymbol{v}_{k-1}\ , \quad (2)$$

$$\dot{\boldsymbol{q}}_{ref} = J^{-1}\boldsymbol{v}_k\ , \quad (3)$$

where $\dot{\boldsymbol{q}}_{ref}$: angular velocity (target value), $\boldsymbol{f}_{ext}$: external force, J : Jacobian, M : virtual mass, K : virtual spring constant, D : virtual viscousity coefficient, I : unit matrix, T : sampling time, and $\boldsymbol{x}$: position.

8 PIN INSERTION EXPERIMENT

A pin insertion experiment was carried out according to the proposed method. The flow chart of the experiment is shown in Figure 11. The transition of pin insertion is shown in Figure 12.

This task was successfully achieved, provided that the error in the initial roughly positioning by the CCD camera was comparatively small.

9 CONCLUSIONS

A pin insertion task to a hole without chamfering and with small clearance is focused on. A human pin insertion task was observed in detail for 10 testees. It was proven that (s)he first explores the contact state between the pin and the hole, by inclining, shifting, and rotating the pin. Finally, (s)he can recognize the relative position of the pin to the hole.

Imitating this observed motion, the relative position of a hole to a pin is recognized by the exploring motion as follows: 1) the pin is moved downward till it becomes contact with the surface, on which the hole exists. 2) The pin is inclined by 10 deg and shifted 1 mm. 3) Fuzzy rule for shifting the pin to the hole is constructed mainly considering the moment M_X . 4) If the antecedent part is not satisfied, i.e., the direction of shifting is wrong, the pin is rotated like wobbling motion, and the correct direction of shifting is searched. 5) The pin is set up to the vertical, i.e., upright, orientation from inclination. 6) The virtual compliance control is carried out while inserting a pin downward.

An experimental system is composed of a 7DOF robot equipped and 6 axes force/torque sensor. A pin insertion task was successfully achieved, provided that the error in the initial positioning by the CCD camera was comparatively small.

REFERENCES

[1] H.Hirabayashi, K.Sugimoto, S.Arai, M.Sakagami, 1986, Virtual compliance control of multiple degree of freedom robot, SICE JCMSI, 22/3:343-350

[2] T.Yoshikawa, 1988, Practical Control of Robot Manipulators, CORONA PUBLISHING CO.,LTD., 117-143

[3] K.Hirota, 1990, Fuzzy Reasoning and Fuzzy Control, Jour.JSME, 93/856:24-30

[4] K.Hara, R.Yokogawa, 1991, Recognition of State in Peg-in-hole by Fuzzy Schema, JSME Journal, 57/533 :124-129

[5] S. Okabe, Y. Kamiya, 1993, Recent Trend on Assembly Robot, Journal of RSJ, 11/1:58-61

[6] H.Mayeda, 1993, Intelligence of Robots for Manipulation, Journal of RSJ, 11/2:172-17

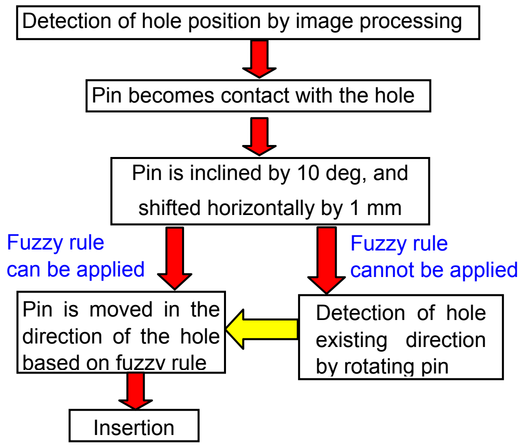


Figure 11: Flow chart of pin insertion expriment.

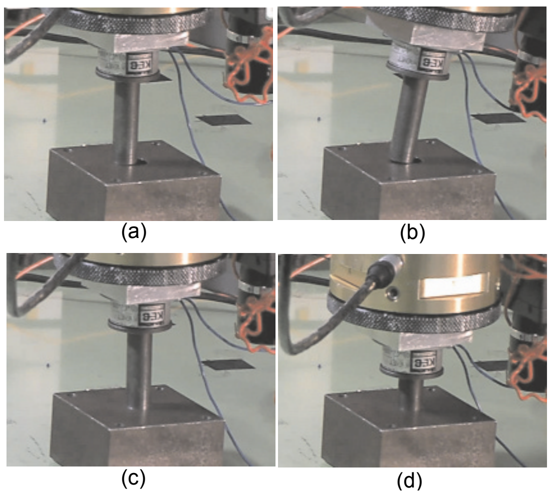


Figure 12: Transition of pin insertion; a) pin is moved downward, b) pin is inclined and shifted based on fuzzy rule, c) pin is set to upright orientation, d) pin is inserted.

Disassembly Support System for Used Products Considering Destruction of Their Parts

Hidefumi Wakamatsu[1], Minoru Matsuishi[1], Eiji Morinaga[1], and Eiji Arai[1]

[1] Dept. of Materials and Manufacturing Science, Graduate School of Eng., Osaka University, Osaka, Japan

Abstract

A disassembly support system which suggests efficient disassembly/destruction sequences is proposed. First, a precedence constraint graph for disassembly is applied to derivation of candidates for efficient destruction procedures. Using such graph, a product destruction of which is not efficient can be determined. Second, the access graph is introduced. It represents connection of empty spaces existing in/around a product. Using the access graph, possible removing sequences of parts can be derived. Next, evaluation of such sequences according to the number of required operations is proposed. Furthermore, parts grouping using the access graph is proposed to reduce the number of removing operations. After that, determination of destroying position is shown. Finally, the effectiveness of our proposed system is demonstrated with a case study. Using our proposed system, workers in recycling manufactures can dispose used products efficiently without knowledge about each product.

Keywords:

Disassembly; Destruction; Reuse; Recycle; Precedence Constraint Graph; Access Graph

1 INTRODUCTION

Recently, 3R becomes more important for society with low environmental load and sustainability. For that, it seems to be crucial that disassemblability or recyclability of products is considered.

Generally, products can be disassembled in the reverse order of assembly sequence[1]. However, it may be difficult to retrieve a valuable/harmful part efficiently when a used product is disassembled in such order. We do not need to disassemble parts to be recycled/discarded maintaining their original shape. This implies that it is possible to retrieve valuable/harmful parts more efficiently by cutting or breaking such recycled/discarded parts.

Actually, in Japanese recycling manufactures, valuable/ harmful parts in used products are retrieved efficiently by cutting or breaking other parts and the rest of the products are crushed whole and their shreds are separated according to their material. It seems that this approach for reuse/recycle will be continued for a while. Therefore, it is important to consider disassembly including destruction.

Disassembly/destruction procedures of used products in a recycling manufacture should be determined according to their manufacturer, year of manufacture, and type. However, they are not compiled in a detailed manual and they depend on workers' experience. Moreover, the disassembly/destruction procedure cannot be determined completely in design phase of a product because such procedure may change according to its status of use, technological advancement, ordinances at the end of its life, and so on. Consequently, we develop a disassembly support system which suggests efficient disassembly/destruction sequences including position to be destroyed to workers in recycling manufactures.

First, a precedence constraint graph for disassembly is applied to derivation of candidates for efficient destruction procedures. Next, the access graph is introduced to generate possible sequences of parts to be removed and evaluation of such sequences is explained. Furthermore, parts grouping with the access graph is proposed. After that, determination of destructing position is shown. Finally, the effectiveness of our proposed system is demonstrated with a case study.

2 EFFICIENT DESTRUCTION PROCEDURES DERIVED FROM PRECEDENCE CONSTRAINT GRAPH

First of all, we apply a precedence constraint graph to derivation of candidates for efficient destruction procedures. Figure 1 shows an example of a product. In this case, disassembly of part P_1 precedes that of part P_2 because P_1 restricts move of P_2 geometrically. We define that P_1 has a geometrical precedence constraint on P_2. Such constraint can be represented as a directed graph as shown in Figure 1(b). Nodes in the graph correspond to individual parts and arrows correspond to precedence constraints between parts. In this case, the geometrical precedence constraint is described as an arrow pointing to the node of P_2. We define an arrow from one node pointing to another node as an output arrow of the relevant node and an arrow from another node pointing to the node as an input arrow of the relevant node.

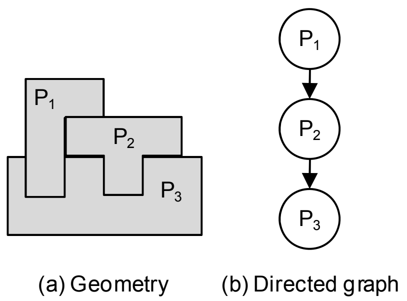
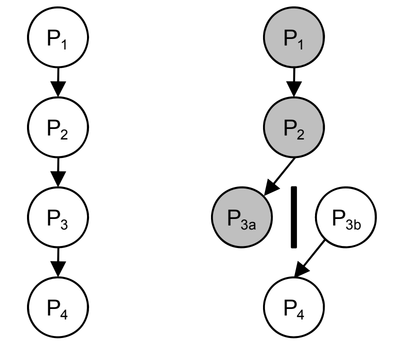


Figure 1: Example of precedence constraint

Figure 2(a) shows an example of a precedence constraint graph. It means that P_1, P_2, and P_3 must be disassembled for removal of P_4. The number of disassembly operations is three. Let us consider cutting of P_3 and its dividing into two pieces. They are referred to as P_{3a} and P_{3b}, respectively. If P_2 does not prevent disassembly of P_{3b} while it still prevents disassembly of P_{3a}, P_4 can be removed after disassembly of P_{3b} alone. In this case, a precedence constraint graph can be described as Figure 2(b). In this figure, a white/gray node corresponds to a part which has to/does not have to be disassembled respectively, and a vertical bar between two nodes represents a cutting operation. It is found that the number of the total operations shown in Figure 2(b) is less than that in Figure 2(a) because two disassembly operations are eliminated while one cutting operation is added. Thus, by cutting some parts, we can reduce the number of operations to retrieve a valuable/harmful part. In this study, we define a procedure with less disassembly/destruction operations as a more efficient destruction procedure. Cutting parts corresponds to changing precedence constraints as shown Figure 2. Furthermore, whether the total number of operations is reduced or not depends on the structure of a precedence constraint graph and which part is cut. So, efficient destruction procedures can be derived from a precedence constraints graph.

Note that we assume that one piece of a cut part, which is indicated by subscript a, inherits all input arrows from the original part in a precedence constraints graph and the other piece, which is indicated by subscript b, inherits all output arrows. Such assumption indicates that by cutting any part, its piece b has no precedence constraint with other parts. Needless to say, the validity of this assumption depends on geometry of individual parts. A method proposed in this section is for derivation of candidates for efficient destruction procedures and geometrical verification is needed to get feasible efficient destruction procedures after such derivation.

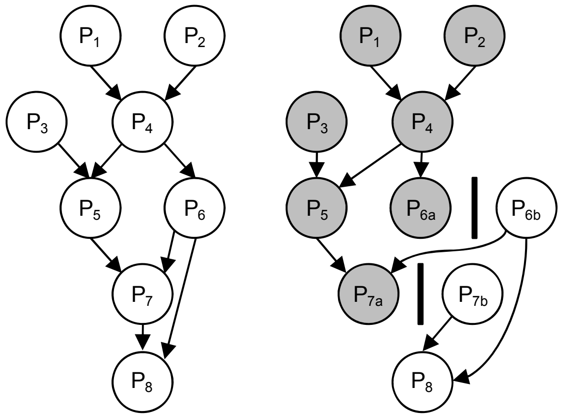


(a) Original (b) After cutting of P_3

Figure 2: Precedence constraint graph

A method to derive candidates for efficient destruction procedures is as follows. Let P_R be a part to be retrieved and let m be the total number of upper parts of P_R. Let n be the total number of upper parts of P_R except topmost parts which can be removed without disassembly of any other part. In the example shown in Figure 2(a), m=3 and n=2 when P_R is P_4. First, all input arrows of ${}_nC_x$ (x=1, ..., n) parts C_i, that is, 1 through n combinations of parts from n parts, is deleted. After that, remained arrows are traced backward from P_R and passed parts are counted. Let n_c be the number of passed parts. Then, m-n_c disassembly operations are reduced by cutting C_i. Let D_j be a part which does not have to be disassembled by cutting C_i and let n_d be the number of D_j. Next, the validity of each combination of C_i is verified. If we cannot reach P_R by tracing remained arrows forward from C_i, it implies that parts which do not have to be disassembled are cut. So, such combination is not appropriate as a candidate for efficient cutting. Thus, invalid combinations are excluded. Finally, combination C_i of which the number n_c is minimum and satisfies the condition $n_c<n_d$ is selected as a efficient destruction procedure.

For example, Figure 3 shows an example of a precedence constraint graph of a product consisting of 8 parts. For retrieving of P_8, 7 disassembly operations are needed. Then, m=7 and n=4. So, ${}_4C_4+{}_4C_3+{}_4C_2+{}_4C_1$=15 combinations of parts are checked. If P_5 and P_6 are cut as shown in Figure 3(b), n_c=2 and n_d=5. P_8 can be retrieved after applying 4 operations, that is, 2 destruction operations and 2 disassembly operations.



(a) Original (b) After cutting of P_6 and P_7

Figure 3: Example of efficient destruction procedure

As mentioned before, to get feasible efficient destruction procedures, geometrical verification is needed. In the case shown in Figure 3, we have to verify that we can cut P_6 so that its piece P_{6b} can be disassembled without disassembly of P_1, P_2, and P_4, for example. However, by using this method, we can determine whether destruction is efficient or not without geometrical verification. Figure 4 shows another example of a precedence constraint graph. In this case, 4 parts have to be disassembled to retrieve P_5 when part cutting is not considered. Figure 4(b) shows one of destruction procedures consisting of 2 disassembly operations and 2 cutting operations. Regardless of its feasibility, the number of operations is not reduced. Applying the proposed method to this example, it is found that the number of operations cannot be reduced even if any part is cut. Therefore, the proposed method in this section, which uses a precedence constraint graph, is useful for determination of a product destruction of which is not efficient.

To apply this method, a precedence constraint graph, that is, a disassembly sequence of a product must be given. A disassembly sequence can be calculated from the geometry of a product[2]. However, such calculation is time-consuming. We tried to reduce such computational cost by introducing some precedence constraints except those with respect to geometry[3]. Furthermore, if we cut a part, product geometry changes. This implies that re-calculation for generation of updated disassembly sequences is needed. Therefore, we propose another method to generate in the next section.

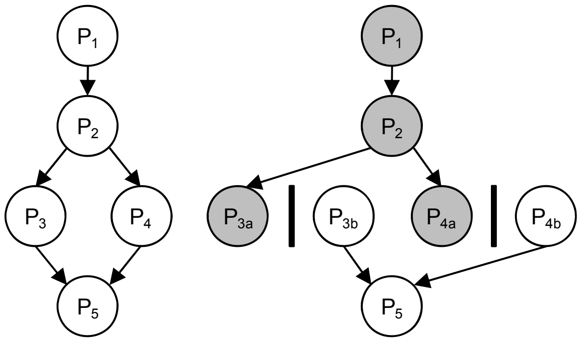


Figure 4: example of inefficient destruction procedure

3 INTRODUCTION OF ACCESS GRAPH

Before explaining another method, we assume that each product has ID tag by which geometry and material of its all parts can be identified. Then, such information can be used for disassembly/destruction by referring ID tags when products are collected and carried to recycling manufactures[4]. Which part can/cannot be destroyed is determined by recycling manufactures according to its material and the current ordinances. If a part is reused, it cannot be destroyed. But if a part is recycled/discarded, it can be destroyed.

In this section, on these assumptions, we propose a method to generate removing sequences of parts by focusing on the accessibility of parts. To remove a part, some tools including a human hand must be able to access it. This indicates that a space around the part must connect to another space where the tools exist. If they don't connect, some parts separating them must be removed. A valuable/harmful part should be retrieved efficiently. Then, we have to get an appropriate connecting sequence of spaces, that is, an appropriate removing sequence of parts. In this study, we introduce the access graph which can be generated from CAD data to derive such sequence.

Let the void space around a product be outer space S_0 and void spaces included in the product be inner spaces S_a (a=1, 2, ...). Then, parts open to outer space S_0 can be accessed. These parts are referred to as the 1st shell. Let us assume that part P_{1j} in the 1st shell is also open to inner spaces S_a. Such inner spaces are referred to as surround R_{1jk}. Let a set of parts open to surround R_{1jk} and not included in the 1st shell be the 2nd shell. Thus, we can define the *i* th shell repeatedly. This indicates that all parts can be classified into some groups, that is, shells. Note that inner space S_a may be *k* th surround R_{ijk} of part P_{ij} in the *i* th shell and also be *n* th surround R_{lmn} of part P_{lm} in the *l* th shell. Then, both parts P_{ij} and P_{lm} correspond to the contour of inner space S_a.

If part P_{1j} in the 1st shell is detached, outer space S_0 and inner spaces R_{1jk}, which exist around part P_{1j}, are connected. Then, we can access parts which are the contour of inner spaces R_{1jk}. This implies that if we want to detach part P_{ij}, its surround R_{ijk} must be connected to outer space S_0 by detaching parts repeatedly in order of increasing their shell number. Note that to detach a part is defined as to disassemble a part whole, to cut the part partially, or to remove cut pieces to connect two spaces separated by that part in this study.

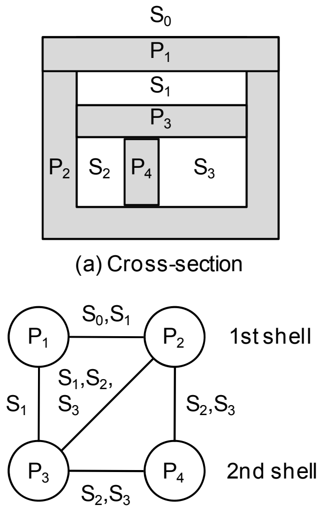


Figure 5: Example of access graph

In order to derive such detaching sequence of parts, we introduce the access graph. In this graph, each node represents each part and each arc between two nodes represents shared spaces by both corresponding parts as their surrounds. Nodes are hierarchized based on shells in which parts are included. Figure 5(a) shows the cross-section of a product consisting of 4 parts and containing 3 inner spaces and figure 5(b) shows its access graph. Part P_1 and P_2 are in the 1st shell and part P_3 and P_4 are in the 2nd shell. To retrieve part P_4, we must connect surround $R_{24k}=\{S_2, S_3\}$ of P_4 to outer space S_0 finally. For this, we must detach part P_2 to connect inner spaces S_2 and S_3 to outer space S_0 or detach part P_3 to connect inner spaces S_2 and S_3 to inner space S_1. For the latter case, part P_1 or P_2 must be detach in advance to access part P_3 because part P_3 is in the 2nd shell and it is not accessible. Then, sequences of parts to be detached are derived as follows:

$$\begin{cases} P_2 \rightarrow P_4 \\ P_1 \rightarrow P_3 \rightarrow P_4 \\ P_2 \rightarrow P_3 \rightarrow P_4 \end{cases} \tag{1}$$

Let us define a sequence of parts to be detached until a target part can be retrieved as a possible detaching sequence. By tracing nodes and arcs from a part in the 1st shell to the target part in the access graph, possible detaching sequences for the target part can be derived.

Furthermore, the access graph can be generated from the CAD data of a product. First, a rectangular solid in which a product is encased is generated and the volume of the product is subtracted from such rectangular solid. As a result, a set of polyhedra which correspond to void inner/outer spaces is remained. Then, we can verify whether a part, which is a real object, contacts with a space, which is an imaginary object, or not. If a space contacts with a part, it is surround of that part. Thus, the access graph and possible detaching sequences can be computed automatically once the CAD data, that is, the geometry of a product is given.

4 EVALUATION OF DETACHING SEQUENCES

In this study, we evaluate possible detaching sequences based on the number of detaching operations. If a part can be disassembled, one disassembling operation is needed. If it cannot be disassembled but can be destroyed, two operations are needed: one is cutting operation and the other is removing operation of cut pieces. Otherwise, the detaching sequence including such part is excluded from candidates for appropriate detaching sequences because at least two disassembling operations are required: one is for the part itself and the other is for a part preventing it from being disassembled.

Note that the minimum number of detached parts in possible detaching sequences for a part in the *i* th shell is *i*. The minimum number of detaching operations of a part is 1 and the maximum number is 2. So, the minimum number of detaching operations for a part in the *i* th shell is *i* and the maximum number is 2*i*. This indicates that we can exclude detaching sequences including more than 2*i* parts from candidates. Possible detaching sequences including no more than 2*i* parts are referred to as feasible detaching sequences. We select sequences with the minimum number of detaching operations from feasible sequences as efficient ones.

A method to derive appropriate detaching sequences is as follows:

1. Generate the access graph and possible detaching sequences.
2. Select feasible detaching sequences.
3. For each feasible sequence, verify the disassemblability of each part in the sequence in order of increasing its shell number.
   - If the part can be disassembled, set the number of its detaching operation to be 1.
   - If the part cannot be disassembled but can be destroyed, set the number of its detaching operation to be 2.
   - If the part cannot be disassembled and cannot be destroyed, exclude the sequence from candidates.
4. Determine sequences with the minimum number of detaching operations as efficient ones.

In this method, we verify the disassemblability of not all parts but only parts included in feasible detaching operations. Moreover, let us assume that the *i*-1 th part in a detaching sequence is determined to be cut. Then, verification of the disassemblability of the *i* th part in the sequence is carried out after regarding the *i*-1 th part as nonexistent. This indicates that the disassemblability of a part is assumed to be independent of how to cut its precedent parts. Then, we can reduce the computational cost to generate efficient disassembly/ destruction sequences.

5 PARTS GROUPING

Two parts made of the same material do not have to be separated if both of them are recycled/discarded. This implies that they can be grouped and can be regarded as one part. And, it leads to reduction of the number of detaching operations.

Conditions to group multiple parts in this study are as follows:

1. They are made of the same material.
2. They contact with each other.
3. If they are grouped, they do not confine a part made of another material to themselves.

If grouped parts confine a part made of another material to themselves, we have to separate them finally to remove the confined part because its treatment may differ from that of the grouped parts. Recall that the objective of parts grouping is to reduce the number of operations. So, condition 3 should be satisfied. Information with respect to condition 1 can be gotten from ID tag, and condition 2 is checked when the access graph is generated. Furthermore, condition 3 can be verified using the access graph.

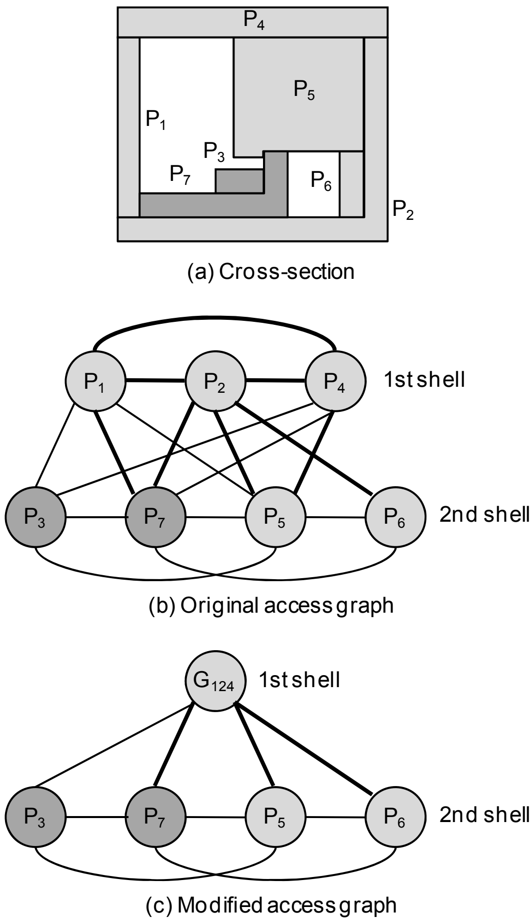


Figure 6: example of parts grouping

Figure 6(a) shows the cross-section of a product consisting of 7 parts made of 2 kinds of material. Part P_1, P_2, P_4, P_5, and P_6 are made of one material while P_3 and P_7 are made of the other material. Figure 6(b) shows its access graph where an arc between two nodes with heavy line means that those two parts contact with each other. Note that space number of each arc is omitted.

If we group part P_1, P_2, and P_4, the access graph changes as shown in Figure 6(c). Then, if we can trace the graph from a part in the 1st shell to a target part without passing of the grouped parts, the target part is not confined and can be accessed by detaching other parts. In this case, the 1st shell consists of the grouped parts G_{124} alone. So, we cannot trace the graph from the 1st shell to part P_3 or P_7 without passing of the grouped parts. Thus, by using the access graph, we can determine whether such parts grouping is feasible or not.

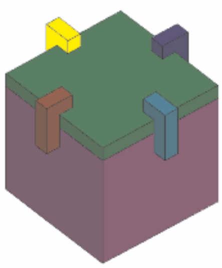
6 DETERMINATION OF CUTTING POSITION

If a part in a detaching sequence cannot be disassembled but can be destroyed, it must be cut. Then, we have to determine the cutting position. By verifying disassemblability of the *i* th part in a feasible detaching sequence, its removing direction can be determined. Let us assume that the *i*-1 th part cannot be disassembled but can be destroyed. Then, we consider two types of cutting of that part to disassemble the *i* th part; contour cutting and horizontal cutting. In the former, the cutting line of the *i*-1 th part is determined by projecting the shape of the *i* th part onto that part in the removing direction. In the latter, the cutting plane of the *i*-1 th part the normal vector of which is parallel to the removing direction is determined so that the plane does not intersect with any part without that part. Which cutting is selected depends on the size of the cutting area in each cutting.

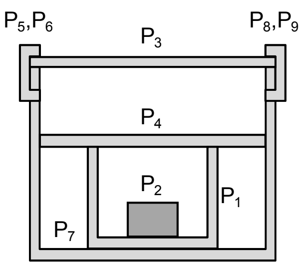
Note that such cutting position cannot be determined in advance, for example, in design phase because detaching sequences/operations, that is, parts to be cut may change according to the situation in disassembly/recycle phase. Utilizing CAD data of a product in disassembly/recycle phase, not only efficient detaching sequences/operations but also the cutting position can be determined.

7 CASE STUDY

In this section, we demonstrate the effectiveness of our method proposed in section 3 through 6 with a pilot system.



(a) 3D shape



(b) Cross-section

Figure 7: Product for case study

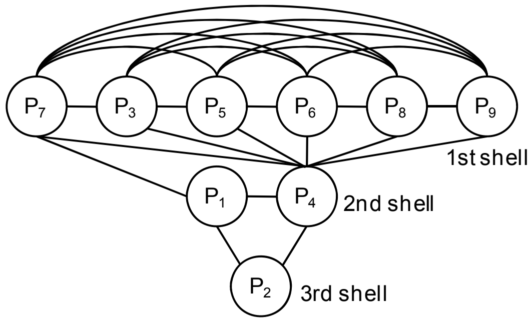


Figure 8: Access graph of product shown in Figure 7

Figure 7 shows a product consisting of 9 parts. Figure 8 shows its access graph. Let us assume that part P_2 in the 3rd shell must be retrieved. Then, the following 14 feasible detaching sequences were derived from the system.

$$\begin{cases} P_7 \to P_1 \to P_2 \\ P_3 \to P_4 \to P_2 \\ P_5 \to P_4 \to P_2 \\ P_6 \to P_4 \to P_2 \\ P_7 \to P_4 \to P_2 \\ P_8 \to P_4 \to P_2 \\ P_9 \to P_4 \to P_2 \\ P_7 \to P_1 \to P_4 \to P_2 \\ P_3 \to P_4 \to P_1 \to P_2 \\ P_5 \to P_4 \to P_1 \to P_2 \\ P_6 \to P_4 \to P_1 \to P_2 \\ P_7 \to P_4 \to P_1 \to P_2 \\ P_8 \to P_4 \to P_1 \to P_2 \\ P_9 \to P_4 \to P_1 \to P_2 \end{cases} \tag{2}$$

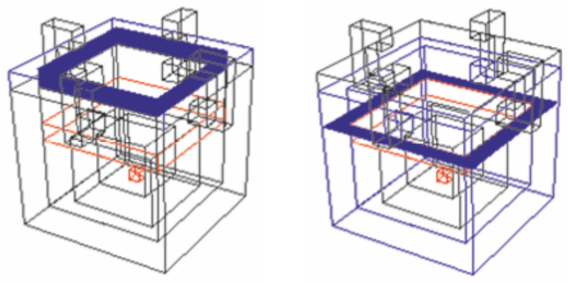
Next, the disassemblability of each part in each sequence was verified. In this case study, it was assumed that all parts can be destroyed. Then, the following detaching sequences requiring 4 detaching operations were selected as efficient ones.

$$\begin{cases} P_3(\text{cut and remove}) \to P_4(\text{disasseble}) \to P_2(\text{disasseble}) \\ P_7(\text{cut and remove}) \to P_1(\text{disasseble}) \to P_2(\text{disasseble}) \\ P_7(\text{cut and remove}) \to P_4(\text{disasseble}) \to P_2(\text{disasseble}) \end{cases} \tag{3}$$

Note that at least 7 disassembling operations are needed to retrieve part P_2 if destruction of parts is not considered.

Figure 9(a) shows the cutting position when we select contour cutting of part P_3 in sequence P_3->P_4->P_2 and Figure 9(b) shows that when we select horizontal cutting of part P_7 in sequence P_7->P_4->P_2. In actual, the cutting position with the minimum cutting area is displayed.

Thus, how to disassemble/destroy a used product can be displayed by our proposed system once constant information such as its geometry and material, and variable information such as its status of use and the current ordinances are input. Then, workers in recycling manufactures can dispose used products efficiently without knowledge about each product.



(a) Contour cutting (b) Horizontal cutting

Figure 9: Output of proposed system

8 CONCLUSIONS

A disassembly support system which suggests efficient disassembly/destruction sequences was proposed. First, a precedence constraint graph for disassembly was applied to derivation of candidates for efficient destruction procedures. Although generation of such graph from CAD data of a product is time-consuming, a product destruction of which is not efficient can be determined using it. Next, the access graph was introduced. It represents connection of empty spaces existing in/around a product and it is generated from CAD data of a product. Using the access graph, possible detaching sequences of parts can be derived. Then, we can select efficient detaching sequences according to the number of detaching operations. At the same time, whether parts to be removed should be disassembled or cut is also determined. Furthermore, the access graph is utilized for parts grouping to reduce the number of detaching operations. The cutting position can be determined from CAD data. So, how to disassemble/destroy a used product can be displayed by our proposed system once information about a product and the current rules for disposal are input. Then, workers in recycling manufactures can dispose used products efficiently without knowledge about each product. Furthermore, by changing information about which part can/cannot be destroyed according to its status of use or the scientific and social environment, appropriate disassembly/destruction sequences can be derived.

9 REFERENCES

[1] Briceno, J. J. and Pochiraju, K., 2007, Automatic Disassembly Plan Generation from CAD Assembly Models, IEEE International Symposium on Assembly and Manufacturing, 64-69.

[2] Arai, E., Uchiyama, N., and Igoshi, M., 1995, Disassembly path generation to verify the assemblability of mechanical products. JSME International Journal, Series C, 38/4:805-810.

[3] Wakamatsu, H., Tsumaya, A., Shirase, K., and Arai, E., 2001, Development of Disassembly Support System for Mechanical Parts and Its Application to Design Considering Reuse/Recycle, EcoDesign 2001: Second International Symposium on Environmentally Conscious Design and Inverse Manufacturing, 372-377.

[4] Tateno, T., and Kondoh, S., 2004, Proposal of Ubiquitous Disassembly System for Realizing Reuse and Recycling in Cooperative Distributed Facilities, International Conference on Machine Automation, 165-170.

The discriminative method of the insertion using a cantilever model for the optical fiber array automatic assemble

Tohru Sasaki[1], Yoshihiro Sakai[1]
[1] Graduate School of Science and Engineering, Toyama University, Gofuku 3190, Toyama, 930-8555, Japan

Abstract
Assembly for miniature mechanism is used in precision mechanics and electronics generally. An optical fiber array is one of the miniature mechanisms and it consists of multi-core optical fibers and the V-grooved substrate. Because the demand for optical fiber arrays are increasing, it needs the automatic assemble of optical fiber array. A fiber must be put in a precise position of the V-groove for automatic assemble of optical fiber array. However, it was difficult that a fiber must be put in a precise position of the V-groove by the conventional method. It is necessary to develop the Insertion method of fiber to the V-grooved substrate for assemble of optical fiber array. This paper deals with a fiber that is inserted like a cantilever model and the developed equipment inserts fiber easily. The relation of a fiber length and positional offset between a fiber and the V-groove was obtained.

Keywords:
Optical Fiber, Automatic Assemble and Vibration

1 INTRODUCTION

This paper deals with insertion method for miniature mechanism assembly. Assembly for miniature mechanism is used in precision mechanics and electronics generally[1], [2]. An optical fiber array is an optical communication device that consists of multi-core optical fibers and the V-grooved substrate [3]. Because the demand for optical fiber arrays is increasing, it needs the mass production of optical fiber arrays. For automatic assemble of optical fiber array, a fiber must be put in a precise position of the V-groove. The friction coefficient of a optical fiber with the V-groove's slope prevents precise positioning. Therefore vibration application is effective in order to reduce the coefficient of friction. It is necessary to develop the insertion method of fiber to the V-grooved substrate for assemble of optical fiber array.

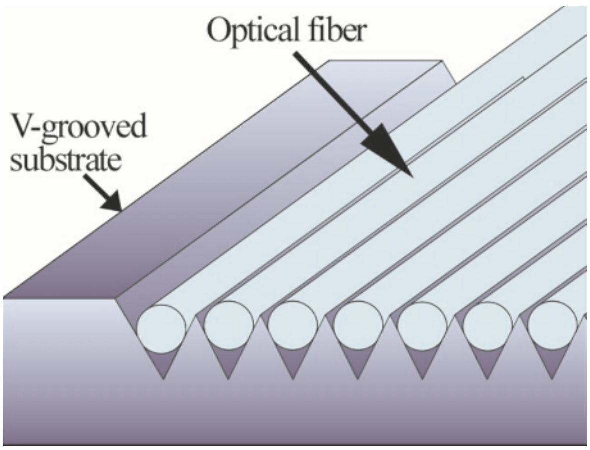
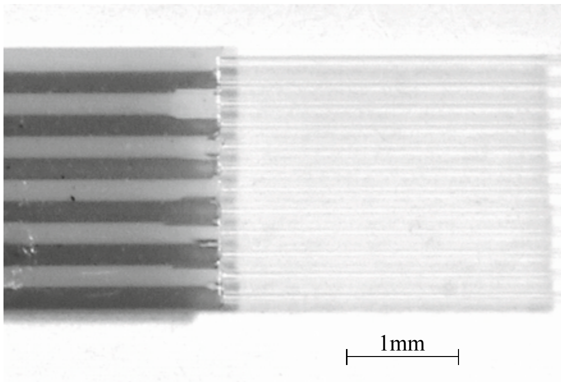
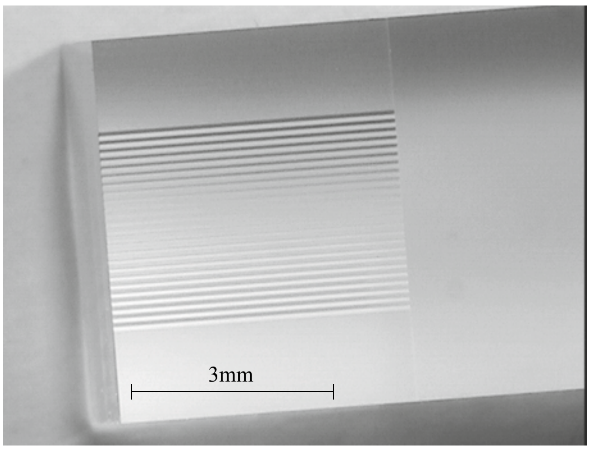


Figure 1: Positioning of optical fiber array



(a) Tape fiber



(b) V-groove substrate

Figure 2: Optical fiber array parts

2 PROBLEM IN THE FIBER ARRAY ASSEMBLY

The optical fiber array consists of a lot of optical fiber and connecting optical parts that it is composed of the V-grooved substrate as shown in figure 1. In this research, to assemble fiber module of 24 cores, two tape fibers of 12 cores shown in figure 2 (a), is piled up and it is pushed into the V-grooved substrate shown in figure 2 (b).

The V-grooved substrate is grinding processed to the glass wafer as shown in figure 3. The cover of the tape fiber is removed and the tape fiber is inserted in the V-groove. The lid is bonded to this, and all processes of the optical fiber array assembly are completed. The part of the optical fiber insertion to the V-grooved substrate is difficult and we study this process.

For assemble the fiber array, It needs that very thin fibers, the diameter is 125μm, are lined up in the V-grooved substrate as shown 4. Because the fiber that was removed the coating has the warp easily, handling of the fiber is difficult. For a small amount comparatively was production, the fiber array has been made by the hand work though is automated many of optical part assemblies today.

As the factor that a defective rate in assembling parts rises, the adhesion of garbage to the V-grooved substrate and damage on the fiber edge are occurred. The occurrence of garbage and the wound is increased because the hand work and a defective rate are raised.

There is the method of handling one fiber is proposed by the piezoelectric element as a method of not using the V-grooved substrate for assembly, but this method is not practical use by the high cost.

3 CHARACTERISTICS OF FIBER IN INSERTION

When the optical fiber is depressed in contact with the V-grooved substrate slope, a fiber has the insertion angle of ϕ is able to consider to be a cantilever model which has come in contact with the slope, that has a lateral offset δ and an angular offset ϕ with valley line of the V-grooved substrate as shown in figure 5. When inserting fiber in the assembly process, the optical fiber assumes that a fixed edge falls downward than the first position because of pressing the optical fiber against the V-grooved substrate slope like fig. 6.

Here, the contact point is assumed not to slip at all until the unti-power by the deformation of fiber is exceeded from the maximum static friction force in the contact point. When the unti-power is exceeded than static friction force, the contact point with the slope and the fiber begins to slip and tip of fiber fall down to bottom of the V-groove.

The pressed part of fiber can be considered to be not causing the deformation and fixation in the slope, from the point where fiber comes in contact with the slope first to the tip. Therefore, the pressed part of fiber contains the contact point pressed against the slope by the bend. This contact model is deformed in the plane including the straight line that connects the contact point of the slope and the fiber with a fixed end.

Moreover, the deformation of the beam in the plane can be considered transform of overhang cantilever beam to which the pin joint was supported with the contact point[4] .

When a fiber has a curve, a lateral offset and an angular offset occurs between the V-grooved substrate with a center

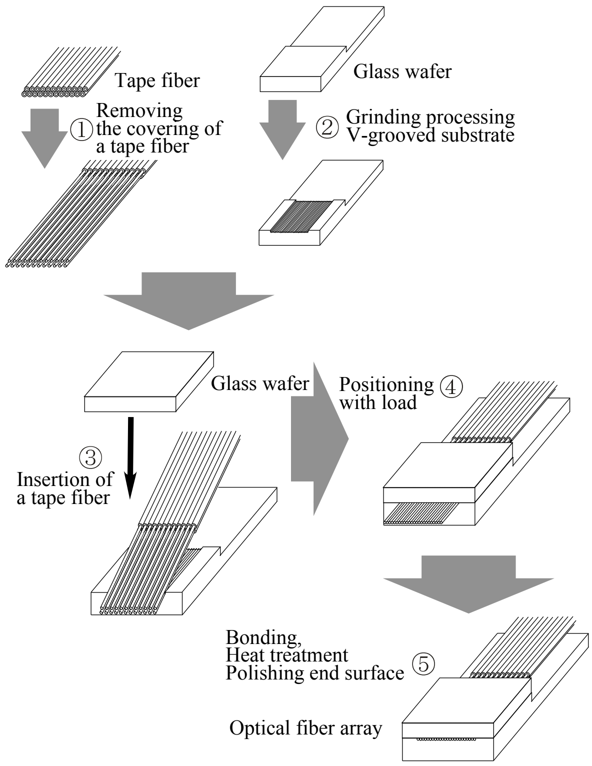


Figure 3: Optical fiber array assemble

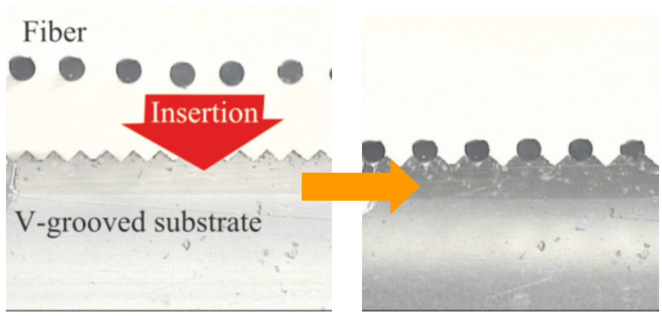


Figure 4: Insertion / lining up on the V-grooved substrate

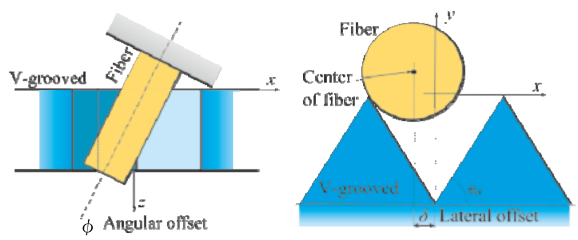


Figure 5: A definition of a position gap / the angular misalignment

of the fiber in each fiber like figure 5. The section A-A' that is perpendicular to the fiber axis makes most quantity of insertion to the V-grooved substrate of the fiber like figure 7 (a). In section A-A', when an insertion angle ϕ is not 0, the angle of the V-grooved substrate θ_a is expressed in

$$\tan\theta_a = \tan\theta \cdot \cos\psi \qquad (1)$$

θ_a decreases by increasing ϕ (Inclining the V-grooved substrate). Moreover the depth of the V-groove in section A-A' decrease and displacement of positioning shrinks at the same time. Because displacement of positioning can shrink, geometric gaps with a fiber and the slope of the V-grooved substrate increase. Therefore it can avoid the impossible insertion state that a fiber touches the slope of the V-grooved substrate with a lateral offset or an angular offset and the friction produces in contact point.

A relative position of the slope of the V-grooved substrate and fiber in section A-A' is expressed the horizontal gap of the slope of the V-grooved substrate and fiber.

When it was assumed that there is not a lateral offset in section A-A', it is r : a fiber diameter, y' : fiber-centered depth at the bottom of the V-groove and γ : angle of a perpendicular line and a point in the circumference of a fiber at the height y like figure 7 (b). The gap Δx of the horizontal direction between the slope of the V-grooved substrate and a point in the circumference of a fiber at the height y is as follows.

$$\Delta x = y\cot\theta_a - r\sin\gamma \qquad (2)$$

Here, it is $\Delta x_{\min}$: the minimum of gap Δx in y direction. When $\Delta x_{\min} > \delta$, a fiber can be pushed down without touching the slope of the V-grooved substrate even if there is a lateral offset. In addition, if fiber fits into only the vertical direction d' in section A-A', even if there is an angular offset, a fiber can be pushed down without touching the slope of the V-grooved substrate as follows.

$$\Delta x_{\min} > \frac{d'\sin\phi}{\sin\psi\cos\psi} \qquad (3)$$

Therefore, if an insertion angle ϕ grows large, the insertion of fiber to the V-grooved substrate makes easy with increasing geometric gaps between a fiber and the slope of the V-grooved substrate like figure 8.

4 INCLINED THE V-GROOVED SUBSTRATE

4.1 Experimental Apparatus and Methods

The factor that positioning to the V-grooved substrate of the fiber is made difficult is the thinness of fiber. Because the diameter of a fiber which removed coating is 125μm, a fiber which have a curve or adsorption not be aligned at the V-grooved substrate. However a fiber is able to be aligned at near the fiber root is coating. Therefore we produced the test device which it can insert a fiber to the V-grooved substrate and slide. At first, the V-grooved substrate that has insertion angle touches to a fiber at near the root. The V-grooved substrate slides to the fiber tip with attaching a fiber next. The V-grooved substrate aligns fiber like combing hair. The constitution of the insertion experimental device is shown in figure 9. Because the V-grooved substrate is fixed on a highly precise stepping motor, an insertion angle of the V-grooved substrate can be variable. Because this mechanism is fixed on a linear actuator, the V-grooved substrate can be slid while changing an insertion angle from a fiber root to the tip of fiber. This mechanism inserts and positions a fiber to the V-grooved substrate precisely.

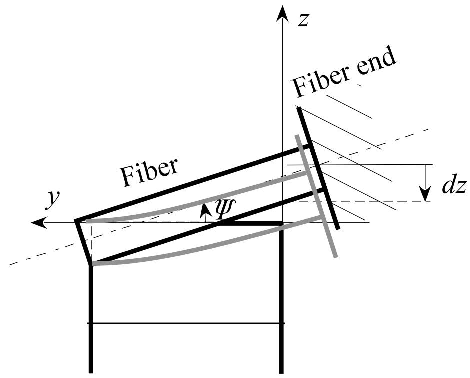
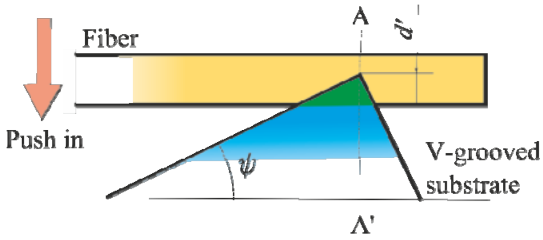
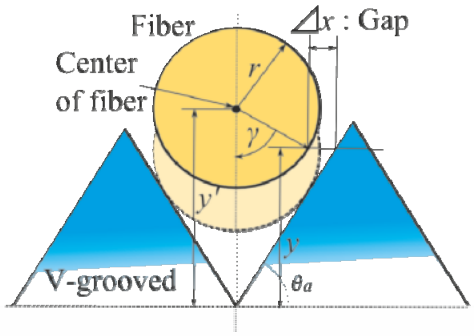


Figure 6: Cantilever model of fiber



(a) Fiber and the V-groove substrate



(b) Position relations

Figure 7: Fiber insertion model in section A-A'

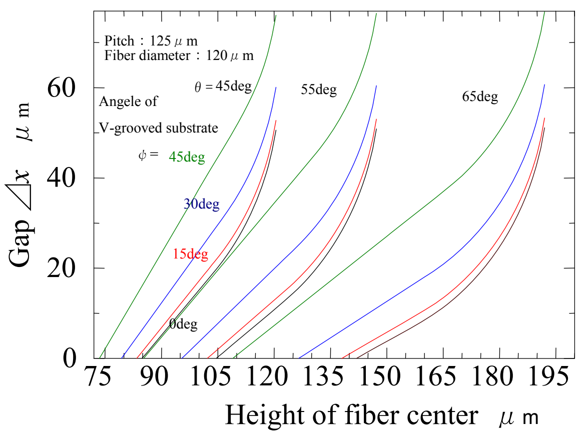


Figure 8: A gap between a fiber and the slope of the V-grooved substrate

Because the fiber root is fixed by a plate on a XY stage, the fiber is able to be positioned in vertical and horizontal direction. Moving the Y stage in the vertical direction, the fixation edge of fiber is pushed down and inserted it in the V-grooved substrate. By a microscope CCD camera the insertion state of the optical fiber is confirm. Figure 10 shows a general view of experimental device.

4.2 Experimental Results

For inspecting the effectiveness of the method that fiber insertion to the V-grooved substrate which has an insertion angle, an insertion experiment carry on with same insertion position and same insertion angle in the test device. It carried with 12 core fibers and the length, a lateral offset and an angular offset ϕ that it is assumed a curve and adsorption is changed. An effect of a lateral offset is shown in figure 11. When an lateral offset δ was large, an displacement until insertion was large. An effect of angular offset is shown in figure 12. When an angular offset ϕ was more than 15 degrees, it was impossible positioning if an insertion angle ψ was 0. An effect of insertion angle is shown in figure 13. The quantity of evaluation is how many fibers (whole 12 core) were positioned into the V-groove. The ratio of insertion success

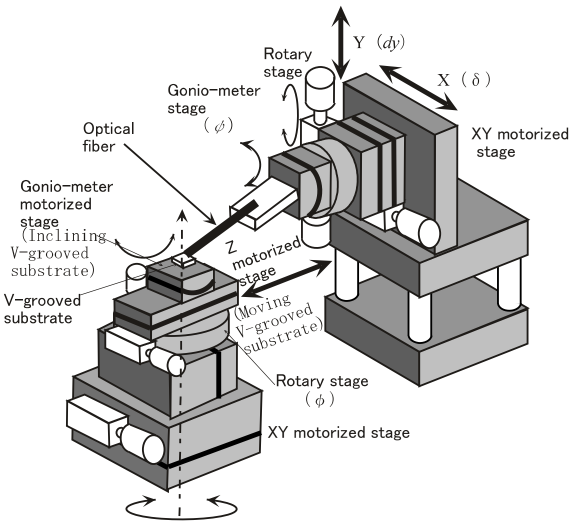


Figure 9: An insertion experimental device

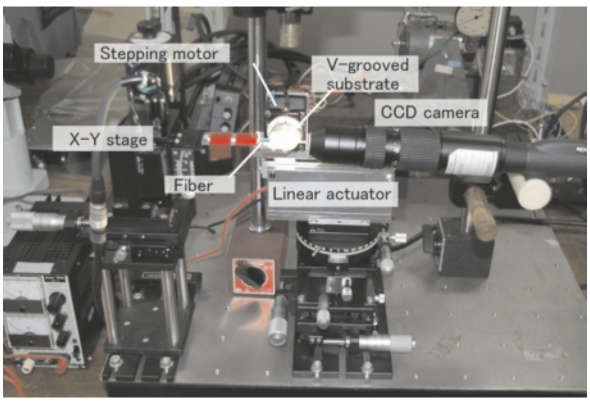


Figure 10: A general view of experimental device

increases with increase of ψ. When an insertion angle ψ is more than 20 degrees, all 12 fiber could be positioned into the V-groove. When an angle offset ϕ is less than 10 degrees, all fiber was able to position it even an insertion angle ψ is 0. As these result, it is effective that the V-grooved substrate has an insertion angle when there is angular offset.

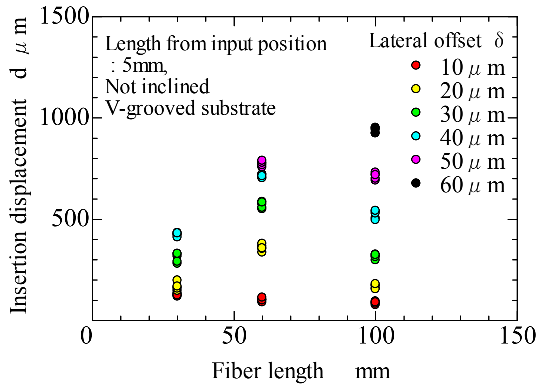


Figure 11: An effect of lateral offset

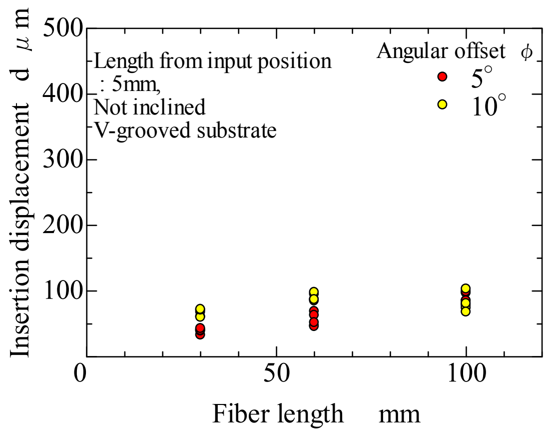


Figure 12: An effect of angular offset

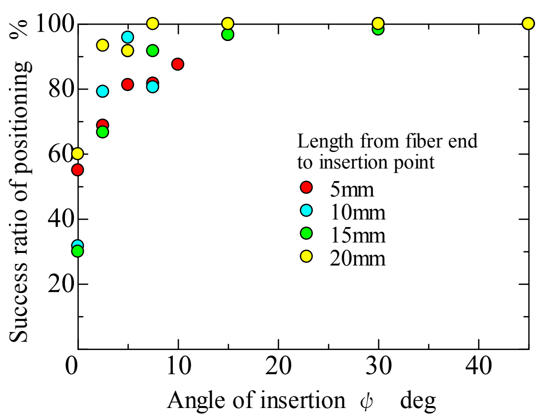


Figure 13: An effect to positioning with an insertion angle

5 SLIDING INSERTION

When there was an angular offset, it is effective that the V-grooved substrate has an insertion angle, but it is not effective when there was a lateral offset.

If fiber becomes long, lateral offset and angular offset of fibers grows large. This reason is fibers being close together by static electric force. Figure 14 shows that unevenness grows large if fiber becomes long.

Fiber insertion is difficult when a fiber insertion position is far from fiber root and exceeds an effective range by a method that the V-grooved substrate inclines. Therefore our device let the V-grooved substrate slide from the insertion position to the near position by the fiber root after inserting fiber into the V-groove.

Because the fiber with coating is fixed in line, coating removed fiber where is near by the fiber root is aligned comparatively. However, because an adsorption and the curve of fiber occur, an unevenness of the fiber grows large and it is impossible of insertion. Figure 15 shows that an unevenness of fiber suddenly grows large where the length from fiber root to insertion position is more than 20mm. On the contrary, a fiber can be inserted into the V-grooved substrate where the length from fiber root to insertion position is less than 20mm easily.

Therefore, like figure 16, the insertion position where the fiber is touched to the V-grooved substrate for the first time is had to near to a coating fiber root. If the V-grooved substrate must be positioned in the fiber tip for assembling, the V-grooved substrate or a fiber slide to the axial direction of fiber from the first insertion position with touching fiber with the V-grooved substrate and move to the necessary position. Finally the V-grooved substrate is positioned in a tip of fiber with aligning.

It is examined whether fiber could slide in a state inserted in the V-grooved substrate under the condition that there was a position gap by this slide insertion method. A lateral offset of the maximum is 60μm.

The first insertion length from coating fiber root is 5mm. The V-grooved substrate has an insertion angle and slide to the fiber axial direction. A quantity of movement of the limit that could keep the state that fiber stood in line on the V-grooved substrate is measured. The insertion angle of the V-grooved substrate changed from 0 degrees to 45 degrees. The slide speed is 1.9mm/sec. In addition, the experiment sample used is as follows.

Test fiber has 12 cores, diameter of fiber is ϕ 125μm, pitch of fibers is 125μm, fiber length is 100mm.

The V-grooved substrate has 24 V-groove, pitch of the V-grooved substrate is 125μm, the V-groove angle is 55 degrees, depth of V-grooved substrate is 90μm.

When there was a lateral offset 60μm like figure 17, the V-grooved substrate was able to slide to 50mm in condition that have been aligned fiber. Because a lateral offset that is slide to the fiber axial direction don't increase, it can be keep a quantity of lateral offset when an first insertion state.

On the other hand, when there is an angle offset, a lateral offset increased by the angle of the fiber's axial direction and the slide direction of the V-grooved substrate. Therefore it is impossible that the state that fiber is aligned on the V-grooved substrate be keep by increasing lateral offset.

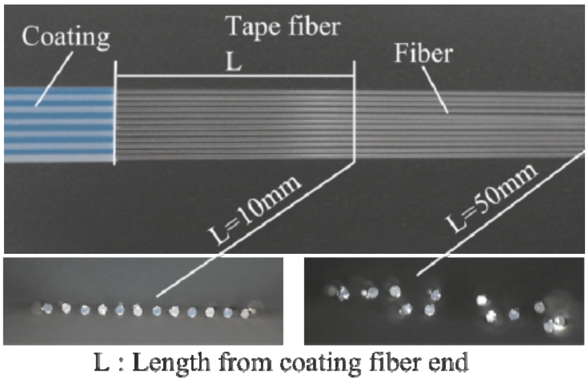


Figure 14: The change of the lining up state by the fiber length

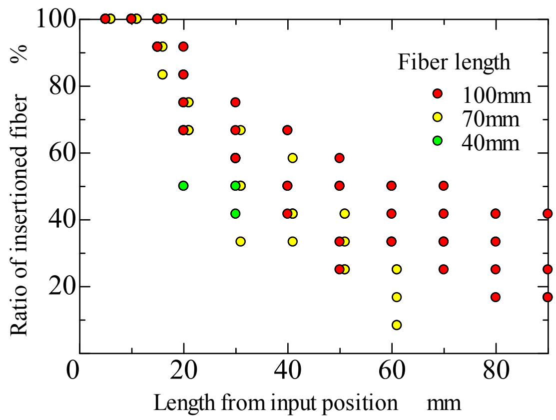


Figure 15: The change of the lining up state by the fiber length

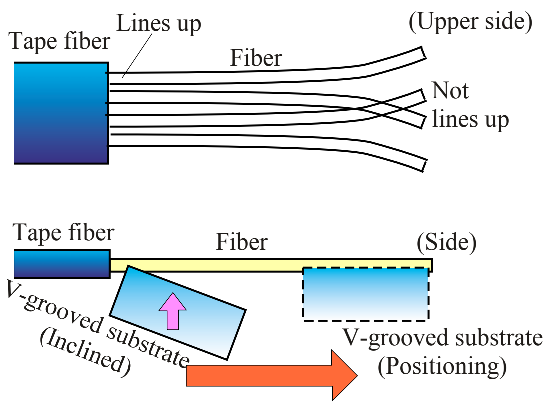


Figure 16: The insertion method

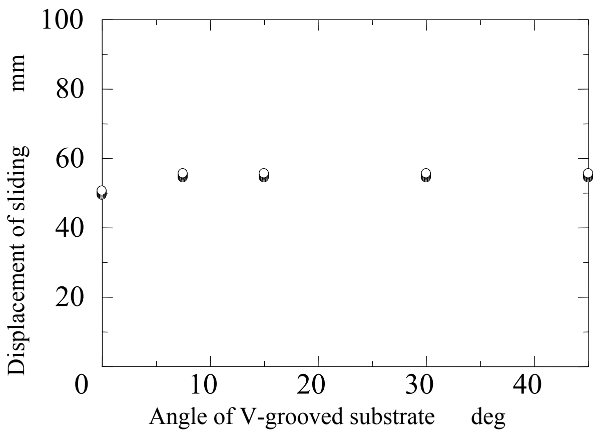


Figure 17: Sliding distance

6 CONCLUSIONS

1) The effectiveness of the method that the V-grooved substrate has an insertion angle and slide to the fiber axial direction for optical fiber array assembling is examined.
2) In the model that considered a fiber to be a cantilever, It is confirmed that the effectiveness of the insertion method to incline the V-grooved substrate.
3) The effectiveness of the method that the V-grooved substrate slides with insertion fiber into the V-groove is confirmed.
4) The maximum length that is enabled insertion and positioning when the fibers have a curve is 50mm.

7 ACKNOWLEDGMENTS

We extend our sincere thanks to the 2007's programs of suzuki foundation contributed to this study.

8 REFERENCES

[1] Tanaka M. Study of Micro-Factory. Science of Machine 2001;53(3):339-7.

[2] Arai F, Andou D, Fukuda T. Micro Manipulation Based on Physical Phenomena in Micro World (1st Report, The Reduction Method of Van Der Waals Force). Transactions of the Japan Society of Mechanical Engineers, Series C 1996; 62(603): 4286-8.

[3] Miyagawa O, N Sornsuwit, Takechi Y, Kawai T, Sawada K, Sata T. Direct Bonding of Small Parts by Means of V-Shaped Microgrooves. The Japan Society for Precision Engineering 1999; 65(7): 987-5.

[4] Sasaki T, Hosaka H, Kondo Y, Yamasaka H, Itao K, Koizumi K. Measurement of Friction Coefficient for Automatic Assemble of Optical Fiber Array. The Japan Society for Precision Engineering 2004; 70(7): 951-5.

Application of the Active Flexible Fixture to a Peg-in-hole Task

Tomomi Yamaguchi[1], Naomichi Furushiro[1], Masahiro Higuchi[1]

[1] Dept. of Mechanical Engineering, Faculty of Engineering Science, Kansai University, Osaka, Japan

Abstract

This paper describes the application of prototype of the AFLEF to a peg-in–hole task. This AFLEF is an active fixture and can fix any hole-work and position the fixed work at short range. We have combined the prototype of it and a 1-DOF inserting device with the RCC device into a new assembly system for such assembly task as a peg-in-hole task. However, there are two problems to realize this task. One is the “incomplete contact condition,” and another is the modification of location of a hole-work to the insertion point. In this paper, in order to solve the former, the AFLEF has been newly equipped with two touch sensors to judge the contact condition and it can modify the contact condition to the complete contact. Moreover, in order to solve the latter, we propose the modification method without any vision sensor. In order to confirm the effectiveness of our proposing modifications of the contact condition and the hole-work’s location, we have tried a peg-in-hole task with the assembly unit. As a result of this trial, although it took a little long time to finish a task, a task has been realized without failure.

Keywords:

Mechatronics; Robotics; Assembly; Fixture; Handling

1 INTRODUCTION

Recently, the automated assembly system needs two flexible functions mainly, i.e. one is the flexible fixing to fix a several types of works and another is the flexible assembly to assemble a work under any deviation caused by the location error or deformation of it. Many fixture systems such as “rigid complex fixture [1],” “reconfigurable assembly fixtures [2]” and “modular fixture [3]” have been developed in order to be adapted to the former function, but, since most of these fixtures are passive, they do not have the latter one. It needs the active fixture system and some active systems have been developed by Chan and Lin [4], Hazen and Wright [5], Grippo *et al*. [6], and Yashima and Kimura [7]. However, these are adaptive surface-fitting type, and they can only fix a work rigidly, except for Yashima and Kimura’s one. Therefore, our goal is the development of “active functional fixture (AFFIX)” having those two functions.

We have started the development of the active flexible fixture (AFLEF) as the first step of the development of AFFIX. The AFLEF can fix any work rigidly and actively by only position control and also position the fixed work at a few millimeters in order to correct the location of fixing point into the assembling point. Thus the AFLEF needs not to be as dexterous as advanced robot-hands but it needs to be more practical than them. We presented the structure and the indispensable functions (fixing and positioning at short range) of the AFLEF on plane level as the prototype [8]. This prototype, hereafter referred to simply as the AFLEF, is composed of four contact fingers, which grip the sides of a work by two driving joints: translating and rotating ones.

This paper describes the application of the AFLEF to a peg-in-hole task. We have combined it and a 1-DOF inserting device with a RCC device into a new assembly unit for the task. In this unit, the inserting device controls a peg only to go down vertically at the insertion point. On the other hand, the AFLEF fixes a polygonal hole-work actively and positions it to the insertion point by position control. Here, there are two problems to realize this task.

One problem is the incomplete plane contact. One of characteristics of the AFLEF is the plane contact between a polygonal work and each contact finger in order to fix the work hardly by large frictional force. However, when four fingers grip a polygonal work, it is very rare that the plane contact is complete at all contacts. Naturally, the condition where the plane contact is incomplete does not exhibit the functions of the AFLEF. In this paper, we present the method to modify the incomplete plane contact to the complete plane contact. Two touch sensors has been newly equipped each finger in order to judge the incomplete plane contact or the complete one. If the contact condition is incomplete, we can modify the contact condition to the complete plane contact by easy control of the driving joints on the basis of an output of each touch sensor, the angle of rotating driving joint and so on.

Another problem is the method of positioning of a hole-work to the insertion point. In this case, the position of work is generally controlled with feedback from a visual position sensor. Naturally, this method can be applied to our assembly system. However, in this paper, we propose the other method taking advantage of the plane contact without vision sensor, because we can find the deviation of the location of the fixed work from the insertion point on the basis of the location of contact point, the length of translating joint, the angle of rotating driving joint and so on.

Moreover, this paper also describes the performance of a peg-in-hole task with the proposal unit that is improved in the abovementioned problems.

2 MECHANISM AND STRUCTURE OF THE AFLEF

The schematic drawing of the AFLEF is shown in Figure 1. It is composed of four contact fingers. Each contact finger comes in contact with the side of a work and grips it.

The schematic drawing of the contact finger is shown in Figure 2. The contact finger has two driving joints: translating and rotating driving joints, in order to move independently on

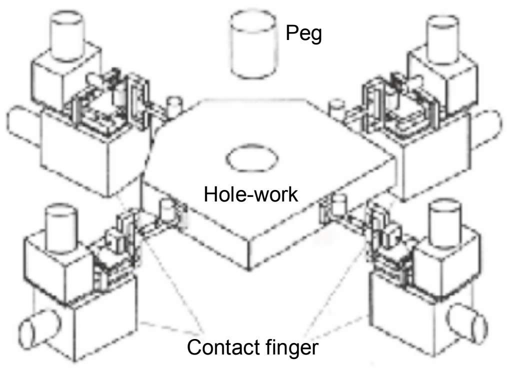


Figure 1: Schematic drawing of the AFLEF.

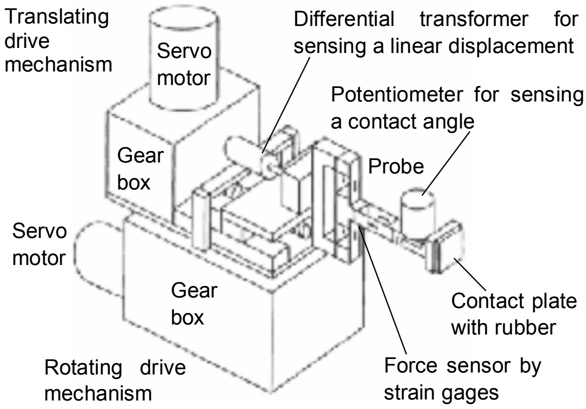


Figure 2: Schematic drawing of the contact finger.

the platform. In the manufactured AFLEF, the movable range of each driving joint is as follows: translating joint: 85 - 105 mm and rotating joint: 0 - 360 degree.

The contact finger is also composed of the probe and the contact plate in addition to the driving joints. The contact plate is equipped with a rubber-slab to cause large friction in the contact point and comes in contact with the side of a work under plane contact condition. Moreover it joined to the probe can rotate freely around the vertical axis. As can be seen from Figure 2, the probe is equipped with the force sensor to measure each contact force in the radial or the angular direction and the potentiometer to measure the angle to the contact plate.

The schematic diagram of position control of the contact finger is shown in Figure 3. The position of translating or rotating driving joint in each contact finger is controlled by inputting the individual reference position data into it from a computer simultaneously and the location of each contact plate is controlled. Each translating driving joint is usually rigid because of the reduction gear, but it can be made elastic by the feedback of a force sensor's signal in addition to a displacement sensor's signal.

In order to exhibit both fixing and positioning at short range functions, two translating driving joints need to be set elastic on the basis of theorem proved by Osumi *et al.* [9]. Moreover, under grip of a work, the contact plane of contact finger with an elastic joint should not be parallel to each other and that with a rigid joint should not also be so. In the manufactured

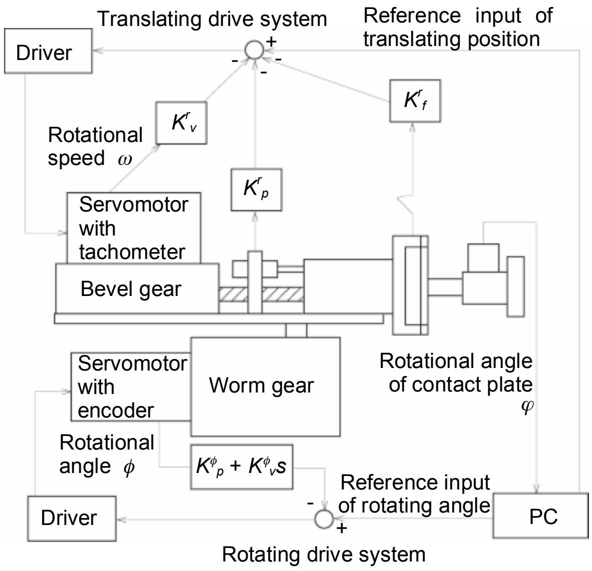


Figure 3: Position control system of the contact-finger.

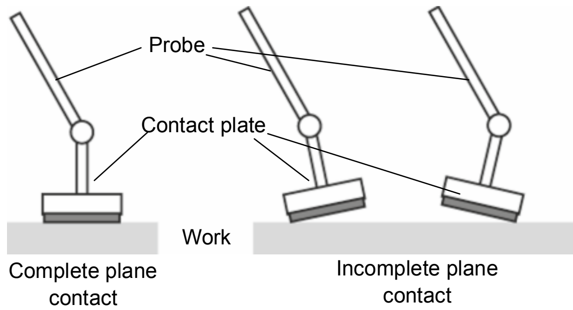


Figure 4: Contact conditions between the contact plate and a work.

AFLEF, the translating joints of contact finger 1 and 2 have been set rigid and that of contact finger 3 and 4 have been set elastic [8].

3 IMPROVEMENT OF THE AFLEF

3.1 Modification of "incomplete contact" to "complete contact"

One of characteristics of the AFLEF is plane contact between a polygonal work and each contact plate in order to fix the work hardly by large frictional force. However, when four fingers with the contact plates grip a polygonal work, in most cases, the contact condition becomes incomplete, that is, the plate keeps in contact with the work at only either end of the plate, not both end of it, as shown in Figure 4. This is caused by the deformation of the rubber-slab of contact plate and the free joint of the plate. Naturally, this incomplete plane contact condition, hereafter referred to simply as the incomplete contact, does not exhibit functions of the AFLEF. Thus, we have improved the AFLEF to modify the incomplete contact to the "complete contact."

First, in order to judge the incomplete contact or the complete one, we have newly equipped the contact plate with two touch sensors at both ends of it, as shown in Figure 5. The touch sensor is a touch switch of push type. Thus the output

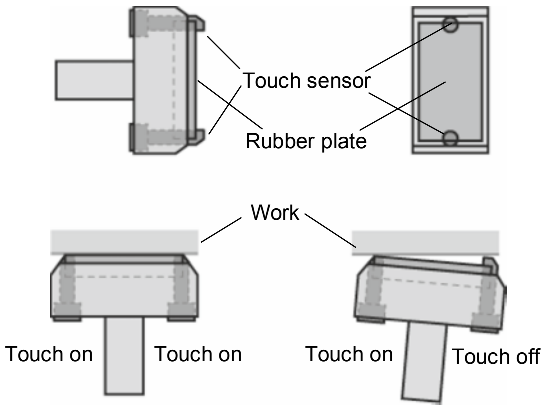
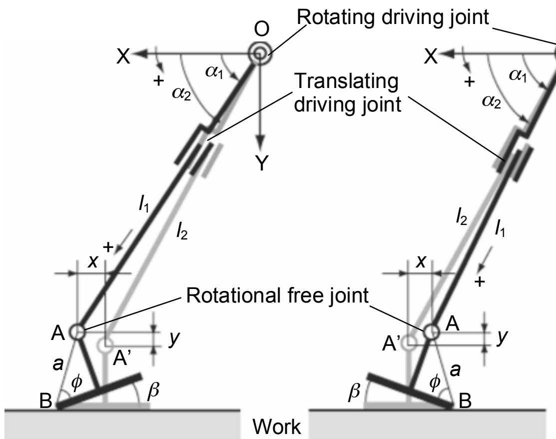


Figure 5: Schematic diagram of the contact plate with touch sensors.



(a) Condition I (b) Condition II

Figure 6: Locations of the contact finger under the incomplete contact and the complete one after the modification.

of sensor keeps ON in contact with a work but it turns to OFF the moment the sensor separates from it. As shown in Figure 5, when only one sensor is ON (another is OFF), the contact condition is judged as the incomplete contact and, when both sensors are ON, it is judged as the complete contact.

Secondly, we have found the method to modify the incomplete contact to the complete contact by the control of translating and rotating driving joints of the contact finger. Figure 6 shows two types of the incomplete or complete conditions between the contact finger and a work. In this figure, the rotating driving joint is set to the origin of the two dimensional Cartesian coordinates system and the line parallel to the surface of a work is also set to X-axis of it. Moreover, the notations used in the figure are as follows: point A is the position of rotational joint of the contact plate under the incomplete contact; point A' is the position of that under the complete contact; point B is the contact point between a work and the end of contact plate; a is the constant distance between point A and point B; l_1 and l_2 are the length of translating driving joint under the incomplete contact and the complete contact, respectively; x and y are the distance between point A and point A' for X-axis and Y-axis, respectively; α_1 and α_2 are the angle of rotating driving joint under the incomplete contact and the complete contact, respectively; β is the angle between the plate of contact plate and a work and ϕ is the constant angle between the contact plate and line AB. As can be seen from Figure 6, in both conditions of (a) and (b), the incomplete contact is modified to the complete contact by positioning of the joint of contact plate from point A to point A' along circular arc AA' with center B. Here, since practical experiment has shown that real value of β is as small as 5 degree in maximum, circular arc AA' can be approximated to the other trajectory such as line AA'. Thus we find the control values of translating and rotating driving joints for the modification on the basis on this fact. In the condition I, x and y are founded by:

$$\begin{cases} x = a\{\cos\phi - \cos(\phi+\beta)\} \approx a\beta\sin\phi \\ y = a\{\sin(\phi+\beta) - \sin\phi\} \approx a\beta\cos\phi \end{cases} \tag{1}$$

Thus, the control value of translating driving joint Δl (= l_1 - l_2) is founded by:

$$\begin{aligned} \Delta l = l_2 - l_1 &= \sqrt{(l_1\cos\alpha_1 - x)^2 + (l_1\sin\alpha_1 + y)^2} - l_1 \\ &\approx \sqrt{l_1^2 + 2l_1 a(\sin\alpha_1\cos\phi - \cos\alpha_1\sin\phi)\beta} - l_1 \\ &\approx l_1\left\{1 + \frac{a\sin(\alpha_1-\phi)}{l_1}\beta\right\} - l_1 \\ &= \frac{a\sin(\alpha_1-\phi)}{l_1}\beta \end{aligned} \tag{2}$$

Similarly, the control value of rotating driving joint $\Delta\alpha$ (= α_1 - α_2) is founded by:

$$\begin{aligned} \Delta\alpha = \alpha_2 - \alpha_1 &= \tan^{-1}\frac{l_1\sin\alpha_1 + y}{l_1\cos\alpha_1 - x} - \alpha_1 \\ &\approx \tan^{-1}\frac{l_1\sin\alpha_1 + a\cos\phi\cdot\beta}{l_1\cos\alpha_1 - a\sin\phi\cdot\beta} - \alpha_1 \\ &\approx \alpha_1 + \frac{l_1 a(\cos\alpha_1\cos\phi + \sin\alpha_1\sin\phi)}{(1+\tan^2\alpha_1)(l_1^2\cos^2\alpha_1)}\cdot\beta - \alpha_1 \\ &= \frac{a\cos(\alpha_1-\phi)}{l_1}\beta \end{aligned} \tag{3}$$

On the other hand, in the condition II, the control values of both driving joints: Δl and $\Delta\alpha$ can be founded by similar process, that is,

$$\begin{aligned} \Delta l &= \sqrt{(l_1\cos\alpha_1 + x)^2 + (l_1\sin\alpha_1 + y)^2} - l_1 \\ &\approx \frac{a\sin(\alpha_1+\phi)}{l_1}\beta \\ \Delta\alpha &= \tan^{-1}\frac{l_1\sin\alpha_1 + y}{l_1\cos\alpha_1 + x} - \alpha_1 \approx \frac{a\cos(\alpha_1+\phi)}{l_1}\beta \end{aligned} \tag{4}$$

As can be seen from Equation (2) to Equation (4), approximately, each control value is directly proportional to the angle β because the other notions: a, l_1, α_1 and ϕ are known. However, since the angle β is unknown, a specific control value cannot be founded directly. Therefore, after judging the condition I or II, we have set control value to β = 0.3 degree to each driving joint as one unit and have modified the contact condition by inputting every unit to both driving joints simultaneously until the outputs of both touch sensors turn to ON.

3.2 Modification of location of a fixed work to insertion point

In the AFLEF, each contact finger is controlled to come contact with the edge of a work at the reference point given beforehand. However, the real grip point usually deviates from the reference one and the grip location of a work also deviates from the initial location. Hence, the AFLEF has the function of positioning at short range but no function of finding a control value to modify a work to the assembly location. It is popular that this function is realized by the control with feedback from a vision sensor. Naturally, this method can be applied to our assembly system. However, we propose another method taking advantage of the plane contact as a characteristic of the AFLEF in this paper. That is, our proposal method needs no vision sensor.

This modification can be applied on the following conditions: a work is polygon on plane level; its shape and size are known; the insertion point is given; the location (position and attitude) of work set at the insertion point is given and the edge of work in contact with each contact finger is given. Figure 7 shows an example of relationship between the location of a work at the insertion point and that at a real fixing point. Here, the contact finger 3 and 4 are left out. In this figure, the Cartesian coordinates system *o-xy* set on a work stand and the insertion point is set to the origin. Moreover, the notations used in the figure are as follows: l^i and α^i are the length of translating driving joint and the angle of rotating driving joint of contact finger i (i = 1, 2), respectively; ψ^i is the angle between the contact plate and the probe in contact finger i (i = 1, 2); r_0 is the constant length between the center of plate and the rotational joint in the contact plate; (X^i, Y^i) is the coordinates to the rotating driving joint of contact finger i (i = 1, 2) and (x_c^1, y_c^1) and (x_{c0}^1, y_{c0}^1) are the coordinates to the center of the plate of the contact plate of contact finger 1 under a real fixing of a work and the fixing of it at the insertion point, respectively. Furthermore, L_0^i shows the line including the edge of a work in contact with the contact finger i (i = 1, 2) under the fixing at the insertion point; L^i shows the line including the edge of a work in contact with the contact finger i (i = 1, 2) under a real fixing; $P_0(a_0, b_0)$ and $P(a, b)$ are the intersection and its coordinates of the lines L_0^1 and L_0^2 and the lines L^1 and L^2, respectively and φ_0^1 is the angle of lines L_0^1 to the x-axis.

Since the translating joints of contact finger 1 and 2 have been set rigid, the line L_0^1 is not parallel to the line L_0^2. Moreover, since the equations of the lines L_0^1 and L_0^2 can be determined from the abovementioned preconditions, the coordinates (a_0, b_0) and the angle φ_0^1 can be founded from these equations. On the other hand, the equation of the line L^i (i = 1, 2) can be expressed as:

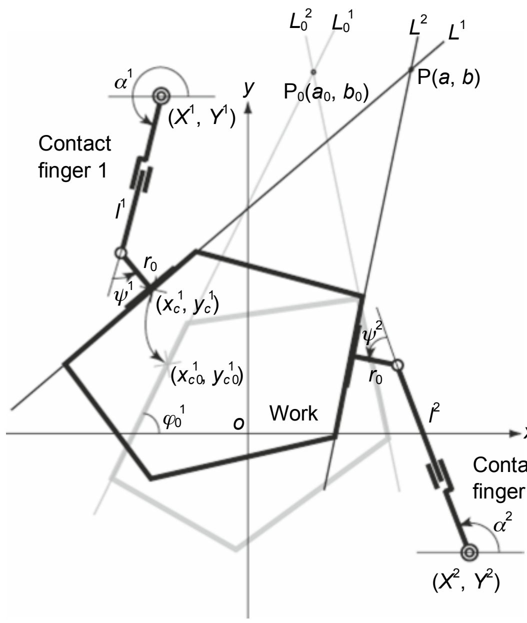


Figure 7: An example of relationship between4the location of a work at the insertion point and that at a real fixing point.

$$
\begin{aligned}
& x+\tan(\alpha^i+\psi^i)\cdot y \\
& = \tan(\alpha^i+\psi^i)\{l^i\sin\alpha^i + r_0\sin(\alpha^i+\psi^i)+Y^i\} \\
& -\{l^i\cos\alpha^i + r_0\cos(\alpha^i+\psi^i)+X^i\} \quad (i=1,\ 2)
\end{aligned}
\tag{5}
$$

Thus the coordinates (a, b) can be founded from Equation (5). The intersection and either slope of a couple of lines (L_0^1, L_0^2) and (L^1, L^2) correspond to the location of work at the insertion point and a real fixing one, respectively. Here, the transfer of a work from a real fixing point to the insertion one is considered as translation after rotation about the origin. Thus this transfer can be expressed as:

$$
\begin{aligned}
\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} &= \begin{pmatrix} \cos\left(\Psi^1+\dfrac{\pi}{2}\right) & -\sin\left(\Psi^1+\dfrac{\pi}{2}\right) \\ \sin\left(\Psi^1+\dfrac{\pi}{2}\right) & \cos\left(\Psi^1+\dfrac{\pi}{2}\right) \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \xi \\ \varsigma \end{pmatrix} \\
&= \begin{pmatrix} \sin\Psi^1 & \cos\Psi^1 \\ -\cos\Psi^1 & \sin\Psi^1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \xi \\ \varsigma \end{pmatrix}
\end{aligned}
\tag{6}
$$

where $\Psi^1 = \varphi_0^1 - (\alpha^1 + \psi^1)$; ${}^t(x \;\; y)$ is the coordinates to any point on a work at the real fixing point; ${}^t(x_0 \;\; y_0)$ is the coordinates to the point transferred ${}^t(x \;\; y)$ to and ${}^t(\xi \;\; \zeta)$ is given by:

$$
\begin{pmatrix} \xi \\ \varsigma \end{pmatrix} = \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} - \begin{pmatrix} \sin\Psi^1 & \cos\Psi^1 \\ -\cos\Psi^1 & \sin\Psi^1 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}
\tag{7}
$$

On the other hand, the coordinates of contact point with the contact finger 1 (x_c^1, y_c^1) can be determined as:

$$\begin{cases} x_c^1 = l^1 \cos\alpha^1 + r_0 \cos(\alpha^1 + \psi^1) + X^1 \\ y_c^1 = l^1 \sin\alpha^1 + r_0 \sin(\alpha^1 + \psi^1) + Y^1 \end{cases} \tag{8}$$

Thus the coordinates of contact point at the insertion point (x_{c0}^1, y_{c0}^1) can be founded from Equations (6), (7) and (8) as:

$$\begin{pmatrix} x_{c0}^1 \\ y_{c0}^1 \end{pmatrix} = \begin{pmatrix} \sin\Psi^1 & \cos\Psi^1 \\ -\cos\Psi^1 & \sin\Psi^1 \end{pmatrix} \begin{pmatrix} x_c^1 \\ y_c^1 \end{pmatrix} + \begin{pmatrix} \xi \\ \varsigma \end{pmatrix} \tag{9}$$

In the other contact fingers, similarly the coordinates of contact point at the insertion point can be founded. Therefore, the position of a fixed work can be modified to the insertion point by inputting each control value calculated from these coordinates to translating and rotating driving joints in each contact finger.

4 ASSEMBLY EXPERIMENT

We experimented with our new assembly unit in a peg-in-hole task. This unit is composed of the AFLEF and a 1-DOF inserting device. Naturally the AFLEF fixes and position a hole-work. On the other hand, the inserting device controls a peg only to go down vertically at the speed of 35 mm/sec.. The peg was a steel cylinder having a diameter of 19.98 mm and a height of 35 mm. The hole-work was an aluminum rectangular parallelepiped whose size was 70 by 70 by 30 mm and its weight was 1.63 N. The hole having a diameter of 20 mm, a depth of 25 mm and a chamfer of 1 mm was drilled in the upper face of the work. Here, the inserting device is equipped with a RCC device because the clearance between the peg and the hole (0.02 mm) is very smaller than the positioning accuracy of the AFLEF (±0.3 mm).

Figure 9 shows phases of fixing conditions of a hole-work with the AFLEF in a peg-in-hole task. First, as the initial condition, a hole-work was set in condition where the center of hole corresponds to the origin of the coordinates system on a work stand and each side of the work was parallel to the x- or y-axis. Secondly, the AFLEF gripped the work. Here, contact condition is incomplete in most cases. Thirdly, the contact condition was modified to the complete contact to fix the work by the method mentioned in Chapter 3.1. Fourthly, the location of the fixed work was modified to that on the insertion point by the method mentioned in Chapter 3.2. Lastly, the peg was inserted into the hole with the inserting device.

We tried this sequence of peg-in-hole. Figure 10 shows the changes of translational and rotatory angular displacements of hole-work and reaction force of each contact finger for the sequence of peg-in-hole task. Here, the former changes (a) are shown as the displacements to the *x*-, *y*- and θ-axes directions measured with a CCD camera (resolution: 0.03 mm/pix.) and the latter change (b) is done as the force measured with force sensor equipped each contact finger with.

5 DISCUSSIONS

The AFLEF finishes the sequence of task with keeping contact with the hole-work, i.e. without failure, as can be seen from Figure 10 (b), because each reaction force does not

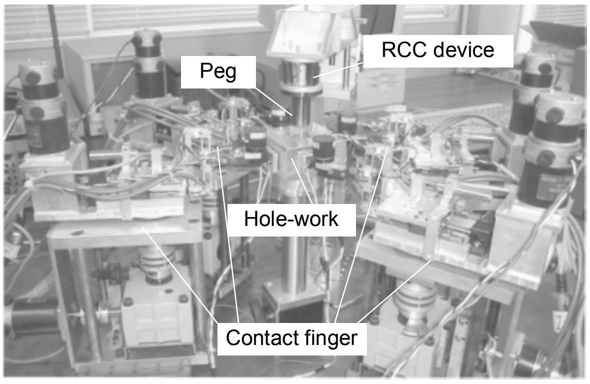
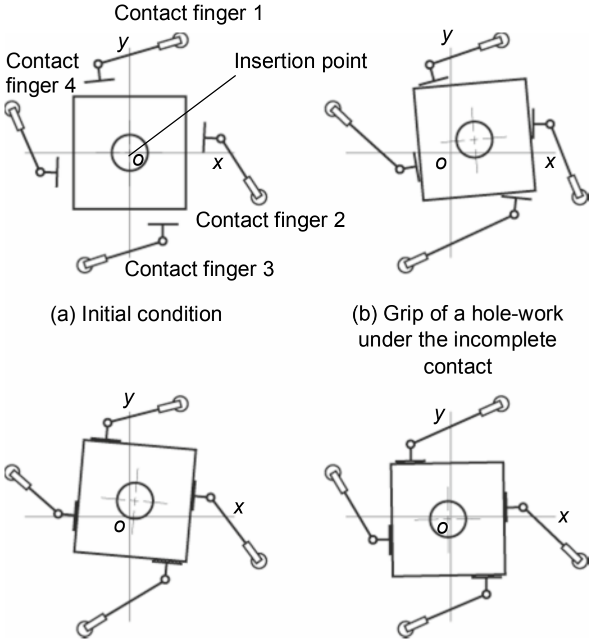


Figure 8: Photograph of assembly unit with the AFLEF for a peg-in-hole task.



(a) Initial condition

(b) Grip of a hole-work under the incomplete contact

(c) Fixing of a hole-work under the complete contact

(d) Modification of the location of a hole-work to insertion point

Figure 9: Phases of fixing conditions of a hole-work with the AFLEF in a peg-in-hole task.

become zero. However, the final displacements to each direction are as follows: 0.5 mm to the *x*-axis direction, 0.7 mm to the *y*-axis direction and 1.2 degree to the θ-axis direction. The Accuracy of positioning on each direction at ±3.0 mm or ±3.0 degree was ±0.3 mm or ±0.3 degree in the previous experiment in positioning at short range [8]. The decline of accuracy is considered to be caused by the modification of location of a hole-work to insertion point with low accuracy, and this low accurate modification is caused by the estimation of real fixing point with low accuracy because the accuracy of potentiometer as the angle sensor measuring the angle of contact plate is low, i.e. the accuracy is ±1 degree. Thus, if the accuracy of potentiometer is more than ±0.05 degree, the location of the work at the insertion point can be estimated to be modified within ±0.3 mm in the *x*- and *y*-axis directions and ±0.3 degree in the θ-axis direction.

On the other hand, the modification of contact condition succeeds but it takes a little long time to finish that. The time needs to be shortened for practical modification level.

6 CONCLUSIONS

We have combined the AFLEF and a 1-DOF inserting device with the RCC device into a new assembly system for a peg-in-hole task. However, to realize this task, there are two problems. One is the incomplete contact, and another is the modification of location of a hole-work to the insertion point. In this paper, to solve the former, each contact finger of the AFLEF has been newly equipped with two touch sensors in order to judge the incomplete contact or the complete contact and we propose the control method to modify the contact condition to the complete contact. Moreover, to solve the latter, we propose the modification method without any vision sensor by taking advantage of the plane contact as one of characteristics of the AFLEF. In order to confirm the effectiveness of our proposing modifications, we have tried a peg-in-hole task with the assembly unit. As a result of this trial, the task has been realized without failure. However, in order to realize the task with the AFLEF on practical level, it becomes clear that higher accurate modification of location and faster modification of contact condition.

REFERENCES

[1] Xiong, YL., Xiong, XR., 2007, Algebraic Structure and Geometric Interpretation of Rigid Complex Fixture Systems, IEEE Transactions on Automation Science and Engineering, 4/2:252-264.

[2] Kong, Z., Ceglarek, D., 2003, Rapid Deployment of Reconfigurable Assembly Fixtures Using Workspace Synthesis and Visibility Analysis, CIRP Annals-Manufacturing Technology, 52/1:13-16.

[3] Mervyn, F., Kumar, AS., Nee, AYC., 2005, Automated synthesis of modular fixture designs using an evolutionary search algorithm, International Journal of Production Research, 43/23:5047-5070.

[4] Chan, KC., Lin, CS., 1996, Development of a Computer Numerical Control (CNC) Modular Fixture - Machine Design of a Standard Multifinger Module, International Journal of Advanced Manufacturing Technology, 11/1:18-26.

[5] Hazen, FB., Wright, PK., 1990, Workholding Automation: Innovations in Analysis, Design and Planning, Manufacturing Review, 3/4:224-236.

[6] Grippo, PM., Thompson, BS., Gandhi, MV., 1988, A Review of Flexible Fixture Systems for Computer-integrated Manufacturing, International Journal of Computer-Integrated Manufacturing, 1/2:124-135.

[7] Kimura, H., Yashima, M., 1994, Intelligent Jig with Function of Manipulation –1st Report, Basic Theory for Manipulation, Journal of the JSME, Series C, 60/575:2248-2254 (In Japanese).

[8] Yamaguchi, T., Higuchi, M., Nagai, K., 2004, Assembly System by Using Prototype of Active Flexible Fixture, Proceedings of the 5th International Conference on Machine Automation (ICMA2004), 149-154.

[9] Osumi, H., Arai, T., 1994, A Method for Introducing Industrial Robots to Cooperative Tasks, Journal of the Robotics Society of Japan, 12/8:1192-1197 (In Japanese).

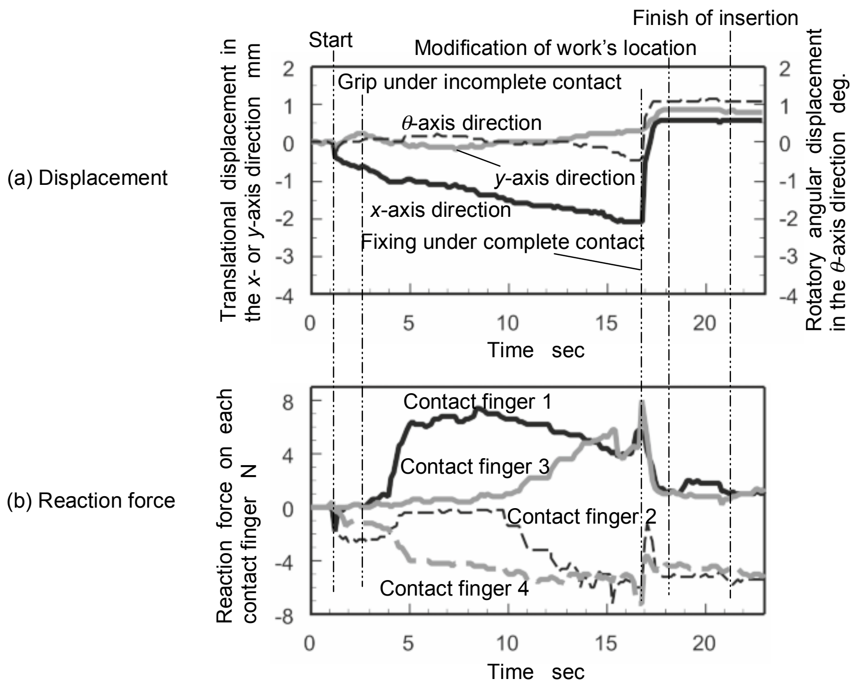


Figure 10: Displacement in each axis direction and reaction force on each contact finger for a peg-in-hole task by using new assembly unit with the AFLEF.

Comparison of Tracking Controllers of Hydraulic Cylinder Drive by Simulations

Tapio Virvalo

Department of Intelligent Hydraulics and Automation, Tampere University of Technology, Tampere, Finland

Abstract

When mechanical cams are replaced with hydraulic drives in relatively fast applications the dynamics of hydraulic drives is very important. From the tracking error point of view the delay between the reference and the load motion is the most serious problem. Due to large and expensive real systems a real animation system and simulations are used in this study. The model of the animation system is validated and simulation is used to study the performance of different controllers. The best results are achieved with the State Feedback Controller, velocity forward loop, and the delay compensation.

Keywords:

Hydraulic servo, racking control, lead time compensation

1 INTRODUCTION

There are some applications where good programmable position tracking capability is important. One application field is where mechanical cam tracking systems are replaced with some kinds of servo drives. Such applications are for instance short serial production by milling using mechanical model and follow-up system. Both electro-mechanical and electro-hydraulic methods have been studied by simulations and experiments in replacing the mechanical cams in machine tools and combustion engine applications [1], [2], [3], and [4]. The application behind this study is replacing cam shaft of gas exchange valves in big Diesel engines with hydraulic servo systems. There are studies of hydraulic applications where on/off-valves are used to control opening and closing of gas exchange valves [5], [6]. In these cases the strokes length and opening and closing timing of gas exchange valves can be controlled. However then the basic working principle is to open and close gas exchange valves with the timing of the control signal not to replace the real mechanical cams. The real replace of mechanical cams with hydraulic servo systems have been studied, for instance in [7], [8].

2 MOTIVATION

In practice it is very difficult, time consuming and expensive to make tests with real applications, with big and expensive machine tools or diesel engines. Simulation is often a powerful tool to find out influences of basic design and parameters on performance of systems. When an application is new and somehow close to the performance limits of considered system, some support to validate simulations is needed. One possibility might be, in certain cases, to support simulation with real time animation with somehow scaled real system.

The well-verified model of hydraulic cylinder and servo valve combination of previous studies is used as basic model in simulations [9]. The structure of the model is kept mainly the same, but there are some uncertainties of parameters. That is why experimental animations of a real application are made. The basic idea is to arrange the laboratory tests with normal commercial components, but defining the performance specifications and ratio of the servo valve and drive dynamics the same as in the real applications. In most hydraulic servo applications the dynamics of servo valves is remarkably higher than cylinder drives. This situation has been often also assumed in design processes. However, quite often it is difficult to estimate application date because authors do not tell the dynamics of actuators and servo valves. In this case both actuator and servo valve dynamics are estimated and the values are verified with basic experimental tests.

3 SPECIFICATION OF SYSTEMS

The basic idea is to replace traditional mechanical cam shaft driving systems with a hydraulic servo system. The motion realized with the cam and spring system is replaced with a hydraulic cylinder. In order to fully replace the cam produced motion the hydraulic cylinder is controlled by a servo valve and closed loop controller. Quite often it is considered that hydraulic systems are strongly non-linear. It is partly true. They are really non-linear if compared to electric drive systems. However, the linearized model of hydraulic cylinder drive with some experience based knowledge of parameters has proved to be useful in performance analyzing and design of controllers [10].

The third order model is used in position and motion control applications. There are three system parameters in the model; natural frequency, relative damping factor, and velocity gain. The effective bulk modulus is the most uncertain factor in the estimation of the natural frequency according to the basic formula [11]. In practice much lower value than of the basic oil has to be used. A practical estimation of the effective bulk modulus is 1000 MPa (basic oil has around 1500 MPa). Other parameters influencing estimation of the natural frequency is estimated quite straight forward. In this special case the minimum natural frequency of the animated system is estimated to be around 900 - 950 rad/s.

The choice of the damping factor is also based on some experience. There are estimation formulas for the damping factor [11]. Experimental tests show that there are three main relatively difficult estimated parameters influencing on the damping factor [12]. A good guess is 0.15 in cases where is no external damping elements.

The maximum velocity of the animated system with the full stroke is estimated to be around 3.5 m/s. This means according to preliminary estimations that the nominal volume flow of the servo valve should be around 70 l/min, when supply pressure is 250 bar. However, the practical commercial choice seems to be the valve with the nominal volume flow 100 l/min. The real velocity gain depends on the flow rate of the servo valve but also the pressure drops in the control notches of the valve. However the velocity gain used in the linear model is based on the maximum velocity and the maximum control signal of the servo valve. The velocity gain is in this case 0.35 m/Vs.

Some cases the closeness of the cylinder natural frequency and the bandwidth of the servo valve might have the significant effect on dynamic performance of whole system. In this special case the bandwidth of the servo valve is around 1500 rad/s. It has to be noticed that the dynamics of this size of the valve one of the highest on the market.

In order to animate the system in the laboratory the following test system is used. The available servo valve and cylinder have the following specifications.

- Nominal volume flow 40 l/min
- Nominal pressure drop/control notch 35 bar
- 90^o bandwidth 500 rad/s with 10% control signal
- Cylinder 32/20-1000 (piston/rod-stroke [mm])

The main idea in the design of the animation system is to make the animating system dynamically and from control point of view as demanding as the animated real case. Because the dynamics of the available servo valve is one third of the animated system servo valve the dynamics of the cylinder drive is designed to be one third, too. The natural frequency should be 330 rad/s. This means that the inertia load should be 20 kg.

Following these ideas the time for the back-and-forth motion should be 50 ms with the stroke of 16 mm. The maximum velocity is around 1.2 m/s. All above estimated parameter values lead to the supply pressure 100 bar.

4 MODEL VALIDATION

The following experimental set-up is made based on estimated parameters in previous chapter and according to components exciting in the laboratory.

- Cylinder 32/20-1000 (piston/rod-stroke [mm])
- Inertia load 20 kg
- Supply pressure 100 bar
- Servo valve
  * Nominal volume flow 40 l/min
  * Nominal pressure drop/notch 35 bar
  * -90^o bandwidth 500 rad/s
* Position sensor
  * Incremental encoder, resolution 0.0025 mm
* Controller
  * PC with dedicated DSP-card

The basic hydraulics is modeled according to principles presented in [1]. This modeling includes the following parts.

- Dynamics of cylinder and load
- Friction forces modeled based on [13]
- Servo valve
  * Static characteristics including dead band and hysteresis [14] and [15]
  * Dynamics with spool velocity saturation
* Pipes and hoses
* Bulk module and reduced mass
* Controller
  * Discrete realization
  * Sampling and executing times of control algorithm

Open loop measurements are made to validate the estimated natural frequency and damping. As an example the velocity responses from both directions with the same strokes and the end point are shown in Figures 1 and 2.

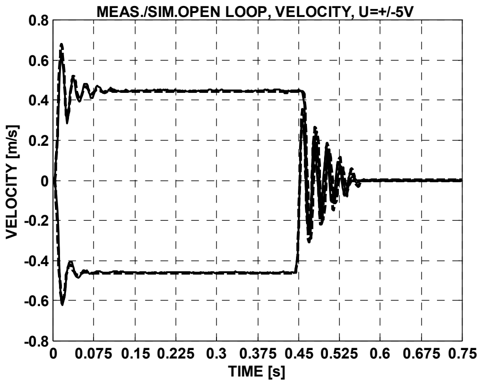


Figure 1. Open loop velocity step responses.

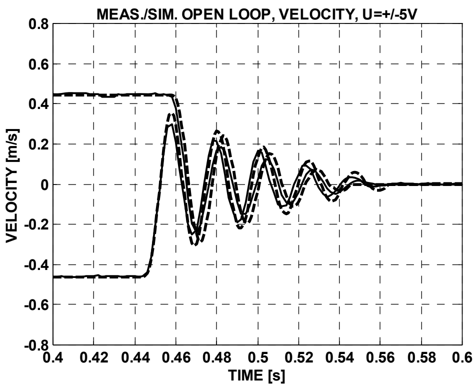


Figure 2. Zoomed velocity step responses at end of stroke.

The natural frequency and damping match well to the estimated natural frequency and damping. The influence of the cylinder asymmetry on the velocity is compensated by the piston area ration scaled control signal of the valve. The same principle is used in the velocity gain and valve control signal saturation in the closed loop case. This way also the open loop gain is the same in both directions. It also means that the position loop gains can be the same in both movement directions.

Slight tuning has to be made in simulation process. The simulated rise time matches well with measured one, if the supply pressure is 100 bar. The nominal flow rate in the manufacturer's specifications of the servo valve is 40 l/min with 35 bar pressure drop in the control notch. In the model a little bit higher value has to be used in this case. The suitable value for different control signals is 42 l/min, which is inside the tolerances (±10%) inform typically for servo valves. Another a little bit uncertain parameter is the effective bulk modulus which influences the natural frequency. It can be figured that the bulk modulus of the normal hydraulic oil is 1500 MPa and typically in preliminary estimations the value 1000 MPa is used, but in this case the suitable value seems to be 1350 MPa. The friction forces influence, among other things, the damping. The following fixed values are used in this case: static friction force 400 N, Coulomb friction force 250 N, and viscous friction coefficient 150 Ns/m.

As an example, the position error and velocity step responses from both directions to the minimum natural frequency point are shown in Figures 3 and 4, when P-controller is used. The open loop gain is 30 1/s. The tuning is based on the responses without overshoots. The same parameter values are used as in the open loop responses.

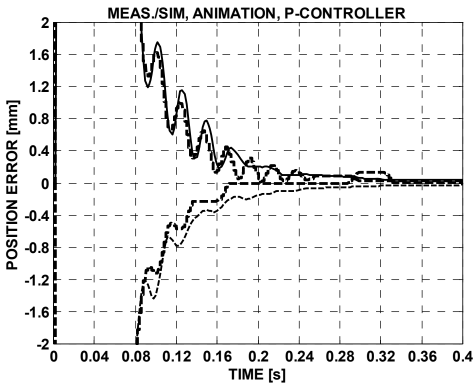


Figure 3. Zoomed position error responses in both directions, P-controller.

The responses are typical for P-controlled position servo systems. Due to low natural damping step responses oscillate which can easily be seen in velocity responses. The results of the comparisons between simulated and experimental responses show that the model is good enough to be used in the design and tests of different controllers.

5 EXPECTED PROBLEMS

Comparing the dynamics of the hydraulic system the required dynamics of responses is very high. The required position profile is shown in Figure 5

The position profile in Figure 5 can be considered as one whole cycle of sin-wave with the frequency of 20 Hz. When the system natural frequency is 50 Hz, there should not be any big problem to track the reference signal accurately enough. In practice the reference signal is not a sin-wave but the profile repeated in a certain intervals. In practice there is a significant delay at the beginning of the position response as can be seen in Figure 6. There are the zoomed initial stages of the step responses of 16 mm to the point of 600 mm from both directions. The important feature is the significant delays of responses, around 8 ms.

The reference signal is smoother in the application and so expected delay is bigger.

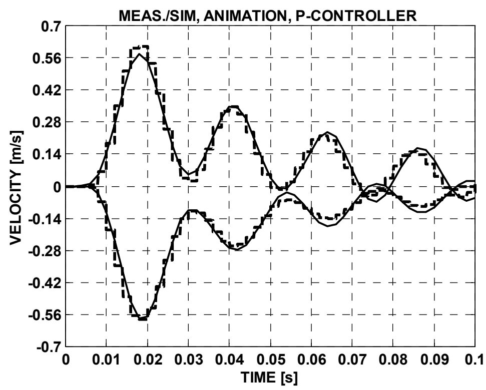


Figure 4. Zoomed velocity responses in both direction, P-controller.

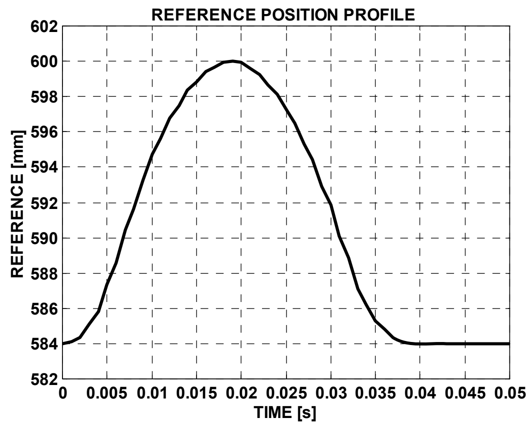


Figure 5. Position profile.

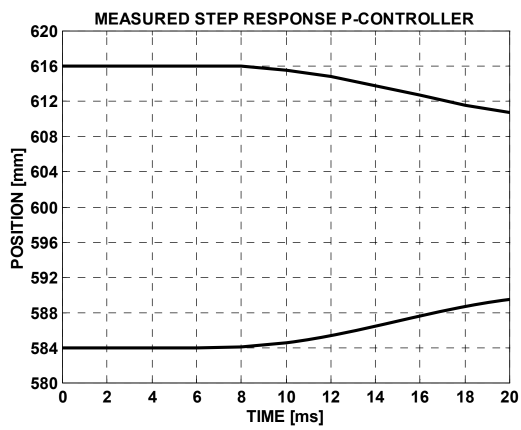


Figure 6. Zoomed position responses.

6 CONTROLLER DESIGN

Because the main goal is to study possibilities to realize hydraulic drive in these kinds of applications the control system has to fulfill some features:

- It should be as simple as possible
- It should have as few tuning parameters as possible
- It should be robust against variations of conditions and parameter changes
- It should have as few sensors as possible
- Dynamics of servo valve and drive could be close
- Commercial components should have been used as much as possible

The P-controller is a simple useful controller, especially in position control systems. It is also relatively robust when it is tuned suitably. However, in fluid power applications, where damping is typically quite low there are often problems in matching the dynamics, response time, and positioning accuracy of control systems. When fast responses, high positioning and tracking accuracy as well as smooth dynamics are required the P-controller is not good enough. For instance in this case the settling time of the step response is around 0.2 s with P-controller mostly due to low damping of the drive. In practice the State Feedback Controller has proved to be a good solution on hydraulic cylinder drives when high performance is required and the parameters of an application do not change much [16] and [17]. In this special application only position sensor might be used. The required stroke is short and available space is small, but quite high dynamics is necessary. The specification of the positioning accuracy is not extra high but the reliability and life time should be high. All these mean that some non-contacting analog position sensor is a good choice into the real application.

The following controllers are tested in the animation system by simulation:

- State Feedback position controller with forward loops.
- Cascaded State Feedback position and velocity controllers.
- Motion controller, combination of State Feedback position and velocity controllers.
- Computed force controller.
- State Feedback position controller with forward loops and lead learning compensation.

The State Feedback position and velocity controllers are first design separately with the linear third order model using the pole placement method. Then the tuning is improved by the simulation using well validated non-linear models. The tuning is based on step responses. The position control is tuned so that the position error responses are fast with small overshoots and oscillations. The velocity control is tuned with different velocity reference values so that the velocity step responses are fast and with small overshoots and oscillations.

In the cascaded and motion control methods the tuned State Feedback position and velocity controllers are used with the gains achieved in their tuning. The velocity forward loop gains are based on the velocity gain. The gain of the velocity forward loop is fine tuned so that the tracking error is small with small oscillations.

It was noticed, that the cascaded position and velocity controller does not work properly. There are two main problems when a linear integrating term is used in hydraulic cylinder drive positioning applications. It is very difficult to avoid overshoots and low frequency hunting (traveling). The same phenomenon has been found also in many other cases [18].

The motion control system which is based on both position and velocity controllers and some switching strategy between them does not work well. This kind of controller is very good in positioning type application where smooth motion and good tracking accuracy are desired [19], [20].

In practical applications of the computed force controller of hydraulic cylinder drives pressure controllers are needed. Two 3-way servo valves are also required for high performance realization [21]. This controller solution requires many sensors and an extra valve and on the other hand simulations show some problems with oscillations and the settling time.

Because of low damping and bad non-linearities the best performance is achieved with the position State Feedback Controller and the velocity feed forward loop, Figure 7. It seems that the acceleration forward loop does not improve the tracking control at least in this case.

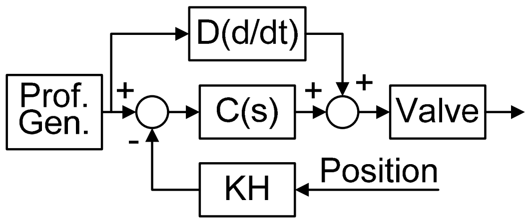


Figure 7. Position control with velocity forward loop.

7 RESULTS

The validated non-linear model is used in simulations. In simulations used basic controller is the State Feedback position controller with the feed forward velocity loop. The position profile shown in Figure 5 is used as a reference signal in all simulations. The position response to one reference cycle is shown in Figure 8. The basic problem from tracking error point of view can be seen in that Figure 8. The shapes of the responses match quite well to each other but the phase shift is remarkable. The delay due to the dynamics of the servo valve and cylinder drive is big; around 8 ms. The delay is around the same as in the responses in Figure 5 in spite of the higher open loop gain. The open loop gain of the State Feedback controlled case can be tuned around three times higher than with the P-controller (Figure 3) due to higher damping. The achieved settling time of the step-wise 16 mm stroke is 25 ms compared with 200 ms of the P-controlled system.

Tracking error is big due to system dynamics which limits significantly the controller gains, Figure 9. Because the tracking performance cannot be improved by tuning the controller gains one way is to improve tracking accuracy is to modify the reference signal.

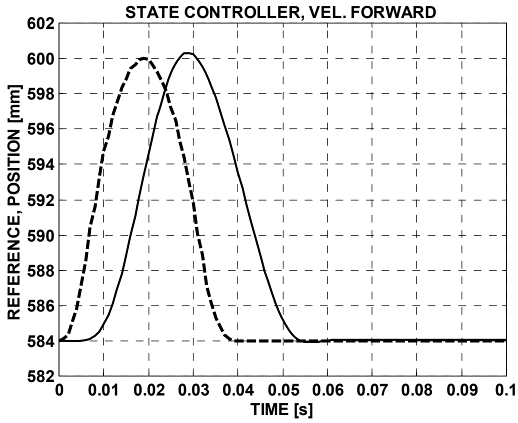


Figure 8. Reference (dashed) and position responses.

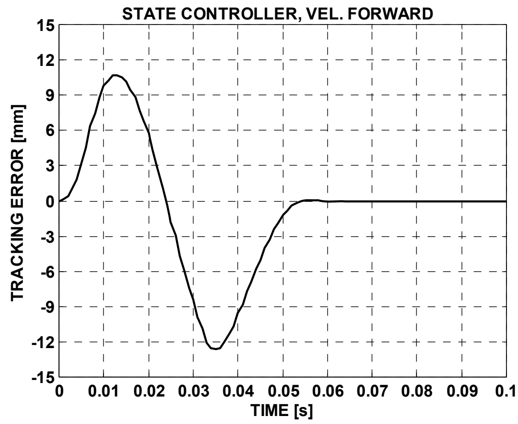


Figure 9. Tracking error.

8 ITERATIVE LEARNING CONTROL

When the reference signal is cyclic different methods have been studied in order to improve tracking accuracy. Both the Iterative Learning Control and Repetitive Control have been used many studies [22], [23]. Adaptive control has also applied in some these studies [24], [25].

Because the delay at the beginning of the response is almost independent on the shape of the reference signal its influence on the tracking accuracy has to be eliminated in some other way. If the frequency and amplitude of the reference are constant the lead time to the reference can be found or trained. In this case the learning process is studied in order to adjust the lead time during initial phase. The response of tracking error is shown in Figure 10 during the learning process.

During the learning process the tracking errors are measured at the certain stages of the reference and then scaled to the lead time of the reference. When the scaling is at reasonable level the system stays stable. The lead time improves the tracking accuracy, but because the tracking error at the rising part of the response is used the error remains quite big at the downward part of the response.

Applying the Iterative Learning Control method further improvements of the tracking accuracy can be achieved as can be seen in Figure 11. As an example of the design parameter changing can be seen in Figure 12, when the size of the servo valve is changed from 40 l/min to 50 l/min.

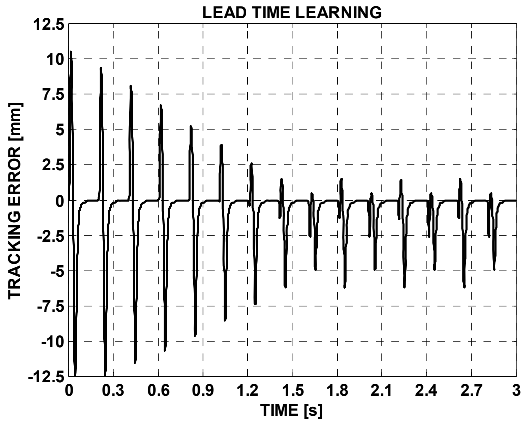


Figure 10. Tracking error, lead time compensation.

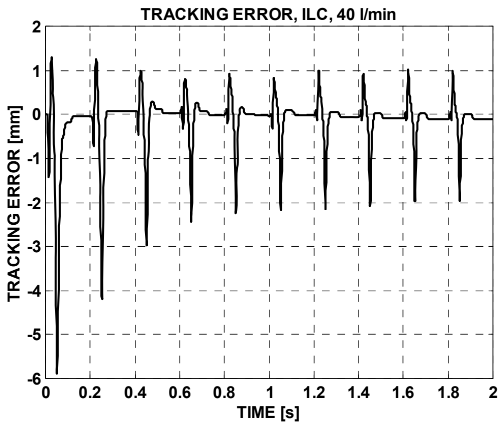


Figure 11. Tracking error, Iterative Learning, Q=40 l/min.

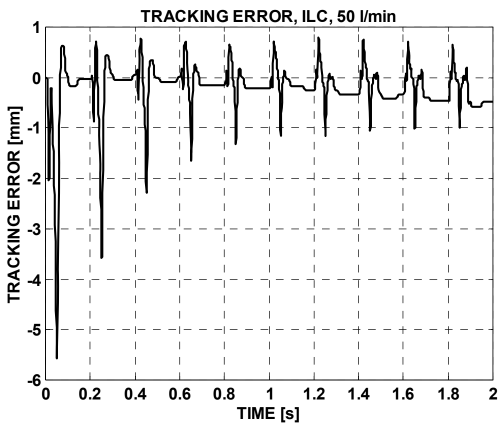


Figure 12. Tracking error, Iterative Learning, Q=50 l/min.

9 CONCLUSIONS

The following conclusions can be made based on this study

- The State Feedback Controller with the velocity feed forward loop is the best controller from dynamics and tracking accuracy point of view
- The response time delay has the biggest influence on the tracking accuracy in fast response hydraulic cylinder drive.
- The lead time of the reference helps remarkably the tracking accuracy.
- The combination of lead time compensation and the use of the Iterative Learning Control improve the tracking accuracy significantly.
- The right size of the servo valve is important.
- The dynamics of the servo valve should be at least three times higher that the dynamics of the cylinder.

10 REFERENCES

[1] Kim H. and Tsao T-C. 1998. Identification and Control of Electrohydraulic Actuator Modeled as a Linear Periodic System. Proceedings of the American Control Conference, Philadelphia, Pennsylvania, June, 1998, pp. 823-927.

[2] Tai C. and Tsao T-C. 2002. Control of Electromechanical Camless Valve Actuator. Proceedings of the American Control Conference, Anchorage, AK, May, 2002, pp. 262-267.

[3] Herranen M., Huhtala K. and Vilenius M. 2007. The Electro-hydraulic Valve Actuation (EHVA) for Medium Speed Diesel Engine – Development Steps with Simulation and Measurements. SAE Technical Paper Series 2007-01-1269, SAE World Congress Detroit, Michigan, April 2007, 8 p.

[4] Stewart P., Gladwin D. and Fleming P. 2007. Multiobjective analysis for the design and control of an electromagnetic valve actuator. Pro. IMechE vol 221Part D, J. Automabile Engineering, 2007. pp. 567-577.

[5] www.sturmanindustries.com

[6] Misovec K and Johnson B. 1999. Digital Valve Technology Applied to the Control of an Hydraulic Valve Actuator. SAE Technical Paper Series 1999-01-0825. International Congress and exposition, Detroit, Michican, March 1999. 5 p.

[7] Kim H. and Tsao T-C. 1998. Robust Performance Control of Electrohydraulic Actuators for Electronic Cam Motion Generation. Proceedings of the American Control Conference, Philadelphia, Pennsylvania, June, 1998, pp. 220-227.

[8] Sun Z. and Cleary D. 2003. Dynamics and Control of An Elertro-Hydraulic Fully Flexible Valve Actuation System. Proceedings of the American Control Conference, Denver, Colorado, June 2003. pp. 3119-3124.

[9] Virvalo, T. 1993. Modelling hydraulic position servo realized with commercial components. 3rd ICFP, Zhejiang University, Hangzhou China, 13-16 September 1993. 6 p.

[10] Virvalo, T. 1999. On the motion control of hydraulic servo cylinder drives. 4th JHPS, Tokyo, Japan, November 15-17, 1999.

[11] Herbert E. Merritt, 1991. Hydraulic Control Systems. Wiley, 368 pp

[12] Virvalo, T., On the damping of a hydraulic cylinder drive. The sixth Scandinavian International Conference on Fluid Power, SICFP'99, May 26-28, 1999, Tampere, Finland, pp 499-518

[13] Olsson, H, 1996, Control Systems with Friction. PhD thesis, Lund Institute of Technology, University of Lund, 1996.

[14] Virvalo, T. 1997. Nonlinear model of analog valve. SICFP 97, Linköping, Sweden, 28-30 May 1997. p. 199-214.

[15] Virvalo, T. 2005. Compensation of dead band and hysteresis of proportional valve. 50th National Conference on Fluid Power, March 16-18, 2005, Las Vegas, Nevada USA pp. 233-241.

[16] Virvalo, T. & Linjama, M. 2001. Experimental study of robustness of some controllers in a hydraulic position servo drive. Sbornik konference, Hydraulika a Pneumatika, June 5-6, 2001, Ostrava, Czech Republic pp.199-208.

[17] Linjama, M. & Virvalo, T. 2001. Robustness of some low-order controllers in an electrohydraulic position servo drive. (PTMC 2001 pp. 309-323, Bath, UK, 2001.

[18] Virvalo, T. 2001. PI and PID-controllers in a Hydraulic Position Servo System- What it is all about. ICFP´2001, April 3-5, 2001, Hangzhou, China pp.130-134.

[19] Virvalo, T. 2002. Comparing controllers in hydraulic motion control. PTMC 2002, Bath, UK, 2002 pp. 215-228.

[20] Virvalo, T. 2002. Hydraulic Motion Control - Robustness Against Inertia Load Changes. 5th JFPS, Nara, Japan, 2002, pp. 561-566.

[21] Virvalo, T., Lammila, M. & Lehto, E. 2003. Can bus applied on hydraulic computed force control? ICOM 2003, London and Bury St Edmunds, UK. Professional Engineering Publishing Limited. pp. 215-221.

[22] Hoffman W. 2001. Iterative Learning Control of Electromechanical Camless Valve Actuator. Proceedings of the American Control Conference, Arlington, VA , June 2001, pp. 2860-2866.

[23] Wang J. and Tsao T-C. 2004. Repetitive Control of Linear Time Varying Systems with Application to Electronic Cam Motion Control. Proceedings of the American Control Conference, Boston, Massachusette, June 2004, pp. 3794-3799.

[24] Anderson M. and Tsao T-C. 1997. Adaptive Lift Control for a Camless Electrohydraulic Valvetrain. SAE technical paper 981029. 27th International Conference on Electromechanical Systems, Lake Tahoe, Nevada, June 1997.

[25] Hanks T. and Lumkes J. 2005. Adaptive Control System for Electrohydraulic Camless Engine Gas Valve Actuator. Proceedings of the American Control Conference, Portland, OR, June 2005, pp.4369-4374.

Development of a Hybrid Position/Force Controlled Hydraulic parallel Robot for Impact Treatment

Y. Liu, H. Handroos , O. Alkkiomäki*, V. Kyrki* and H. Kälviäinen*

Laboratory of Mechatronics and Virtual Design
*Laboratory of Information Processing
Lappeenranta University of Technology, Finland
(E-mail:yong@lut.fi and heikki.handroos@lut.fi)

Abstract

This paper presents a study of a hybrid controlled robot system for impacting targets made of different materials. The robot is 5-Dof parallel structure with hydraulic actuators, which is going to be applied in the field of Ultrasonic Impact Treatment. A hybrid control approach is proposed in order to implement the application, which one direction to the impact target is applied by a force control and other directions are applied by the position control. The overall control scheme consists of three modules, position control, position increment control and force control. The first one generates candidate actions to drive a tool, which is attached on the end-effector, as accurate as possible directly in the same direction of a target the second module can drive the tool moving to the target with a constant speed, so that the third module could impact the target with a constant force. The proposed strategy is presented in some detail and further discussed using a few test-case experiments. The experimental results support the claim that it could be successfully applied to track and impact a target with a constant force.

Keywords: hybrid control, force control, parallel robot, impact target.

1 Introduction

Robot manipulator contact task execution represents an important problem as many tasks exist in which the robot is required to make intentional contact with fixed objects in the robot's work environment. In the case of repairing equipments in hostile environments, it is difficult to teach the shape of the work piece in advance. Even if it is possible to do that, a large amount of labor is required. To apply robot manipulators to a wider class of tasks such as deburring, grinding and grasping [1]-[4], it will be necessary to control not only the position of a manipulator but also the force exerted by the end-effector on an object. Basically, a sensor position and force feedback can enhance the performance of robotic manipulator in those industrial applications. Since real robotic manipulation tasks imply interaction with the environment, the robot end-effector motion is constrained at the contact. Proper execution of constrained motion tasks can be achieved using compliant control systems which attempt to accommodate external forces.

Traditional industrial robots that have been used as general-purpose positioning devices are open chain mechanisms that generally have the links actuated in series. These kinds of manipulators usually have a long reach and large workspace, but are inherently not very rigid and have poor dynamic performance at high speed and high dynamic loading under operating conditions. Compared with open chain manipulators, parallel mechanisms have high stiffness, high accuracy and high force /torque capacity in a reduced workspace, and have found many applications in manufacturing systems. Since there are no commercial solutions applicable to a certain environment, a new robot system, using hydraulic drives to achieve the required force density, has been designed and a prototype procured for testing. parallel robot Penta-WH, which has five degrees of freedom driven by hydraulic cylinders. The mechanism has a large, singularity-free workspace and high stiffness. It was used to act as a transport device for welding, machining and inspection end-effectors inside the ITER vacuum vessel [5]. We will use the same robot to carry out initial step in impact treatment application. It is well known that the principle of Ultrasonic Impact Treatment (UIT) is based on the instrumental conversion of harmonic oscillations of an acoustically tuned body into resonant impulses of ultrasonic frequency. The acoustically tuned body is brought to resonance by energizing an ultrasonic transducer. It could improve the quality and effectiveness of the application by using the parallel robot due to its high stiffness and accuracy

In this paper, a hybrid controlled robot system for impacting targets is studied and proposed. The method combines simplicity control with the ability of simultaneously handling both position and force variables typical of the hybrid control approach. This goal is achieved by using different independent controllers and managing conflicting situations by means of a priority strategy with dominance of the force controller over the position controller; automatic recovery from impacts is made possible by exploiting sensor measurements. The study is going to be applied in the field of Ultrasonic Impact Treatment. The experiments on impacting targets made of different materials have been carried out in order to check the control performance. A proposed hybrid control is selected based on a few test-case experimental results, which the overall control scheme consists of three modules, position control, position increment control and force control.

2 The Structure of hydraulic parallel robot and its model

The Penta-WH has been set up in the laboratory of Mechatronics and Virtual Engineering in Lappeenranta University of Technology. It is a parallel robot Penta-WH, which has five degrees of freedom driven by hydraulic cylinders. A centre beam is connected to the base platform via a universal joint, and fixed to the mid platform at one end. The hydraulic drivers are connected to the based platform by the universal joints, another end of three of them to the mid platform, two of them to the output flange platform by the ball joint.

It is well known that the structure of parallel mechanisms and closed kinematic chains make it difficult to derive its forward kinematics. Moreover, there may exist multiple solutions to the forward kinematic equations. On the other hand, it is possible to obtain its inverse kinematics with a unique solution. The inverse kinematics is the mapping from the generalized coordinate to the length of cylinders in this structure. The structure of the hydraulic parallel robot and its coordinate system are shown in Fig. 1. The robot is a five-degrees of freedom parallel robot with actuation by five hydraulic actuators. One end of telescopic beam is linked by a universal joint at the based platform, which provides α_1 and β_1 degrees of mobility in x and y axis. While another end is connected by a universal joint at the moving platform, which provides α_2 and β_2 degrees of mobility in x and y axis. Joint rotation angle of the based platform and the moving platform about z-axis is assumed to be zero. $L(t)$ is assumed as the displacement of telescopic beam. The vector of generalized coordinates of the robot is defined as

$$Z(t) = \left[L(t), \alpha_1(t), \beta_1(t), \alpha_2(t), \beta_2(t)\right]^T \quad (1)$$

The origin of the global coordinate frame $oxyz$ is assigned to the cross point between one end of the telescopic boom and the based platform, which is linked by revolute joint. The origin of one local coordinate frame is assigned to the same point of $oxyz$, while another is assigned to the cross point between another end of the boom and the moving platform.

The pose of the end effector in Cartesian coordinate system is defined as follows

$$r = \left[x, y, z, \alpha, \beta\right]^T \quad (2)$$

where (x,y,z) and (α, β) are described as the position and orientation of the end effector in the global frame, respectively. The orientation of the end effector has relationship with the universal joints at the based platform and the end platform. It can be expressed as follows

$$\alpha = \alpha_1 + \alpha_2 \quad (3)$$

$$\beta = \beta_1 + \beta_2 \quad (4)$$

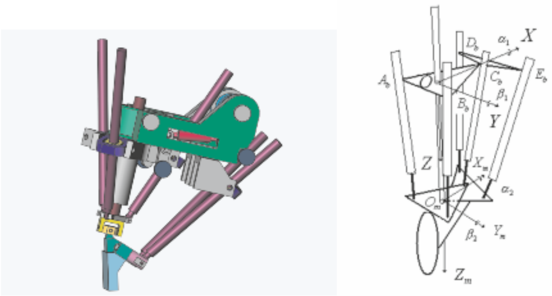


Fig. 1 The structure of the parallel robot and its coordinate system

The position vector of cylinders $l = \left[l_1 \quad l_2 \quad l_3 \quad l_4 \quad l_5\right]^T$ is expressed as follows

$$l = l(Z) \quad (5)$$

Where $l_i (i = 1,\ldots,5)$ is position of cylinders. The dynamic equation of the manipulator is achieved in the following form:

$$M(Z)\ddot{Z} + V(Z, \dot{Z}) + G(Z) + KZ = Q \quad (6)$$

where M is the inertia matrix of the manipulator which is symmetric and positive def ninite, K the stiffness matrix of the manipulator, V a vector of generalized Coriolis and centrifugal forces and the vector of the gravity torques. Q is the vector of generalized force, which can be determined as follows

$$Q = J_h^T F_H \quad (7)$$

where J_h^T is transpose Jacobian matrix from actuator space to generalized space. Thus it is possible to achieve hydraulic force vector in Cartesian coordinate in the following

$$F = J_f^T Q \quad (8)$$

where J_f^T is transpose force Jacobian matrix from generalized space to task space. More detail process sees [6].

3 Control design for the robotic system

3.1 Position control for the hydraulic cylinder

From the point of view of control, the system has five control inputs and five outputs. A proportional control is simplest way to control the position of hydraulic cylinders. The pressure load feedback is applied in the control process in order to attenuate the vibration. Therefore, a proportional control with pressure load feedback is proposed to set up for the robot system. The diagram of control system is illustrated in Fig. 2.

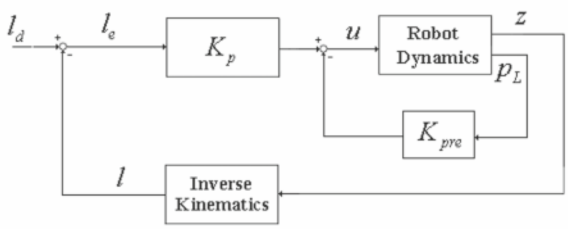


Fig. 2 Diagram of the position control for the hydraulic cylinders

The cylinder position vector l is used as the feedback because it could be measured by encoders. The controller is expressed as

$$u_p = K_p(l_d - l) + K_{pre} p_L \tag{9}$$

Where l_d is the desired actuator length vector, p_L is pressure load, K_p is the proportional gain, and K_{pre} is pressure load feedback gain.

3.2 Force Control

Force control is widely considered really useful strategy to control interaction with environment [8]-[9]. Control of interaction between a hydraulic robot and environment is crucial for successful execution of a number of practical tasks where the end effector has to manipulate an object or perform some operation on the surface. A proportional-integrated control is proposed as follows

$$u_f = K_p(F_d - F) \tag{10}$$

Where F_d and F are the desired and measured force vector, K_p is suitable positive definite matrix gains. Regulation of contact force to the desired values can be obtained, provided that the control gains are properly chosen so as to ensure stability of the closed loop system.

3.3 Position and position increment control

An increment type position control apparatus is adapted to evaluate a numerical value concerning a positional point by making subtraction between the data in a position feedback counter for counting a feedback pulse from an incremental encoder and the data in a commanded position buffer from a processor. Upon depression of an origin return command button provided on a control panel, the position feedback counter and the commanded position buffer are each loaded with a predetermined value, whereupon a pitch is given, whereby an object being controlled is controllably moved in the direction toward the origin. If and when the object being controlled exceeds the origin position, the same is stopped and the counter and the buffer are each again loaded with an appropriate numerical value, whereupon a given pitch is provided, whereby the object being controlled is moved in the direction toward the origin and comes to a point to almost exceeding the origin position. Just at the timing when the object being controlled is about to pass the origin position, the position feedback counter is set, whereby the data concerning the origin is regained in the apparatus. The block of position increment in x axis is shown in Fig. 3.

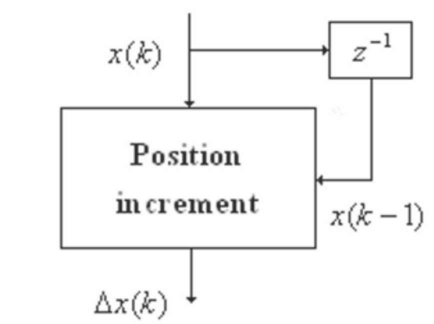


Fig. 3 Position increment in x axis

The position control in task space is simply proposed as

$$u_r = K_r(r_d - r) \tag{11}$$

where r_d is the desired trajectory of the end effector, r is the actual position, and K_r is the control gain..

The position increment control in x axis, the direction which is going to impact a target, is proposed as

$$u_i(k) = K_i(x(k) - x(k-1)) \tag{12}$$

where $x(k)$ and $x(k-1)$ is the position and the position unit delay in x axis, K_i is a gain.

4. Overall Control

In order to implement the application, hybrid control is proposed to carried out the process, which the direction to a target is applied by the combined control, for example, position or position increment control is used for approaching the target firstly, then force control is used after impacting to the target, the other directions are used by position control for tracking a certain trajectory. Therefore it is possible for a tool to work on a certain trajectory in the target surface with constant force.

The overall controller in the impact direction consists of the position or position increment controller and the force controller. The transition from position control to force control is based on force treshold. Let f_t be the force threshold. The overall controller can be expressed by

$$u_1 = \begin{cases} u_r, F < f_t \\ u_f, F \geq f_f \end{cases} \tag{13}$$

or

$$u_2 = \begin{cases} u_i, F < f_t \\ u_f, F \geq f_f \end{cases} \tag{14}$$

Diagram scheme of the hybrid control of robotic manipulator is shown in Fig. 4.

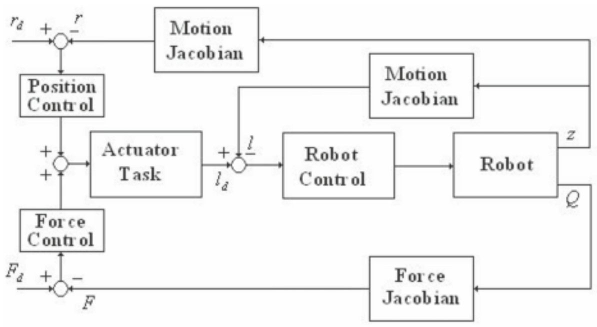


Fig. 4 Hybrid control system of robotic manipulator

4 Experimental results

In order to implement the application and analyze interaction between a tool and target, a prototype was built in IMVE at Lappeenranta University of Technology. The robot is fixed on a frame. With this experimental device several experiments have been done, like calibrating, workspace investigating, position accuracy and repeat accuracy testing. The end-effector of robot can reach the required workspace. 300 mm in z-axis, ±100 mm in x and y axis and The orientations about x and y axis are ±20° . The position accuracy is 0.1 mm and repeatability is 0.04 mm after calibrating.

Position sensors were installed at the hydraulic cylinders for position measurement, and ATI Multi-Axis Force/Torque Sensor system Omega160 IP60 was installed at end-effector to use for force/torque measurement. The position of the hydraulic cylinders can be achieved by using inverse kinematics solutions when the position and orientation of end tip are given. Omega160 IP60 consists of a transducer, shielded high-flex cable, and intelligent data acquisition system or stand-alone controller, which measures all six components of force and torque. The data acquisition card reads the signals and ATI software converts those signals to force and torque output. It is necessary to remove the influences of the inertial forces from the measured forces by the force/torque sensor so that the force control stability can be achieved. A target is installed on the slide bar, which is actuated by electric motor. The position is measured by a position sensor. Prototype of hydraulic parallel robot and a target are shown in Fig. 5.

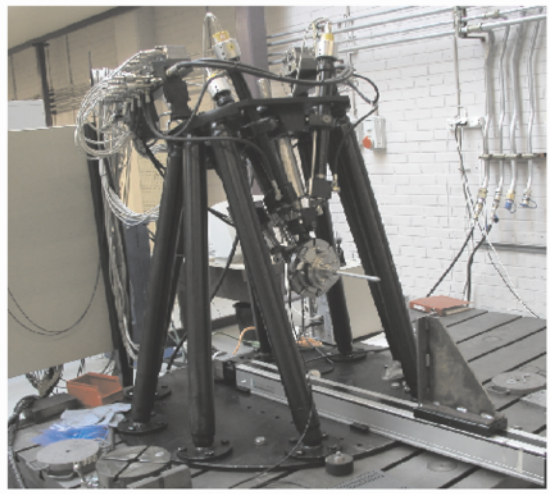


Fig. 5 Prototype of hydraulic parallel robot and impact target

Block diagram of measurement of the robot and DSP system is shown in Fig. 6. Excitation and response can be achieved in DSP system.

The proposed hybrid control, which integral force control and position or its increment control in the impact direction, and other directions by using position control after impacting the target, are carried out in the experiment. The robot was started from the position or its increment control, then switch to force control after touching the object. The desired contact force is set to 20 N. The experimental results are shown in Fig. 7- Fig. 11.

It is integrated force and position control for impacting targets made of different materials experiment in Fig. 7. The target material is made of wood and steel, respectively. The experimental result is shown that higher stiffness of material could be excited higher noise response by the impact and contact process.

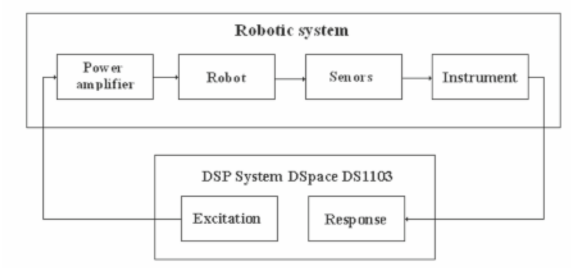


Fig. 6 Block diagram of measurement of the test bench and the DSP system

The proposed hybrid control, which integral force control and position or its increment control in the impact direction, and other directions by using position control after impacting the target, are carried out in the experiment. The robot was started from the position or its increment control, then switch to force control after touching the object. The desired contact force is set to 20 N. The experimental results are shown in Fig. 7- Fig. 11.

It is integrated force and position control for impacting targets made of different materials experiment in Fig. 7. The target material is made of wood and steel, respectively. The experimental result is shown that higher stiffness of material could be excited higher noise response by the impact and contact process.

Two different velocities and its corresponding contact forces are shown in Fig.8. The results of the speed for 200 mm/s and 300mm/s are shown in Fig. 11. The force overshoot can be higher if the target is impacted by higher speed. Higher velocity could lead to higher force overshoot when velocity control is switched to force control after impacting the target.

The hybrid force and position control for impacting target made of wood is shown in Fig. 9. The robot starts to move to the target 300mm/s by the position control, contacts the target with 20 N by the force control after impacting, keep the force control in x axis and start to make a movement in y axis at the same time, then the robot returns to the initial position after finishing those task. The response of the force, the position in x axis and y axis are shown in Fig. 9 (a), Fig. 9 (b) and Fig. 9 (c), respectively. The force overshoot in this case is excited by the impacting the target and moving on the surface of the target. The same process is also carried out for the target made of steel, the results is shown in Fig. 10.

The hybrid force and position increment control for impacting target made of wood is shown in Fig. 11. The robot starts the movement from the initial position to the impacting position by using the position increment control, then switch to the force control after touching the target with the constant force 20 N. The force response is shown in Fig. 11 (a) and the position response in x axis is shown in Fig. 11(b). The force overshoot in this case has been improved compared with the last case. The robot could carry out the different trajectories on the surface of the target depend on applications. The result of the circle response on the surface of wood target is shown in Fig. 12. It includes the force the circle response in Fig. 12 (a) and Fig. 12(b).

The experiment on the different robot hybrid controls for impacting targets made of different materials has been carried out in this section. The results demonstrated that the hybrid force and position increment control could reduce or remove the force overshoot. The trajectory error is also in the valid range of the coming future impact treatment. Therefore the hybrid control is proposed to be applied in the application.

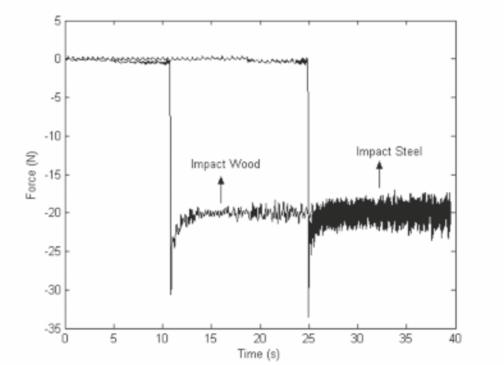


Fig. 7 Force response of impact test for the different materials

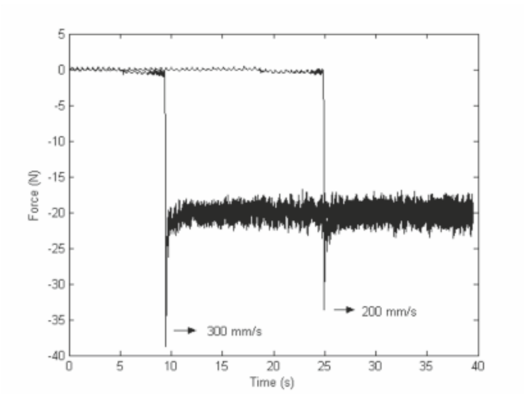
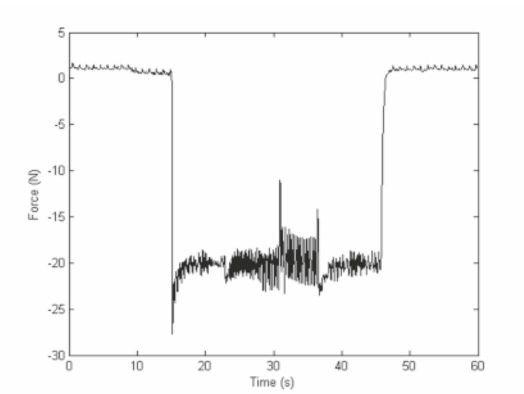
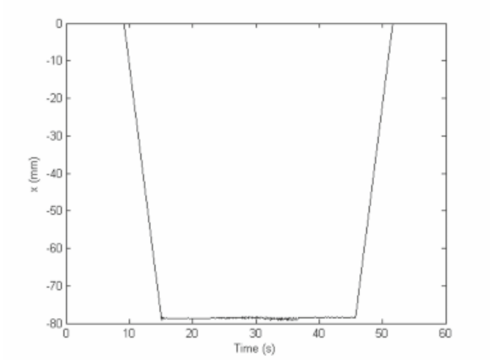


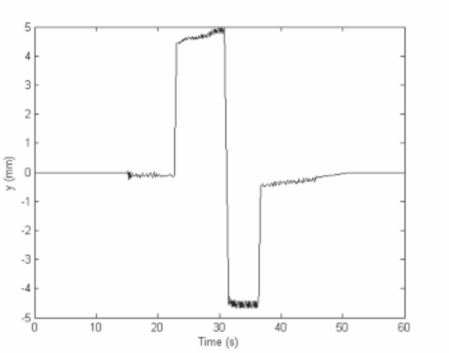
Fig. 8 Force response of impact test with the different speeds



(a) Force response

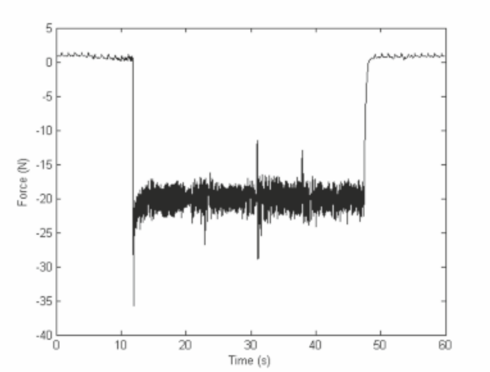


(b) x-axis position

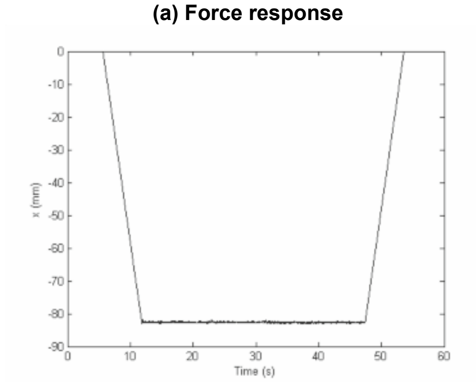


(c) y-axis position

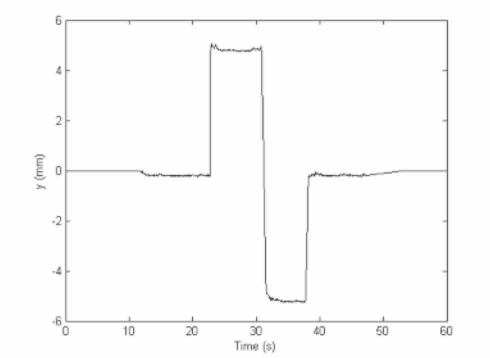
Fig. 9 The response for the movement after impact wood target



(a) Force response

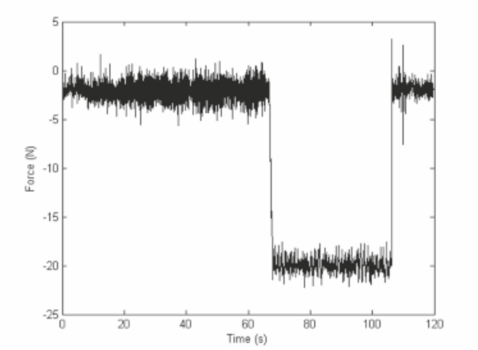


(b) x-axis position

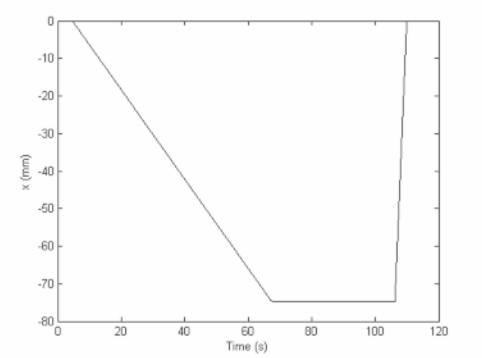


(c) y-axis position

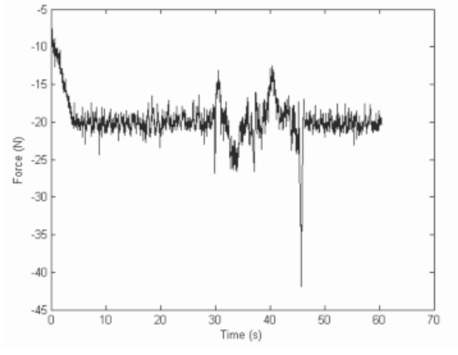
Fig. 10 The response for the movement after impact steel target



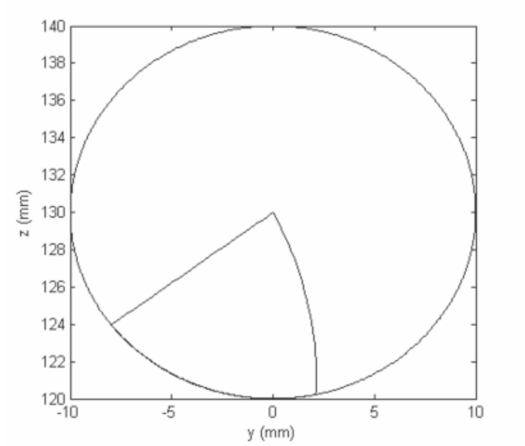
(a) Force response



(b) x-axis position
Fig. 11 Position increment/Force control with Wood target.



(a) Force response



(b) Circle response
Fig. 12 The circle response during contacting with wood

5 Conclusions

A study of a hybrid controlled robot system for impacting targets made of different materials in order to successfully perform impact treatment task in the future. The basic consideration is to reduce the impact force overshoot of a force controller in the initial contact phase and keep a tool movement in the valid error range on the surface of work piece. The end effector position and force can be simultaneously controlled as long as the robotic manipulator does not fall into singular configuration. The control strategy is proposed in some detail and further discussed using a few test-case experiments. The experimental results support the claim that it could be successfully applied to track and impact a target with a constant force.

The future work is to combine a vision system into the hybrid controlled robot system in treatment impact application. The vision system can carry out online trajectory, and guide the robot to finish some tasks on the surface of work piece. The initial work on a force/vision control has been done in a servo bench system [7].

References

1. Jinno, M., Ozaki, F., Yoshimi, T., and Tatsuno, K., Development of a Force Controlled Robot for Grinding, Chamfering and Polishing, IEEE Internatlonal Conference on Robotlcs and Automatlon, pp. 1455-1460, 1995.
2. M. H. Raibert and J. J. Craig, Hybrid Position/Force Control of Manipulation, Trans. ASME, J. of Dynamic System, Measurement, and Control, Vol. 102, pp.126-133, 1981.
3. L. Liu, B. J. Ulrich and M. A. Elbestawi, Robot Grinding Force Regulation, Design, Implementation and benefits, 1990 IEEE Conference on Robotics and Automation, pp.258-265, 1990.
4. M. Jinno, T. Yoshimi and A. Abe, Force Controlled Grinding Robot System for Unstructured Tasks, Proceedings of the 1992 IEEE/RSJ International Conference on Intelligent Robots and Systems, Raleigh, NC. pp. 1117-1123, July 7-10, 1992.
5. H. Wu and H. Handroos, Design of Parallel Intersector Weld/Cut Robot for Machining Processes in ITER Vacuum Vessel, Fusion Engineering and Design, 69, 2003, pp. 327-331.

6. Y. Liu and H. Handroos, “Modeling and control of a parallel redundant manipulator with hydraulic actuators,” Proc. IMechE Vol. 220 Part I: J. Systems and Control Engineering, 220(I3), 211–222.
7. Olli Alkkiomäki, Ville Kyrki, Heikki Kälviäinen, Yong Liu, and Heikki Handroos, Smooth transition from motion to force control in robotic manipulation using vision, International Conference on Control, Automation, Robotics and Vision, Singapore, December 2006.
8. Siciliano B., Villani L., Robot Force Control, Kluwer Academic Publishers International Series in Engineering and Computer Science, Boston, MA, 1999
9. Olsson T., Feedback Control and Sensor Fusion of Vision and Force, PhD Thesis, Department of Automatic Control, Lund Institude of Technology, 2004.

Modeling and Control of Water Hydraulic Driven Parallel Robot

Huapeng Wu, Pekka Pessi, Yongbo Wang, Heikki Handroos

Lappeenranta University of Technology, Lappeenranta . Finland

Abstract

This paper presents a novel mobile parallel robot, which is able to carry welding and machining processes from inside the international thermonuclear experimental reactor (ITER) vacuum vessel (VV). The kinematics design of the robot has been optimized for ITER access. Kinematics modeling is given in the paper, and control algorithm is designed for water hydraulic.

Keywords:

Kinematics modeling; Parallel robot; ITER; Water Hydraulic; Control

1 INTRODUCTION

The international thermonuclear experimental reactor (ITER) is a joint international research and development project that aims to demonstrate the scientific and technical feasibility of fusion power. The reactor tokomak (vacuum vessel) is made of stainless steel, and contains nine sectors welded together; each sector has about the size of 10 meter high and 6 meter wide. The sectors of ITER vacuum vessel (VV) (Fig. 1) require more stringent tolerance (±5mm) than normally expected for the size of the structure involved. The walls (inner wall and outer wall) of ITER sectors are of 60mm thick and are joined by high quality leak tight welds. In addition to the initial vacuum vessel assembly, the sectors may have to be replaced for repair. Meanwhile, the machining operations and lifting of a possible e-beam gun column system require extreme stiffness property and good accuracy. The payload to weight ratio also has to be significantly better than it is in the commercial industrial robots.

The conventional robots, providing a high nominal payload, are lack of stiffness and accuracy in such machining condition. Since commercially available machines capable of handling large payloads require floor mounting and their workspaces are insufficient for reaching the cross section at a single mounting position, a special robot is needed. Parallel robots have high stiffness, high dynamic performance and good payload to weight ratio in comparison with the conventional serial robots. Stewart [1] presented the novel idea of six-degree-of-freedom parallel robot in 1960's. A remarkable number of research articles and books about parallel manipulators have been published during the last two decades. There are also a number of successful industrial applications developed [2], [3], [4], [7]. The parallel manipulators have potential advantages compared with the conventional serial link manipulators. The parallel manipulators are closed-loop mechanism presenting good performances in terms of accuracy, rigidity, high speed, and ability to handle large loads. They are becoming popular in applications such as machining, welding, assembly, flight and vehicle simulators, mining machines, and pointing devices [2], [3], [4], [7]. In some industrial applications, it is necessary to move the end-effector of manipulator from point to point rapidly. On the other hand, some applications require extreme stiffness properties. Each of these properties can be achieved by an appropriate design of parallel structure. The most important drawback of parallel robots is the small workspace, which can be made larger by adding additional serial axes in the robot.

For the assembly of the ITER vacuum vessel sector, the precise positioning of welding end-effectors at some distance in a confined space from the available supports will be required, while it is not possible using conventional machines or robots. The parallel robot presented in this paper is able to carry out welding and machining processes from inside the ITER vacuum vessel, consisting of a six degree-of-freedom parallel mechanism mounted on a carriage driven by electric motor/gearbox on a rack. The robot carries both welding gun (such as a TIG, hybrid laser or e-beam welding gun to weld the inner and outer walls of the ITER vacuum vessel sectors) and machining tools (to cut and mill the walls with necessary accuracy). It can also carry other heavy tools and parts to a required position inside the vacuum vessel. As the requirement of environment in assembling of ITER, the oil hydraulic may course some pollution inside the surface which is not allowed. Therefore water hydraulic has been applied in robot driven.

The robot offers not only a device but also a methodology for assembling and repairing VV. For assembling, an on-line six-degree-of-freedom-seam-finding algorithm has been developed. The algorithm enables the robot to find welding seam automatically in a very complex environment. For machining, the multi flexible machining processes carried out automatically by the robot have also been investigated, including edge cutting, smoothing, and defect point milling. The kinematic design of the robot has been optimised for the ITER access and a hydraulically actuated prototype has been built. Finally the experimental results are presented and discussed. The earlier development phases of the robot are presented in [8] and [9].

2 STRUCTURE OF VV AND ASSEMBLING PROCESS

The inner and outer walls of the VV of ITER are made of 60mm-thick stainless steel 316L and are welded together through an intermediate, so-called “splice plate”, inserted between the sectors to be joined. The splice plates have two important functions: (*i*) to allow access to bolt together the thermal shield between the VV and coils; and (*ii*) to compensate the mismatch between adjacent sectors to give a good fit-up of the sector-sector butt weld. The robot's end-effector will have to pass through the inner wall splice plates opening to reach the outer wall. As shown in Fig.1, the assembly and repairing processes have to be carried out from inside the vacuum vessel.

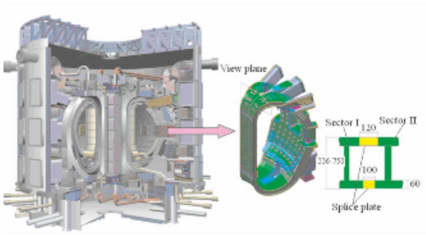


Fig.1 ITER and VV sectors to be welded

The assembly or repair will be performed according to four phases: cutting, edge machining and smoothing, welding, and non-destructive tests (NDT) control. The robot will carry out welding, machining, and inspecting inside the VV. The maximum robot force arises from cutting. It can be up to 3kN.

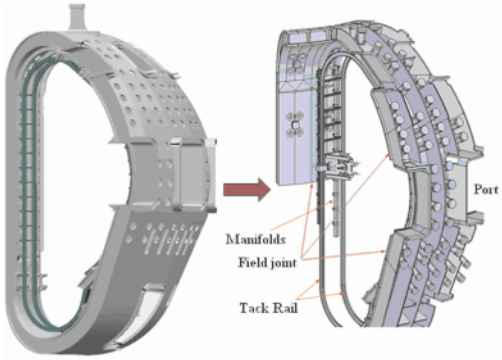


Fig.2 The track rail mounted inside VV and robot on the track

In order that the robot can operate in the cross section of the vessel, a track is assembled inside the sector. The track has a rack on one side of the rail and it is supported by manifolds and beams (shown in Fig. 2). The robot driven on the rail carries out welding or machining along the edge of the sector. After finishing the assembly of two sectors, the robot has to be moved to the next sector where there is also a track assembled. After finishing the assembly of all the sectors, the robot can be taken out via the port of VV.

3 KINEMATICS MODEL OF PARALLEL ROBOT

3.1 Structure of parallel robot

The proposed parallel robot has ten degrees of freedom (Fig.3). It consists of two relatively independent sub-structures: (*i*) carriage, which provides four additional degrees of freedom, i.e., rotation, linear motion, tilt rotation and tracking motion, and these four degrees are added to enlarge the workspace and to offer high mobility; and (*ii*) the Hexa-WH parallel mechanism driven by six hydraulic cylinders contributes six degrees of freedom for the end-effector. Thus the robot is a hybrid redundant manipulator with four extra degrees of freedom provided by serial kinematic axes.

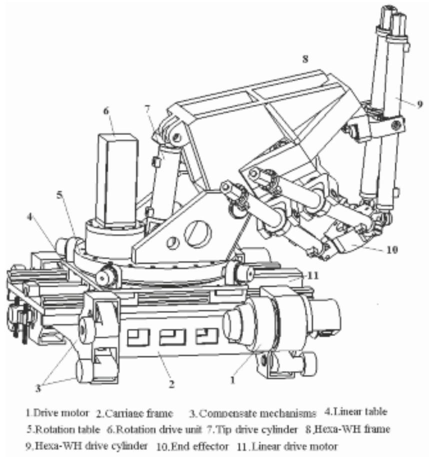


Fig. 3 Parallel robot

a). Carriage mechanism

The carriage mainly consists of 5 units. *i) Carriage frame*: The carriage frame is a complex structure welded by multi-steel-plates, and it is able to carry high payload and offers enough room to maintain mechanisms. Stiffness and weight are the most important considerations in the design, and they have been optimized to achieve necessary stiffness with light weight. *ii) Tracking drive unit:* The tracking drive unit consists of electric motor, reduction unit CYCLO, V-shape bearing, and driving gear. The electric servo motor with position feedback controller offers the high accurate motion. In order to output large torque to drive the heavy mass and payload, the reduction unit CYCLO is added to the motor to reduce speed and to transmit high torque to the driving gear. Two V-shaped wheels keep the carriage on the tracking rail at right position to avoid the cross motion. Two drive units are used in the carriage to offer enough torque in order to drive the robot and payload around inside the VV. *iii) Compensation mechanism:* The compensation system is an important unit that limits the backlash caused by the inaccurate assembling of the tracking rail and compensates the distance changing between the wheels in bending area. As the shape of the VV is very complicated, it is difficult to keep the tracking rails lying on the VV surface in the accurate position. The position tolerance can be up to ±2mm. The distance of the coupled wheels has to be adjusted to follow the changes of the rail, and all the wheels must touch the parallel rails with certain force during motion; hence an adaptive distance compensation system is needed and it should be able to undertake the summed weight of the robot and the payload, when the robot is upside down at the top position insider VV. Since the total payload is very heavy, a hydraulic cylinder is applied to justify the compensation force according to the position where the robot is located. Fig. 4 shows the compensation system, where the upside is the tolerance adaptive mechanism that passively follows the changes of track rail and the downside is the hydraulic distance compensation system that ensures a constant force is applied to the rails. *iv) Linear drive unit:* The linear drive unit enlarges the workspace of the robot, and consists of five parts: ball screw drive unit, servo motor, rails, linear bearings, and a table. Two parallel rails are fixed on the carriage frame to offer the motion crossing the frame and to extend the distance of the robot in direction Y. In direction Y, the distance from the inner board to the outer board can be 900mm in one VV sector, i.e., the robot needs longer reach in this direction, and the linear drive unit helps the robot end-effector to reach the farther border of the VV. *v) Rotation drive unit:* This unit offers a rotation motion about the Z axis, so that the robot can machine the flexible houses on the inner wall at any position. The rotation drive unit consists of slewing bearing, drive gear, reduction unit CYCLO, and servo motor. The slewing bearing integrates bearing and gear together, leading to a compact structure with light weight. The rotation of the unit can reach ±180°.

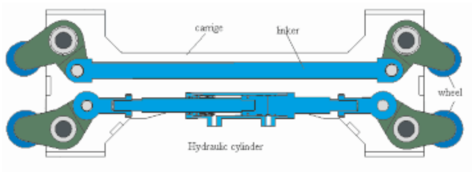


Fig. 4 Compensation mechanism

b). Hexa-WH

A Stewart based mechanism, driven by six servo control water hydraulic cylinders, offers six-degree freedom to the end-effector, where the machining head and welding gun are mounted. Because of the special shape of the VV, a full six-degree freedom motion for tool is needed to enable the robot to carry out welding and machining. The Hexa-WH can offer the required accuracy and the high force capacity due to its novel configuration and the hydraulic drive.

3.2 Kinematics model

The kinematics model is very important for the robot motion control. As the robot has redundant degree freedom, it is difficult to find the kinematics solution directly. The kinematics models can first be set up for the carriage and the Hexa-WH separately, and then be combined together by using an optimization algorithm in solving the redundant problem [4], [5].

a). Forward kinematics

As described above, the carriage offers the robot the four-degree freedom: two linear motions and two rotations; while the Hexa-WH offers the end-effector the full six-degree freedom. The transformation matrix of the robot can be defined as:

$$T_c = T_1 \cdot T_2 \cdot T_3 \cdot T_4 \cdot T_5 \tag{1}$$

where

$$T_1 = \begin{bmatrix} 1 & 0 & 0 & X_1 \\ 0 & 1 & 0 & Y_1 \\ 0 & 0 & 1 & Z_1 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$T_2 = \begin{bmatrix} 1 & 0 & 0 & X_2 \\ 0 & 1 & 0 & Y_2 \\ 0 & 0 & 1 & Z_2 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$T_3 = \begin{bmatrix} c\phi & -s\phi & 0 & X_3 \\ s\phi & c\phi & 0 & Y_3 \\ 0 & 0 & 1 & Z_3 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

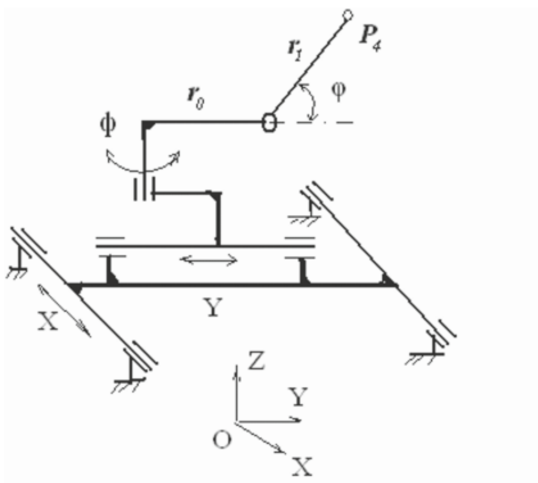
$$T_4 = \begin{bmatrix} 1 & 0 & 0 & X_4 \\ 0 & c\varphi & -s\varphi & Y_4 \\ 0 & s\varphi & c\varphi & Z_4 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$T_5 = \begin{bmatrix} c\alpha c\beta & c\alpha s\beta s\gamma - s\alpha c\gamma & c\alpha s\beta c\gamma + s\alpha s\gamma & X_5 \\ s\alpha c\beta & s\alpha s\beta s\gamma + c\alpha c\gamma & s\alpha s\beta c\gamma - c\alpha s\gamma & Y_5 \\ -s\beta & c\beta s\gamma & c\beta c\gamma & Z_5 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

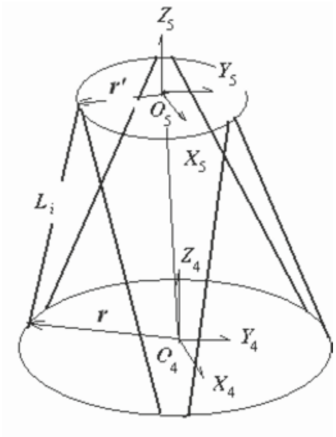
Once the parameters of joints are given, the forward kinematics of the robot can be defined as

$$\bar{P} = T \bullet \bar{P}_0 = T_1 \cdot T_2 \cdot T_3 \cdot T_4 \cdot T_5 \cdot \bar{P}_0 \tag{2}$$

To solve the forward kinematic model of the Hexa-WH, the numeric iterative method can be employed and it can be solved from its inverse kenamatic model given later in the chapter.



(a)



(b)

Fig. 5 a) Mechanism of carriage, b) Hexa-WH

b). Inverse kinematic model of robot

As the robot has four-degree freedom of redundancy, we give an inverse kinematic model first to the carriage, then to the Hexa-WH.

Inverse kinematic model of carriage: The inverse kinematic model of the carriage is defined to find the values of the four actuators with respect to the frame ***o*** for a given position and an orientation of P_4 on the Hexa-frame. The principle of the carriage mechanism is shown in Fig. 5. In the application, rotation angle ϕ is fixed only at a few values, 0°, ±90°, and 180°, and we can calculate the values of other actuators by fixing ϕ, i.e., for a given position $\boldsymbol{P_4}(x, y, z)$, the centre of the Hexa-Frame, we have

$$X + (r_0 + r_1 \cos\varphi)\cos\phi = x, \tag{3}$$

$$Y + (r_0 + r_1 \cos\varphi)\sin\phi = y, \tag{4}$$

$$r_1 \sin\varphi = Z. \tag{5}$$

Then

$$\varphi = \arcsin(Z / r_1), \tag{6}$$

$$X = x - (r_0 + r_1 \cos\varphi)\cos\phi, \tag{7}$$

$$Y = y - (r_0 + r_1 \cos\varphi)\sin\phi. \tag{8}$$

Inverse kinematic model of Hexa-WH: The inverse kinematic model for the Hexa-WH is defined to find the values for each cylinder at a given position and an orientation of the end-effector with respect to the Hexa-frame. Here O_4 is coincided with P_4 on the carriage side. Fig.5 b) demonstrates the coordinates of the Hexa-WH.

The inverse kinematics model for the Hexa-WH is

$$\overline{L}_i = \overrightarrow{O_4 O_5} + R \cdot \overline{r}_i' - \overline{r}_i \quad (i = 1, 2, \ldots, 6) \tag{9}$$

where

$$R = \begin{bmatrix} c\alpha c\beta & c\alpha s\beta s\gamma - s\alpha c\gamma & c\alpha s\beta c\gamma + s\alpha s\gamma \\ s\alpha c\beta & s\alpha s\beta s\gamma + c\alpha c\gamma & s\alpha s\beta c\gamma - c\alpha s\gamma \\ -s\beta & c\beta s\gamma & c\beta c\gamma \end{bmatrix},$$

$\overline{r}_i$ denotes the vector of the joint of the i^{th} cylinder on the Hexa-frame with respect to frame $\boldsymbol{o_4}$ and $\overline{r}_i'$ is the vector of the joint of the i^{th} cylinder on the end-effector with respect to frame $\boldsymbol{o_5}$.

The length of each cylinder can be found, when (x, y, z, γ, β, α) is defined with respect to frame $\boldsymbol{o_4}$.

$$l_i = \left|\overline{L}_i\right| = \sqrt{(\overrightarrow{O_4 O_5} + R \cdot \overline{r}_i' - \overline{r}_i) \bullet (\overrightarrow{O_4 O_5} + R \cdot \overline{r}_i' - \overline{r}_i)}. \tag{10}$$

There are two ways to combine the two inverse kinematic models to get the solution of the whole robot. One simple way is to calculate the coordinates (x, y, z, γ, β, α) of the end-effector with respect to {O_4} and the values for each actuator from equations (3-10) for the given coordinates of the end-effector with respect to frame {**o**}, while fixing {O_4} at a certain position with respect to frame {**o**} according to experience. The other way is to use an optimization algorithm to find redundant solution, which is subjected to minimize the deflection of the robot during motion, i.e., $\min_{l} f(\overline{q}, \overline{l})$, where f is the deflection model of the robot, $\overline{q}$ is the position vector of the end-effector, and $\overline{l}$ is the value vector of ten actuators. For a given $\overline{q}$ we can find $\overline{l}$ by solving the optimization problem $\min_{l} f(\overline{q}, \overline{l})$.

4 HYDRAULIC SYSTEM

4.1 Hydraulic control system

The parallel robot is mainly driven by water hydraulic actuators, first because of hydraulic system can offer high power density so that the robot can be make light, and secondly water hydraulic is clean for the environment inside of VV. But at the same time it is a big challenge for water hydraulic driven because of limit the flow rate of servo valves. The responsible speed can not be very high (over than 3m/min) .otherwise the speed error will be grate than required and the robot can not follow the curve very accuracy when speed is high.

Fig.6 shows the water hydraulic system. Pressure servo control is applied in locking cylinders. Position servo controller is used in cylinders 1-7. There are three loops in the servo position control: the position loop together with the speed loop that provide an accurate and fast trajectory tracking; and the load pressure feedback loop that is applied to damping the self-excited oscillations, which normally occur in the natural frequency. The speed loop can eliminate the speed error, while the pressure feedback damps the vibration of hydraulic actuators (Fig. 7).

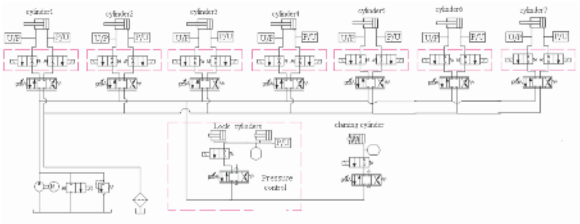


Fig.6 Water hydraulic circle

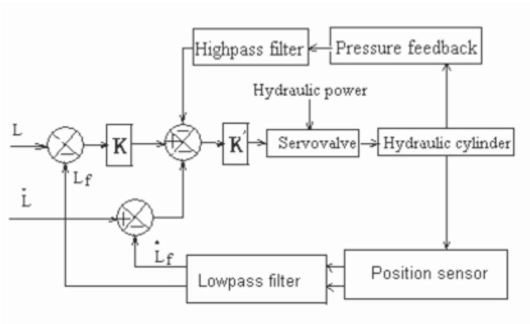


Fig.7 Hydraulic servo position control

The hydraulic cylinders normally lack damping that makes the cylinder control difficult by using conventional PID-controllers. The damping can effectively be increased by means of load pressure feedback. The major drawback in using pressure feedback is its negative effect on the static stiffness of the actuator. To overcome this problem, the high pass filters are used in the load pressure feedback loops. The high pass filter removes the negative effect of pressure feedback at low frequency.

4.2 Motor Control

Two drive motors contribute effort for the tracking motion of the robot. As the tracking rails are not always straight, the speeds of the two motors are not the same when the robot is moving. The torque control together with a position feedback algorithm is implemented. Fig. 8 shows the control principle.

In this method, one motor works as master, and another one works as slave. For the master motor, the position control plus the speed control is applied to guarantee the required speed and position accuracy of the carriage on the tracking rail. For the slave motor, the torque control is applied, which contributes the driving torque for the robot.

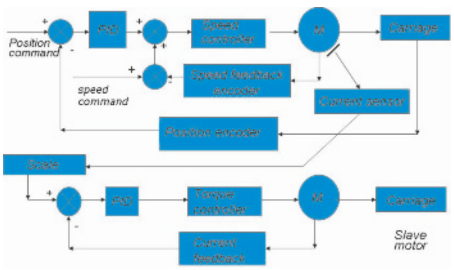


Fig.8 Tracking motor control

4.3 Control of hardware and software

Because there are no commercial controller and software available for the special functions of the parallel robot, an open architecture of hardware and programmable software are being developed. Fig. 9 shows the structure of hardware control system. The controller is an industrial-PC-based motion controller. It provides a reliable and easy-at-use environment for controlling the robot because Earthnet bus is used in the connection of iPC and I/O interfaces.

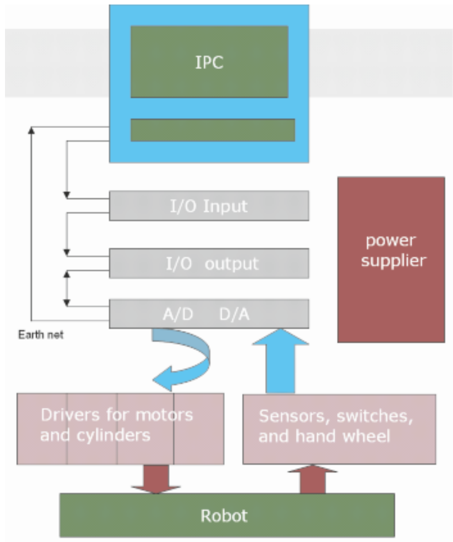


Fig.9 Structure of robot controller

The software is defined in Fig.10, including graphical interface, trajectory planning, forward and inverse kinematics models, interpolator, controller, and I/O interface functions. And those functions have to be integrated with the program offered by iPC and run completely in real time.

Graphical interface is a high level program, it includes parameter setting, condition monitoring, and graphical visualization functions. User can easily exchange information with this program.

Trajectory planning is also a high level program. As the robot has redundant actuators, the trajectory planning is much more difficult than usual, so an optimization algorithm, which is subjected to minimize the deflection of the robot during motion, has been employed.

Forward and inverse kinematics models and interpolator are real time functions, which generate data for motion controller.

Controller is a real time function including water hydraulic controller and motor controller. As the robot has two tracking motors and the speed of the motors are not always the same at some positions, a master–slave control algorithm has been used.

I/O interface functions are real time functions, which enable transferring date from sensor to controller and from controller to driver.

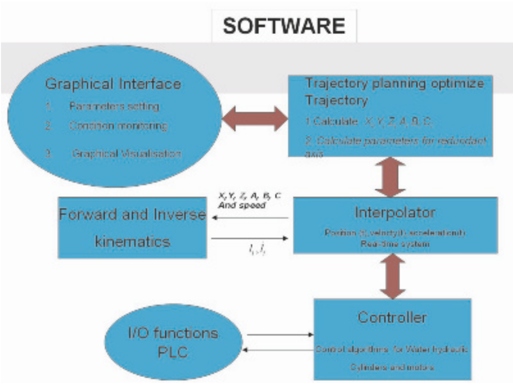


Fig. 10 Structure of software

5 CONCLUSION

A hybrid parallel robot with four additional serial motion axes is developed for carrying out the necessary machining and welding tasks in the assembling and repairing of the ITER Vacuum Vessel. The robot is capable of holding all necessary machining tools and welding end-effectors in all positions accurately and stably. The kinematics analysis of the robot is presented. The models are complex because of the redundant structure of the robot. The models are separately derived for the Hexa-WH and the carriage mechanism. An optimization algorithm finds the solution in the trajectory planning, ensuring the maximum stiffness during the robot motion.The entire design and testing process of the robot is a very complex task due to the high specialization of the manufacturing technology needed in the ITER reactor. The results demonstrate the applicability of the proposed solutions quite well.

6 ACKNOWLEDGEMENT

This work, supported by the European communities under the contract of Association between EURATOM/Finnish TEKES, was carried out within the framework of the European Fusion Development Agreement, and the views and opinions expressed herein do not necessarily reflect those of the European Commission.

7 REFERENCES

[1] D. Stewart (1965), A platform with six degree of freedom, Proc. Inst. Mech. Eng., London, Vol. 180, pp. 371-386.

[2] T. C. Arai, K. Homma, H. Adachi, and Nakamura (1991), Development of parallel link manipulator for underground excavation task, Proc. Int. Symposium on Advanced Robot Technology, pp. 541-548.

[3] C. Gosselin and J. Hamel (1994), The agile eye: A high-performance three degree of freedom camera –orienting device, Proc. IEEE Int. Conference on Robotics and Automation, pp. 781-786.

[4] L. W. Tsai (1999), Robot analysis: The mechanics of serial and parallel manipulators, A Wiley-Interscience Publication, John Wiley & Sons Inc.

[5] G. Lebret, K. Liu, and L. Lewis (1993), Dynamic analysis and control of a Stewart platform manipulator, J. Robotic Systems, Vol. 5, No. 10, pp. 629-655.

[6] G. Zheng, L. S. Haynes, J. D. Lee and R. Carroll(1992), On the dynamic model and kinematic analysis of a class of Stewart platforms, Robotics and Autonomous Systems, Vol. 9, pp. 237- 254.

[7] K. H. Häfele, H. Haffner, and P. Spencer (1992), “Automatic Fettling Cell- An Example for Applying Computer- Aided Robotics”, Industrial Robot. Vol.19 No. 5. pp.31-34.

[8] H.Wu, H. Handroos, P. Pessi, J. Kilkki, L. Jones (2005), “Development and control towards a parallel water hydraulic weld/cut robot for machining processes in ITER vacuum vessel”, Fusion Engineering and Design.

[9] H. Wu, H. Handroos (2006), Mechatronics design and development towards a heavy-duty waterhydraulic welding/cutting robot in Book: Mechatronics for Safety, Security and Dependability in a New Era. Elsevier.

Development of Intelligent Ankle-Foot Orthosis (i-AFO) with MR Fluid Brake and Control System for Gait Control

Takehito KIKUCHI[1], Sousuke TANIDA[1,2], Kikuko OTSUKI[1], Taigo KAKEHASHI[1] and Junji FURUSHO[1]

[1] Dept. of Mechanical Engineering, Graduate School of Engineering, Osaka University, Osaka, Japan

[2]Shiga School of Medical Technology, Shiga, Japan

Abstract

Ankle-Foot Orthoses (AFOs) are orthotic devices supporting movements of ankles of disabled people for example hemiplegia, polio, peroneal nerve palsy, etc. In our research, we have developed intelligently and passively controllable AFOs (i-AFO) which can control its ankle torque with a compact Magneto-rheological fluid brake (MRB). In this paper, we describe a design method and a control method of the i-AFO. In the final part of this paper, we present experimental results in which the i-AFO was applied to gait control for a polio patient.

Keywords:

Ankle-foot orthosis, controllable brake, magneto-rheological fluid, gait control

1 INTRODUCTION

Locomotion is a most important skill of the activity of daily living (ADL) for human beings. Therefore gait trainings have been made a high priority in rehabilitative trainings. Normal gait is cyclic and can be characterized by timing of foot contact with the ground; an entire sequence of functions by one limb is identified as a gait cycle [1], [2], [3] as shown in Fig.1. Each gait cycle has two basic components: "stance phase," which designates the duration of foot contact with the ground, and "swing phase," the period during which the foot is in the air for the purpose of limb advancement. The swing phase can be further divided into three functional subphases: "initial swing," "mid swing" and "terminal swing." In the same manner, the stance phase can be partitioned into five functional subphases: "initial contact," "loading response," "midstance," "terminal stance" and "preswing" [1], [4], [5]. In the normal gait, initial contact becomes "heel-contact (or heel-strike)," appropriately. And in the initial swing (or "toe-off"), normal subjects can keep appropriate clearance from the ground to prevent a malfunctional interaction with the ground.

For patients who have dysfunction of ankles, for example polio and peroneal nerve palsy, it is difficult to control their ankle by themselves. This problem causes "drop-foot" or a lack of ankle dorsiflexion during the swing phase. In many case, they can not prevent from stumbling their toe with even small steps on the ground. Additionally, they have a tendency to incline their upper body more than healthy persons because of the rough motion to prevent from stumbling of the toe. It causes undesired energy-loss in walking.

An orthosis is defined as a device attached or applied to the external surface of the body to improve functions, restrict or enforce, or support a body segment [1], [6]. In order to improve the gait of the patients, lower limb orthoses are applied to them. In order to assist their ankle function, ankle-foot orthoses (AFOs) are often used to restrict their involuntary plantarflexion and so on.

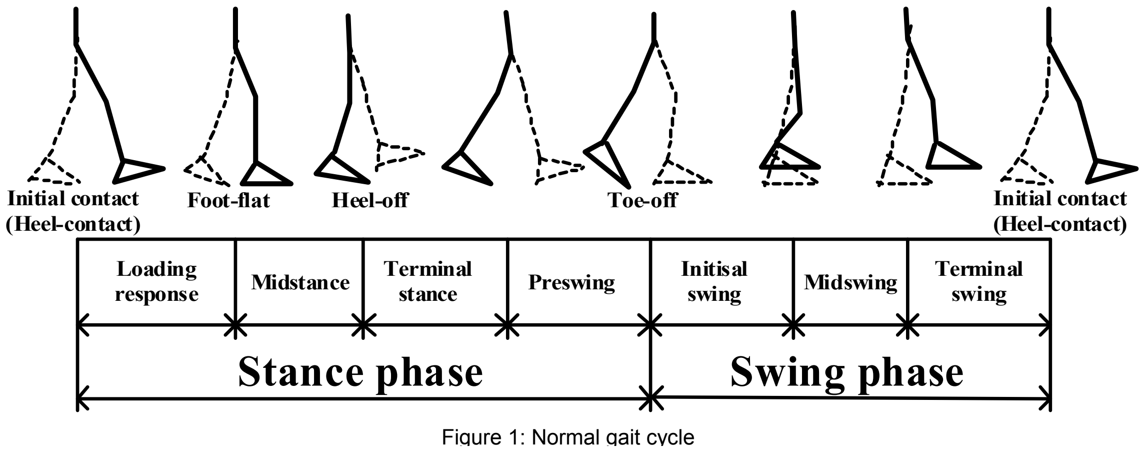


Figure 1: Normal gait cycle

From a long time ago, many researches focusing on powered dynamic-controllable lower extremity prostheses with robotics technologies are reported [7]-[11] and some kind of products are released (Otto Bock C-Leg [12], Ossur RHEO KNEE [13]). These prostheses are designed to be patient-adaptive or environment-adaptive. These reports give us grate amount of useful information about gait control [14] for the idea of powered and dynamic-controllable AFOs.

Actually, powered AFOs are focused on in recent years. Some kinds of the powered AFO were reported [15] by using several types of actuators, e.g. pneumatic actuation system (Keith E. Gordon and Daniel P. Ferris [16]), ball screw drive system (Bashir M. Y. Nouri and Arafat Zaidan [17]), series elastic actuator (Joaquin A. Blaya and Hugh Herr [18]), and so on. However, developments of such a powered AFO device are more difficult than that of the powered prostheses because we have to consider dynamics of a paralyzed leg or foot (inertia, viscosity, elasticity and voluntary/involuntary force from them). Additionally, weight saving is required strongly.

In this study, we suggest passive controllable AFOs with compact brake devises for dynamic gait controls [19]. In particular, the control of drop-foot can be realized by only using passive devices. The passive controllable AFO also have a great advantage for cost, safety and downsizing.

For the beginning of this paper, we describe development of a compact MR fluid brake. Next a passive controllable AFO system with this brake is explained. Then, we discuss a control method. And finally experimental results of gait control for a polio patient are presented.

2 DEVELOPMENT OF MR FLUID BRAKE

2.1 MR fluids (Magneto-Rheological fluid) and brake devices using this fluids

MR fluids (MRFs) are kinds of functional fluid, which is composite materials of non-colloidal solution and magnetic metal particles (e.g., iron particles). The diameter of the particle is 1~10 micrometers. Its rheological properties can be controlled by applying a magnetic field [20]. The response of its viscous change is very rapid (about several milliseconds) [21], [22] and repeatable with large range. By utilizing the viscosity change of the MRFs with share flow mode, we can develop a brake device with good responsibility, higher torque/inertia ratio than any other conventional brakes, e.g. powder brakes and electromagnetic brakes [21]-[23]. In recent years, a prosthetic knee using the MR fluid braking device was reported [11] and has produced by a prosthesis maker [13].

2.2 Basic Structure of Compact MRF Brake (CMRFC) with Multi-Layered Disks and Narrow- Gaps

A conceptual drawing of a compact MRF brake (CMRFC) is shown in Fig.2. A coil is rolled round an output shaft and it generates the magnetic flux shown by dashed lines in the drawing. Multi-layered disks are fixed on the stator-parts and rotator-parts. The MRF is filled between these disks.

As you see in this figure, multi-layer structure is utilized for enlargement of the brake torque. However, we have to impress a magnetic flux density of about 0.5~1.0 tesla on the MRF layers in order to make use of the viscous change of the MRF effectively. Because the MRF is a material which has a

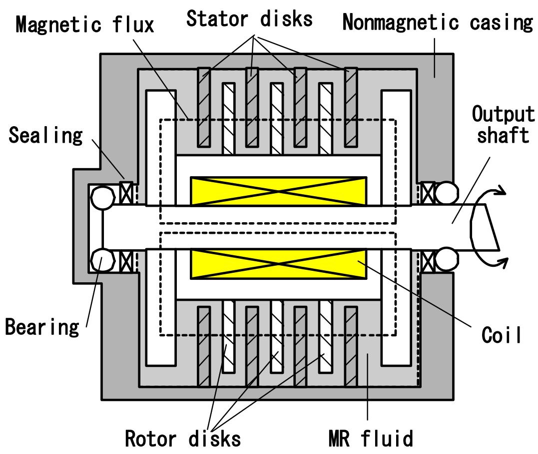


Figure 2: Basic structure of a compact MRF Brake with multi-layered disks

low magnetic permeability, it is important to reduce the total gap of the MRF layers for the reduction of electric power consumption.

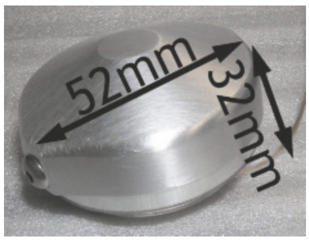
Based on this restriction of the gap-size, total thickness of the MRF layers was decided to be less than 0.5~1.0 mm, empirically. Therefore, we suggest to utilize narrow-gaps of 10~100μm for each gap. The gap-size, the diameter of disks, the number of the layered disks and the number of turns of the magnetic coil should be estimated based on results of magnetic analyses and processing (or assembling) accuracy of the multi-layered disks.

2.3 Design Method of Compact MRF Brakes

We formulated the way to estimate output torques of the CMRFC. The analysis flow is shown as follows;

(1) Geometric design of a CMRFC with 3-D CAD software,

(2) Magnetostatic analysis for the estimation of the magnetic



(a) Front view

(b) Rear view

Figure 3: Compact MRF Brake (CMRFC)

TABLE I
SPECIFICATIONS OF CMRFC

Total thickness [mm]	32
Outer diameter [mm]	52
Number of disks	9
Number of MR fluid layer	18
Turning number of coil	191
Idling torque [Nm]	0.15
Maximum torque [Nm]	6.0
Weight [g]	237

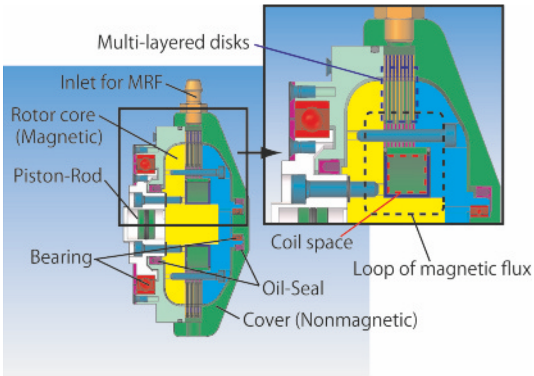


Figure 4: Cross section of CMRFC

flux density with CAE (FEM) software,

(3) Decision of a yield stress of the MRF depending on the results of the process (2),

(4) Calculation of a maximum torque depending on the results of the process (3) and other size-parameters.

In the process (2), we need to input nonlinearity data (B-H curve) of ferromagnetic materials (silicon steal etc.) and the MRF. Commercially produced MRF (140CG, Lord Corp.) was used as an MRF in this study.

In the process (3), we referred to the characteristics data between the yield stress and the magnetic flux density of 140CG, which is presented by Lord Corp. [24].

Figures 3 show a developed compact CMRFC. Specification data of this clutch is shown in table 1. Multi-layered disks are fixed with gaps of 50μm accurately. The MRF (140CG, Lord Corp.) is filled between these gaps completely. Figure 4 shows a cross-sectional view of the CMRFC. In this figure, black broken line means a loop of the magnetic flux. The condition of filled fluid can be checked through the inlet hole made direct above multi-layered disks as shown in the sectional view of Fig.4. The number of turns of the coil is 191 turns. According to the result of a magnetostatic analysis mentioned above, application of the electric current of 1A to the coil generates magnetic flux density of 0.55~0.65 tesla in the MRF layers.

2.4 Characteristics of the Compact MRF Brake

Torque characteristics of this device were measured by an experiment setup as shown in the Fig.5. The casing of MR brake was fixed on a turning table and rotated at a constant speed of 1 rad/s. A lever-arm of which length is 250 mm was fixed on the output shaft of the brake. We measured braking torque with a force sensor connected with this arm.

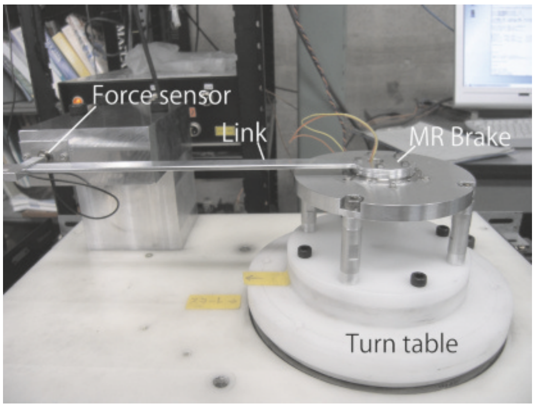


Figure 5: Experimental setup of torque tests

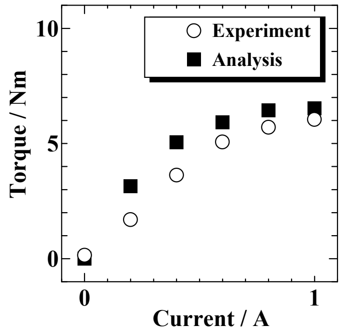


Figure 6: Static torque of CMRFC

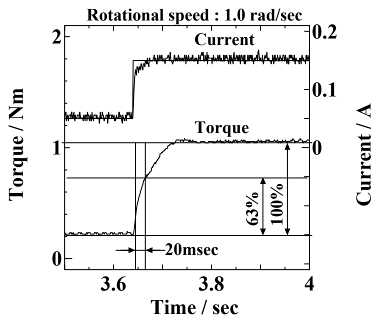


Figure 7: Step response of CMRFC

At first, static torque tests were conducted at constant speed of 1 rad/s. The experimental results are shown in Fig.6. White circles are experimental data. Black squares mean the results of the analysis mentioned in the previous session. As shown in this figure, the experimental result under the static condition indicates good similarity to the analytic result. Therefore, we can predict the maximum torque of the CMRFC with the suggested method.

Secondly, we also conducted dynamic torque tests (step-response). Time constant (63% response) of step-response was about 20 milliseconds as shown in Fig.7. This results

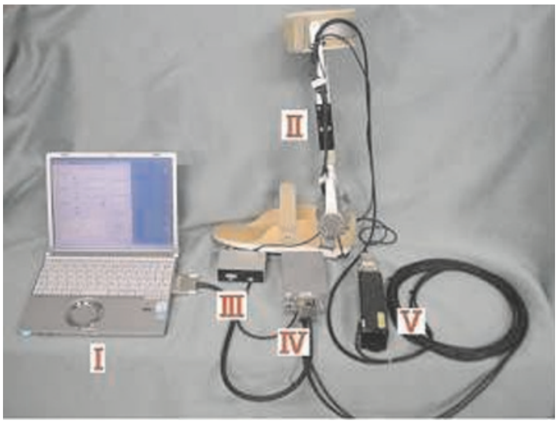


Figure 8: Control system of i-AFO

shows that the developed CMRFC has a better response time than that of conventional powder clutches.

3 PASSIVE CONTROLLABLE AFO SYSTEM USING COMPACT MRB

Figure 8 shows a system of the controllable AFO with the developed compact MRF brake. In this section, we describe this AFO system and the control method.

3.1 Sensing components for measuring gait

This AFO system has three types of sensing devices as follows;

(a) A potentiometer at the ankle position,

(b) Foot switches under the toe and the heel,

(c) A bending moment sensor built in a brace.

A commercially produced potentiometer was used to measure the angle of the ankle joint. Foot switches detect contacts of the ground and a toe or a heel. The bending moment sensor developed by BL Autotech, Ltd., Japan was utilized to measure bending moment straining against a brace.

3.2 Control system for the AFO

As shown in Fig.8, this system has five components as follows;

I: A controller and a user-interface (Lap-top PC embedded a multi-functional card of A/D and D/A),

II: The AFO with the previous MRB (i-AFO),

III: A battery box (power supply to amplifiers and sensors),

IV: A current amplifier,

V: A sensor amplifier (to amplify the signal of the bending moment sensor).

Braking torque is controlled with an electric current applied from the current amplifier. Reference signals for the current amplifier are outputted from the D/A card. The signals from four sensors (a potentiometer, toe and heel switches and a bending moment sensor) are sent from the A/D board to the controller.

3.3 Control method for the passive controllable AFO

Figure 9 shows a signal form of braking torque depending on the gait phases (see Fig.1). A screen copy of the control panel is shown in Fig.10. Alphabets in Fig.9 and Fig.10 are

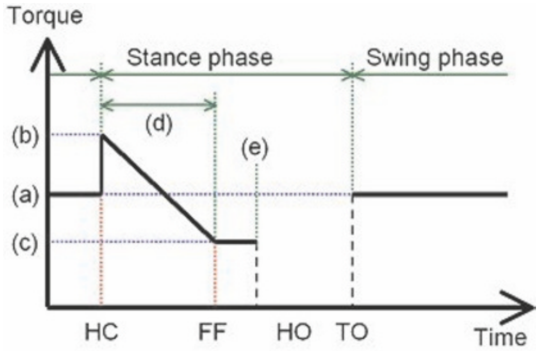


Figure 9: Signal form for braking torque

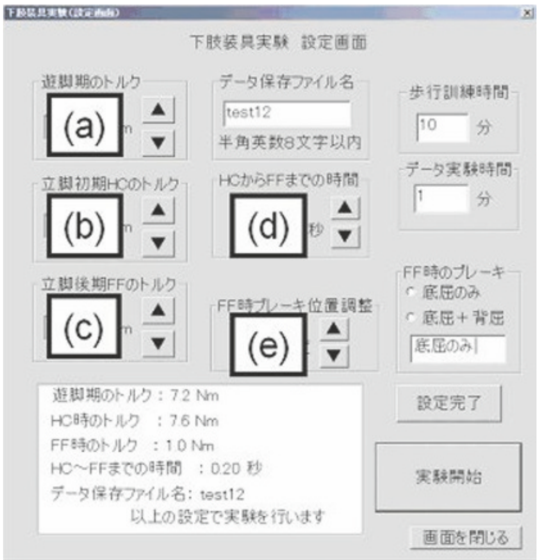


Figure 10: User interface view

parameters that can be varied for each subject. Each parameter can be set in the panel shown in Fig.10.

The meanings of each parameters are as follows;

(a) Torque of swing phase,

(b) Torque of heel-contact,

(c) Torque of foot-flat,

(d) Period from heel-contact to foot-flat,

(e) An ankle angle for free torque.

During swing phase, the brake outputs torque (a). Then, it outputs torque (b) at heel contact, and decreases linearly until foot-flat (torque (c)). Period from heel-contact to foot-flat (d) is estimated through experiment. Torque (c) is kept until a certain ankle angle (e). These parameters are manually determined by repeated experiments.

4 GAIT EXPERIMENT FOR A POLIO PATIENT

4.1 Experimental condition

We tried a clinical test with this AFO system for a polio patient. Table 2 shows a data table of the subject. This polio patient has a flaccid paralysis on his right ankle, but in a daily life he lives without any orthosis. The drop-foot is main problem in his gait. We applied our AFO system to his right

TABLE II
DATA TABLE OF A SUBJECT

Sex	Male
Age	59
Symptom	Polio Right ankle flaccid paralysis, sensory paralysis
Height	157 [cm]
Weight	44 [kg]

Figure 11: Appearance of the experiment

ankle. Figure 15 shows an appearance of the AFO system and his right leg. Working length of this trial was set about 5 m. The experiment was started with the subject standing on a neutral position shown in Fig.11.

4.2 Results

Figure 12 shows an experimental result without control. And figure 13 shows an experimental result with control mentioned above. Figure 12 and 13 include three data of ankle angles, reaction forces, and bending moments. Periods with zero-GRF must be swing phases and others must be standing phases.

4.3 Discussion

As shown in Fig.12, the gait without control has features as follows;

(1) In swing phase, the subject could not maintain the dorsal flexion because of the drop-foot.

(2) At the moment of reaching ground, GRF lacked smoothness.

(3) The subject could not contact at heel in the initial contact. Therefore, he can not walk smoothly.

On the other hand, the gait with control (Fig.13) has features as follows;

(1) In swing phase, the subject could maintain the dorsal flexion, and prevent the drop-foot.

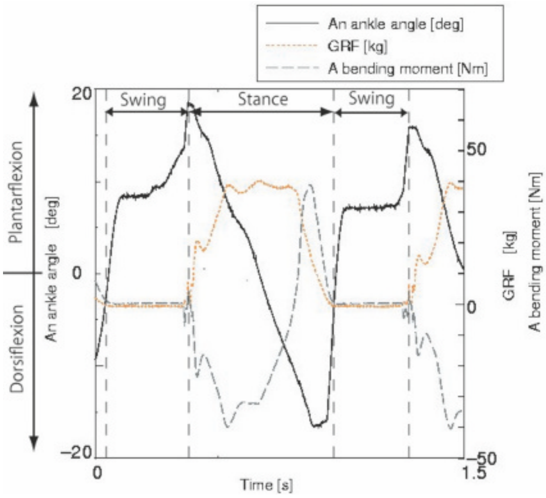


Figure 12: Experimental result without control

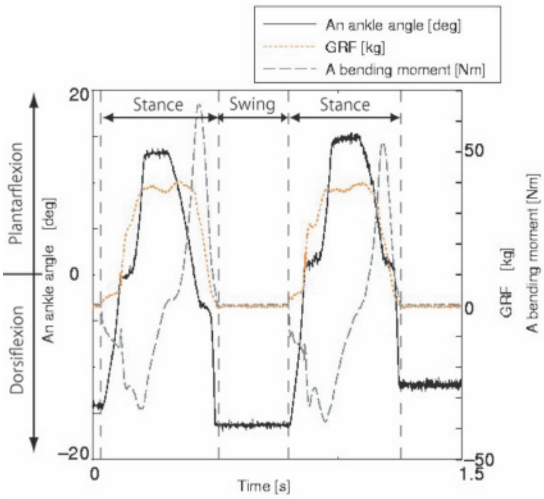


Figure 13: Experimental result with control

(2) The subject was able to contact to the ground at heel.

(3) The maximal bending moment with control was larger than one without control. Therefore, subject could transfer his weight to forward smoothly.

(4) Walking cycle was shorter than one without control.

As above, we realized to control the gait of the polio patient with developed passive controllable AFO system. It prevented the drop-foot in the swing phase and the slap-foot at the heel-strike.

5 CONCLUSION

In this study, a compact and high-torque MR fluid brake was developed. Design method with magnetostatic analyses for the MRF brake was described. We developed a passive controllable AFO with the compact MRF brake. This AFO system has three types of sensors, a potentiometer for the angle, a bending moment sensor for the bending moment on AFO and foot switches to detect contacts against the ground. By using this AFO system, we tried a gait experiment for a polio patient. A control method for this AFO system was developed. Consequently, we could prevent the drop-foot in swing phase, and slap-foot at the heel-strike. The experimental result shows that the gait of the polio subject was improved.

6 ACKNOWLEDGMENTS

A part of this study was financially supported by the Japanese grant, new energy industrial technology development organization (NEDO), "Research and development in NEDO for practical application of human support robot" in 2006~2008. Other part of this study was also supported by Global COE Program, "in silico medicine" at Osaka University. And gait experiments are conducted with a close support of Hashimoto Artificial Limb Manufacture Co. Ltd.

7 REFERENCES

[1] Physical Medicine & Rehabilitation (Third edition): chapter 5 Gait Analysis, Saunders Elsevier, 2007, pp93-110.

[2] Cappozzo, A., 1984, Gait analysis methodology, Hum. Mov. Sci., 3:27-50.

[3] Inman, V., Ralston H. and Todd, F., 1981, Human walking, Baltimore, Williams & Wilkins.

[4] Bampton, S., 1979, A guide to the visual examination of pathological gait, Philadelphia, Temple University-Moss Rehabilitation Hospital.

[5] Esquenazi, A. and Hirai, B., 1991, Assessment of gait and orthotic prescription, Phys Med. Rehabil. Clin. North Am., 2:473-485.

[6] Redford, JB., 1984, Orthoses, In: Basmajian, JV., and Kirby, RL., eds., Medical rehabilitation, Baltimore, Williams & Wilkins, 101.

[7] Herr, H., 2003, Cyborg Technology-Biomimetic Orthotic and Prosthetic Technology, In: Bar-Cohen, Y., and Breazeal, C., eds., Biologically Inspired Intelligent Robots, Washington, SPIE PRESS, 103-143.

[8] Popovic, D. and Sinkjaer, T., Control of movement for the physically disabled, Springer-Verlag London.

[9] Stein, J., Rolf, R. and Tepavac, D., 1990, Active suspension abov-knee prosthesis, Goh JC 6th Int Conf Biomech Eng, 317-320.

[10] Wilkenfeld, A., 2000, An auto-adaptive external knee prostheis, Ph.D Thesis, MIT.

[11] Herr, H., Wilkenfeld, A., and Olfa, B., 2003, Speed-adaptive and patient-adaptive prosthetic knee, U.S. Patent No.823931.

[12] Otto Bock, http://www.ottobock.com/

[13] Ossur, http://www.ossur.com/

[14] Zlatnik, D., Steiner, B., and Schweithzer, G., 2002, Finite-State Control of a Trans-Femoral (TF) Prosthesis, IEEE Transactions o control systems technology, 10/3:408-420.

[15] Dollar, A.M., Herr, H., 2007, Active Orthoses for the Lower-Limbs: Challenges and State of the Art, Proceedings of the 2007 IEEE 10the International Conference on Rehabilitation Robotics, 968-977.

[16] Gordon, K.E., and Ferris, D.P., 2007, Learning to walk with a robotic ankle exoskeleton, Journal of Biomechanics, 40:2636-2644.

[17] Nouri. B., M.Y. and Zaidan, A., 2006, Conputer control of a powered two degree freedom reciprocating gait orthosis, ISA Transaction, 45/2:249-258.

[18] Blaya, J.A., and Herr, H., 2004, Adaptive Control of a Variable-Impedance Ankle-Foot Orthosis to Assist Drop-Foot Gait, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 12/1:24-31.

[19] Furusho J., Kikuch, T., Tokuda, M., Kakehashi, T., Ikeda, K.,, Morimoto, S., et al., 2007, Development of Shear Type Compact MR Brake for the Intelligent Ankle-Foot Orthosis and Its Control (Research and Development in NEDO for Practical Application of Human Support Robot), Proceedings of IEEE International Conference on Rehabilitation Robotics 2007, 89-94.

[20] Carlson, J. D., and Jolly, M.R., 2000, MR fluid, foam and elastomer devices, Mechatronics, 10:555-569 .

[21] Yamaguchi, Y., Furusho, J., Kimura, S. and Koyanagi, K., 2005, Development of High Performance MR Actuator and Its Application to 2-D Force Display, International Journal of Modern Physics B, 19/7-9:1485-1491.

[22] Takesue, N., Furusho, J., Kiyota, Y., 2004, Fast Response MR-Fluid Actuator, JSME International Journal, Series C, 47/3:783-791.

[23] Kikuchi, T., Furusho, J., Yamaguchi, Y., and Kimura, S., 2006, Design of the high-performance MR brake and its characteristics, Proceedings of the 10th International Conference on ER Fluids and MR Suspensions, 132.

[24] LORD TECHNICAL DATA, MRF-140CG Magneto-Rheological Fluid, http://www.lord.com/

Development of Force-Measurable Grip and Software for "PLEMO", Rehabilitation System for Upper Limbs Based on Physical Therapy

Takehito Kikuchi[1], Takuya Ozawa[1, 2], Kazuki Fukushima[1], Junji Furusho[1]

[1] Dept. of Mechanical Engineering, Graduate School of Engineering, Osaka University, Osaka, Japan

[2] Rehabilitation section, Kano general hospital, Osaka, Japan

Abstract

In recent years, many researchers have studied the potential of using robotics technology to assist and quantify the motor functions for neuron-rehabilitation. Some kinds of haptic devices have been developed and evaluated its efficiency with clinical tests, for example, upper limb training for hemiplegic patients. However, almost all the devices are active-type (motor-driven) haptic devices and they basically require high-cost safety system compared to passive-type (brake-based) devices. In this study, we developed a new practical haptic device "PLEMO-P1"; this system adopted two ER brakes as its force generators. In the previous system, we could not detect symptoms of abnormal movement of patients, for example synergy patterns of stroke patients, because of a loss of sensors. In this paper, we developed new sensing device for detecting such abnormalities of patient's movements. The purpose of this study is to detect abnormal movements with new PLEMO system and notice it to users. As a preparation, we conducted reaching tests with this device in Lab-level tests. By comparison with Demo-stroke movement and a normal movement, we recognized some differences of wrist angle and gripping forces between these two movements.

Keywords:

Upper limbs rehabilitation, Synergy pattern, Haptic device, Force feedback, Rehabilitation robotics

1 INTRODUCTION

The increasing numbers of the aged people and their physical deteriorations have been one of the most serious problems in many countries. Many stroke patients suffer from disabilities which restrict Activities of Daily Living (ADL), e.g., reaching actions. Therefore, sufficient rehabilitative training is necessary for such patients.

In general, therapists make rehabilitation programs based on their inspections and measurements of each patient. However, it is difficult to adopt appropriate rehabilitative programs for all patients, because the evaluation method is based on experiences of each therapist. Nowadays, Evidence Based Medicine (EBM) is required strongly in the field of rehabilitation. Therefore, rehabilitation systems that utilize robotics technologies are expected for; 1) quantification of the effect of the rehabilitative training; 2) enhancement of the motivation for patients with creating new training methods (many patients are giving up rehabilitative training because of its boredom); 3) improvement of the efficiency of physical therapist's works.

As examples of the therapeutic robots, Krebs H.I., et al. [1] have developed active (motor-driven) haptic device (MIT-MANUS) and conducted many clinical tests for upper limbs rehabilitation. However, the motor-driven robots basically require high-cost safety system compared to passive (brake-based) devices. Book W.J., et al. [2] have developed passive haptic devices. In their system, conventional powder brakes were used as haptic generators. Generally, the response time of the powder brake is more than 100ms and it causes lack in quality of force feedback.

To solve these problems, we have developed passive haptic devices using ER Fluid Brakes (ER Brake) [3]. Due to the rapid response of the ER Brake (2~3ms), our haptic device presented good performance as a force display. In this study, we improved this system (ER brake-based system) as a practical rehabilitation system for upper limbs. To meet demands for the rehabilitative training in 3-D space, we developed a new haptic device "Quasi-3-DOF Rehabilitation System for Upper Limbs" or "PLEMO-P1"[4]. PLEMO-P1 has 2-DOF force-feedback function in its working plane but the working plane can be adjusted its inclination.

However, the previous systems could not detect symptoms of abnormal movement of patients, for example synergy patterns of stroke patients. If rehabilitative trainings are continued with such abnormal movements, trainer possibly learns bad pattern of movements and it prevent from proper learning of movements.

In this study, the symptom of the synergy movement pattern is detected with automatic operation by a new sensor grip device. This device consists of three sensors; 1) a gripping force sensor, 2) a reaction force sensor against the table and 3) an angle sensor for a wrist. The purpose is to detect abnormal movements by using this device and give patients a warning and additionally present a better movement direction. In this paper, we describe basic structure of the sensor grip device and lab-level experiments for reaching tests as a preparation of clinical tests.

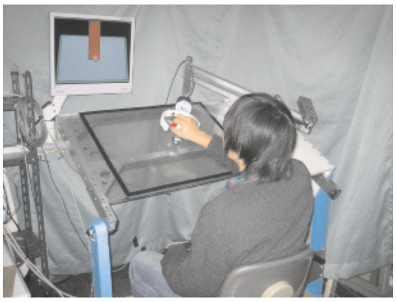


Figure 1: Quasi-3-DOF rehabilitation system "PLEMO-P1"

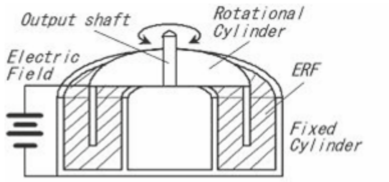
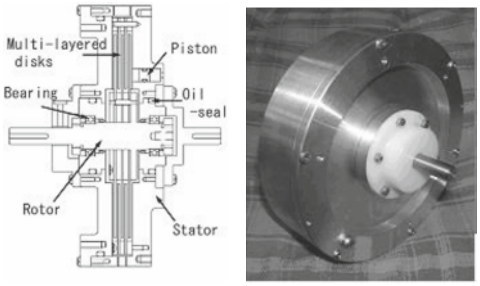


Figure 2: Basic structure of ER Fluid Brake



Figures 3: Cross section (left) and picture (right) of ERB

2 QUASI-3-DOF REHABILITATION SYSTEM FOR UPPER LIMBS, "PLEMO"

We have developed a Qasi-3-DOF rehabilitation system for upper limbs, "PLEMO-P1" [4] as shown in the figure 1. PLEMO-P1 is a stand-alone robot and a passive-type robot: it can present passive-type force display which can output several kinds of virtual force, for example resistance, viscosity, vibration etc. This machine has two active degrees of freedom (DOF) in a working plane and one passive DOF of the inclination of the working plane. It has some application software for rehabilitation.

In the training, an operator grasps a handle of PLEMO, watches a display and plays application software as rehabilitation training and evaluation test. This system can control and measure the position of the hand of the operator, and feedback force sense to the operator depending on the situations defined by the application software. In this section, we describe the mechanism of both the brakes and the PLEMO system.

2.1 ER Fluid Brake (ERB)

ER fluid is one of functional fluids, whose rheological properties can be changed by applying an electrical field [5]. To use this fluid as a working fluid, we can make electrically controllable brake (ER Fluid Brake, or ERB, simply) with high-performance (good rapidity and repeatability of brake torque) [6]. We use this brake as force generators of a new rehabilitation system (force-feedback system) in this paper.

TABLE 1
SPECIFICATIONS OF ER BRAKE

Total weight	2.3 kg
Diameter	15cm
Height	4cm
Maximum brake torque	4.0Nm (at 3.0kV/mm applied)
Idling torque	0.1Nm (at 0.0kV/mm applied)
Num. of rotor disks	3
Disk gap	1mm
Time constant of response	2~3 msec

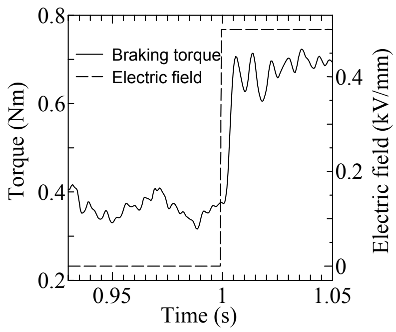


Figure 4: Step response of ERB

Figure 2 shows a basic structure of a cylindrical-type ER brake. It consists of fixed cylinders and rotating cylinders with the ER fluid between them. The two groupes of the cylinders also play the role of pairs of electrodes. The rotating cylinder is fixed on the output shaft and driven by external forces through this shaft. When a voltage is applied between the pair of cylinders, the electric field is generated in the ER fluid between them, and then the viscosity of the fluid increases. This increase of viscosity generates the braking torque and reduces the rotational speed.

Figures 3 show a cross-sectional view and appearance of the brake developet for the PLEMO system. As shown in the left drawing of Figs. 3, this brake consists of multi-layered disks (three pairs of stator-disks and rotor-disks). ER fluid is filled between the rotor-disks and stator-disks. As a result, six layers of ER fluid generate the brake torque with change of the fluid. Piston mechanism works for the prevention of liquid spill with the expansion of the fluid.

Table 1 shows specifications of the brake. We can control the brake torque from 0.1Nm to 4.0Nm with the electric field from 0.0kV/mm to 3.0kV/mm, respectively. As shown in Fig. 4, the time constant of response is 2~3 ms [6]. In this figure, a solid line is the torque response of the ER brake, and a dashed line is an input of the electric field to the brake. Thanks to this rapidity, a haptic device using this ER brakes

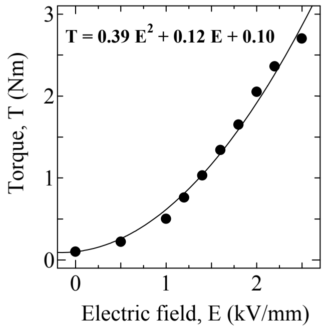


Figure 5: Brake torque vs. Electric field of ERB

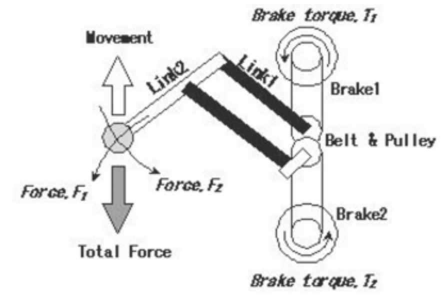


Figure 6: Parallel linkage of PLEMO

can realize a high frequency response (e.g. an impact force of virtual hockey).

Figure 5 shows the brake torque of the ER brake depending on the electric field applied to it. This characteristic has good repeatability and we can formulate this relation as follows;

$$T = 0.39E^{2} + 0.12E + 0.10 \tag{1}$$

where, T represents the brake torque [Nm] and E represents the electric field [kV/mm]. The resolution of force depends on other components described as the next section.

2.2 Mechanism of Force Control

Force control unit consists of two ER brakes and these brakes generate output-force on a handle by a parallel linkage. Figure 6 shows parallel linkage of this system. As shown in this figure, T_1 or brake torque of brake1 generates force F_1 and similarly torque T_2 generates force F_2. If the movement of the handle is vertical direction like this figure and this system generates force F_1 and F_2, the user feel resistance against his movement direction. If these forces are constant, users feel constant resistance. In other case, if these forces are proportion with velocity of the movement, the users feel viscous force. Necessary torques T_1 and T_2 are calculated from required forces F_1 and F_2 based on the inverse kinematics. Lengths of the link 1 and the link 2 are 450mm. This value is designed on the basis of the manipulability-analysis.

2.3 Control System

Figure 7 is a schematic drawing of the structure of PLEMO and signal flow chart of this system. Absolute encoders (FA Coder, TS566N320, Tamagawa Seiki Inc., Japan, resolution: 17bits) measure the rotational angle of brakes. We can

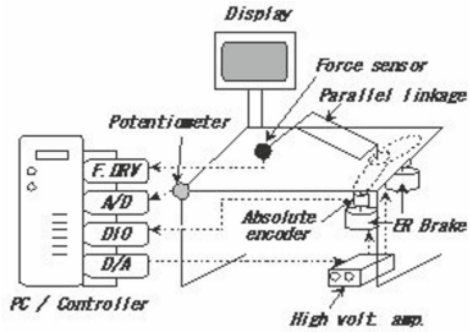


Figure 7: Structure and signal flow of PLEMO

calculate the position and the velocity of the handle depending on each angle of brakes. Digital Input/ Output (DIO) board (PCI-2154C, Interface Inc., Japan) loads this information to a controller (personal computer). Operating handle includes a force sensor (IFS-70M35A, Nitta Inc., Japan), and operating force is measured by this sensor. A potentiometer (CP-2F, Midori Precision Inc., Japan) measures the inclination of the worktable and the angle is loaded by Analog/Digital converter (A/D) board (PCI-3165, Interface Inc., Japan, resolution: 16bits). The brake torque of the ER brake is controlled by applied voltage from high voltage amplifiers (HEOP-3P10-LS, Matsusada Precision Inc., Japan). Digital/Analog converter (D/A) board (PCI-3338, Interface Inc., Japan, resolution: 12bits) outputs the reference signal to the amplifiers.

Depending on the above specifications of the components, specifications of the PLEMO system are as follows;

1. Resolution of the angle is 4.8×10^{-5} rad and resolution of the displacement is 0.02 mm for each DOF.
2. Resolution of the braking torque is 6.5×10^{-3} Nm and resolution of the force at the end-effecter is 1.4×10^{-2} N for each DOF.
3. Range of motion on the table is 600 mm (W) * 500 mm (D).
4. Adjustable angle of the inclination is from -30 to 90 degree.

PLEMO changes the angle of table from vertical plane to horizontal plane. So, we can conduct from vertical training to horizontal training by only one system. Total size of the system is 1000 mm (W) * 600 mm (D) * 700 mm (H), except for the display. This is similar to the size of general-purpose office desks. This passive system is more compact, simple, and reasonable for the cost than conventional actuator-type systems.

3 DEVELOPMENT OF SENSOR GRIP DEVICE FOR DETECTION OF ABNORMAL MOVEMENT

3.1 Concept

In the previous PLEMO system, we could not detect symptoms of abnormal movement of patients, for example synergy patterns of stroke patients, because of a loss of sensors. Furthermore, in reaching tests with the previous system, we observed unfavourable movements of stroke patients as follows;

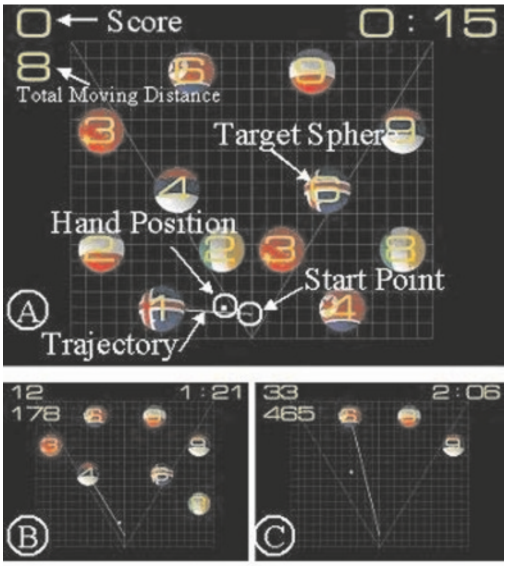
1. Excessive reaction force against the table,
2. Excessive gripping force,
3. Abnormal flexion of the wrist.

The "synergy movements" mentioned above must have caused such symptoms. In other words, if we can measure such information, we have a chance to detect their synergy movements and give patients a warning for their abnormality. It becomes great help for their effective recovery.

3.2 Basic structure and its performance of the sensor grip device

Figures 8 show mechanisms of our former and current handle. The former handle mechanism is shown in Figs. 8(A): it is gimbals-structure with no sensors. On the other hand, the current handle consists of three kinds of sensors (shown in Figs. 8(B)): a grip force sensors; a potentiometer for yaw rotation of patient's wrist; an axis force sensor for detecting the reaction force against the table (simply called "reaction force" in this paper). The size of grip is 150mm in height and 30mm in radius, which allow the operator to bend his/her wrist in ranges of - 90 to + 90 deg.

The mechanism of measuring the grip forces is that strain gages attached to four props can measure the applied forces: these four props are installed on the foundation of the grip (shown in Figs. 8(D)). Figs. 8(C) shows a part of these props. In order to measure yaw rotation angle of patient's wrist, we incorporated a potentiometer into this handle system (shown in Fig. 8(D)). Moreover, we installed load-cell on the bottom side of this grip to detect the reaction force against table during robot therapies. When a patient moves and pushes the robot hand strongly against the working plane, this sensor would react to his/her movement and alert him/her in real-time, which would help him/her to perceive his/her abnormal synergy pattern. The specifications of current handle mechanism can be summarized in Table II.



Figures 9: Reaching software

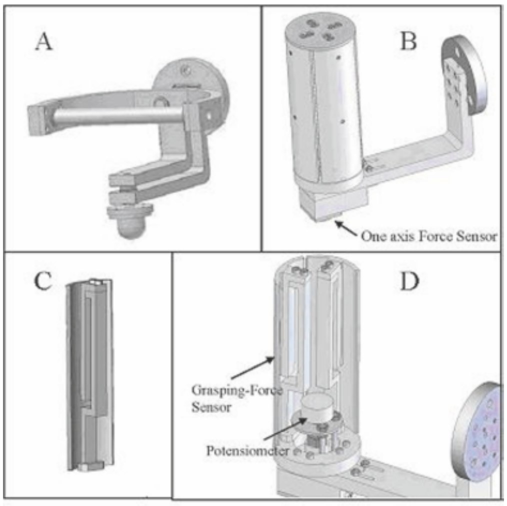
TABLE II
SPECIFICATIONS OF CURRENT HANDL

Height	15 cm
Radius	3 cm
Yaw Rotation	- 90 to + 90 degree
Grasping-Force Range	0 to 12 N

4 REACHING TEST FOR RECOGNIZING SYNERGY MOVEMENT

4.1 Software

We have developed a new application to detect patient's synergic movements (shown in Fig. 9 (A)).



Figures 8: Sensor grip device: (A) Former handle (passive gimbals), (B) Current handle (sensor grip), (C) A part of grip force sensor, (D) inner mechanism of the grip device

In this program, the patient is instructed to the position of target point (sphere shape in this figure) with each numbers in a monitor. The position of the hand is shown as a white dot, which is moved depending on the movement of the end-effecter. The task starts when a user operates his hand to the start position displayed as bottom side of a center of Fig.9.

Each target sphere has the following parameters: 1) position; 2) score; 3) area number; 4) priority level. When one of the fixed target spheres is activated by ascending order of the target's priority level, a straight trajectory connects to the starting point and the target sphere is displayed as a white line in the monitor (Fig.9 (B)). When the patient moves his/her hand along the trajectory and reach a target sphere, the score of the target is added to the patient's score, which displayed on upper left of the monitor. Total distance [cm] of the movements of patient's hand is displayed just beneath this score. After returning his/her hand to the starting position, a next target sphere is activated (Fig.9 (C)).

The gaps between the robot desired position along the trajectory and the patient's hand position are accumulated depending on the target's area number: after finishing the training, the average trajectory error of each area number is calculated and evaluated, which enables us to more clearly identify the patient's weak area. In general, stroke patients tend to move their upper limbs following synergy pattern; therefore, right side hemiplegic patients will have difficulty with their movements toward right-front direction more than

Figure 10: Explanation of targets

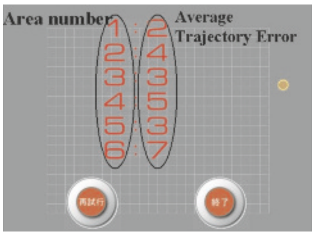


Figure 11: Evaluation view

any other direction. Figure 10 shows a certain combination of area numbers. Figure 11 is a result of demonstration with the healthy volunteer. In center of this figure, left numbers correspond to the area numbers and right numbers indicate the error of the trajectory (the score is calculated with division of the error area by the length of the trajectory). As shown in Fig. 11, the patient can check results of the training immediately just after each training. Moreover, this system can record data of the movement; therefore, this system can evaluate a recovery degree quantitatively by other evaluation method. Optional features of this training are;

1. to present virtual force against the correct direction. This resistance activates patient's intention to move their own arm,

2. to alert the patient by sound, telop and virtual force according to the griping force. The patient is required to keep his/her griping force constant in a certain level while performing the required task.

Depending on the disability of the patient, physiotherapist can select these optional features above. Moreover, a physiotherapist can set the parameters of target spheres: 1) position; 2) score; 3) area number; 4) priority level by himself/herself, which enables us to administer various kinds of effective training to stroke patients.

4.2 Method

The subject is a healthy volunteer of 40 years old (physical therapist). PLEMO was used to measure the features of the movement of the healthy person and a stroke patient (right paralysis) demonstrated by this subject (we call it "Demo-stroke" in this paper). All experiments were executed with sitting on the chair and fixing the back on a back rest. The reaching program mentioned above was used as shown in Fig.12. The start position was defined as X = 0cm, Y = 0cm. Two target positions were selected as follows;

Left-side target: X = -10cm, Y = 20cm,

Right-side target: X = 10cm, Y = 20cm.

Reaching process was defined as a process from the start position to a target position. On the other hand, pulling process was defined as a process from the target to the start position.

4.3 Results & Discussion

In this paper, we show experimental results of gripping forces (see Figs.13) and wrist angles (see Figs.14). According to these results, we can see big differences between normal movements and abnormal ones in both results of griping force and wrist angle.

Generally, Demo-stroke takes a lot of time for reaching actions. At the same time, its motions are demonstrated with strong grip forces, and high reaction forces against the table. Additionally, small vibrations are seen in the Demo-stroke. These features can be explained depending on the symptom of the stroke.

In terms of the gripping force, as shown in Figs.13, the force of the Demo-stroke is larger than that of healthy person in all directions. These may be caused by an exaggerated contraction of muscle, or abnormal muscular tonic (spasticity), depending on the synergic pattern of hemiplegia.

In terms of the wrist angle, at the same time, synergy movements works during Demo-stroke's movements and these wrong actions caused different pattern from normal tendencies. In normal movements, wrist angles tend to be constant during both reaching and pulling actions. However, in Demo-stroke movements, the wrist angle changed dynamically.

Of cause, we can not say the results hold true to real patients. But the difference will not probably be so large. At least, we can detect abnormalities with this device. For the next stage, we are planning to conduct clinical tests for real patients and investigate how we can detect their abnormalities with PLEMO system.

5 CONCLUSION

In this paper, we describe the motivation for the detection of patient's abnormality, mainly synergy movement of stroke patients. In order to detect these symptoms, we developed a new sensor grip device. In this report we attached this device on the PLEMO system, which is a quasi-3DOF rehabilitation system for upper limbs previously reported.

We also developed new reaching program for PLEMO and experiments were conducted. The subjects were a healthy person and Demo-stroke, which is a stroke patients demonstrated by the healthy person (experienced physical therapist). The experimental results show big difference between the Demo-stroke and the healthy person. We

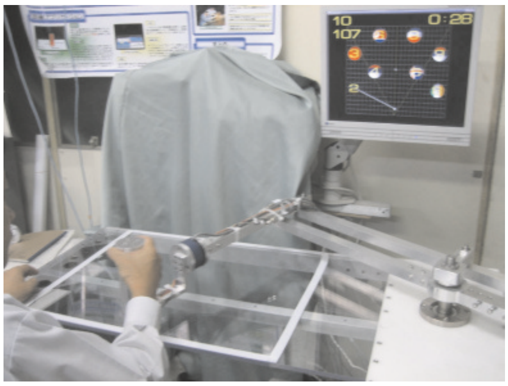
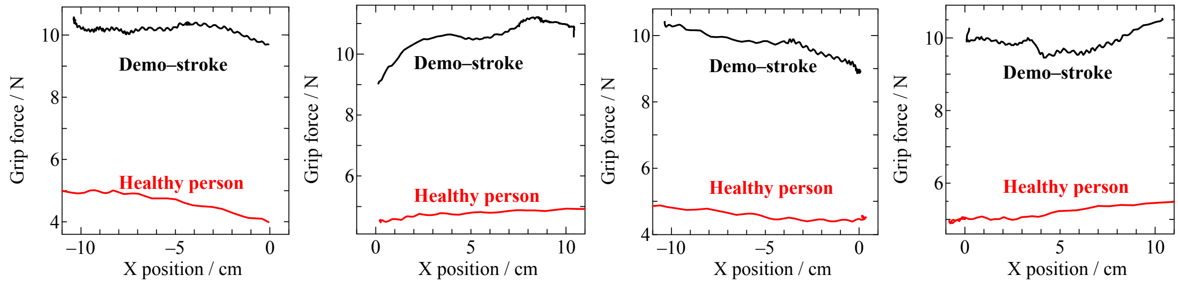
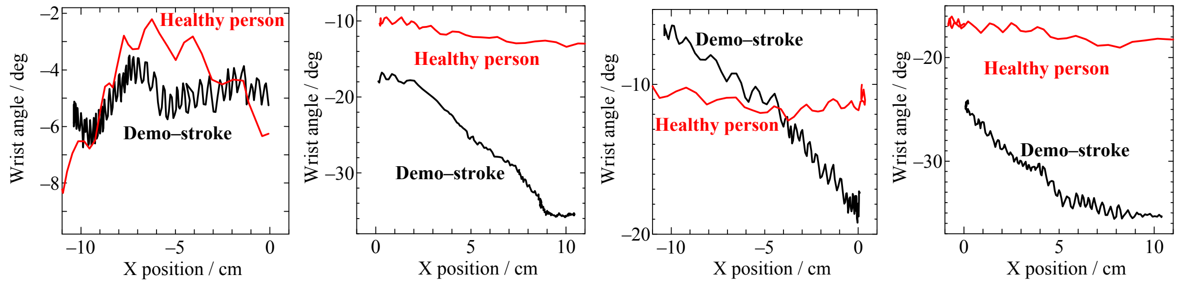


Figure 12: Scene of an experiment



(a) Left reach (b) Right reach (c) Left pull (d) Right pull

Figures 13: Experimental Results of the grip force



(a) Left reach (b) Right reach (c) Left pull (d) Right pull

Figures 14: Experimental Results of the wrist angle

construed the difference is based on the synergy movement of stroke.

6 ACKNOWLEDGMENTS

This work has conducted with close cooperation of medical staff of the Kano general hospital.

7 REFERENCES

[1] Krebs, H.I., Volpe, B.T., Aisen, M.L. and Hogan, N., 2000, Increasing productivity and quality of care: Robot-aided neuron rehabilitation, Journal of Rehabilitation Research and Development, 37/6:639-652.

[2] Munir, S., Tognetti, L. and Book, W.J., 1999, Experimental Evaluation of a New Braking System for Use in Passive Haptic Displays, Proceedings of the American Control Conference, 4456-4460.

[3] Kikuchi, T., Furusho, J. and Oda, K., 2003, Development of Isokinetic Exercise Machine Using ER Brake, Proceedings of 2003 IEEE International Conference on Robotics and Automation, 214-219.

[4] Kikuchi, T, Xinghao, H, Fukushima, K., Oda, K., Furusho J. and Inoue, 2007, A., Quasi-3-DOF Rehabilitation System for Upper Limbs: Its Force-Feedback Mechanism and Software for Rehabilitation, Proceedings of IEEE International Conference on Rehabilitation Robotics 2007, 24-27.

[5] Winslow, W.M., 1949, Journal of Applied Physics, 20: 1137-1140.

[6] Kikuchi, T., Furusho, J., 2002, Velocity Control of Brake Using Particle-Type ER Fluid and its Application, the Sixth International Conference on Motion and Vibration Control 2002, 837-842.

Mechanism Design and Software of Quasi-3-DOF Active-Passive Rehabilitation System for Upper Limbs,“Hybrid-PLEMO”

Ying Jin[1], Takehito Kikuchi[1], Junji Furusho[1], Hiroki Akai[1]
[1] Dept. of Mechanical Engineering, Graduate School of Engineering, Osaka University, Osaka, Japan

Abstract
Until now, some rehabilitation systems for upper limbs were developed and clinical effectiveness was reported in several studies for the aged people or stroke patients. It is important for designing a rehabilitation system which trains in the 3-DOF space because the upper limbs of humans works in 3-DOF space even expect for the wrist. However, there was few rehabilitation system for 3-D training with high safety. We developed the quasi-3-DOF rehabilitation system Hybrid-PLEMO, which has 2-DOF force-feedback function in working plane but its working plane can be adjusted the inclination. Hybrid-PLEMO is a compact, low-cost rehabilitation system for upper limbs with high safety by using ER brakes or ER actuators and can be switched between active type and passive type. Additionally, in Hybrid-PLEMO, we take direct-drive linkage mechanism by adding sub links. In this paper, we describe the mechanism of Hybrid-PLEMO.

Keywords:
Human robot co-existance ; Rehabilitation engineering ; Safety

1 INTRODUCTION

Depending on the aging society, stroke patients are increasing in many countries, and many of them develop ataxia: paralysis caused by brain stroke, or asynergia. Early detection of functional deterioration and sufficient rehabilitative training are necessary for such patients. In recent years, the needs for rehabilitation support systems are increasing, which use robot technology and virtual reality technology. These technologies are efficient for evaluating rehabilitation from a quantitative viewpoint.

Several rehabilitation robots for upper limbs have been proposed so far, and clinical effectiveness was reported in several studies for the aged people or stroke patients [1][2][3]. Although the upper limb of human works in 3-D space even except for the wrist, there was few rehabilitation support robot system for 3-D training with high safety. And if any, most of them have disadvantages in cost or ease of maintenance. Practical systems should be required to be more compact and better for maintenance.

In our current research, to meet these demands above, we developed a new haptic device "Quasi-3-DOF Rehabilitation System for Upper Limbs" or "PLEMO-P1" [4][5]. PLEMO-P1 has 2-DOF force-feedback function in working plane but its working plane can be adjusted the inclination. And PLEMO-P1 is a passive-type system using ER Brake, which can output several kinds of virtual force. Although that system meets some demands like that cost, high safety, and compact, it can't be adapted to the patient who has severe disabilities and who has difficulties in moving voluntarily, due to passive system.

Therefore, based on development of PLEMO-P1, we developed a new type of PLEMO. This new system has the same features as PLEMO-P1. Moreover, this system can be switched from active type to passive. We named this device "Quasi-3-DOF Active-Passive Rehabilitation System for Upper Limbs" or "Hybrid-PLEMO" (Figure 1).

Hybrid-PLEMO was developed to realize the difference of the rehabilitation training between active force feedback and passive force feedback. And it is adopted new ER-actuator we developed which consists of double-shaft ER-clutch.

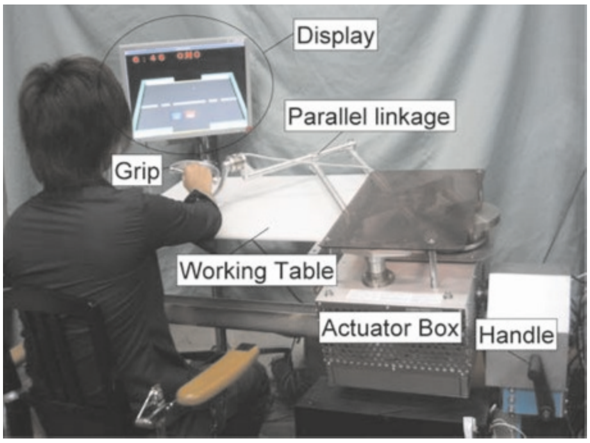


Figure 1: Hybrid-PLEMO

2 ER-CLUTH

2.1 ER-fluid

ER fluid is one of the functional fluids of which rheological properties can be changed by applying electrical fields [6]. In this paper, particle-dispersed-type ER fluid is used. Its shear stress depends on the application of electric field from 0.0kV/mm to 2.0kV/mm and does not depend on shear rate.

By using this fluid as working fluid, we can construct electrically controllable brake (ER brake) or clutch (ER clutch) with high-performance: highly responsive (several msec) and repeatability of brake torque. And we can construct high performance clutch-type actuation system with ER clutch (ER actuator, Figure 2). We use this brake and actuator system for force generators of our rehabilitation system.

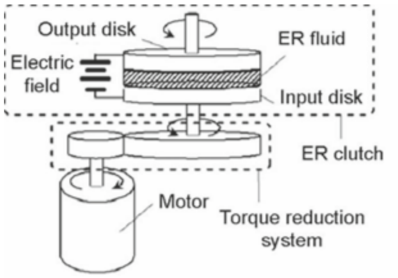


Figure 2: ER-clutch-type actuation system

2.2 Double shaft ER-clutch

Figures 3 show a sectional view of double shaft ER-clutch device developed in this study. This device has two groups of pair of multilayered disks in its package. Input disks of the two groups are fixed on the casing. When the casing is rotated by motor, input disks are rotated simultaneously and the device works as a clutch. When the casing is stopped, input disks are stopped and the device works as a brake. Two groups of output disks are fixed on inner shaft and outer shaft of output shaft, respectively. The particle type ER fluid is filled between each disk and we can control 2-DOF output torque independently.
Diameter and height of this device are 192mm, 225mm, respectively. Diameter, thickness and gap of disks are 155mm, 1.0mm, 1.0mm, respectively. Numbers of disks are 4 for each group. Figure 4 shows output torque of this device.

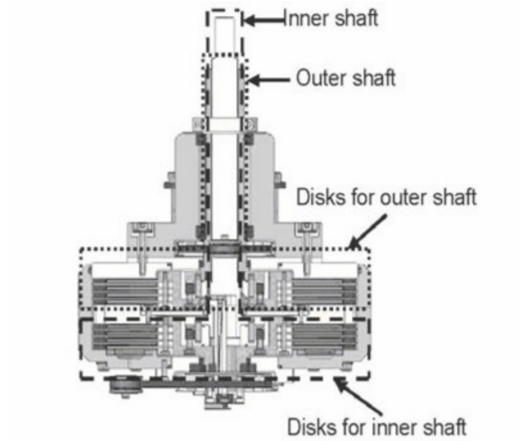


Figure 3: Double shaft ER-clutch

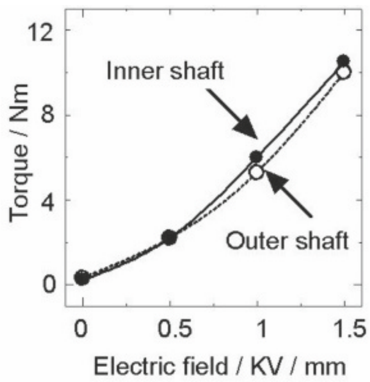


Figure 4: Output torque of double shaft ER-clutch

3 HYBRID-PLEMO

3.1 Quasi-3-DOF mechanism

"PLEMO" has 2 controllable DOF on a working plane and 1 passive DOF of the inclination of the working plane as shown in Figure 5. We defined this working space as a "Quasi-3-DOF Workings pace". An operator grasps a handle on the end-effecter of its arm, watches visual information on a display and plays application software as rehabilitative training and evaluation test.

3.2 Actuator Box

Haptic force on the end-effecter of Hybrid-PLEMO is controlled by torque-control-unit with ER actuators mentioned above. In Figure 2, the motor, which works for an input part of the clutches, is rotated by simply constant voltage in order to assure high safety of the clutch-type actuator. Therefore, the rotation direction of the ER actuator is one way. We need two actuators of CW (clockwise) and CCW (counterclockwise) direction for one controllable DOF.
To realize two controllable DOFs of PLEMO, we utilized two double shaft ER-clutch devices described above. The one is rotated in CW, the other is rotated in CCW. Input parts of the ER actuators are shown in Figure 6. As shown in this figure, both CW and CCW direction are generated by gears from one way rotation of a DC servo-motor. Each CW and CCW rotation is transmitted by belt-pulley system to "ER clutch1" and "ER clutch2". Additionally, when the motor is stopped, each clutch works as a brake.

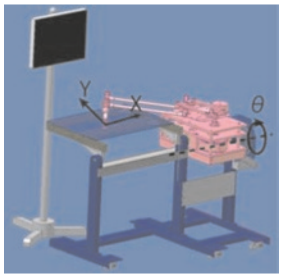


Figure 5: Quasi-3-DOF mechanism of Hybrid-PLEMO

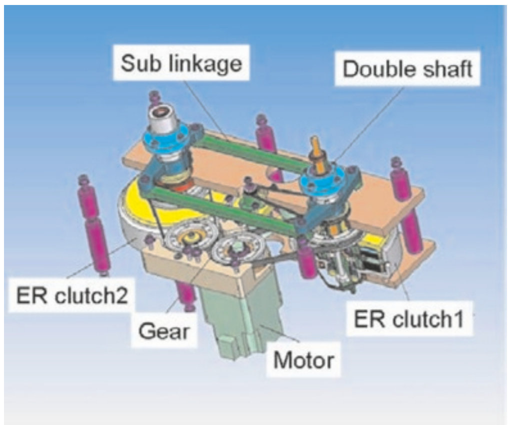


Figure 6: Input parts of actuator box

3.3 Stopper mechanism

We take stopper mechanism (Figure 7) under counter weight. With this mechanism, end point position doesn't exceed walking table (height 500mm, width 600mm). Figure 8 shows stopper position calculated from end point position.

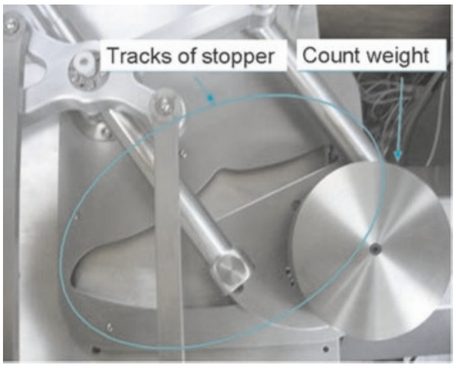


Figure 7: Stopper mechanism

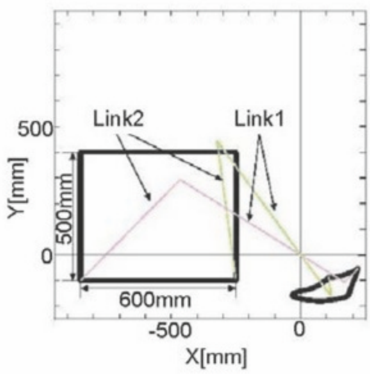


Figure 8: Stopper position

3.4 Parallel linkage mechanism

A parallel linkage mechanism of Hybrid-PLEMO is shown in Figure 7. “ER clutch1” and “ER clutch2” have a pair of two controllable shafts, which are a pair of an outer shaft and an inner shaft. Two outer shafts with opposite rotations are connected with “sub link1”. In same manner, two inner shafts are connected with “sub link2”. By using sub link1 and sub link2, we realize two controllable DOF for haptic control. These two DOFs are converted to orthogonal two directions of the end-effecter by using main parallel linkage, which consists of “link1” and “link2”.

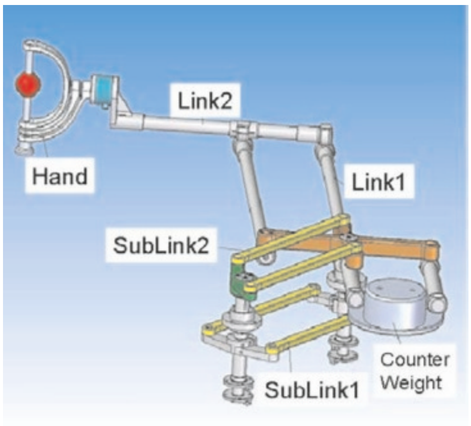


Figure 9: Parallel linkage mechanism of Hybrid-PLEMO

3.5 Active -Passive switching mechanism

Hybrid-PLEMO can switch Active/Passive mode very easily. Table 1 shows the comparison between active force feedback and passive force feedback.At passive mode, ER-clutch is fixed by disk-brake. If we use Hybrid-PLEMO as active mode, we set disk-brake off and servo-motor on. ER-clutch rotates, and we electrically control 4 shafts (inner shaft and outer shaft of ER-clutch 1 and 2). If we use Hybrid-PLEMO as passive mode, we set disk-brake on and servo-motor off. ER-clutch stops and we control 2 shafts (inner shaft and outer shaft of ER-clutch 1 or 2).

Table 1: Comparison between active force feedback and passive force feedback

Feedback mode	Active	Passive
Force generator	Actuator	Brake
Subject	More broadly applicable than passive force feedback	Only patient with voluntary movement
Safeness	Less safer than passive force feedback	Safe in mechanism
Cost	Expensive	Less expensive than active force feedback

3.6 Control system

Figure 10 shows control system for Hybrid-PLEMO. Absolute encoders (FA Coder, TS566N320, Tamagawa Seiki Inc., Japan, resolution: 17bits) measure the rotational angle of ER actuators or brakes. We can calculate the position and the velocity of the handle depending on each angle. Digital Input/ Output (DIO) board (PCI-2154C, Interface Inc., Japan) loads this information to a controller (personal computer). Operating handle includes a force sensor (OPFT-220N, Minebea Co. Ltd., Japan), and operating force is measured by this sensor. A potentiometer (CP-2F, Midori Precision Inc., Japan) measures the inclination of the worktable and the angle is loaded by Analog/Digital (A/D) converter board (PCI-3165, Interface Inc., Japan, resolution: 16bits). The brake torque of the ER brake is controlled by applied voltage from high voltage amplifiers(High voltage amplifier, MAX-ELECTRONICS, Co. Ltd., Japan). Digital/Analog (D/A) converter board (PCI-3338, Interface Inc., Japan, resolution: 12bits) outputs the reference signal to the amplifiers.

A controller is a personal computer (DOS/V), and an operating system (OS) is Vine Linux 3.2 and ART-Linux (kernel 2.4.31) as a real-time OS. Open-GL and Glut3.7 are used for the graphic library. Graphic process and control process are both executed by one PC. Multi-process programming is used to realize it. The control process is repeated by 1[ms] exactly.

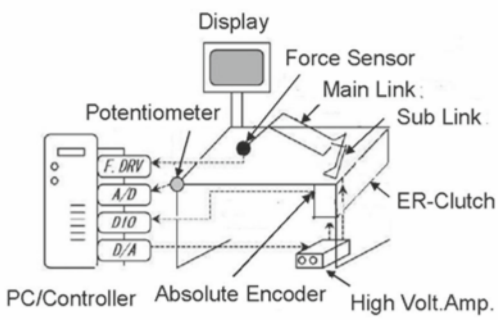


Figure 10: Control system of Hybrid-PLEMO

4 APPLICATION

Reaching motion is one of the most important tasks for rehabilitative training. In physical therapy or occupational therapy, reaching training is done without pulling patient's hand in the correct direction, but with giving resistance against correct direction. This resistance activates patient's intention to move their own arm.

In this section, we suggest two types of reaching program with Hybrid-PLEMO system. These two programs are based on the reaching training with force guidance or resistance mentioned above. However, the methods of force feedback are different each other. The one utilizes active force feedback mode of PLEMO system, and the other utilizes passive mode. The aim of this program is to operate the handle, search target trajectory with only force information and trace it. In active mode, PLEMO system generates outgoing- vector force of 5N from target trajectory. On the other hand, in passive mode, the system generates a distribution of resistant force (0N or 3N or 5N). The nearer to the target the hand position is, the stronger the resistance is. The start position was same (X = 0cm) in every experiments. The target changes its X position (X = -20, -10, 0, 10, 20cm) in a random manner. In order to evaluate simply the effect of force information, we hide any visual information on display. Figures 12 show experimental results of same target position (X = 10cm) with active/passive mode. An operator is healthy person. Broken lines show the target trajectory, and black dots show the starting position of the handle. As shown in the left side of Figure 12, operator can recognize the target position smoothly with active-type force guidance. On the other hand, as shown in the right side of Figure 12, it took more time to recognize the target position with passive-type force guidance than active mode. The reason of this delay is thought that, in passive mode, the operator needs more time to understand the distribution of force field with his own motion, and recognize correct direction toward the target. In future works, we must clarify effects and roles of active / passive force feedback for human's sense, motion and rehabilitation.

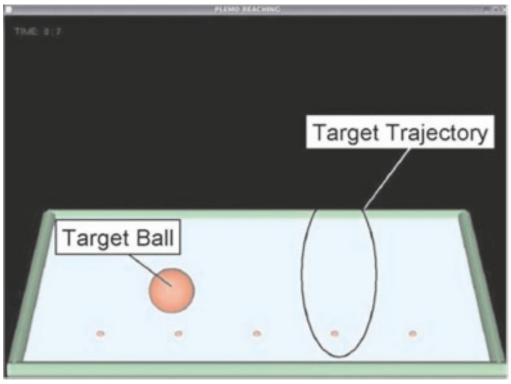


Figure 11: Application "reaching"

5 CONCLUSION

This paper described the development and evaluation of "Hybrid-PLEMO", which is one of the rehabilitation systems for upper limbs with quasi-3DOF mechanism and switchable mechanism between active/passive modes. We expect to use this system to clarify the effects and roles of active / passive force feedback in rehabilitative training. It will be possible to measure the effect of rehabilitation with EEG or NIRS for same subjects under same environments by using this system.

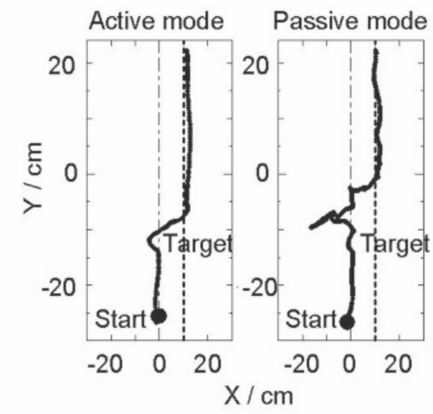


Figure 12: Experimental Results
(Left: active mode, right: passive mode)

6 REFERENCES

[1] Charles G. Burgar, Peter S. Lum, Peggy C. Shor, H.F. Machiel Van der Loos: Development of robots for rehabilitation therapy: The Palo Alto VA/Stanford experience, Jurnal of Rehabilitation Research and Development, Vol.37, No.6, pp.663--673 (2000).

[2] H. I. Krebs, B. T. Volpe, M. L. Aisen, N. Hogan: Increasing productivity and quality of care : Robot-aided neuro rehabilitation,Jurnal of Rehabilitation Research and Development, Vol.37, No.6, pp.639--652 (2000) .

[3] Li-Qun Zhang,Hyung-Soon Park, and Yupeng Ren : Developing an Intelligent Robotic Arm for Stroke Rehabilitation, Proceedings of IEEE International Conference on Rehabilitation obotics, pp984-993. (2007).

[4] Takehito Kikuchi, Hu Xinghao, Kazuki Fukushima, Kunihiko Oda, Junji Furusho and Akio Inoue:"Quasi-3-DOF Rehabilitation System for Upper Limbs: Its Force-Feedback Mechanism and Software for Rehabilitation", Proceedings of IEEE International Conference on Rehabilitation Robotics, pp.24-27 (2007).

[5] Takehito Kikuchi, Kazuki Fukushima, Junji Furusho, Takuya Ozawa, Study on Development of Its Mechanism of Rehabilitation System for Upper Limbs: PLEMO-P1, Proceedings of 11th International Conference on Electrorheological Fluids and Magnetorheological Suspensions, p18 (2008).

[6] G. Bossis, Ed., Proceedings of the Eighth International Conference on Electrorheological Fluid and Magnetorheological Suspensions, World Scientific, (2002).

6-DOF Stand-Alone Rehabilitation System 'Robotherapist' and Its Applications to Stroke Patients

Kunihiko Oda[1], Yuuki Ohyama[2], Junji Furusho[2], Takehito Kikuchi[2], and Shiro Isozumi[2]

[1] Dept. of Physical Therapy, Faculty of Medical and Welfare Engineering, Osaka Electro-Communication University, Osaka, Japan

[2] Dept. of Mechanical Engineering, Graduate School of Engineering, Osaka University, Osaka, Japan

Abstract

There is a need for quantitative diagnosis and evaluationin rehabilitation. This can be realized by means of rehabilitation robots system. This study is aimed at developing a rehabilitation robot capable of controlling 6 degrees-of-freedom (DOF) for upper limbs. Therefore we have developed a 6-DOF stand-alone rehabilitation system for upper limbs, named 'Robotherapist'. Furthermore we developed application software for stroke patients in order to evaluate states and train exercises on the basis of physical therapy. This paper presents the mechanism of Robotherapist and its applications to stroke patients.

Keywords:

Rehabilitation engineering, Human robot coexistence, Safety, Stroke patients

1 INTRODUCTION

Many people have a physical disorder due to aging, stroke, etc. In the future, the number of such people will increase because of the aging society. For example, in Japan more than two hundred and fifty thousand people have stroke every year, and many of them suffer from the aftereffects of paralysis. On the other hand, movements of the upper limbs, such as eating and operating appliances, are indispensable for daily activities. Therefore, rehabilitation for upper limbs is very important in order not only to decrease the numbers of them, but also to enable them to take an active part in society.

But the present rehabilitation has some problems as follows:

1. quantitative diagnosis and evaluation are not still established, so there is a need for both evaluation and training based on quantitative evaluations
2. shortage of therapist
3. training skills are depended on experiences of therapists
4. some of the patients give up rehabilitation because of its boredom

In recent years, there is a need for quantitative diagnosis and evaluation based on Evidence Based Medicine (EBM) in rehabilitation. Therefore, a need for rehabilitation support systems are increasing, which use robot and virtual reality technology.

Using these systems not only provides supports to therapists but also creates new training and exercises methods in rehabilitation [1]-[4]. Additionally feeding back the quantitative evaluations by these computerized system can enhance the motivation of an operator and the qualitative effect of training.

Some rehabilitation systems for upper limbs exist that use these technologies. Furusho, et al. of Osaka University developed a 3-DOF rehabilitation support system 'EMUL' [2]. Krebs, et al. of MIT developed a rehabilitation robot for wrists [5]. There are devices which has high capabilities and multi-DOF. However, there are few efficient applicaton software to treat stroke stroke patients on the basis of physical therapy.

This study is aimed at developing a rehabilitation support system capable of controlling 6-DOF for upper limbs, with the following integrated features:

1. high safety
2. quantitative diagnosis and evaluation
3. efficient training

Therefore, we have developed a 6-DOF stand-alone rehabilitation system for upper limbs, named 'Robotherapist' (Figure 1) [6]. It is used both robot and virtual reality technology. In addition, we two set of cameras for motion capture, and then we developed two sorts of application software for stroke patients on the basis of physical therapy. It enables efficient rehabilitation training and quantitative evaluation test.

This paper presents Robotherapist and its application software.

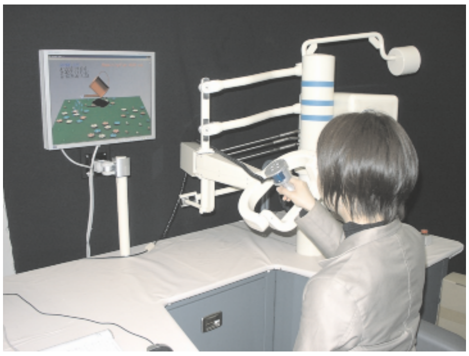


Figure 1: Robotherapist

2 ROBOTHERAPIST

Robotherapist is a stand-alone system for rehabilitation. Including other studies, there are two types of rehabilitation robots: stand-alone type and wearable one[7]. The stand-alone system is safer than the wearable one even when it goes out of control.

This robot is able to: control and measure both the position and the posture of the operator's hand ; display application software, for example, games, evaluation test, and rehabilitation training; and present force sense including the wrist torque to the operator according to situations of application software.

An operator grasps a handle of Robotherapist, watches a display and plays application software as rehabilitation training and evaluation test.

In this section, we describe the mechanism and control.

2.1 Mechanism for position of the Handle

The mechanism controlling the position of the handle part is shown in Figure 2. The mechanism has 2-DOF for the horizontal rotation (Figure 3) and 1-DOF for the vertical movement in the arm part (Figure 4). The motion range is about 0.90*W* × 0.54*D* × 0.50*H* [m]. The maximum axial force at the handle is about 25[N].

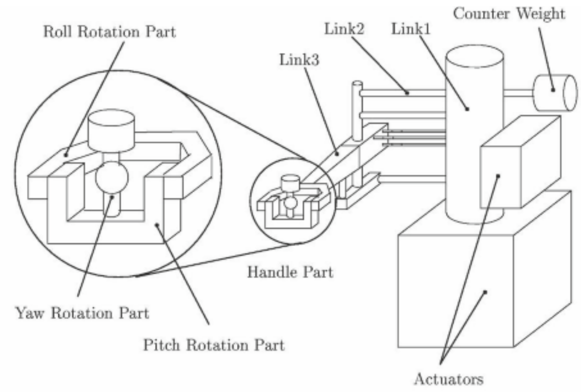


Figure 2: Mechanism of Robotherapist

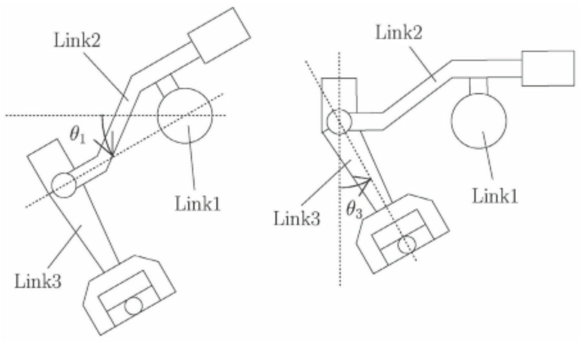


Figure 3: Mechanism of Horizontal Rotation

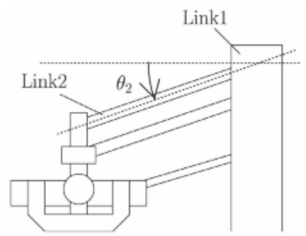


Figure 4: Mechanism of Vertical Rotation

2.2 Mechanism for posture of the Handle

The mechanism controlling the posture of the handle is shown in Figure 2. The handle has 3-DOF; that is, roll rotation, pitch rotation and yaw rotation.

The motors and encoders for a posture of the handle place near Link1, and the torque of each motor is transmitted to the handle by driving shafts and wire-pulleys system. The maximum torque at the handle is about 1[Nm].

2.3 Control System with Motion Capture

The overall image of Robotherapist is shown in Figure 5 The system has three sets of PCs, and these PCs exchange data through LAN.

The PC1 is used for control. It always records the position, posture, and force of the handle; sends those data to the PC2; calculates the force that should be generated; and control the output to actuators.

The PC2 is used for graphics. It receives the information from the PC1 and PC3; send the data from PC3 to the PC1; creates graphics; and shows them on the display.

The PC3 is used for motion capture. It records the position of each marker which, for example, attatched the the elbow and shoulder of an operator; send those data to the PC2.

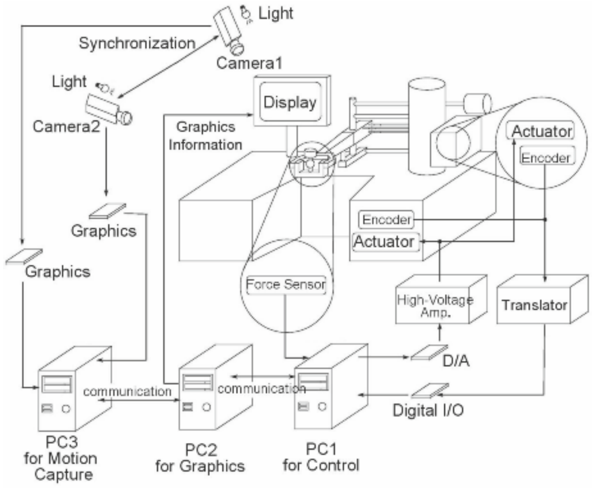


Figure 5: Control system

3 FORCE SENSE: GUIDANCE UTILIZING REACTION

Many stroke patients cannot move their hands actively as they want. Though they know a desired direction, they do not know how to move their hands in that direction. However, they can resist the force passively as they want. In physical therapy, there are techniques utilizing these reaction. One of the techniques is called Proprioceptive Neuromuscular Facilitation (PNF).

Techniques of PNF are used to place specific demands in order to secure a desired response [8]. One of the techniques of PNF is the following: in order to guide a hand of a patient in one direction, a therapist adds force to the hand in the opposite direction ; the patient resists it (Figure 6), and then his brain and body catch on how to move his own hand in the correct direction; the patient moves his own hand by itself resisting the force. Utilizing the reaction against the force toward the opposite direction, the patient can tell correct movements to his brain and body. The force that the therapist adds depends on the maxim power of the patient. By repeating training with this technique, their brain learns the correct pattern of movements.

We name this technique 'Guidance Utilizing Reaction (GUR)'. Futhermore we introduce GUR technique into application software of Robotherapist.

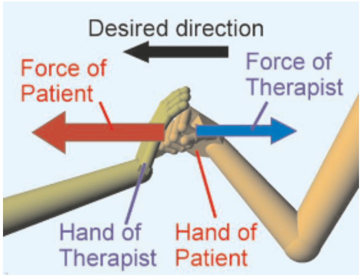


Figure 6: Guidance Utilizing Reaction

4 APPLICATION TO STROKE PATIENTS

We have developed some application software so far. Some of them are emphasized on amusement and quantity of movements. In this study, we developed two sorts of application software, which are emphasized on quality of movement. Quality of movements is considered because wrong movements are the contrary effects for rehabilitation. Then we introduce physical therapy into the application software for stroke rehabilitation.

4.1 Linear Exercise

Hand movement between two spots directly is difficult for stroke patients because that needs a separate movement. Therefore, they need to train separate movements. The Linear Exercise is used GUR and capable of training the movements.

The procedure of Linear Exercise program is the following.

In the display of Robotherapist (Figure 7), there are three balls and transparent cylinder. An operator controls the green ball and moves it toward a target ball by means of the handle of Robothrapist. After the green ball touched the target, he moves it toward the other ball, which become a new target ball. A present target ball shines red. The Cylinder is designed a desired course. He moves the green ball toward the red target ball along the course. When the green ball touches the present target, Robotherapist tells him that the green ball touches the target by three ways: that the target turns yellow and another ball turns red; sounds; a force sense like touching a ball actually. Next, he moves it to the new red target ball. When the green ball separates the previous target ball in yellow, the yellow ball turns gray. He repeats these movements.

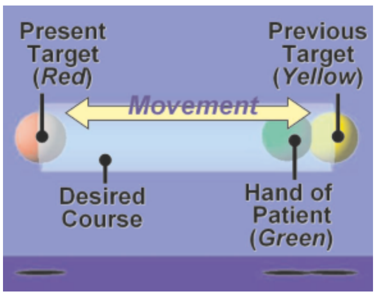


Figure 7: Display in Linear Exercise

When he moves the green ball to the target, Robotherapist gives force of GUR, that is the force in the opposite diredtion, to the handle. In addition, when the green ball get away from the course, Robotherapist gives the force of GUR in order for him to retuen to the course by himself. On the other hand, the system shows a warning and stops force when he moves the handle faster than the speed that we set to before he starts to play this training.

4.2 Arc Exercise

We have focused on one of the synergy pattern movement. When stroke patients raise their own arm, they often bend their elbow. This is because the flexion of their shoulder effects on the flexion of elbow. They need to learn the separate movements. The Arc Exercise is also used GUR and capable of training the movements .

The procedure of Arc Exercise program is the following.

In the display of Robotherapist (Figure 8), there are a green ball with cylinders on the top and buttom, a series of rings in an arc, and a shape of arm. The shape of hand is showed by means of the motion cature system in addition to the encorders.

An operator controls the ball. The rings form a desired arc course. He moves the ballis just on the arc along the course without flexing his elbow and wrist.

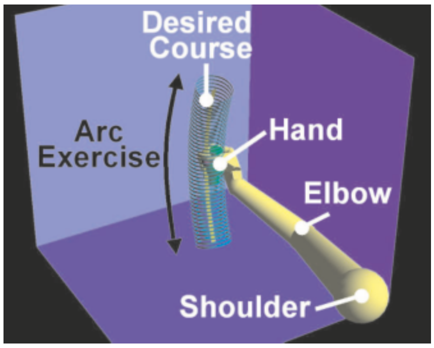


Figure 8: Display in Arc Exercise

When he moves it along the course, there is no force. However, when he moves it away from the course, Robotherapist gives force of GUR to the handle for the sake of returning the course.

5 SUMMARY

We have developed a 6-DOF stand-alone rehabilitation robot for upper limbs. The system is used robotics, virtual reality technology, and physical therapy. The main features of the system are as follows:

1. The system is able to measure the position and the posture of the wrist of an operator, and present a force sense to the operator.
2. We add motion capture system, and then it enables us to measure the position of a whole limbs.
3. Two sorts of application software are developed on the basis of PNF. It enables stroke patients to train separate movements efficiently.

6 ACKNOWLEDGMENTS

The development of ‘Robotherapist’ was supported financially by NEDO (New Energy and Industrial Technology Development Organization as a semi-governmental organization under the Ministry of Economy, Trade and Industry of Japan).

7 REFERENCES

[1] Furusho, J., Kikuchi, T., 2007, A 3-D Rehabilitation System for Upper Limbs "EMUL", and a 6-DOF Rehabilitation System "Robotherapist", and Other Rehabilitation Systems with High Safety, *Rehabilitation Robotics* (Edited by Sashi Skommu), I-Tech Education and Publishing, Chapter 8, 115-136.

[2] Krebs, H. I., Volpe, B. T., Aisen, M. L., Hogan, n., 2000, *Increasing productivity and quality of care: Robot-aided neuro rehabilitation,* Journal of Rehabilitation research and development, Vol. 37, no.6, 639-652.

[3] Furusho, J., Koyanagi, K., Imada, Y., Fujii, Y., Nakanishi, K., Domen, K., Miyakoshi, K., Ryu, U., Takenaka, S., Inoue, A., 2005, *A 3-D Rehabilitation system for Upper Limbs Developed in a 5-year NEDO Project and its Clinical Testing,* Proceedings of the 9th ICORR, 53-56.

[4] Takahashi, Y., Terade, Y., Inoue, K., Ito, Y., Lee, H., Komeda, T., 2003, *Upper-Limb Rehabilitation Exercises Using Haptic Device System,* International Journal of Human-friendly Welfare Robotic Systems, Vol. 4, no.2, 18-22.

[5] Krebs, H.I., Volpe, B.T., Williams, D., Celestino, J., Charles, S.K., Lynch, D., Hogan, N., 2007, Robot-Aided Neurorehabilitation: A Robot for Wrist Rehabilitation, Neural Systems and Rehablitation Engineering, IEEE Trans., Volume 15, Issue 3, 327 – 335.

[6] Furusho, J., Shichi, N., Hu, X., Kikuchi, T., Nakayama, K., Li, C., Yamaguchi, Y., Inoue, A., Ryu, U., 2006, *Development of a 6-DOF Force Display System with High Safety and its Application to Rehabilitation*, Proceedings of the 2006 IEEE ICMA, 962-967.

[7] Li-Qun Zhang, *IEEE Sr. Member*, Hyung-Soon Park, and Yupeng Ren, 2007, Developing an Intelligent Robotic Arm for Stroke Rehabilitation, ICORR.

[8] Voss, D. E., Ionta, M. K., Myers, B. J., *proprioceptive neuromuscular facilitation, patterns and techniques*, Third Edition, Sedgwick Mead.

Development of Grip Mechanism Assistant Device for Finger Rehabilitation

Shahrol Mohamaddan and Mohd Shahril Osman
Dept. of Mechanical and Manufacturing Engineering, Faculty of Engineering,
Universiti Malaysia Sarawak (UNIMAS), Sarawak, Malaysia

Abstract
This paper describes the work being carried out to develop a mechanism assistant device that can be employed for finger rehabilitation. Two design concepts were proposed. The initial design employed three L-shaped rod and a single column mechanism. However, due to some limitations, the initial design was modified. The final design employed four L-shaped rod and a single column. Simulations were conducted to check the extension and flexion of the two designs. From the simulation, it was found that the rotation of either motor can be 180 degrees clockwise for finger extension and anti-clockwise for finger flexion.

Keywords:
Finger rehabilitation; Grasping; Extension; Flexion

1 INTRODUCTION

The usage of finger is important in human daily life. However, circumstances such as an injuries, disabilities, diseases and deformation of the hand, may affect normal daily life.

In Malaysia, finger injuries and disabilities are basically caused by road accident and stroke. Based on the research conducted as in reference [1] [2] the road accident in Malaysia is increasing every year due to the increasing number of vehicles. Although there are no specific data between the road accident and the finger injuries, 42,856 cases related to road accident injuries or death was reported in 2001 [1]. On the other hand, stroke is one of the third largest cause of death in Malaysia behind heart diseases and cancer [3]. There are about 17,909 stroke victims reported in 2005 and the figures are expected to exceed 25,000 every year by 2020 [4].

One of the solutions for the finger injuries and disabilities are finger rehabilitation. The finger rehabilitation is a therapy which has the intention to recover partially or totally the finger motor abilities of the patient. The rehabilitation therapy is based on the manipulation of the paretic limb supported by a specialist therapy. The rehabilitation therapy may be accomplished with daily frequency for up to several months depending on the severity of the finger and condition of the patient. In order to regain the normal life, the patient needs timely and persistent rehabilitation assisted by the therapist. However, since the number of the therapist is inadequate, it will not be easy for the patient to have the rehabilitation period supported by the therapist.

The objective of this project is to develop a device that can assist the finger rehabilitation exercises. Due to the inadequate numbers of therapist and the timely and persistent rehabilitation period that required, there are needs to create a rehabilitation system where the patient can have the rehabilitation exercises unaided [5]. This project, however is a preliminary work in order to enhance research on the finger rehabilitation. The project starts with a design of a grip mechanism assistant device. The main idea of the device is to perform an extension and flexion of the finger or gripping based on a mechanism that can transfer the forces from the actuators.

In this paper the development process, design concept, simulation and the fabrication of the device will be discussed. The initial result of the device is also included. Since the device only perform an extension and flexion by the mechanism, the improvement of the device will be conducted in order to qualify as finger rehabilitation device. Subsequent section will details out the work in this project.

2 THE DEVELOPMENT PROCESS

In this project the development process of the device is shown in Figure 1. The development process is based on the methodology of designing a robotics hand with multiple fingers as in reference [6]. The development process flow is useful in order to identify the strength and weakness of the design of the device.

In reference [7][8][9][10], most of the finger rehabilitation device is focus on the stroke patient. However, the usage of the finger rehabilitation can be expanded to help the elderly and disabled people. This can be done by providing the elderly or disabled people with a device that can increase the power of the grasping. Therefore, in this project the idea is to develop a device that can be used for wide range of end user.

Since hand or fingers are the major focus of the project, the analysis of human hand anatomy is the first step in the development process. The understanding of the human hand anatomy is important because it can help the designer to focus in which part of the finger is involve in the finger movement. Beside that, the analysis of related work or research is also been done to give an overview of this field. The detail of the process will be explained in Section 3 of this paper.

After analyzing the human hand anatomy and related work, design and fabrication of the device will be conducted. At this stage, the work are divide into mechanical and electronics development. The mechanical development involve designing the mechanism of the device, simulation, material selection and the fabrication process. The electronics development involve the control system of the device. This will be explained in subsequence sections of this paper.

The next step in the development process is the device analysis. In this process, experiments will be conducted in order to check the mechanism and the control system. Besides that, the discussion with therapist will be conducted to have an input of the device.

After the analysis or revision process, the optimization process will be involved. At this stage, the redesigning and improvement on the device will be conducted. All these process are identified as design and fabrication phase as shown in Figure 1.

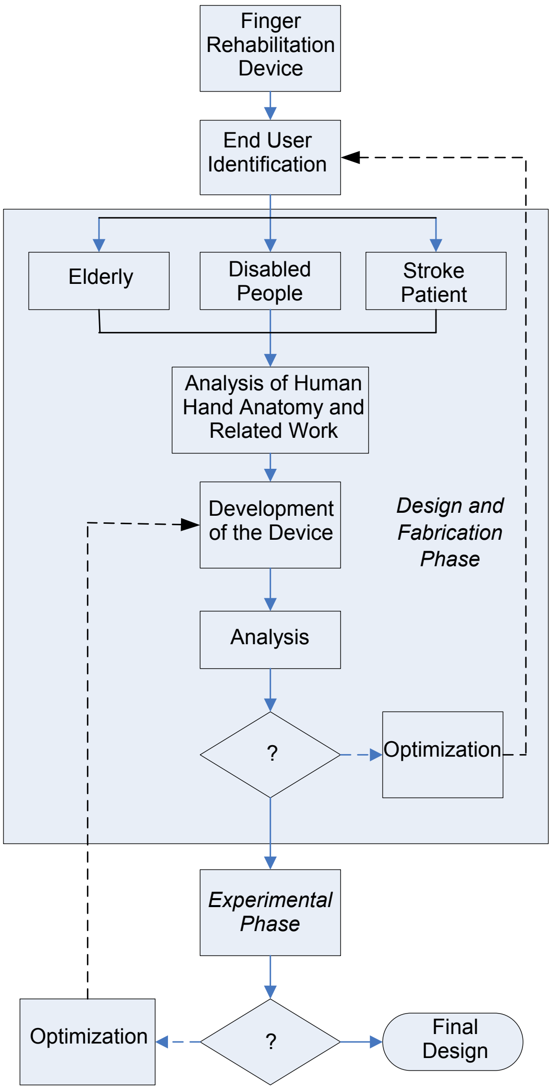


Figure 1 : Design process for finger rehabilitation device

After the design and fabrication phase, the development process will enter the experimental phase. At the experimental phase, a clinical experiment with the elderly, disabled people or stroke patient will be conducted. A feedback from the elderly, disabled people or stroke patient will be gathered to further improve the device. The assistant from the therapist will be needed as input from them is also important for designing the device. Finally, the performance of the device will be analyze and accepted design will be fabricate and will be used as a finger rehabilitation device.

3 HUMAN HAND ANATOMY AND RELATED WORK

3.1 Human Hand Anatomy

The anatomy of human hand is shown in Figure 2. The skeleton consists of 27 bones where eight of them are carpals bone arranged in two rows. The metacarplas (MET) or palm consists of five fingers while the remaining fourteeen fingers are digital bones which is proximal phalanges (PP) with five fingers, middle phalanges (MP) with four fingers and distal phalanges (DP) with two fingers.

Figure 2: Anatomy of the Human Hand [11]

Human hand motion is divided into two main categories ; global hand motion and local finger motion. Global hand motion represent the large rotation and translation movement of the palm. Local finger motion involve basic movement such as finger flexion, extension, abduction and adduction. Local finger motion is articulated and self-occlusion with a total of 21 degree of freedom (DOF) [12][13]. In this project the main focus is only on the extension and flexion movement of the finger.

3.2 Related Work

Research on finger rehabilitation related to injuries, disabilities, diseases or deformation of the hand had been conducted widely especially in Japan and Europe. The fundamental idea for developing the finger rehabilitaion device is by understand the human hand kinematics, includes human hand anatomy and kinematics model of the hand as stated in the reference [11]. This is important in order to design the device to ensure which part of the fingers involve in the design.

Beside the human hand anatomy and kinematics model of the hand, research on finger rehabilitation is focusing either force or moments of the hand. In reference [14] an extension finger-force measurement device was developed and the intra-individual repeatability was investigated. The idea of the device is to determine the weakness and mobility difficulties of the disabled people with rheumatoid arthritis (RA) disease in order to evaluate the functional impairment. The device measures the isometric force at every single phalange around the metacarpophalangeal joint (MPJ) of the finger.

There are several device that had been developed focus on the stroke patient such as Electromyographic signal (EMG) [5][7], Lateral Symmetric Master Slaved Exoskeleton [8], Hand-Wrist Assisting Robotic Device (HOWARD) [10][15], Virtual reality (VR) [9][16][17] and passive training [18].

The Electromyographic signal (EMG) is the signal derive by the patient impaired limbs which reprocess into command that actuated the exoskeleton for the desired movement. Practically, EMG signal was measured through the surface of skin of affected limb by attaching special electrode and transferring valuable data related to the neuromuscular activities of limb. These signal commands the assisting device to response by bending the related joint of fingers, and hence guiding the impaired hand to achieve the desired motion. As examples, electrode attached to the muscles of Flexor Digitorum Superficialis, Flexor Pollicis Longus which position at the mid-forearm will generate signal of flexing four fingers together and flexion of thumb [5] when the patient was trying to do grasp or fist making task. The main goal of this device is to obtain individual finger motion, active fingers extension and flexion.

Lateral Symmetric Master Slaved Exoskeleton devices establishing the concept of self-motion control. The technology enabled patient self therapy by using the normal hand as references motion (master) to actuate the devices and reproduce the desired motion (slave) for impaired hand activities assistance. The device consists of an input data glove which act as a master system and a corresponding slaved exoskeleton which assist the patient throughout the rehabilitation process. The rehabilitation process is further enhanced with the supportive virtual reality program which improved the effectiveness of rehabilitation process. Main advantage of this self-motion control device is that patient is able to adjust the device motion base on their own decision. Thus exercise can be stopped whenever they feel pain or motion out of range [8].

Hand-Wrist Assisting Robotic Device (HOWARD) is another robotic based rehabilitation device. The device consists of a platform which rests the forearm and the splints which clamps the fingers. The overall device function is supported with computer program which guide the patient throughout the physical rehabilitation tasks. Mainly, HOWARD will guide the patient in performing grasp and release exercise. During treatment, patients will initiate the hand motion. Since, patient impaired hand is usually weak and can only budge, therefore, robot will detect this motion and actuated for assistance. Clinical trial had shown that, patient's gripping, grasping, pinching ability had improved around 10% and 17% of greater range of motion had developed after three weeks of training session with HOWARD [10][15].

Virtual reality (VR) is the technology which simulates the real-life environment to provide physical, occupational, and sensory therapy in post stroke rehabilitation. Generally, robotic assisting device or rehabilitation gloves will be supported with this graphical interfaced VR programming. VR rehabilitation process is just similar with playing a computer game. Patient is required to move their hand around, performing specific tasks such as opening bottle cover or filling the cup with water. A score will be given as an evaluation for the patient at the end of the training session. Interesting graphical feedback of VR program increases the excitement and motivation for the rehabilitation therapy which is usually painful and boring. In short, VR based stroke rehabilitation encouraged active learning through interaction of visual sense, brain thinking, and limb action to achieve the rehabilitation goal [19][20]

Clinical facts show that rapid improvement in hand function will be observed in the first three months after stroke [15]. After this optimum period, recovery will be obviously slow down or even maintained as partial disabilities if untreated well. Thus, home based treatment was crucial for further rehabilitation treatment of outpatient.

Dynasplint system is an example of home based rehabilitation device which supported patients self treatment along recovery period. Dynasplint system designed to improve the fingers mobility by increasing the contracted finger's range of motion due to stroke or others injuries. It provides patients a passive finger exercise by permanently stretching and bending the finger joint into range. Patient is required to wear the splint with certain constraint degree of bending until finger tolerances for a pain free wear. Splint will then be tighten or increased degree of constraint gradually until the normal finger mobility is achieved. The system improve the range of motion by creating permanent, non-traumatic tissue elongation and remodeling, thus virtually eliminating the range of motion rebound effect [18]

The design developed in this project is trying to simplified other device that had been used for the rehabilitation process. The device incorporated design based on local needs for patient in Malaysia. It is mainly employ mechanical and electronic parts which is based on extension and eflexion movement. This will be further explained in subsequence sections.

4 THE DESIGN CONCEPT

The design of the grip mechanism assistant device is based on the simplicity, easy attachment and fit for all concepts. The initial device consists of the index and the thumb finger. Since the index and the thumb have a different extension and flexion movement, and different number of bones, different mechanism for each finger had been developed.

The index finger had been selected to represent the other fingers; middle, ring and little finger which have the similar arrangement between the bones in the hand anatomy. There are two design concepts that had been conducted. The final design concept is based on the improvement that had been made due to the problems encountered during the simulation process and the limitation of the machine for fabrication process. The design and simulation of the device had been conducted using the computer aided-design software.

4.1 Initial design concept

The mechanism of the index finger is shown in Figure 3. The mechanism is focusing on three places which are the Proximal Phalanges (PP), Middle Phalanges (MP) and Distal Phalanges (DP). Each of the places is mounted with three main columns. For column (1) and (2), both the front and the back of the column have a rolling mechanism to allow movement from top to the bottom of the column. For column (3) the rolling mechanism is only on one sided. The figure shows that all three columns are connected by an L-shaped rod (a), (b) and (c) with different length.

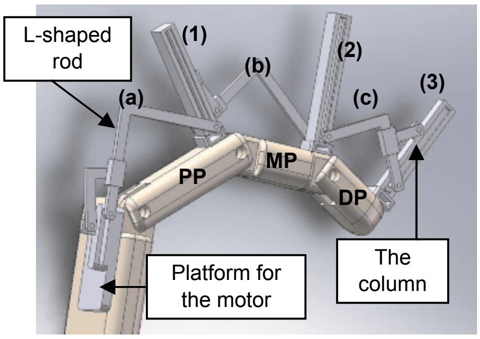


Figure 3: Initial design concept of the index finger

The L-shaped is attached to a free moving block which allows the rolling up and down along the column. In the design, the motor is placed at the platform which is on the metacarpals. The rolling mechanism is using a bearing to move. In the mechanism, when L-shaped (a) move, L-shaped (b) and (c) move along at the same time. The mechanism allows the finger to conduct extension and flexion task based on the motor rotation.

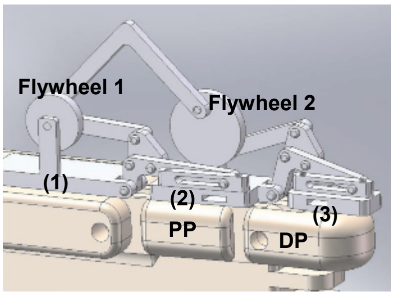


Figure 4: Initial design concept of the thumb finger

The mechanism of the thumb is shown in Figure 4. Instead of using columns on each phalanx, flywheels are used in the thumb mechanism. Since the thumb does not have the MP, the mechanism only involves movement in the PP and the DP. The thumb mechanism consists of three different platforms on each phalanx. The first (1) and second (2) platforms are connected to a flywheel where the flywheel 1 (15mm diameter) is smaller than the flywheel 2 (20mm diameter).

Both flywheels are connected together by L-shaped rods which ensure that both flywheels rotate together at the same direction. When the motor is rotate the flywheel 1 and the flywheel 2 will also rotate. However, since the flywheel 2 is bigger, the angular distance travel is greater even though the rotational angular is the same.

When the flywheel 1 rotates, the flywheel 2 will also rotate and drive the L-shaped rod forward. Therefore, the L-shaped rod will bend downward pushing the PP down. Since the flywheel 2 rotates together, it creates a chain reaction with the same mechanism driving the DP downward at the same time. The rotation will continued based on the motor rotation.

4.2 Final design concept

Based on the software simulation, there are problems encountered in the first design concept especially on the three columns for the index finger shown in Figure 2. The movement of each column in the design are not smooth and the force from the motor is insufficient to move column three which is on the DP. Therefore the device cannot perform the extension and flexion of the index finger. Some of the mechanism is redundant and the design and can be simplified. An improved design of the device had been carried out.

The major changes of the new design is on the index finger which is shown in Figure 5. Instead of using three L-shaped rod and three columns, the design had been changed to four L-shaped rod with single column. From the simulation, the new design can perform the extension and flexion movement smoother than the first design. Besides that, the usage of four L-shaped rod makes the force from the motor to be distributed accordingly and the extension and flexion movement can easily be performed.

The design of the thumb finger did not have any major changes. Since the mechanism can perform the extension and flexion movement smoothly, the flywheels concept is maintained. However, in order to have a lighter device, some of the mechanism had been simplified. Several holes also created on the flywheels to make sure the end user can adjust the device according to the size of the hand. This is to ensure a universal fit for the device.

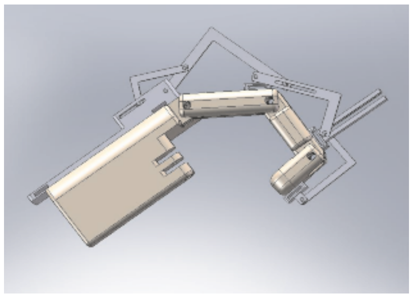


Figure 5: Final design concept of the index finger

5 FABRICATION, MATERIAL SELECTION AND DEVICE COMPONENT

5.1 Fabrication and Material Selection

The fabrication of the device includes several machining processes due to relatively small component size of the device had been performed. The device is made from aluminum material in order to make it light in weight. Portable size and light weight is very important to make sure that the maximum impact of the rehabilitation process can be achieved and the patient can have comfortable rehabilitation exercise. As for the attachment to the hand, Velcro straps and rubber mats will be used.

The Velcro straps are used to attach the device onto the patient hand whereas the rubber mats are placed between the device and patient hand for comfort and to avoid any discomfort and slippery.

After the fabrication process, the device will be controlled by the DC (Direct Current) servo motor. There will be two motors used by the device and each finger is controlled by one motor. The power and torque from the DC servo motor is expected to drive the finger device mechanism. The final design of the device with motor placement and Velco strap holes is shown in Figure 6.

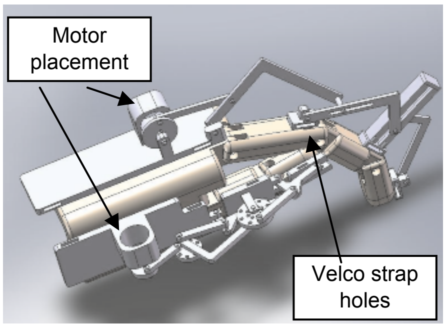


Figure 6: Final design of the device

5.2 Device Component

The grip mechanism assistant device for index finger and thumb finger are shown in Figure 7 and Figure 8 respectively. The index finger in Figure 7 is 232 mm length and 30 mm width at the back and 10 mm width at the front side. On the other hand, the thumb finger in Figure 8 is 154 mm length and 29 mm width at the back and 10 mm width at the front side. The flywheel 1 is 15 mm in diameters and the flywheel 2 is 20 mm in diameters.

The back side of the device is wider than the front in order to make it easier to attach at the finger. Pin is used as a bearing to connect the rod and to ensure the movement of the joint can be performed.

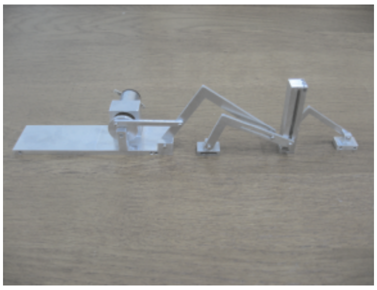


Figure 7: The index finger

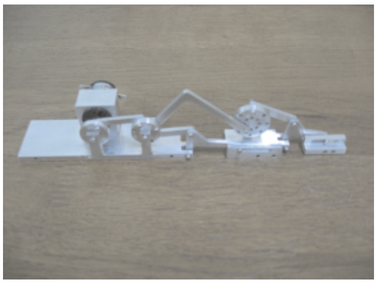


Figure 8: The thumb finger

6 CONTROL SYSTEM

In order to conduct a testing for the device, all signals had been processed in the AT89C51 microcontroller and powered by +5V regulated voltage. A programme had been employed and loaded into the microcontroller to rotate the motors.

The rotation of the motors have to be programmed 180 degrees clockwise to perform extension of the finger and 180 degrees anti-clockwise to perform the flexion movement. It is important to ensure that the device is able to perform the movement and a smoother extension and flexion can be achieved.

However, the usage of AT89C51 microcontroller is subjected to be changed because of the limitation of not being able to control to the desired position. A new control system for the device will be more user-friendly and easy to operate by the end users. The usage of grapichal user interface (GUI) using the Visual C++ is expected to be developed.

7 DEVICE TESTING

The device is fabricated and a simple test is performed. From the preliminary testing, the device is able to perform the extension and flexion movement similar with the simulations. The design however, is relatively large as the mechanism is a preliminary prototype. Further fabrication work with actual size is expected for testing as shown in Figure 9.

Although the mechanism move accordingly, the development of the feedback system such as sensors will be employed. This is to ensure that the extension and flexion movement can be measured simultaneously.

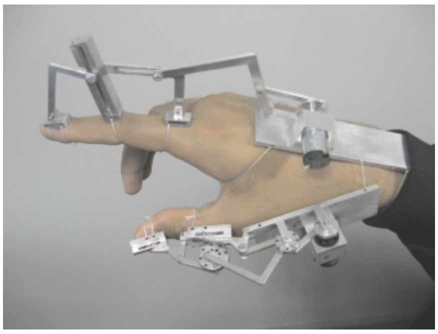


Figure 9: Overview of the device

8 CONCLUSION AND FUTURE WORK

Development of grip mechanism assistant device is based on simple design, easy attachment and universal, had been presented in this paper. Based on the simulation, there are several problems encountered in the design. The simulation features in the software provide an overview and errors that can be rectified. Aluminum material had been selected as it is a light weight material. The device consist of the index and the thumb finger. L-shaped rod mechanism had been implemented in the device in order to have a smoother flexion and extension movement for the finger.

Although the device can perform the extension and flexion movement there are many areas that need to be in order to understand the finger rehabilitation. The size of the device and the feedback control system is the major improvements that have to be done for the future work. Further discussions with the therapist for suggestion are also required for modification.

9 ACKNOWLEDGMENTS

This research was partially funded by Osaka Gas Foundation for research Development of Grasp Mechanism Assistant, 2007-2008.

10 REFERENCES

[1] Abdul K, 2003, Review of Global Menace of Road Accidents with Special Reference to Malaysia - A Social Perspective, Malaysia Journal of Medical Science, Vol.10, No.2, 31-39

[2] Saidon N., Baguley C.J., 1994, Road Accident in Malaysia, Proceeding Seminar on Crash and Injury Control, University Malaya

[3] National Stroke Association of Malaysia, http://www.nasam.org

[4] The Star, http://thestar.com.my

[5] Marcello M., 2005, An EMG-controlled Exoskeleton for Hand Rehabilitation, Proceeding of IEEE International Conference on Rehabilitation Robotics, 371-374

[6] Jorge E.P.P., Nestor E.N.R., Marco C., 2008, A Methodology for the Design of Robotics Hands with Multiple Fingers, International Journal of Advanced Robotic Systems, Vol. 5, No. 2, 177-184

[7] EMG-Controlled Robot Hand for Hand Rehabilitation, http://www.dlr.de/rm-neu/en/desktopdefault.aspx/tabid-3817//6237_read-9003/

[8] H. Kawasaki, H. Kimura, S. Ito, Y. Nishimoto, H. Hayashi and H. Sakae, 2006, Hand Rehabilitation System Based on Self-Motion Control with a Clinical Case Report, World Automation Congress, 1-6

[9] Heidi C.F., Kathy S., Tiffany K., Xun L., Robert V.K., Derek G.K., Hand Rehabilitation Following Stroke : A Pilot Study of Assisted Finger Extension Training in a Virtual Environment, 2007, Top Stroke Rehabilitation www.thomasland.com, 14(1), 1-12

[10] C.D. Takahashi, L. Der-Yeghiaian, V.H. Le, S.C. Cramer, 2005, A Robotic Device for Hand Motor Therapy After Stroke, International Conference on Rehabilitation Robotics, 17-20

[11] Robert N.R., John M.H., 1994, Modeling and Parameter Estimation of the Human Index Finger, Proceeding of IEEE International Conference on Robotics and Automation, Vol.1, 223-230

[12] Marzke M.W., Marke R.F., 2000, Evolution of the Human hand : Approaches to Acquiring, Analysing and Interpreting the Anatomical Evidence, Journal of Anatomy, 121-140

[13] John Lin, Ying Wu, TS Huang, 2000, Modeling the Constraints of Human Hand Motion, Proceedings Workshop on Human Motion, 121-126

[14] Olandersson S., Lundqvist H., Bengtsson M., Undahl M., Baerveldt A.J., Hilliges M., 2005, Finger-force Measurement Device for Finger Rehabilitaion, Proceedings IEEE, International Conference on Rehabilitation Robotics, 135-138

[15] Robotic Therapy Helps Restore Hand Use After Stroke http://www.today.uci.edu/news/release_detail.asp?key=1569

[16] Atif A., Rosa I., Mohamad E., Abdulmotaleb E.S., Shervin S., Edward L., 2007, Haptic Exercises for Measuring Improvement of Post-Stroke Rehabilitation Patients, IEE International Workshop on Medical Measurement and Applications, 1-6

[17] Xun L., Robert V.K., Tiffany K., Heidi C.W., Derek G.K., 2005, An Augmented Reality Training Environemnt for Post-Stroke Finger Rehabilitation, International Conference on Rehabilitation Robotics, 329-332

[18] Dynasplint System : Finger, PIP and MCP Extension and Flexion, http://www.dynasplint.com/type/finger.php/

[19] J.A. Edmans, J.R.F. Gladman, S. Cobb, A. Sunderland, T. Pridmore, D. Hilton and M.F. Walker, 2006, Validity of a Virtual Environment for Stroke Rehabilitation, American Heart Association, http://stroke.ahajournals.org/cgi/content/full/37/11/2770

[20] R.W. Teasell, L. Kalra, 2004, What's New in Stroke Rehabilitation, American Heart Association, http://stroke.ahajournals.org/cgi/content/full/36/2/215

Development and Evaluation of a Centaur Robot

Satoshi Tsuda[1], Kuniya Shinozaki[1], Ryohei Nakatsu[3]

[1] School of Science and Technology, Kwansei Gakuin University, Sanda, Japan

[2] Interactive & Digital Media Institute, National University of Singapore, Singapore

Abstract

Recently various types of robots are being studied and developed, which can be classified into two groups: humanoid type and animal type. Since each group has its own merits and demerits, a new type of robot is expected to emerge with greater strengths and fewer weaknesses. In this paper we propose a new type of robot called the “Centaur Robot” by merging the concepts of these two types of robots. This robot has a human-like upper body and a four-legged animal-like lower body. Due to this basic architecture, the robot has several merits, including human-like behaviors and stable walking even on non-smooth ground. We describe its hardware and software architectures. Also we describe the experiments to evaluate its walking capability.

Keywords:

Humanoid Robot; Four legged Robot; Centaur Robot; Walking Stability; Physical Support

1 INTRODUCTION

In recent years, various robots are being studied and developed in research institutes and companies that can be classified into two groups: a humanoid robot with two legs [1][2], an animal type robot with four or more legs [3][4][5]. Also a humanoid robot can be classified into those with two legs and those with wheels [6]. Each of these types has its own merits. The design of a humanoid robot with two legs is based on humans and can mimic such human motions as walking. Since this robot’s behavior resembles human behavior, it might easily be introduced into society. In the future, such robots are expected to support us in various aspects of our daily life. At the same time, however, its walking capability still lacks stability, and it sometimes falls down, restricting its area of activity. Also tt has difficulty maintaining its balance on ground that is not flat. On the other hand, the merit of an animal type robot is its four legs, which allow it to walk stably even on uneven ground. Since it can also basically stand on three legs, it can adopt to various ground pattern changes. So far, however, the robot has mainly been developed as a pet to which useful applications have rarely been applied. A humanoid robot with wheels for locomotion, which we call a wheel type robot, can move very smoothly and stably on the ground. It rarely falls down. It can even move on slightly uneven ground. On the other hand, it has no ability to move on stairs, which greatly restricts its area of activity since houses usually contain stairs and other types of height differences.

One approach to overcome these problems is to develop new types of robots by merging the strengths of existing robots. In this paper we propose a new type of robot with a human-like upper body and an animal-like lower body that we call a “Centaur Robot.” In the following sections, we describe its basic concept and then its detailed software/hardware architectures. Also we describe the experiments to evaluate how this robot can achieve a waking capability on non-smooth ground.

2 RELATED WORKS

Recently, especially in Japan, various kinds of robots have been studied and developed, particularly humanoid robots that are expected to support our daily life. For example, HONDA has developed a humanoid robot called ASIMO that has sophisticated walking capability [1]. For animal types of robots, on the other hand, most have been studied and developed as pets instead of supportive robots, including AIBO developed by Sony [3].

Although much research/development continues on humanoid and animal types of robots, little research has integrated these two types for several reasons. One reason is that since there are so many research themes for new functions and improvements for each of these types of robots, researchers have little incentive to concentrate on new types of robots that go beyond humanoid or animal types. Another is that even myths or folktales only contain a few examples of such creatures as centaurs, mermaids, and sphinxes in which humans and animals are integrated. Thus it is rather hard to imagine the functions and application areas that such a new type of robot might have.

Therefore, we developed a centaur robot because we believed by integrating two types of robots we could create a new type of robot with advantages over conventional robots.

3 HUMANOID ROBOTS

In our work, we are developing a robot that can stably achieve various motions by merging two types of robots: a humanoid and an animal.

There are two approaches for such integration: from the humanoid robot side and from the animal robot side. The former approach tries to realize a four-legged body as well as four-legged walk while maintaining a human-like upper body and achieving human-like motions. On the other hand, the latter approach achieves various human-like motions by adding a human upper body to a four-legged robot. In our study, we chose the former approach and modified the hardware and software of a humanoid robot to realize a centaur robot.

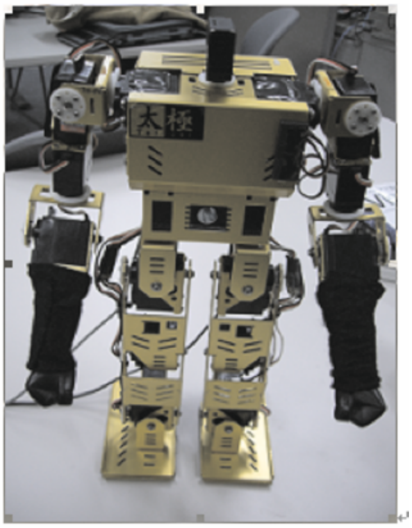


Figure 1: Humanoid robot

Table 1: Specifications of humanoid robot

Size/Weight	34 cm/1.7 kg
Degree of flexibility	22 (12 legs, 8 arms, 1 waist, 1 head)
CPU	SH2/7047F
Motor	KO PDS-2144, FUTABA S3003, FUTABA S3102, FUTABA S3103
Battery	DC6V

We adopted a humanoid robot developed by Nirvana Technology as a platform robot [7]. This robot has 22 servo motors that can express various human-like motions. Figure 1 shows its appearance, and Table 1 shows its specifications. Figure 2 illustrates the construction of its hardware. The control board, on which a microprocessor SH2 is attached, is connected to the servo motors, a gyro sensor, acceleration sensors, PC, and a battery. The program on the controller can achieve autonomous robot behaviors. At the same time, we can send commands to the robot by PC.

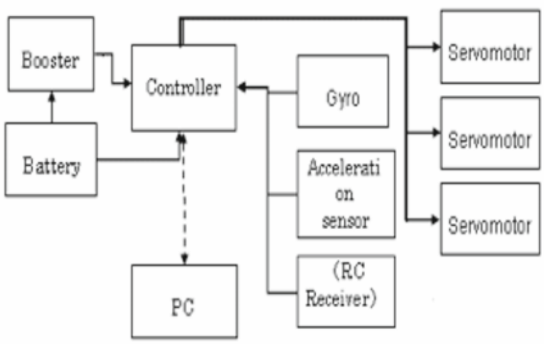


Figure 2: Hardware construction of humanoid robot

Figure 3 illustrates the software construction. The calculation of the commands necessary to move each motor is carried out each fifteen milliseconds and sent to each servo motor. The instructions to the robot from the PC are first analyzed and based on results go through one of two processes: one command for walking and other commands for other motions. For other commands, the motion data corresponding to the command is read from memory and the control data for each motor is calculated, and then the control data is sent to each servo motor. On the other hand, if the input command is a command for walking, then the real time calculation of the control data for each servo motor is carried out and sent to each servo motor. Calculation consists of three processes: trajectory generation calculation, inverse kinematics calculation, and servo motor angle calculation. In trajectory generation calculation, the position of each ankle studied by observing human walking motion is calculated every fifteen seconds. Then by inverse kinematics calculation the rotation angle of each foot joint is calculated for the same timing. Based on these calculations, finally the angle of each servo motor is calculated. Thus the rotation angle to be achieved for each motor is sent every fifteen milliseconds.

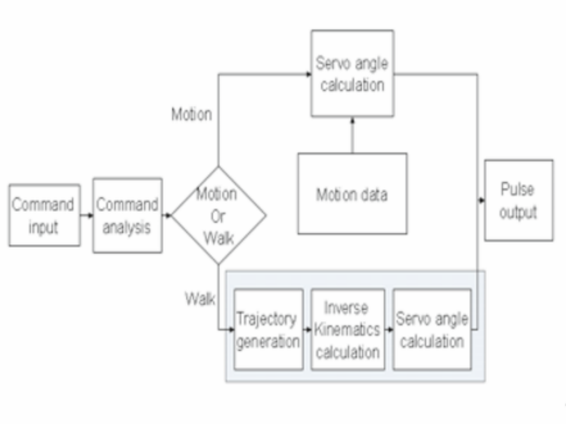


Figure 3: Software construction of humanoid robot

4 CENTAUR ROBOT

4.1 Overview

We developed a centaur robot based on the humanoid robot described in the previous section. We prepared two humanoid robots and used one as a front body. For another robot, we only used its lower body as a back of the centaur robot. Then we connected these two parts by a flat plastic board that functions as the shoulder part. Figure 4 shows the centaur robot's appearance.

4.2 Hardware construction

Now we explain the robot's hardware construction, as illustrated in Figure 6. Apparently for the front the hardware of the original humanoid robot was used, and for the back only the lower body was used. But a comparison of Figs. 3 and 5 shows that this robot's control structure is somewhat different from the original. Two controllers were used for complete control of the robot. One controls the servo motors required for upper body motions. The other controls the servo motors corresponding to the lower body. Since all the sensors are provided for the upper body, the controller corresponding to the upper body manages all sensor feedback. We adopted these two boards for several reasons. One, by using two boards, one of which controls the motions of the upper body and the lower body, it is possible to separately control the behaviors of the upper body as well as the lower body. For the power supply and battery, both controllers are connected to one battery. Also commands from PC are sent to both controllers.

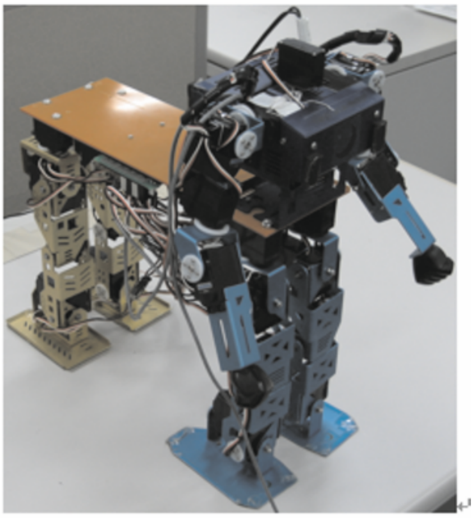


Figure 4: Centaur robot

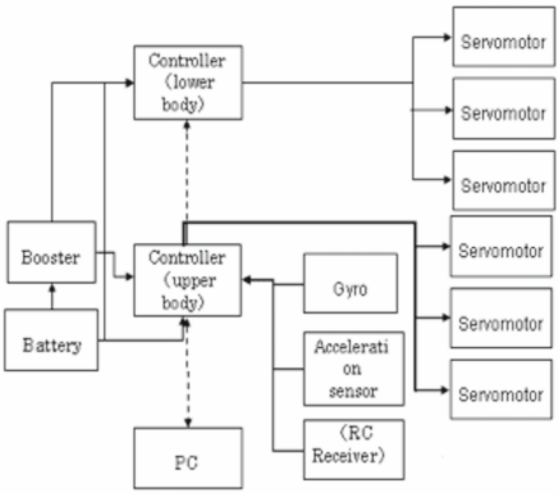


Figure 5: Hardware construction of centaur robot

4.3 Software construction

Next, we explain the robot's software construction, as illustrated in Figure 6. The software of the original humanoid controls both the upper and lower bodies together. For the centaur robot, we checked all the original robot's software and separated the software codes into two groups: one that controls the upper body and another that controls the lower body. Thus we reconstructed the whole software. For the upper body, it is unnecessary to carry out calculations for walking. When commands other than a walking command are sent from the PC, it retrieves motion data stored in the memory and sends the necessary rotation angle data to each servo motor. On the other hand software corresponding to the lower body must treat two types of commands as in the case of the original humanoid robot: a command for walking and other commands for additional motions. Also we adopted a method of inserting an arbitrary phase shift between the servo motor control of the front and back legs so that the robot can adopt the most adequate walking motions depending on the walking speed.

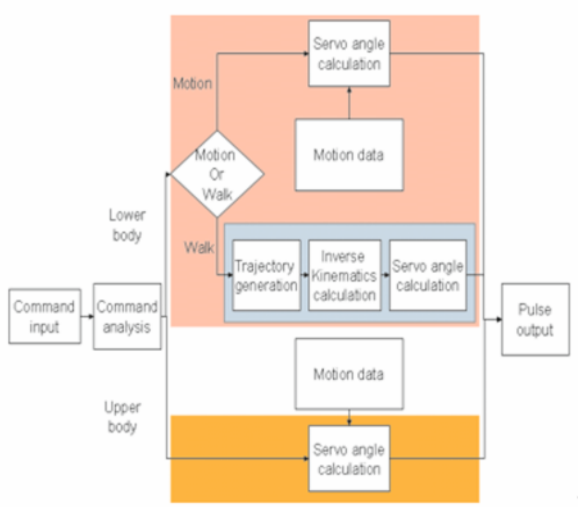


Figure 6: Software construction of centaur robot

By adopting such basic software structure, robot control has the following merits:

(1) The upper and lower body motions can be controlled separately. So far all the motion data developed for achieving various types of humanoid robot motions must be developed to describe the whole body movement. Since the motions of the upper and lower bodies have been separated, we can separately develop two types of motions, and by combining these two types of data, we can generate various kinds of whole body movements for the robot. This idea can easily be applied to the original humanoid robot.

(2) The front and back body movements can be separately controlled. Although it seems natural to let the front lower body and back lower bodies perform identical motions, sometimes it is better to control the two bodies by different body motions. Especially in the case of walking and running motions there would be some differences between these two bodies. For example, for trot type walking there should be a 180° phase shift between the front and the back legs. In the case of gallop running, the front legs and the back legs should move synchronously.

4.4 Evaluation of the robot

We carried out several experiments to evaluate the motion capability of our centaur robot.

(1) Walking capability

We inserted a phase shift of 0 degree, 90 degree, and 180 degree between the walking motion cycle of the front and back legs. These waling styles correspond to those of animals such as “pace,” “gallop,” and “trot.” We confirmed that the robot could move smoothly with almost the same speed by adopting each of these walking styles. Therefore for walking stability and speed all of the three walking styles perform the same capability.

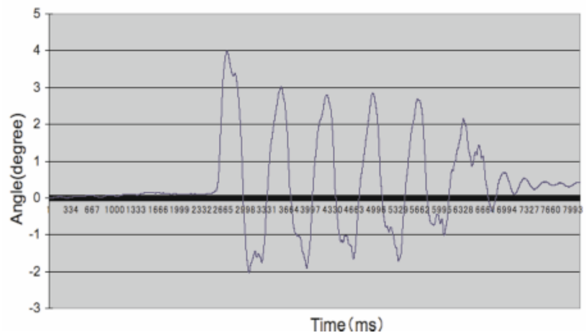


Figure 7: Tilt angle for “pace” walking style

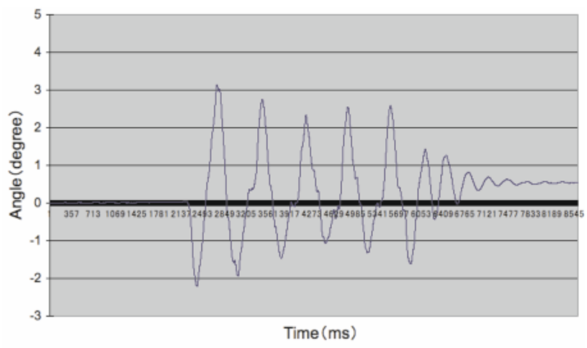


Figure 8: Tilt angle for “gallop” walking style

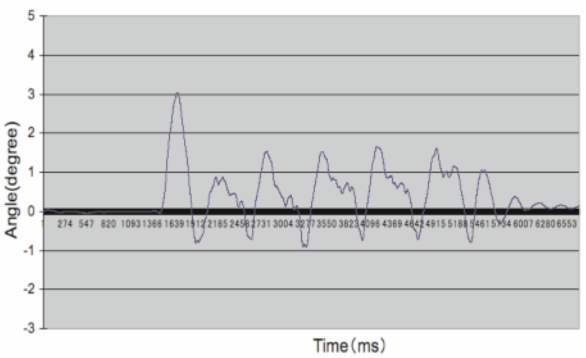


Figure 9: Tilt angle for “trot” walking style

As a next step, as we expect that one of the applications of this robot would to carry light load, we evaluated the walking stability from a point of carrying a load. For this we measured the tilt angle of the shoulder when it walks by fixing a gyro sensor on its shoulder and obtaining tilt data from it. We observed the time sequence data of the tilt angle ten times for each walking style and averaged the data. Figures 7, 8, and 9 show the obtained data for a phase shift of 0 degree (pace), 90 degree (gallop), and 180 degree (trot) for a specific observation. Also Figures 10, 11, and 12 show the same data after carrying out averaging operation.

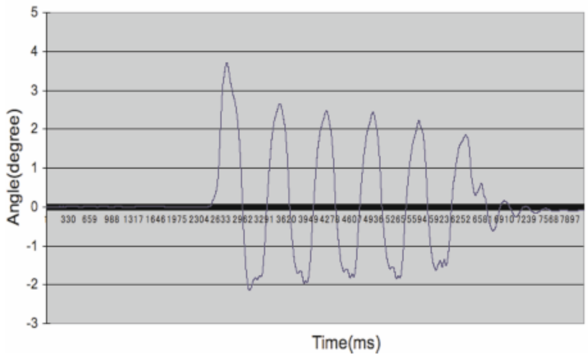


Figure 10: Tilt angle for “pace” walking style (average)

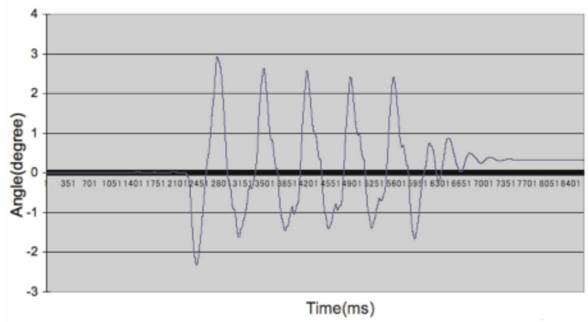


Figure 11: Tilt angle for “gallop” walking style (average)

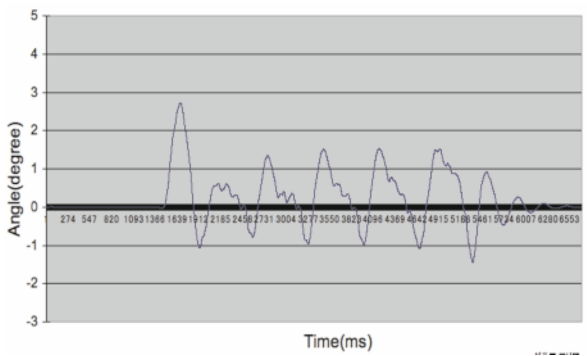


Figure 12: Tilt angle for “trot” walking style (average)

In this case the front and back legs move in opposite modes. For example, when the front left leg moves forward, so does the back right leg. Thus this walking style corresponds to trot style walking of animals. By changing the degree of phase shift, the robot can carry out various walking motions such as galloping. We will further study the relationship between phase shift and motion stability/speed. These results show that the tilt angle for trot walking style is lower than other two and thus more stable when carrying a load. At the same time even in the case of trot, the first step causes a little bit large tilt angle.

(2) Capability for other motions

We developed various types of human-like motions for the original humanoid robot [7]. An interesting question is which of these motions could work well on the centaur robot. We tried to transfer the humanoid robot motions to this robot and found that most of the motions worked fairly well on this robot. On the other hand, motions including such postures as bending and twisting did not work well or needed modifications. One interesting future research theme is automatically transferring the humanoid robot motions to the motions of four-legged robot such as this robot.

5 CONCLUSION

In this paper we proposed a new type of robot that is an integration of two types of robots: humanoid and four-legged. We adopted a humanoid robot with two legs and walking capability as a platform for this new robot. By integrating two of the humanoid robots we easily and successfully developed a centaur robot. We described its software and hardware and also its merits. We confirmed that by inserting a phase shift of 0 degree, 90 degree and 180 degree between the front and back leg motions the robot can stably achieve pace, gallop, and trot walking motions. Then we evaluated these three walking styles from a point of tilt angle of its shoulder and found that the trot walking style is more stable than other styles. Since this robot has merits of both humanoid and four-legged robots, we are also going to evaluate its new capabilities that neither of the two type robots could achieve by themselves.

6 REFERENCES

[1] http://www.honda/co.jp/ASIMO/

[2] Friedmann, M., Kiener, J., Petters, S., Thomas, D., Stryk, O. von, Sakakmoto, H., 2006, Versatile, high-quality motions and behavior control of humanoid soccer robots, Workshop on Humanoid Soccer Robots of the 2006 IEEE International Conference on Humanoid Robots, 9-16.

[3] http://www.jp.aibo.com/

[4] Golubovic, D., Li, B., Hu, H., 2003, A Hybrid Software Platform for Sony AIBO Robots, RoboCup 2003: Robot Soccer World Cup VII, 478-486.

[5] Kerepesi, A., Kubinyi, E., Jonsson, G. K., Magnusson, M. S., Kiklosi, A., 2006, Behavioural Comparison of Human-Animal (Dog) and Human-Robot (AIBO) Interactions, Behavioral Processes, Vol. 73, No.1, 92-99.

[6] Ishiguro, H., Ono, T., Imai, M., Kanda, T., 2003, Development of an interactive humanoid robot 'Robovie' -An interdisciplinary approach, R. A. Jarvis and A. Zelinsky (Eds.), Robotics Research, Springer, 179-191.

[7] Wama, T., Higuchi, M., Sakamoto, H., Nakatsu, R., 2004, Realization of Tai-chi Motion Using a Humanoid Robot, Entertainment Computing, Springer LNCS, 14-19.

Motion Simulation for a Stance Robot by Repeatedly Direct Kinematics

Jining Liu[1], Yoshitsugu Kamiya[2], Hiroaki Seki[3] and Masatoshi Hikizu[4]

[1] Division of Innovative Technology and Science, Kanazawa University, Kanazawa, Ishikawa, Japan

[2] Division of Innovative Technology and Science, Kanazawa University, Kanazawa, Ishikawa, Japan

[3] Division of Innovative Technology and Science, Kanazawa University, Kanazawa, Ishikawa, Japan

[4] Division of Innovative Technology and Science, Kanazawa University, Kanazawa, Ishikawa, Japan

Abstract

The aim of this research is to develop a multi-joint robot can stand up successfully with a luggage. This research requires the operations with two items. First, when the luggage rises, the robot stands up as easily as possible and does not fall over. Second, the load on each joint should be as small as possible. In this article, methods have been tried to evaluate the load torque and rotated angle of each joint in the geometrical limit and find the best algorithm to generate the trajectory of a lift-up motion for a stance robot by repeatedly direct kinematics.

Keywords:

Multi-joint robot; Trajectory Generation; Load Torque; Repeatedly Direct Kinematics

1 INTRODUCTION

With the developments of the society, more and more multi-joint robots have been used to meet the needs of the people and industry. For example, wheeled mobile robots and humanoid robots are placed in dangerous work in some fields such as in medical treatment, architecture, manufacture, and so on. This research focuses on how to control the robot easily and how to generate the better trajectory of the robot with multiple joints.

Concerning trajectory generation, there are two aspects to be considered. One is the goal you will achieve. What kind of trajectory is generated? Is it collision-free or time-optimal or energy-optimal? The other one is the constraint existing in the process of the trajectory generation [1].

Sometime we want the humanoid robot to work as workhorse without damage. But the joints are the parts, which are easy to damage. So how to make the joint load minimum is that we need to consider.

In this paper, we try to generate a trajectory that consists of many link postures, each of which makes the load torque of all the joints as low as possible even with the limitation of no turnover. Based on the idea above, the algorithm is proposed, and simulations are provided. The trajectories of lift-up motion for a multi-joint robot with a load are generated, only considering the static influences under some limitations [2]. And the precision of trajectories depends on the specified motion increment of each link in the calculation.

2 RESEARCH MODEL AND CALCULATIONS

The object of this study in this paper is simplified as the model shown in Figure 1. This model has five joints as a humanoid robot [3], which can move along the horizontal direction. So it has six degrees of freedom, which is a two dimensional model. On one side, we usually research about the robots under dynamic conditions, but which make the research complicated. The joints of robot are mostly drove by reducers in fact, and the motion of the links is very tiny in this research, so the static effect is first considered that the dynamic performances can be neglected.

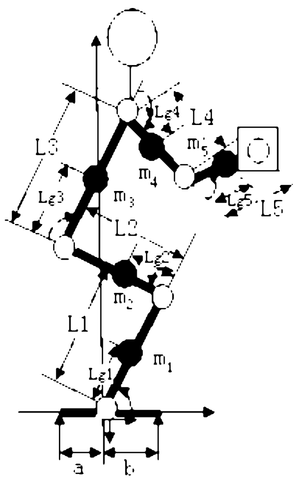


Figure 1: Model of the robot and its geometric parameters

The geometric parameters are showed in Table 1 below.

Table 1: Geometric parameters of the model

Joint	1 ankle	2 knee	3 waist	4 shoulder	5 elbow
***m*(kg)**	9.6	14.4	36	8.5	8.5
***L*(mm)**	463	450	416	300	320
***Lg*(mm)**	180	252	392	162	*x*
T_{imax}(Nm)	1200	500	550	400	400
T_{imin}(Nm)	-1200	-500	-550	-400	-400

Here, L is the length of the link, and Lg is the distance from the *CoM* (Centre of Mass) of the link to the beginning point of the link. The position of the *CoM* of the elbow link follows the mass of the luggage. To realize the lift-up motion for a multi-joint robot with a luggage, the following two requirements or constraints must be satisfied in this procedure.

First, the robot must maintain its stability or keep its balance so that it will not fall over, which is called the balance constraint here. To satisfy this condition, the *ZMP* (Zero Moment Point) or the projection of the *CoM* of the robot onto the ground must remain within the predefined stability region, that is, between the tiptoe and the heel. Because the dynamic performances are neglected, we do not consider the inertial force and influence from external forces. As mentioned above, when the robot is static, the *ZMP* coincides with the projection point of the *CoM* on the ground.

Second, the load torque of all the joints must be as low as possible, which is called the load constraint here. The multi-joint robots usually have many degrees of freedom. For example, the model in this research has 6 degrees of freedom, that is, each posture in the motion procedure will have 3^6 options. Finding the best option becomes a question we must face. So we adopted the load constraint to select a better option, which can evaluate the load situation of each joint in the motion procedure.

According to the model, the load torque of each joint (T_1, T_2, T_3, T_4, T_5) and reaction force (R_A, R_B) are indicated in the Equation below.

$$T_5 = m_5 L_{g5} g \cos(\theta_1 + \theta_2 + \theta_3 + \theta_4 + \theta_5) \tag{1}$$

$$T_4 = (m_4 L_{g4} + m_5 L_4) g \cos(\theta_1 + \theta_2 + \theta_3 + \theta_4) + T_5 \tag{2}$$

$$T_3 = [m_3 L_{g3} + (m_5 + m_4) L_3] g \cos(\theta_1 + \theta_2 + \theta_3) + T_4 \tag{3}$$

$$T_2 = [m_2 L_{g2} + (m_5 + m_4 + m_3) L_2] g \cos(\theta_1 + \theta_2) + T_3 \tag{4}$$

$$T_1 = [m_1 L_{g1} + (m_5 + m_4 + m_3 + m_2) L_1] g \cos\theta_1 + T_2 \tag{5}$$

$$R_A = \frac{T_1 + (m_1 + m_2 + m_3 + m_4 + m_5) g b}{a + b} \tag{6}$$

$$R_B = \frac{(m_1 + m_2 + m_3 + m_4 + m_5) g a - T_1}{a + b} \tag{7}$$

To satisfy the balance constraint, R_A and R_B must be more than 0. According to the Equation (6) and (7), the load torque of ankle joint T_1 has a limitation.

$$\begin{aligned} & -b(m_1 + m_2 + m_3 + m_4 + m_5) g \le T_1 \\ & \le a(m_1 + m_2 + m_3 + m_4 + m_5) g \end{aligned} \tag{8}$$

In fact, each joint of the robot has an overload constraint of the load torque that is initialized as Table 1.

To satisfy the load constraint, the Equation (9) is provided.

$$H_{T_i} = \begin{cases} \dfrac{T_i}{T_{i\min}} \times 100\% & (T_i < 0) \\ \dfrac{T_i}{T_{i\max}} \times 100\% & (T_i > 0) \end{cases} \tag{9}$$

Here, H_{Ti} is the torque rate, which must be less than 1, if the robot does not overload. H_{Ti} is one of the important evaluation parameters in this research, which can show us the situations about the burden of the torque that each joint affords in the movement process. As we known, the knee joint and the waist joint often suffer a pain to the aged, so we can also apply a torque rate limitation on this research to reduce the load torque the joints afford.

Considering the conditions above, we made the following algorithm to simplify and facilitate the analysis in Section 3.

3 ALGORITHM FOR TRAJECTORY GENERATION

An appropriate motion planning is important for the robot. That needs a better algorithm to implement robot's standing up with a luggage.

There must be a datum mark to take the robot to work out the erect motion, so we can select a joint as the datum mark of the motion in the procedure. In this research, the shoulder joint and the wrist joint are selected as the datum mark, because when the shoulder joint, or the wrist joint reaches the peak, the robot can rise to the erect phase. We will discuss respectively.

Because the standing up movements is a series of motions from an initial posture to the erect posture, the RDK (Repeatedly Direct Kinematics) method can be introduced into applying the trajectory generation of the standing movements [4]. In the task, there are three variables to the five joints (θ_1, θ_2, θ_3, θ_4, θ_5) and the robot maybe move forward or backward to adjust the centre of gravity. According to the RDK method, each joint is given a small increment, which means each posture have 3^6 options. The option that satisfies the aim and the constraint is selected as the result of this posture at this time. Such procedure is reiterated until the erect posture is reached.

The following is the algorithm of the trajectory generation:

1. Set up the limitations of joint torques T_{imax} (i=1, 2, 3, 4, 5).
2. Because there are four joints rotations and one foot motion, each motion part is given a small increment ($\Delta\Phi$, Δs) as below:

   $\theta_i=\theta_i+\Delta\theta_i$, $\Delta\theta_i=\{-\Delta\Phi, 0, +\Delta\Phi\}$ (i=1, 2, 3, 4, 5);

   $S_{foot}= S_{foot} +\Delta S_{foot}$, $\Delta S_{foot} = \{-\Delta s, 0, +\Delta s\}$

   There are 3^6 options available to each posture in the simulation.
3. Choose the options that the datum mark raised. If $R_A \times R_B$ <0 in all the options, the *ZMP* will move out of the predefined stability region. To prevent the robot from falling over, we should choose the option that R_j (j = A or B) is increased which is minus.

   Go back to 1.
4. Choose the options that R_A>0 and R_B>0. Find the best one from the options that must satisfy the following conditions:

It will not cause to fall over, having no violation of the joint torque limits.

It has a small torque rate as possible.

Go back to 1.

5. If the datum mark can not rise, end the program.

4 SIMULATION UNDER VARIOUS INITIAL POSTURES

Different trajectories result from different initial posture and different conditions. The trajectory will change, when the datum changes, even if a same initial posture. The usual simulation has 3 situations: no motion, moving backward and moving forward, as shown in Figure 2～7. The Figure 2～4 show the simulations of the shoulder joint as the datum. Figure 2 shows a simulation that an erect procedure of the robot having no horizontal motion, because the *ZMP* moves within the predefined stability region. As the 3rd graph shows, the *ZMP* is between the tiptoe (200mm) and the heel (-60mm), so the robot does not need to move. Figure 3 shows a simulation that an erect procedure of the robot having backward motion. As the 3rd graph shows, the *ZMP* is behind the heel (-60mm), that is, the *ZMP* moves out of the predefined stability region. At this time, the load torque that the ankle joint affords has exceeded the balance torque limitation, so the robot moves backward to make the *ZMP* move into the safe area. Figure 4 shows a simulation that an erect procedure of the robot having forward motion. As the 3rd graph shows, the *ZMP* is in front of the tiptoe (200mm), so the robot moves forward to adjust the position of the *ZMP*.

The Figure 5～7 show the simulations, when the wrist joint is the datum.

The model robot can move forward or backward under some conditions. It depends on the position of the *ZMP*. According to these simulations, we know that different trajectories can be generated when the datum is different, even if a same initial postures. While as Figure 2 and Figure 6 shown, although the robot has a same initial posture and a same datum, different mass of the luggage cause different motion of the robot and thus different trajectories.

According to the results of the simulations, we find that the end position of the luggage is not often on the top, when the shoulder joint is the datum, which is different from the simulations when the wrist joint is the datum. As Figure 2～4 shown, the final position of the luggage has 3 possibilities, above the shoulder joint, near the shoulder joint and under the shoulder joint, which depends on the initial posture of the robot.

For the same initial posture, whether the model robot can stand up or not also relates to the mass of the luggage. If overloaded, the robot will not be able to stand up with the luggage, that is, a certain joint cannot afford too heavy torque. For example, if the mass is more than 60kg, the robot cannot stand up with this luggage when the initial posture is θ_1=60°, θ_2=90°, θ_3=-80°, θ_4=-90°, θ_5=45°, which is same with Figure 4. At this time, the load torque of the knee joint exceeds the torque limitation, so the robot cannot stand up with a luggage as usual, unless the torque limitation of the knee joint is enlarged.

If a different motion increment is given to one of the joints, its trajectory will also change.

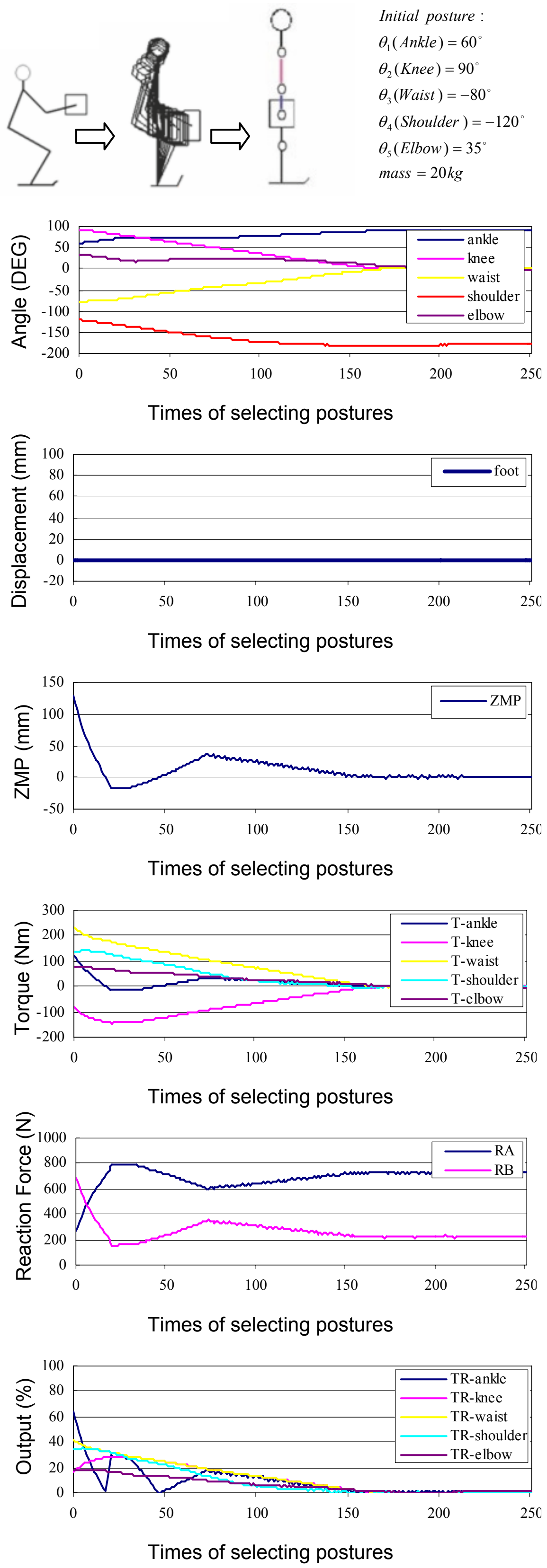


Figure 2: Simulation under the shoulder joint as the datum mark ①

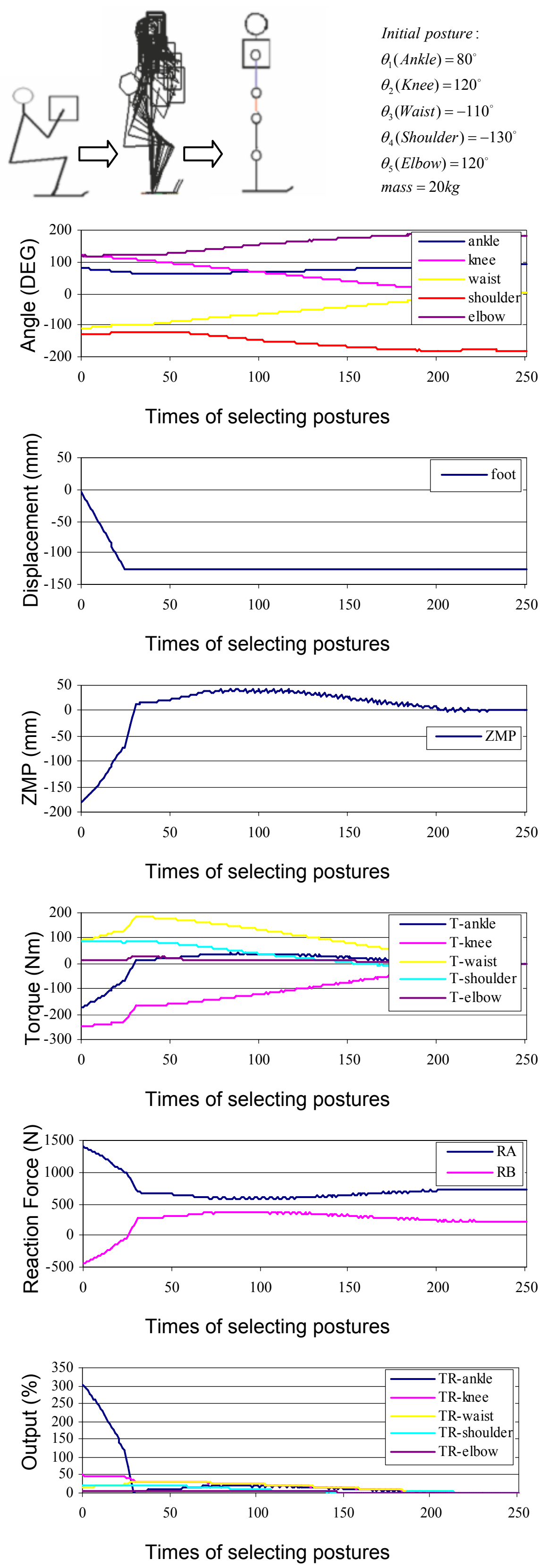


Figure 3: Simulation under the shoulder joint as the datum mark ②

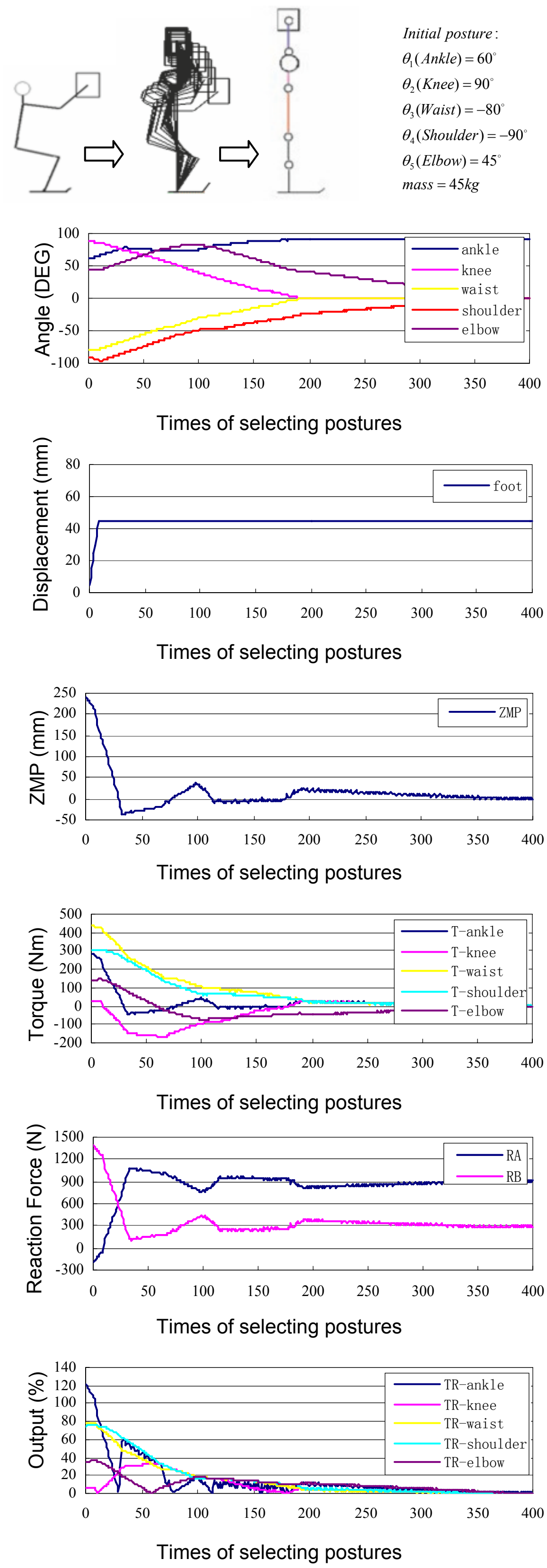


Figure 4: Simulation under the shoulder joint as the datum mark ③

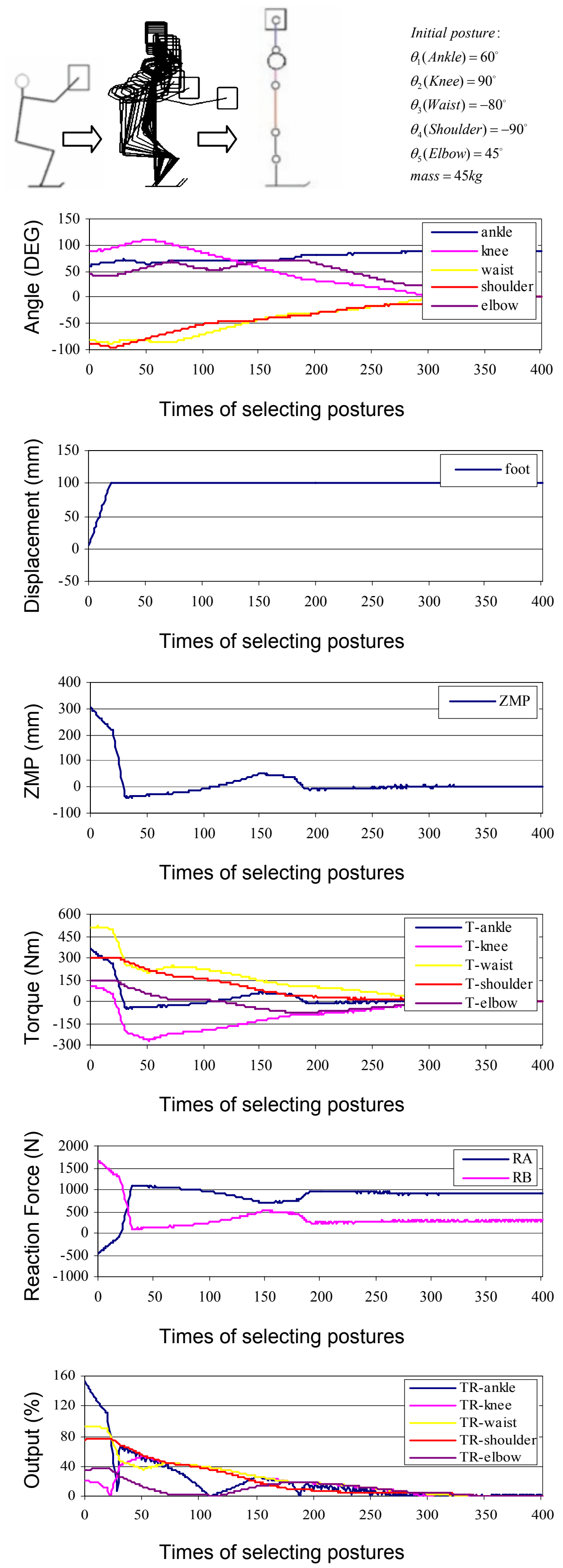


Figure 5: Simulation under the wrist joint as the datum mark ①

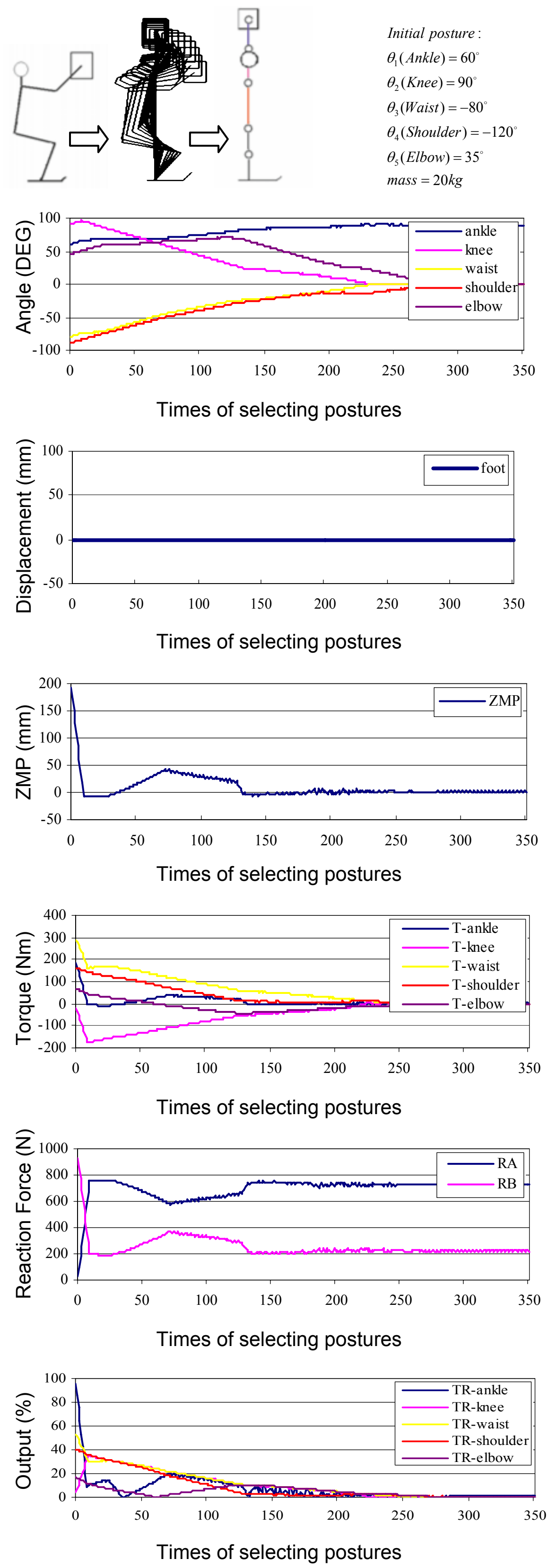


Figure 6: Simulation under the wrist joint as the datum mark ②

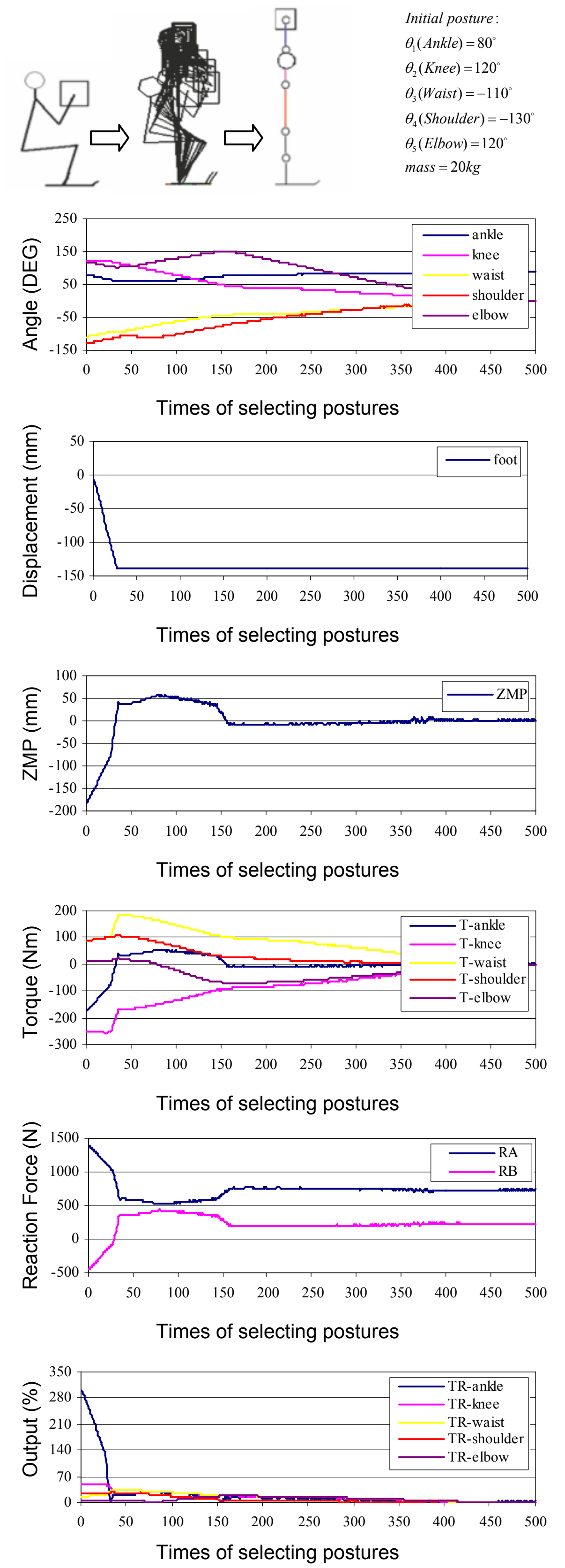


Figure 7: Simulation under the wrist joint as the datum mark ③

5 SUMMARY

According to the algorithm, we made simulations successfully under some predefined conditions and generated the trajectory of a lift-up motion for a stance robot by repeatedly direct kinematics. This simulation helps us gain the situations about the burden of the torque that the joint affords in the movement process.

This algorithm can benefit the research about the humanoid robot movement and helping the aged and the handicapped.

It is not enough for which the research is based on the static in simple environments. The research will be proposed in the dynamic and complex environments in future.

6 REFERENCES

[1] E. Bayo and B. Paden, On Trajectory Generation for Flexible Robots, Journal of Robotic Systems, 4(2), pp229-235, 1987.

[2] R.Kamnik: Standing-up robot, Journal of Medical Engineering & Technology, 28, 74/80 (2004)

[3] Lubo Song, Yoshitsugu, Kamiya, Hiroaki Seki, Sakiichi Masatoshi Hikizu and Qin Zhang, Solutions of Inverse Kinematics of Articulated Robot by Using Repeatedly Direct Kinematics, Journal of the Japan Society for Precision Engineering, 67(6), pp971-976, 2001

[4] Yoshitsugu Kamiya, Tetsuya Kubo, Seiji Aoyagi and Sakiichi Okabe, Control of Robotic Manipulators Using the Solutions of Repeated Direct Kinematics, Transactions of the Japan Society of Mechanical Engineers (Series C), 59 (564), pp125-130, 1993

Manipulation Method Using Wheeled-Hand Mechanism for Humanoid Robot

Takuya Fukuda[1], Tomohito Takubo[1], Yasushi Mae[1], Tatsuo Arai[1]

[1] Dept. of System Innovation, Graduate School of Engineering Science, Osaka University, Osaka, Japan

Abstract

A new manipulation method by using wheeled-hand mechanism for humanoid robots is proposed. The manipulation enables long range reaching tasks by putting hands on the table. The wheeled-hand mechanism can change continuously the contact point of the hand and the table by wheel. The wheeled-hand motion is controlled considering the stable whole body motion. We define multipoint support regions defined by two legs and one hand, and the center of mass (CoM) of the robot is controlled to be located in them. The support polygon is defined so that the robot keeps stability even if one hand lifts off to catch a target object. In this method, the robot can catch the object in far place on the table without discrete step motion with arm. We develop a prototype wheeled-hand and implement it to humanoid robot HRP-2. The validity of the proposed method is confirmed by experiment.

Keywords:

Humanoid Robot; Manipulation; Multipoint Support; Wheeled-Hand

1 INTRODUCTION

The advanced robot technology will introduce robots to various fields for helping human workers. The human-robot cooperative work could be achieved easily by using humanoid robot which has two arms and two legs since it can use human like motion. On the other hand, robots have characteristic mechanism like crawler driven and wheel locomotion. They have many advantages: conservation energy, easy control and simple mechanism. Thus, hybrid Leg-wheel robots provide the advantage of stable, efficient movement over rough terrain in walking and speeding over flat terrain on wheels [1]-[6]. In the previous study, the hybrid motions are utilized for locomotion so that the travelling speed and power efficiency could be better. Legged mechanism has great adaptability for the discontinuous terrain like steps. Wheeled mechanism can travel fast, stably, and efficiently on the flat terrain and the small undulation. Based on the advantages, we propose a new manipulation motion with wheeled-hand which has a active wheel mechanism. It provides continuous modification of the contact region, and also it can support manipulation force by the drive force of the wheel.

Manipulation is one of the feature points of humanoid robot. Harada [7] proposed a discrete approach method which plans stable whole body motion based on multipoint support. It enables to approach an object in far place on a table from the standing position by using stepping motion with arms. However, the planning of contact point is difficult and it requires many steps to approach a target according to the stability of the robot. Our proposed hand-wheel hybrid motion makes the problem easy by using continuous modification of support region with wheeled-hand motion. The approach is achieved by simultaneous motion of arm and wheeled hand, and the contact between the hand and the environment can be kept even if the hands-legs support region should be changed. For performing whole body manipulation with wheeled-hand, we define stable support region to estimate the robot stability and manipulation strategy. The stable region consists of the projection of each contact point of arms and legs. When the projection of the center of mass (CoM) locates in the stable support region, humanoid robot can figure out well positioned. When the robot returns to the upright posture from the leaning posture, it can use the drive force of the wheel mechanism to get reaction force from table. The specification of the wheeled-hand and its implementation are illustrated, and the feasibility of the wheeled-hand manipulation with whole body motion is shown.

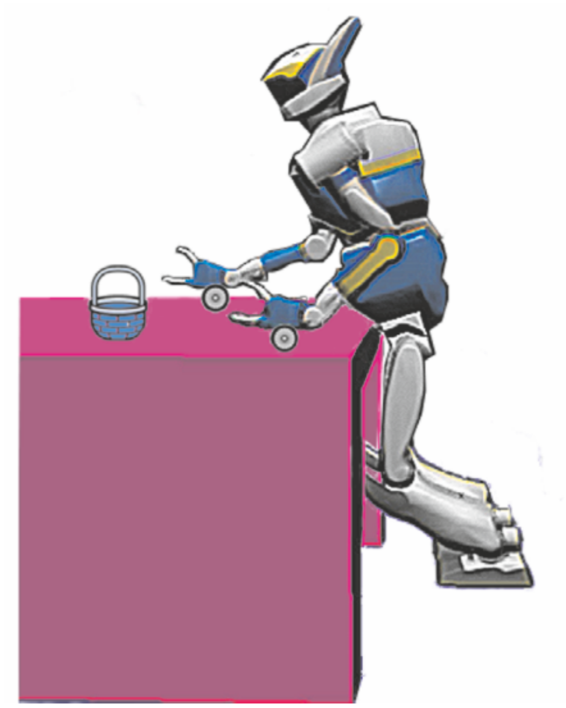


Figure 1: Manipulation task using wheeled-hand motion.

This paper is organized as follows. In section 2, we describe the manipulation strategy with wheeled-hand. Section 3 explains a developed wheeled-hand mechanism. Section 4 shows an experiment using wheeled-hand motion and the effectiveness of the proposed method. Finally, we conclude this paper in section 5.

2 MULTIPOINT SUPPORT MOTION USING WHEELED-HAND

2.1 Multipoint support motion

Whole body motion for humanoid robot using multipoint contact with environment is proposed by Harada [7], [8] and Nishiwaki [9]. Hauser [11] and Escande [12] proposes the motion planner to select a contact point which provides the target posture. These methods are complecated and a lot of procedures are required. Usually, when the robot changes the contact point of hand and environment, the contact should be broken. In this case, the robot should make premilinary motion to keep stability by another hand. To simplify the motion, we propose wheeled-hand and new manipulation strategy.

Figure 1 shows proposed manipulation concept using wheeled-hand. The wheeled-hand has an active wheel mechanism on the back of hand. The wheel makes contact with environment to move the contact point without releasing. The condition provides advantage to keep stability since the support polygon is modified continuously and it makes large area. In addition, the wheel can generetes drive force to support manipulation force, even if support state consists of three points during manipulating by one hand.

In this paper, we consider a manipulation task for an object put on a table. This task includes reaching and grasping. If the object is put on the far place from the edge of the table, it is difficult to move the hand to the object without using multi-point support motion with hands. In this situation, humans make the contact of their hand with the table to keep the stability. It also improves stability and work space of manipulation that humanoid robot makes multipoint support state with arm and leg.

In this manipulation task, we assume as follows;

- Position and size of the table are known.
- The table is not deformed by the force from the robot.
- Table does not slip on the floor.
- Friction between the table and the wheel is enough.

2.2 Stability criterion on multipoint support

When the position of the object is far from the edge of table, the robot makes multipoint support state. Figure 2 shows a projection of contact points between the robot and environment on X-Y plane. F1 and F2 indicate contact points of each foot, and H1 and H2 indicate contact points of each hand. We define two support polygons which are F1-F2-H1 and F1-F2-H2 by assuming both feet keep contact with ground. Now, we consider two types of multipoint support state as follows;

4-point-support state

Both feet come into contact with the floor, and both hands touch the table.

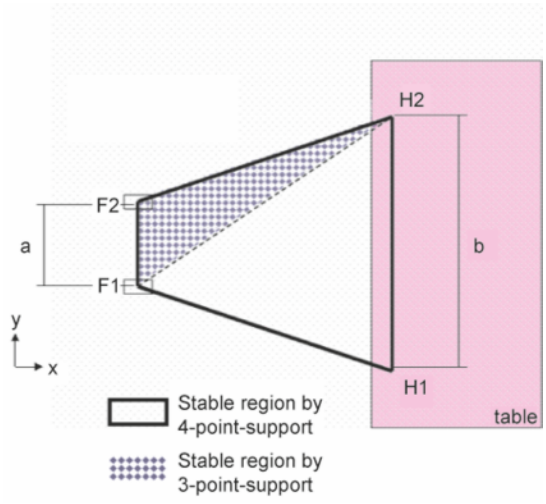


Figure 2: Support region by multipoint contact.

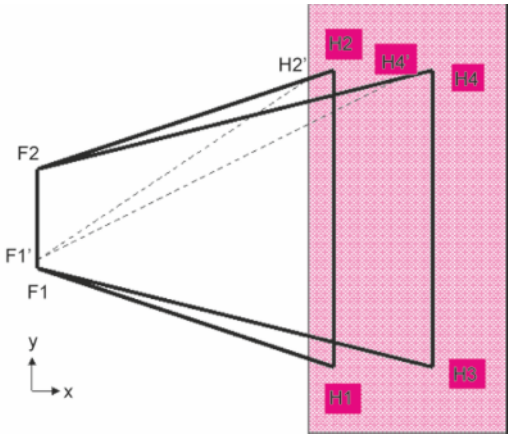


Figure 3: Transition of support polygons.

3-point-support state

Both soles come into contact with the floor, and one hand touches the table. Another hand is lifted for manipulation and it does not support the weight of the robot.

In either condition, the CoM position should be located in support polygon, but the 4-point-support state is stable rather than the 3-point-support state. Thus, 4-point-support state is used for approach motion, and it changes to 3-point-support state for manipulation.

Here, manipulation strategy is defined as follows;

1. Double leg support phase: The robot is standing by using only both feet
2. Transition to multipoint support: Both hands are put on the table, and the CoM moves forward to be supported by both hands and both feet for making stable posture.
3. Manipulation: By using drive force of wheel, both hands move forward to approach the object on the table. When the object is in the manipulation area, one hand grasps and lifts it.

4. Return to initial phase: By using wheel locomotion, another hand moves back on the table. The drive force of wheel supports to move the object. Then, the CoM is moved into the support polygon defined by the feet, and the hand lifts off from the table.

In this strategy, the CoM moves twice. The target CoM position is defined to provide enough work space for both hands. Therefore, the target CoM should be controlled into the common region of support polygons of F1-F2-H1 and F1-F2-H2. Figure 3 and Table 1 show support polygons and contact points corresponding to the each motion. The phase (a) is initial phase of multipoint support motion. The hand can reaches to the object in the phase (b). The phase (a) and the phase (b) employ 4-point support states to keep stability and work space. When the object is located in the work space, one hand grasps it and lifts off from the table. The phase shifts to (c) which is 3-point support state. In phase (d), another hand which touches the table returns to the initial contact point.

In Figure 3, there is common region of each support polygon and the target CoM position is defined in it so that the manipulation strategy becomes easy. However, from the phase (c), the required force for supporting legs and arm becomes larger than the former phase since the weight of wheeled-hand for manipulation and the target object should be supported by the other contact points of environment. The transition ***r*** of CoM position is calculated as follow;

$$\boldsymbol{r} = \frac{m}{M+m}(\boldsymbol{r_2} - \boldsymbol{r_1}) \qquad (1)$$

where M is the weight of the robot, m is the total mass of the object and the wheeled-hand, $\boldsymbol{r_1}$ is the initial position of the CoM, and $\boldsymbol{r_2}$ is the target position of the hand to grasp the object. Hence, the support polygons on 3-point support state should be changed like dashed line in Figure 3.

To locate the CoM into the support polygon, we define the target position of the CoM as following C_{refx},

$$C_{refx} = K\frac{a}{a+b}r_{Fx} \qquad (2)$$

where a is the distance between both feet in y-axis, b is the distance between both hands in y-axis, and r_{Fx} is the position of hand in the leg coordinate system. We can get target CoM position C_{refx} by using arbitrary coefficient K. When $0 < K < 1$, the target CoM is located in the support polygon.

3 WHEELED-HAND

We developed a wheeled-hand module for humanoid robot HRP-2 to implement multipoint support motion. Figure 4 and Figure 5 show the developed hand module. This is implemented to the wrist of the humanoid robot. Wheeled-hand consists of two mechanism: multi-finger hand and active wheel mechanism. Multi-fingered hand works for grasping and manipulating various objects. Wheel is driven by motor, and it makes the contact with environment. Table 2 lists the specific of wheeled-hand.

Table 1: Multipoint support state and contact points.

State	Contact points	Corresponding motion
(a)	F2,F1,H1,H2	Start multipoint support.
(b)	F2,F1,H3,H4	Reach to the object.
(c)	F2,F1’,H4’	Lift the object.
(d)	F2,F1’,H2’	Move hand to the body.

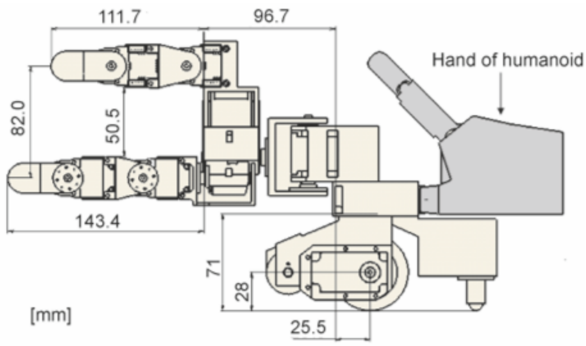


Figure 4: Side view of developed wheeled hand (CAD).

3.1 Wheel mechanism

- One wheel is attached to a powerful motor (Max. Torque is 53kgfcm). The wheel consists of rubber tire to keep large friction with environment.
- Rotation angle can be measured by a rotary encoder connected with timing belt.
- The target angle is commanded via RS485 from PC.

3.2 Multi-fingered hand

- Each joint employs high torque intelligent servo Dynamixel DX-116 developed by the ROBOTIS Co. in which the motor driver is packaged.
- The two fingers with 3 servo-motor and one finger with 2 servo-motor are selected to maintain the grasping function and minimise the number of servo-motor for light weight.

4 EXPERIMENT AND DISCUSSION

4.1 Experiment on HRP-2

The developed wheeled-hand is implemented to the humanoid robot HRP-2 [12], we experiment the manipulation task using multipoint support motion. The object is a cylinder whose diameter is 65mm, and it is put on the table whose height is 0.7m. Figure 6 shows support polygons in the experiment. These support polygons are corresponding to Figure 3. The area enclosed by dashed line indicates the phase (c) and the phase (d), referred to Table 1. And the area enclosed by solid line indicates the state (a) and the state (c). Distance of both feet is 0.19m, and distance of both hands is 0.50m. X-position of the object is 0.4295m from first contact point of wheeled-hand with the table. The coefficient K is set 0.6, so the target x-position of CoM C_{refx} is defined 0.07m in foot coordinate system

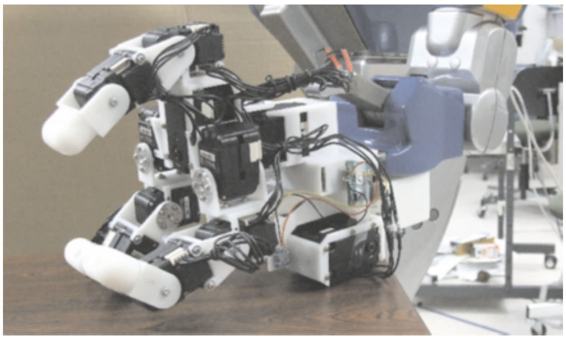


Figure 5: Wheeled hand connected to HRP-2.

from equation (2). When the robot finishes moving the CoM to the reference position, wheeled-hand is controlled as follows;

1. Wheeled-hand moves forward 0.15m.
2. Right hand grasps the object by multi-finger, and lifts it. It makes 3-point support state by both feet and left hand.
3. Left wheeled-hand moves backward 0.15m.

Weight of the humanoid robot M is 58kg, and the total mass m loaded right arm is 2kg.

Figure 7 shows snapshots of the experiment. While multipoint support motion, the CoM keep the constant position but the motion is stable even if the one hand lifts off. Detailed comment is as follows;

- The robot touches table with both hands: (c).
- The CoM shifts forward and it becomes multipoint support state: (d).
- Both hands move forward by wheel and right hand grasps the object: (e).
- Right hand lifts the object: (f).
- Left hand moves backward by wheel at 3-point support state: (g).
- The CoM shifts backward to be located in the support polygon defined by both feet: (h).
- The left hand lifts off from the table: (i).

Figure 8 shows the motion of approaching to the object, and Figure 9 shows the motion of lifting the object. Figure 10 shows the total reaction force on both hands in travelling direction, and Figure 11 shows the reaction force of each hand in vertical direction. These reaction forces are measured by built-in force sensor at the end of arm on HRP-2. From Figure 11, it is confirmed that the force in the vertical direction is required around 40N to 60N for both hands. From Figure 10, around 15N force in travelling direction is required to move the left hand backward. The reaction force is generated by the drive force of the wheel.

4.2 Evaluation of hand module

Strength of hand module

The strength of wheeled hand is necessary to support the weight of the robot at contact point. Developed wheeled hand has a frame which connects multi-fingered hand and wheel mechanism to the arm of humanoid robot. Force from the arm applies the contact point through this frame. Thus, the strength of the frame of wheeled hand should be evaluated in

Table 2: Specific of wheeled hand.

Diameter of wheel	56mm
Number of fingers	3
Number of actuators(fingers)	8
Number of actuators(wrist)	2
Total mass	1.5kg

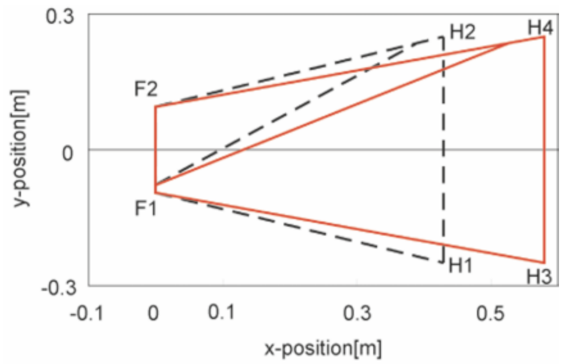
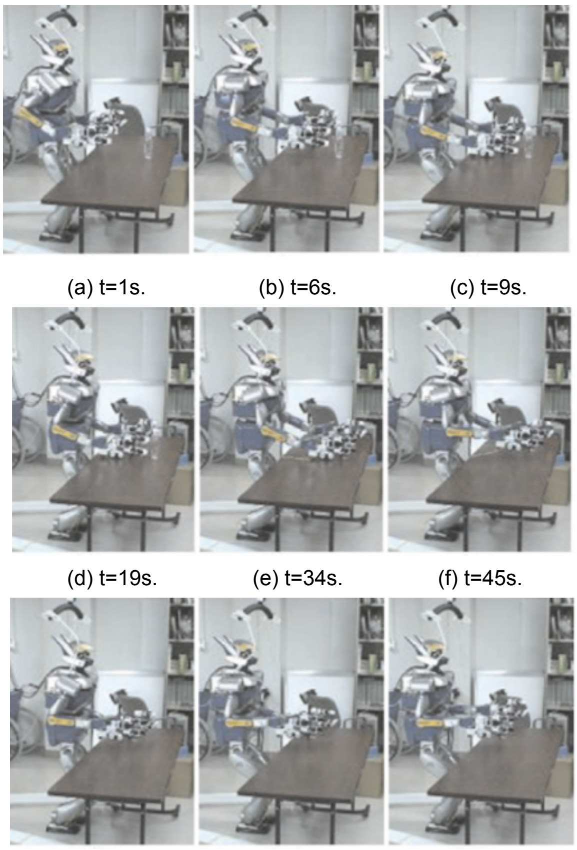


Figure 6: Support polygons in experiment.



(a) t=1s. (b) t=6s. (c) t=9s.

(d) t=19s. (e) t=34s. (f) t=45s.

(g) t=53s. (h) t=64s. (i) t=69s.

Figure 7: Experimental result.

simulation. A structure analysis tool of AUTODESK INVENTOR is used for simulation. Simulation conditions are as follows;

- Force from the arm is 500N, and a load by multi-fingered hand and the manipulated object is 10N.
- The frame is made from ABS plastic.

Property of ABS plastic is below. Young's modulus is 2890MPa. Poisson ratio is 0.38. Tensile strength at yield is 40.33MPa.

Figure 12 shows an analysis result of the stress distribution on the frame, and Figure 13 shows an analysis result of deformation distribution of the frame. The report of analysis is as follows;

- Maximum stress is 7.66MPa.
- Maximum deformation is 0.151mm.
- Safety factor is over 5.265.

Hence, even if all weight of the robot applies one arm, the frame is not broken. Therefore, this frame can be used for more heavy works.

Equilibrium at the contact point

The friction force at contact point between wheel and table is discussed here. The maximum drive force of wheel is calculated by using maximum motor torque and radius of the wheel. In case of developed wheeled hand, maximum torque is 53kgfcm, and radius is 28mm. Thus, the wheel can generate 18.9kgf in the maximum.

However, the maximum reaction force from the environment is defined by friction coefficient. Now, we can define slip condition as follow:

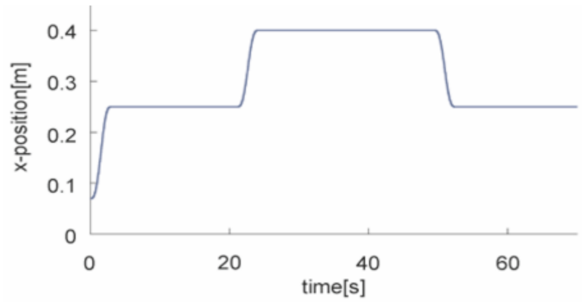


Figure 8: X-position of hand in experiment.

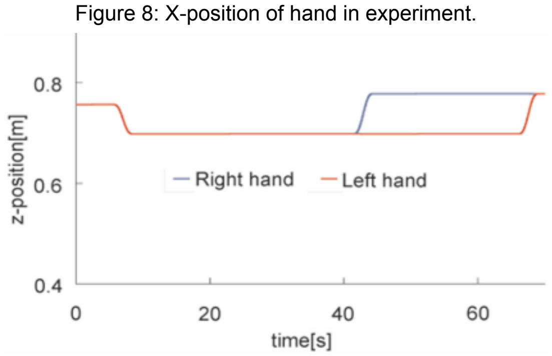


Figure 9: Z-position of hand in experiment.

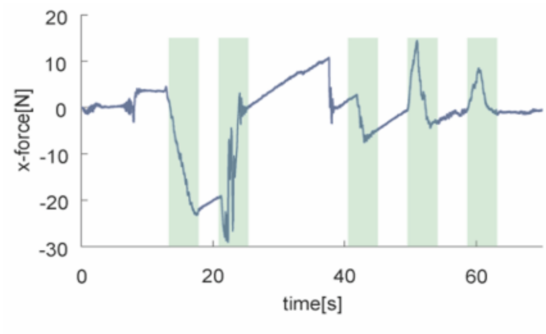


Figure 10: Total reaction force of both hands in x-direction.

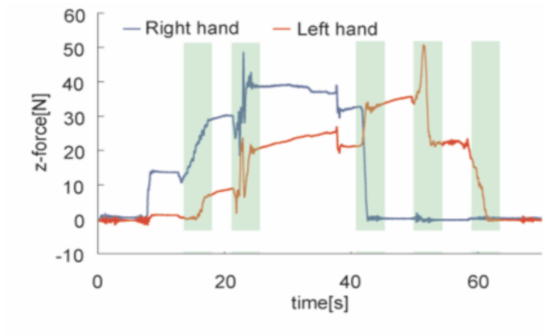


Figure 11: Reaction force of hand in z-direction.

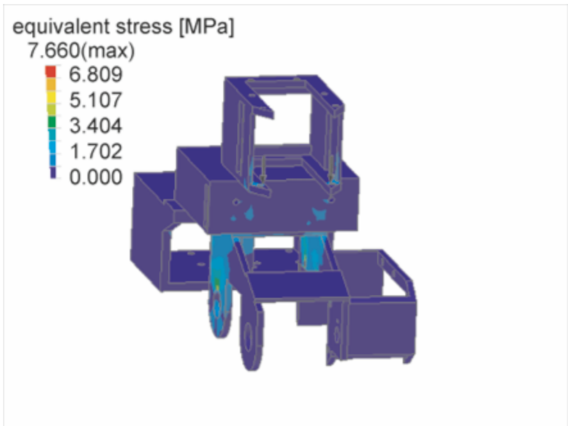


Figure 12: Analysis result of stress distribution.

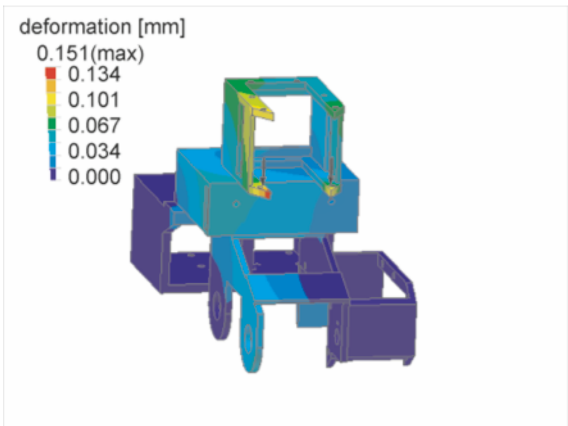


Figure 13: Analysis result of deformation distribution.

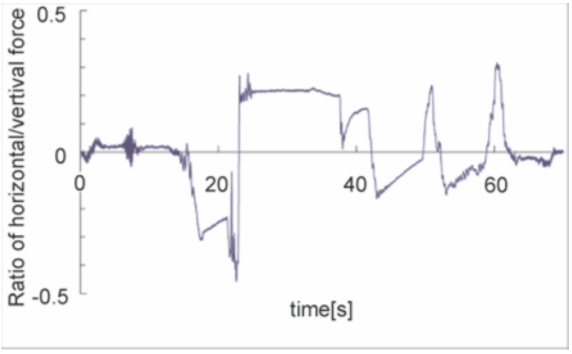


Figure 14: Ratio of horizontal/vertical force at contact point.

In this calculation, W is set to 2kgf. Experimental value of mu is 0.4 to 0.6. Therefore, Figure 14 shows no slips of tire.

$$\frac{f_H}{f_V + Wg} \le mu \tag{3}$$

where f_H is horizontal force, f_V is vertical force, W is mass of wheeled-hand, g is gravity acceleration, and mu is friction coefficient tire with top of the table. If this condition is satisfied, friction force is larger than f_H, and it does not slip. Figure 14 shows the ratio of force at contact point in experiment. This ratio is equal to left side of equation (3).

5 COCLUSION

This paper presents a new manipulation method using wheeled-hand mechanism whose active wheel generates drive force to move the contact point of hand without lifting it. The stable multi-point support motion with wheeled-hand is proposed to manipulate an object in far place on the table. The method allows the humanoid robot to keep the contact between hands and environment. The condition makes the strategy of whole body motion easy. The proposed method is realized on the humanoid robot HRP-2 and the feasibility is confirmed.

6 REFERENCES

[1] Yuan, J., Hirose, S., 2005, Zero Carrier: A Novel Eight Leg Wheels Hybrid Stair Climbing Mobile Vehicle, Journal of Robotics and Mechatronics, Vol.17, No.1 pp.44-51

[2] Grand, C., BenAmar, F., Plumet, F., Bidaud, P., 2004, Decoupled control of Posture and trajectory of the hybrid wheel-legged robot Hylos, Proc. of ICRA2004, pp.5111-5116.

[3] Nakajima, S., Nakano, E., Takahashi, T., 2006, Motion Control Technique for Practical Use of a Leg-Wheel Robot on Unknown Out-door Rough, Proc. of IROS2006, pp.1353-1358.

[4] Hirose, S., Takeuchi, H., 1996, Study on Roller-Walk (Basic Characteristics and its Control), Proc. of ICRA1996, pp.3265-3270.

[5] Adachi, H., Koyachi, N., Arai, T., Shimizu, A., Nogami, Y., 1999, Mechanism and Control of a Leg-Wheel Hybrid Mobile Robot, Proc. of IROS1999, pp.1792-1797.

[6] Takahashi, M., Yoneda, K., Hirose, S., 2006, Rough Terrain Locomotion of a Leg-Wheel Hybrid Quadruped Robot, Proc. of ICRA2006, pp. 1090-1095.

[7] Harada, K., Kajita, S., Kaneko, K., Hirukawa, H., 2003, ZMP Analysis for Arm/Leg Coordination, Proc. of ICRA2003, pp.75-81

[8] Harada, K., Hirukawa, H., Kanehiro, F., Fujiwara, K., Kaneko, K., Kajita, S., Nakamura, M., 2004, Dynamical Balance of a Humanoid Robot Grasping an Environment, Proc. of IROS2004, pp.1167-1173.

[9] Nishiwaki, K., Kuga, M., Kagami, S., Inaba, M., Inoue, H., 2005, Whole-body Cooperative Balanced Motion Generation for Reaching, International Journal of Humanoid Robotics, Vol.2, No.4, pp.437-457.

[10] Hauser, K., Bretl, T., Latombe, J.C., 2005, Non-Gaited Humanoid Locomotion Planning, Proc. of the 2005 IEEE/RAS International Conference on Humanoid Robots, pp.7-12.

[11] Escande, A., Kheddar, A., Miossec, S., 2006, Planning support contact-points for humanoid robots and experiments on HRP-2, Proc. of IROS2006, pp.2974-2979.

[12] Kaneko, K., Kanehiro, F., Kajita, S., Hirukawa, H., Kawasaki, T., Murata, M., Akachi, K., Isozumi, T., 2004, Humanoid Robot HRP-2, Proc. of ICRA2004, pp.1083-1090.

Study on One-legged Robot Jumping

Masafumi Miwa[1], Hiroyasu Sakane[2], Kenji Nagase[2], Yasuhiro Koshimoto[2], and Shigeki Tuchitani[2]
[1] Institute of Technology and Science, The University of Tokushima, Tokushima, Japan
[2] Dept. of Opto-Mechatronics, Faculty of Systems Engineering, Wakayama University, Wakayama, Japan

Abstract
In this study, simulations and experiments were done with experimental one-legged robot, and jump motion results with or without impedance matching design were compared. Jumping motion of the experimental robot was captured by high-speed video camera, and jumping height was measured from captured movie. As the results, jumping motion with impedance matching design was about one and a half times higher than that without impedance matching design in both experiments and simulations.

Keywords:
One-legged robot; Jump motion; Impedance matching design

1 INTRODUCTION

Recently, various types of robots are used in many fields, such as factories (limited environment), barren, disaster site, also human's living environment (non limited environment). Above all, needs of support robot for human are increasing. In such case, these robots require the moving ability in human's living environment, for examples, obstacle avoidance, going up/down the step, moving velocity equal to human, etc. At present, common moving mechanisms of robot are wheel, crawler and legs. Legs have advantage on obstacle avoidance, crawler has advantage on going up/down the step, and wheel has advantage on moving velocity. When leg type robots adapt to human's living environment, improvement of moving velocity is important. And, 'running' operation is necessary for high speed moving of legs type robot.

In this study, we focused on jumping motion that is element feature of running, and tried to design efficient jump motion. Test one-legged robot and simulation program were developed, and the impedance matching index [1][2][3][4][5] was introduced for evaluating transmission efficiency to analyse / design / evaluate the jump motion. The Impedance matching index is an index of torque transmission efficiency estimated from robot attitude.

2 EXPERIMENTAL SETUP

2.1 Experimental robot [6]

Figure 1 shows the experimental robot. Experimental robot is made with the same parts of HOAP-1 (Fujitsu Automation : Humanoid Robot; motor: L0PW20221-B01X; control circuit : L0PW20221-B04X). Robot frames were made from aluminium. This robot has two joints and two motors, and one motor drives knee joint and the other drives heel joint. Motors and joints are connected with timing belt.

2.2 Robot model

Figure 3 shows the dynamic model of the experimental robot. In this study, only movement on x-z plane is considered. Table 1 shows parameters of each links and robot parameters.

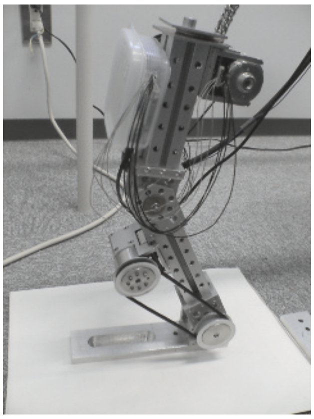


Figure 1: Test One-legged Robot.

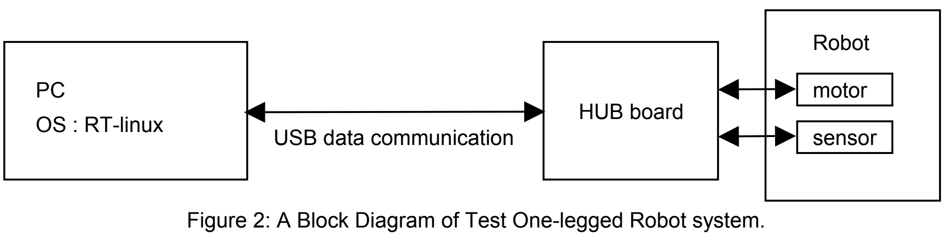


Figure 2: A Block Diagram of Test One-legged Robot system.

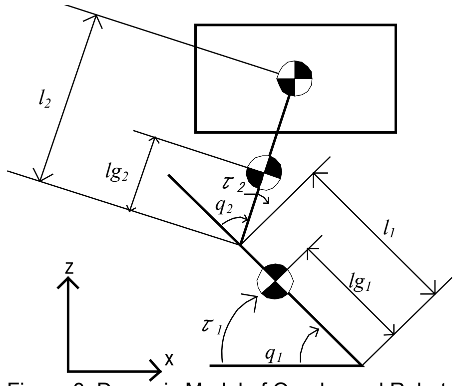


Figure 3: Dynamic Model of One-legged Robot.

3 TRANSMISSION EFFICIENCY ESTIMATION BY IMPEDANCE MATCHING [1][2][3][4][5]

Before experiments and simulations, the transmission efficiency of knee joint and heel joint torque were calculated with the impedance matching index.

Dynamics of 2DOF robot in Figure 3 was derived by Euler-Lagrange equation.

Table 1: One-legged robot specifications

Mass(kg)	1.336(kg)
Width	0.18(m)
Depth	0.04(m)
Height	0.36(m)
l_0（Body link）	0.18(m)
l_1（2nd thigh length）	0.18(m)
l_2（thigh length）	0.13(m)
m_0（body mass）	0.235(㎏)
m_1（2nd thigh mass ）	0.465(㎏)
m_2（thigh mass）	0.450(㎏)
M_p（frame mass）	0.185(㎏)
I_1	2nd thigh inertia
I_2	thigh inertia
lg_1	gravity point of 2nd thigh
lg_2	gravity point of thigh

$$\boldsymbol{M(q)\ddot{q} + C(q,\dot{q}) + G(q) + J(q)^T F_e = \tau} \tag{1}$$

where

$$\boldsymbol{M(q)} = \begin{bmatrix} m_1 l_{g1}{}^2 + I_1 + m_2(l_1{}^2 + l_{g2}{}^2 + 2l_1 l_{g2} c_2) + I_2 & m_2(l_{g2}{}^2 + l_1 l_{g2} c_2) + I_2 \\ m_2(l_{g2}{}^2 + l_1 l_{g2} c_2) + I_2 & m_2 l_{g2}{}^2 + I_2 \end{bmatrix}$$

$$\boldsymbol{C(q,\dot{q})} = \begin{bmatrix} -2m_2 l_1 l_{g2} s_2 \dot{q}_1 \dot{q}_2 - m_2 l_1 l_{g2} s_2 \dot{q}_2^2 \\ m_2(l_{g2}{}^2 + l_1 l_{g2} \dot{q}_1 \dot{q}_2) \end{bmatrix}$$

$$\boldsymbol{G(q)} = \begin{bmatrix} m_1 g l_{g1} c_1 + m_2 g(l_1 c_1 + l_{g2} c_{12}) \\ m_2 g l_{g2} c_{12} \end{bmatrix}$$

$$\boldsymbol{J(q)} = \begin{bmatrix} -l_1 s_1 - l_2 s_{12} & -l_2 s_{12} \\ l_1 c_1 + l_2 c_{12} & l_2 c_{12} \end{bmatrix}$$

s1=sin(q1), s12=sin(q1+q2), c1=cos(q1), c12=cos(q1+q2).

In left part of equation (1), first term is inertia, second term is velocity square term (centrifugal force, Coriolis force, etc.), third term is gravity term, and 4th term is external forces on body part. And number of joints is n = 2, number of considering dimension is m = 2. Dynamics of the body part is represented by

$$M_p \ddot{\boldsymbol{x}} + M_p \boldsymbol{g} = \boldsymbol{F_e} \tag{2}$$

The first term of equation (2) is inertia, and the second term is gravity term. Equation (3) is derived from equations (1) and (2) with equality $\dot{\boldsymbol{x}} = \boldsymbol{J(q)\dot{q}}$.

$$\boldsymbol{\tau = M(q)J(q)^{\dagger}} M_p{}^{-1} \boldsymbol{(F_e} - M_p \boldsymbol{g} - M_p \boldsymbol{\dot{J}(q)\dot{q}) + C(q,\dot{q}) + G(q) + J(q)^T F_e} \tag{3}$$

$$= \boldsymbol{Q(q)(F_e - F_{bias})} \tag{4}$$

where

$$\boldsymbol{F_{bias} = (J(q)^T + M(q)J(q)^{\dagger} M_p{}^{-1})^{\dagger} \times [M(q)J(q)^{\dagger}(g + \dot{J}(q)\dot{q}) - C(q,\dot{q}) - G(q)]} \tag{5}$$

$$\boldsymbol{Q(q) = J(q)^T + M(q)J(q)^{\dagger} M_p{}^{-1}} \tag{6}$$

and $J(q)^{\dagger} \in R^{m \times n}$ is a pseudo inverse of the Jacobian matrix $J(q)$. Equation (4) denotes the relationship between motor torque and force and momentum applied to the body. And coefficient matrix $Q(q)$ indicates the torque-force transfer efficiency, F_{bias} is a bias term related to velocity and gravity.

Equation (7) shows the singular value decomposition of $Q(q)$,

$$Q(q) = U \Sigma V^T \tag{7}$$

where

$$\Sigma = \left[diag(\sigma_1, \sigma_2, \cdots, \sigma_x), 0\right] \quad (\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_x \geq 0)$$

U, V are orthogonal matrix, and σ_i is singular value.

Let multiply all singular value, we get

$$\omega = \sigma_1^{-1} \cdot \sigma_2^{-1} \cdots \sigma_x^{-1} \tag{8}$$

Equation (8) shows the impedance matching index, which expresses the total transmission efficiency from motor torque to force and momentum on the robot body.

3.1 Comparison of impedance matching level

To estimate impedance matching index of each motion pattern of the one-legged robot, $Q(q)$ was calculated by substituting all parameter values to equation (5), then ω was calculated by singular value decomposition on MATLAB.

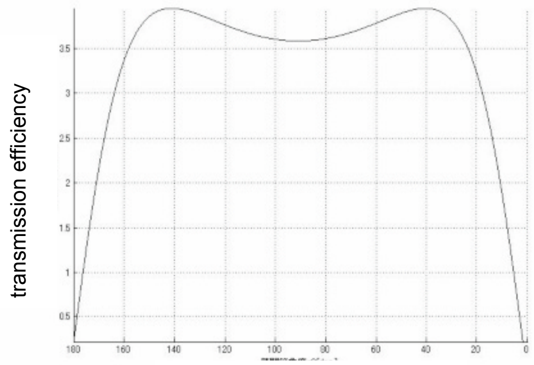


Figure 4: Transmission efficiency of knee joint torque.

Figure 4 shows the transmission efficiency of knee joint torque. In this Figure, impedance matching index gets maximum value when knee joint angle is 40 [deg.]. So we can design high transmission efficiency motion with impedance matching index by changing the parameters such as link length, mass, etc.

Using these results, some motion patterns were designed with/without optimisation of transmission efficiency.

4 JUMP MOTION DESIGN WITH IMPEDANCE MATCHING ELLIPSOID [1][2][4][5]

In previous chapter, force on robot body generated by torque on each joint was calculated by impedance matching index ω.

But it is impossible to determine the force acting direction with this index.

Next, we considered robot jumping motion which maximize the Z-axis component of force on robot body with impedance matching ellipsoid.

4.1 Impedance matching ellipsoid

Assume that each motor torque limit values are symmetrical,

$$-\tau_i^{limit} \leq \tau_i \leq \tau_i^{limit} \tag{9}$$

Let transformation matrix L and normalized torque denote as

$$L = diag(\tau_1^{limit}, \tau_2^{limit}, \cdots \tau_n^{limit}) \tag{10}$$

$$\tilde{\tau} = L^{-1}\tau \tag{11}$$

where $\tilde{\tau}$ is normalized torque. Therefore, when normalized torque with the magnitude of 1 is produced, force and momentum on the body are derived from

$$\tilde{\tau}^T \tilde{\tau} \leq 1 \tag{12}$$

i.e.,

$$(F_e - F_{bias})^T Q^T L^{-2} Q(F_e - F_{bias}) \leq 1 \tag{13}$$

Equation (14) shows the ellipsoid of transmitted force and momentum from joints to the body, and called Impedance matching ellipsoid.

4.2 Coordinate conversion from joint torque space to force space of robot body

We get next equation by substituting equation (7) to equation (4),

$$\tau = U \Sigma V^T (F_e - F_{bias}) \tag{14}$$

And from orthogonal matrix property ($U^{-1} = U^T$),

$$U^T \tau = \Sigma V^T (F_e - F_{bias}) \tag{15}$$

Equation (15) means that the rotated joint torque vector τ by U^T is transformed by Σ^1 to the rotated force-momentum vector acting on the robot body F_e by V^T.

Let coordinate converted joint torque vector and force-momentum vector by U^T and V^T denote as

$\tau' = \{\tau'_i\} = U^T \tau$ and $F_e' = \{F_e'{}_{,i}\} = V^T(F_e - F_{bias})$, respectively.

Using τ' and F_e', Equation (15) is expressed as followings.

$$\tau' = \Sigma F_e' \tag{16},$$

or, equivalently,

$$\tau'_i = \sigma_i F'_{e,i} \tag{17}$$

Now, transformation from equation (12) to equation (13) is

$$\tau^T L^{-2} \tau \leq 1 \tag{18}$$

$$\tau'^T U^T L^{-2} U \tau' \leq 1 \tag{19}$$

$$F_e'^T \Sigma^T U^T L^{-2} U \Sigma F_e' \leq 1 \quad (20)$$

$$(F_e - F_{bias})^T V \Sigma^T U^T L^{-2} U \Sigma V^T (F_e - F_{bias}) \leq 1 \quad (21)$$

Figure 5 shows these formula transformations.

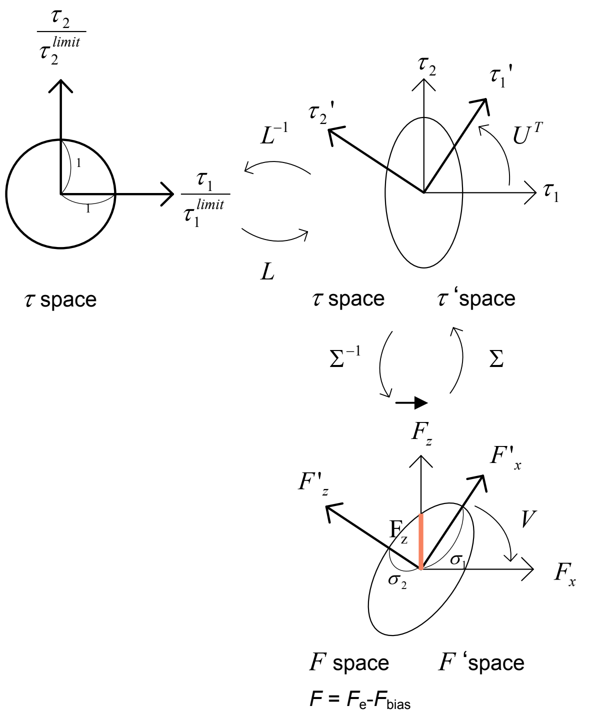


Figure 5: Coordinate conversion to Impedance matching ellipsoid.

4.3 Z-axis component of force on body

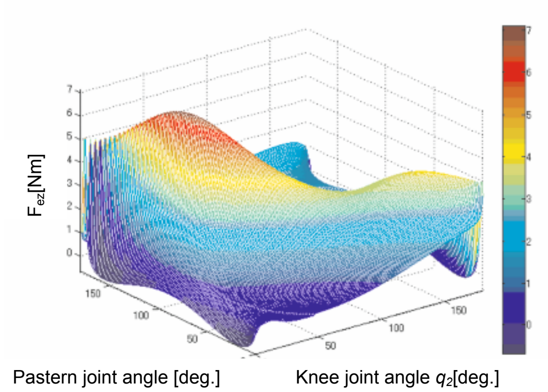
In Figure 5, Fez expresses Z-axis component of force on body.

We calculated the relation between robot pose and force on body using equation (21) by MATLAB. Pastern joint angle q1 was ranging 0~180 [deg.], and knee joint angle q2 was ranging 0~180 [deg.]. Figure 6 shows the results. In Figure 6, Fez becomes large at red part, and becomes small at blue part.

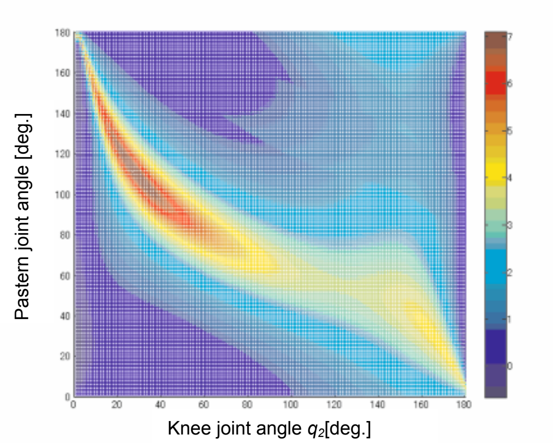
4.4 Jump motion design

It seems that jump height is affected by the Z-axis component of the generated force on the robot body between start and end of jump motion. So, we designed jump motion of the one-legged robot using Figure 6. At first, Jump motion was defined as followings.

- Jump motion is done by rotating each motor from jump start pose to jump end pose.
- Jump start pose : pastern joint angle q_1 : 40~90 [deg.]
- knee joint angle q_2 : 0〜60 [deg.]
- Jump end pose : pastern joint angle q_1 : 90 [deg.]
- knee joint angle q_2 : 0 [deg.]
- Angular velocity of each motors are able to set individually.



(a) Overhead view



(b) Contour map

Figure 6: Relation between robot pose and force on body.

Under these conditions, five jump motion patterns were designed for comparison. These five patters (see Table 2) are designed so that sum of displacement angle on each joint has same value.

Transmission efficiency from joint torque space to force space of robot body is affected by robot pose, and the value is getting large as next.

Jump1 > Jump2 > Jump3 > Jump4 > Jump5

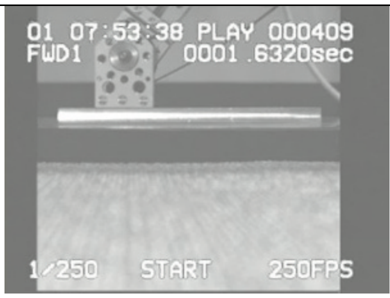
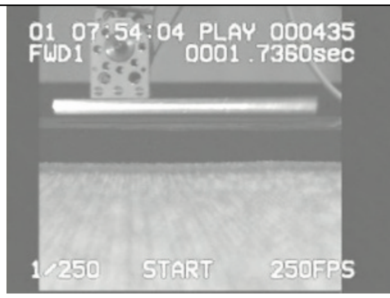
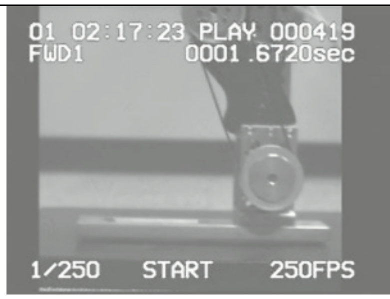
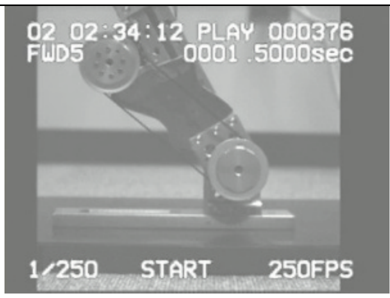
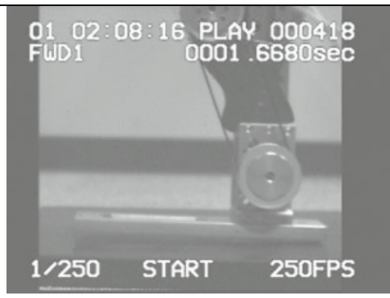
If we assume that jump height changes proportional to force on robot body, jump height will be higher as this order.

If transmission efficiency is not considered, jump3 and jump4 will be higher jump, because two motor are used in a balanced manner.

Table 2 : Designed jump motion patterns

	Jump start pose		Displacement to end pose		
	Pastern joint angle [deg.]	Knee joint angle [deg.]	Displacement angle of pastern joint [deg.]	Displacement angle of knee joint [deg.]	Sum of displacement angle [deg.]
Jump1	80	60	10	60	70
Jump 2	70	50	20	50	70
Jump 3	60	40	30	40	70
Jump 4	50	30	40	30	70
Jump 5	40	20	50	20	70

Table 3: Results of jump motion pattarens

	Jump height[mm]	Jump start pose	Pose at highest point
Jump1	5		
Jump2	6		
Jump3	4		
Jump4	3.6		
Jump5	×	Fail	Fail

5 EXPERIMENTS OF ONE-LEGGED ROBOT JUMPING

5.1 Experimental method

One-legged robot was programmed as table2, and tested the jumping motion. Jumping motion was captured by high-speed video camera (PHOTRON : FASTCAM-Eagle 500-1), and jumping height was measured from captured movie.

5.2 Results

Table 3 shows the experimental results. Jumping height was getting large as next.

Jump2 > Jump1 > Jump3 >Jump4 >Jump5(fail)

With the exception of Jump1, the results are similar to the preconception.

Jump3 is optimum pattern commonly considered, Jump1 and Jump2 are designed with impedance matching ellipsoid so that force transmission efficiency is large, and Jump4 and Jump5 are considered as low force transmission efficiency motion.

From these results, Jump1 height was lower than Jump2 height. Jump1 was prospected as higher than Jump2 because of generated force on body, but the jump height was midpoint of Jump2 and Jump3, in reality. The reason can be considered as followings. Robot jump motion is affected by feet botom shape, feet botom mass, etc. about vaulting condition of feet bottom. But they are not considered in this study. So these contributing factors affected on Jump1 motion.

And reason of Jump5 failure seems to be force transmission efficiency was too law and required torque was larger than the motor output torque. Figure 7 shows the relation between joint angle and required torque for joint driving. In Figure 7, black part means that joints can not support the robot body because motor torque is smaller than required torque to hold the joint angle.

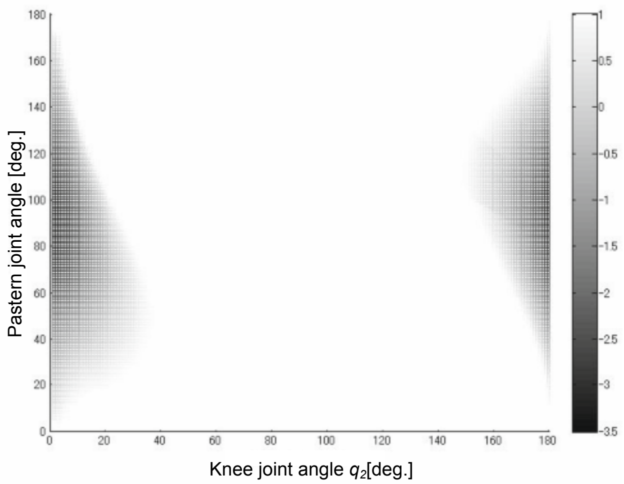


Figure 7: Joint angle and required torque for joint driving.

6 CONCLUSION

In this study, high transmission efficiency jump motion pattern for one-legged robot were designed with impedance matching ellipsoid. Experiments and simulations were done with experimental one-legged robot, and jump motion results with and without impedance matching design were compared.

As the results, height of jump motion with high transmission efficiency was 1.6~1.7 times higher than that of jump motion with low transmission efficiency, and 1.5 times higher than that of jump motion without consideration of impedance matching. Also impossible pose (load torque is larger than motor) were calculated by transmission efficiency.

With transmission efficiency and impedance matching, robot motion design will realize efficient power transmission, increasing the performance, and consuming lower amounts of power.

In future, we will design jump motion with vaulting condition of feet bottom to improve jump motion as actual human jumping.

7 REFERENCE

[1] Ryo Kurazume and Tsutomu Hasegawa, "New index for a robot manipulator based on the concept of impedance matching", proceedings of 9th Robotics Symposia, 1C2, (2004)

[2] Ryo Kurazume and Tsutomu Hasegawa, "Impedance Matching for Serial Link Manipulators", JRSJ Vol.23, No.2, pp.99-107

[3] Ryo Kurazume and Tsutomu Hasegawa, "Experimental Study on Impedance Matching for Serial Link Manipulators", proceedings of the 22nd Annual Conference of the Robotics Society of Japan 1H14, (2004).

[4] Ryo Kurazume and Tsutomu Hasegawa, "Impedance matching for free flying robots", proceedings of the 20nd Annual Conference of the Robotics Society of Japan, 3J16, (2002).

[5] Ryo Kurazume and Tsutomu Hasegawa, "Impedance matching for a serial link manipulator", Proc. IEEE Int. Conf. on Robotics and Automation, pp. 4802-4808, 2004.

[6] Hiroki Ohara, "Jump Operation Analysis of One-Foot Robot", master's thesis of graduate school of systems engineering, wakayama university (2006)

[7] F.Kanehiro,H.Hirukawa,S.Kajita,

Remote Control Support System for R/C Helicopter

Masafumi Miwa[1], Ittetsu Shiraishi[2], Makoto Matsushima[3], and Kiyoshi Minami[3]

[1] Institute of Technology and Science, The University of Tokushima, Tokushima, Japan

[2] Department of Mechanical Engineering, Faculty of Engineering, The University of Tokushima, Tokushima, Japan

[3] Arts Tech Labo Co., Ltd., Wakayama, Japan

Abstract

Recently, radio-controlled helicopters are widely used for crop-dusting, overhead wiring work, aerial photographing, etc. Especially, aerial photographing is used for many purposes, such as construction check, excavation, and disaster site observation. For these cases, skilled operators of radio-controlled helicopter are necessary, and hey have high-level operating technique. In this study, we present the remote control support system consists of microcomputers, various sensors. The aims of this system are to simplify the control operation and safety improvement of R/C helicopter maneuver. This study reports some experimental results on control support in practical situation.

Keywords:

Remote control support system; Attitude control system; Angular support mode

1 INTRODUCTION

Recently, model airplanes are widely used for industrial needs in Japan. For example, radio-controlled helicopters are used for crop-dusting, overhead wiring work, aerial photographing, etc. Especially, aerial photographing is used for many purposes, such as construction check, excavation, and disaster site observation. For these cases, skilled operators of radio-controlled helicopter are necessary. Some autonomous control type radio-controlled helicopters were reported [1]. But they consist of specialized airframe and special control, and they are so expensive that the local governments, companies, and etc. may not operate them. So, we present the remote control support system consists of hobby-class radio-controlled helicopter and microcomputers, various sensors [2]. The aims of this system are simplification of the control operation and safety improvement. This study reports some experimental results on attitude control in practical situation.

2 EXPERIMENTAL SETUP

2.1 2.1 Remote control support system

In this study, the remote control support system (RSCC) was constructed for attitude holding system (roll, pitch, yaw) to keep hovering maneuver. It holds horizontal level, nose direction, altitude, and position when the sticks are released. When the operator tilts the control sticks, tilting angles are treated as targeted inclination angle of the airframe.

According to blade element theory [3], Thrust $\boldsymbol{T}$ of helicopter is calculated by

$$\boldsymbol{T} = \boldsymbol{A}\theta \tag{1}$$

where, θ is the pitch angle of rotor blade, $\boldsymbol{A}$ is proportionality factor. This equation means that thrust T and torque momentums which tilts the airframe around the pitch, roll, and yaw axis are controlled by pitch angle of the rotor blade when rotating speed is constant. Thus, each rotational dynamic equation around 3 axes and dynamic equation of height are same form equations,

$$\frac{d^2 \boldsymbol{y}}{dt^2} = \alpha \boldsymbol{x} \tag{2}$$

where, y is rotational angle around any axis or height, x is a pitch angle of rotor blade or tail rotor, α is proportional constant. Also air drag is assumed to be zero.

From equation (2) and the transfer function of servomotor, PD control system was constructed for each axis and height control.

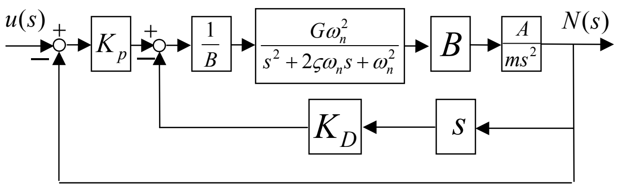


Figure 1 : Block diagram of height control system

Figure 1 shows the Block diagram of height control system for example, where, $\boldsymbol{K_P}$ is P gain, K_D is D gain, $\boldsymbol{m}$ is mass of the helicopter, $\boldsymbol{A}$ is conversion gain (pitch angle to Trust), and $\boldsymbol{B}$ is conversion gain (servo motor output angle to pitch angle). And

$$\frac{G\omega_n^2}{s^2 + 2\varsigma\omega_n s + \omega_n^2} \tag{3}$$

is transfer function of servomotor.

2.2 2.2 Experimental apparatus

We developed the remote control support system consists of microcomputers and various sensors. Figure 2 shows the schematic diagram of remote control support system. Hobby-

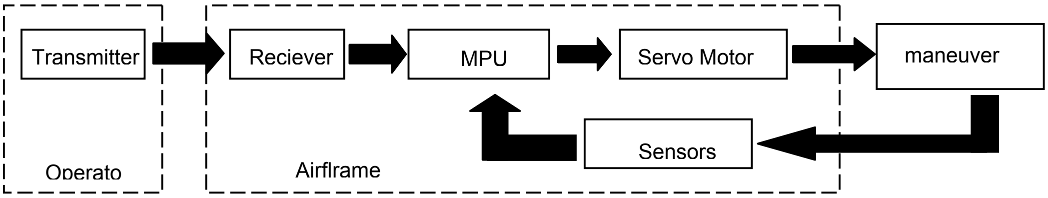


Figure 2: Remote control support system

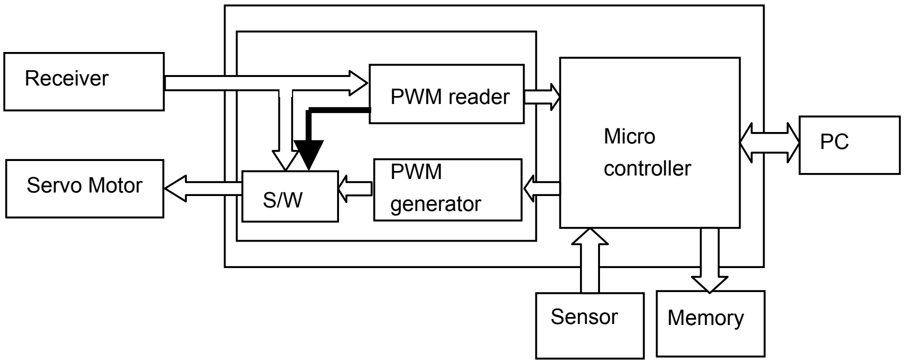


Figure 3: Block diagram of the system board

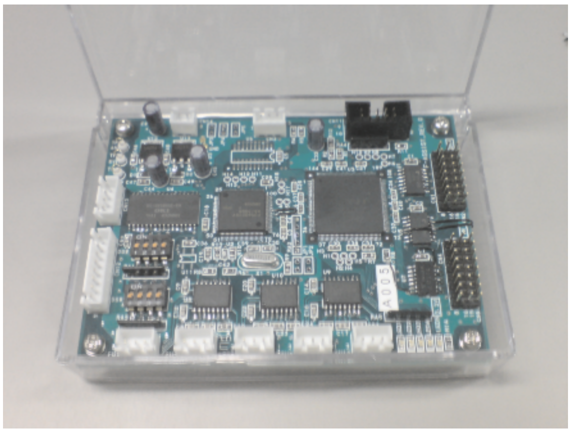
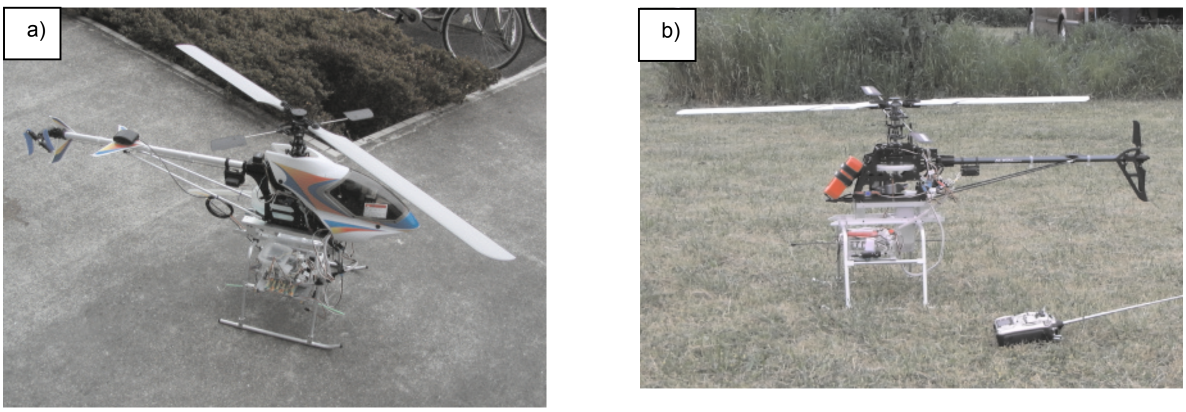


Figure 4: Control board (Flight Assist)



a) Engine type test helicopter (HPM system)

b) Motor type test helicopter (CCPM system)

Figure 5: Experimental RC helicopter

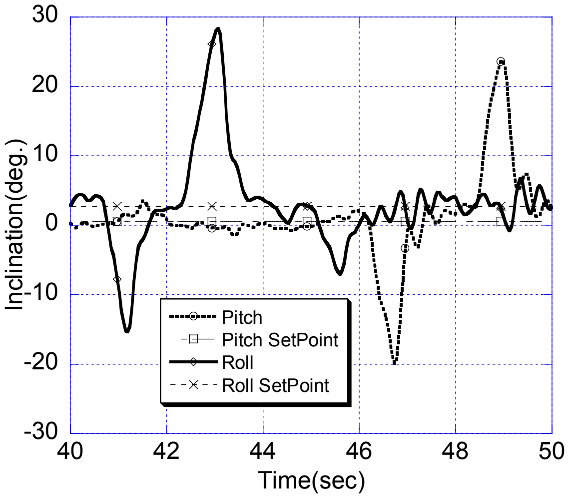


Figure 6: Results of Roll-Pitch control

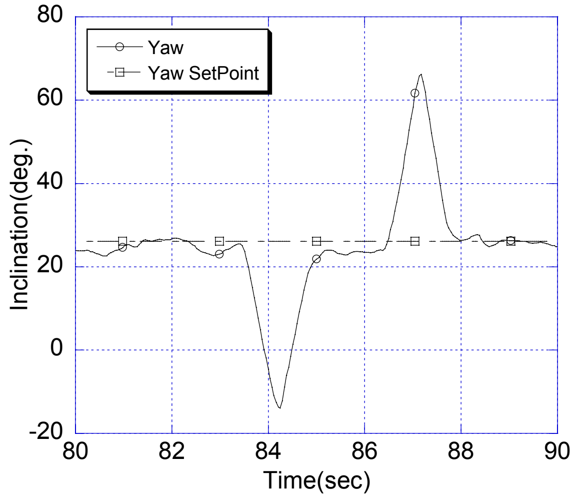


Figure 7 :Results of Yaw control

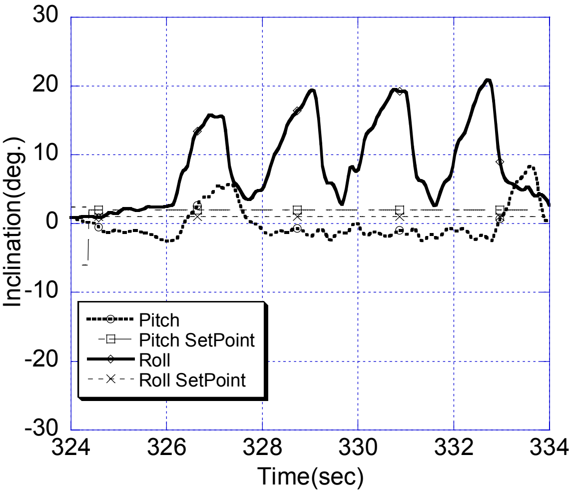


Figure 8 : Result of normal support mode

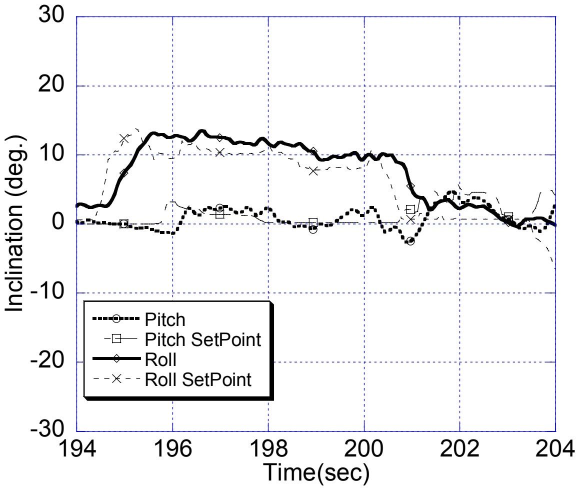


Figure 9 : Result of Angular support mode

class radio-controlled helicopters (HIROBO: Shuttle Sceadu Evolution 50(B)HPM:HPM system, Align: T-Rex600CF: CCPM system), radio system (Futaba: FF9H-T9CHP and 12FG), gyro sensor (Futaba: GY-401), governor (Futaba: GV-1) are used for test airframe. Experimental control board consists of cpu (Renesas: H8/3069R) and PWM (Pulse Width Modulation) signal processor programmed on CPLD IC (Altera: MAX2). The PWM signal processor reads PWM signals from radio control receiver, and it sends the values of PWM signals (pulse width) to the cpu. The cpu sends values of PWM signals to the PWM signal processor, and it outputs PWM signals to servomotors on the airframe. 9th channel signal of the receiver is assigned as switching signal which selects manual mode or support mode. Figure 3 shows the block diagram of the control board, and Figure 4 shows the control board. 3-axis angle sensor (Microstrain: 3DM-GX1) is used as inclinometer, laser distance meter (DIMETRIX: DLS-B15) is used as altimeter, and GPS (Garmin: GPS18-5Hz) is used as horizontal velocimeter. Also Bluetooth units (Inter Solution Marketing: Parani-SD100) were used for communication between PC and cpu to upload control program and to download experimental data. Figure 5 shows the experimental RC helicopters.

3 EXPERIMENTAL RESULTS

3.1 Attitude holding system

Helicopter moves with the component force of rotor trust by tilting its airframe. If a helicopter is forward titled, the trust component push it ahead, and the helicopter moves ahead horizontally while it is kept to be tilted. And if the helicopter is tilted leftward, it moves to left horizontally Attitude control is important for helicopter maneuver and safety hovering. So attitude holding system was developed to hold the helicopter airframe as level against disturbance.

To test the attitude holding system, tilt angle around each axis was added to the airframe by manual control as disturbance. Attitude holding system of the control board was activated, and PD control responses were recorded. P gain and D gain were previously decided by trial and error. Wind speed was 3 ～5m/s.

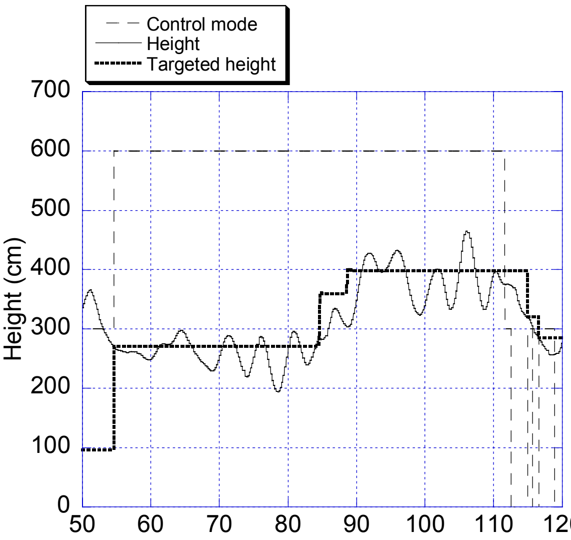


Figure 10: Results of height hold

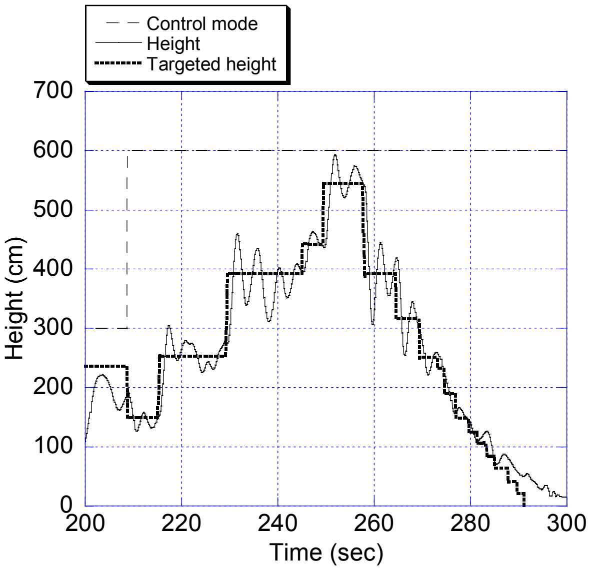


Figure 11: Results of height control

Figure 6 shows the results of Roll-Pitch axis control. In Figure 6, tilt angle around roll axis were added by manual control. Attitude holding system was activated when the control stick was released, and the airframe was returned to horizontal level. Also tilt angle around yaw axis tests were done. Figure 7 shows the results of Yaw axis control.

As the results, the attitude holding system can recover attitude of the airframe immediately from the influence of disturbance, and it can keep hovering maneuver.

3.2 Support mode

RCSS has two support modes. One is called normal support mode. In this mode, an operator can manuever in regular operation way, i.e., control stick inclination is directly send to the servomotors. And when the sticks are released, attitude holding system of RCSS is activated to hold horizontal level, current nose direction, and current altitude.

The other mode is angular support mode. In this mode, control stick Inclinations are always treated as targeted values of each angle or height. Attitude holding system of RCSS reads targeted tilt angles and height from control stick Inclinations that the operator tilts, and holds altitude, nose direction, and height. When the sticks are released, control stick Inclinations are neutral, and attitude holding system keeps horizontal level, current nose direction, and current altitude.

To evaluate these mode, horizontal migration test were done. Figure 8 and Figure 9 show the results of normal support mode and angular support mode, respectively. For normal support mode, operator tilted the control stick to rotate the airframe until the airframe titles to the desired angle, and the helicopter started to move. Then he released the control stick to stop rotating the airframe, and he expected that the helicopter would move continuously. But after that, the airframe returned to horizontal level by attitude holding system, because attitude holding system was activated by control sticks were released, and the helicopter stop. So the operator tilted the helicopter again. These operation and attitude holding system supports appear as wave shapes in Figure 8.

Figure 9 shows the results of angular support mode. For angular support mode, operator tilted the control stick until the airframe titles to the desired angle, and he holds the angle of control stick. Then Helicopter was titled and held at the target angle, and it moved while the operator held the control stick angle. Then attitude holding system held horizontal level, nose direction, and altitude when the sticks are released. These results show that angular support mode is better suited for RCSS.

3.3 Evaluation test

To evaluate RCSS, Evaluation test was done at Tokushima University and Wakayama University. Trial subjects were aerial photograph agency (two persons: all control) and inexperienced people (seven persons: with simple instruction; only aileron control stick using trainer mode). After controlling the experimental helicopter for about 5 minutes, their interviews were done.

After evaluation test, agency people said, “attitude holding is good, but moving velocity is small”. Tilting angle was controlled by attitude holding system, and it was limited within 5 degrees for beginners. So, they felt moving speed slow. Next, inexperienced people said that the concern between stick operation and airframe maneuver was difficult, but it was not so hard as they expected. Indeed, inexperienced people sometimes made overshoot movement, but they did not bring down the airframe. As the results of assessments, RCSS is effective for keep the horizontal level for not only beginners, but also experts. But speed up of moving velocity, safety training for beginners are required.

3.4 Height holding system

Height holding system was constructed with laser distance meter. Also PD control was used for height control. Figure 10 shows the results of height holding with target height changing. In Figure 10, red line shows the control mode. If “control mode” values are 600, it means that support control

Figure 12: Automatic hovering with velocity feedback control

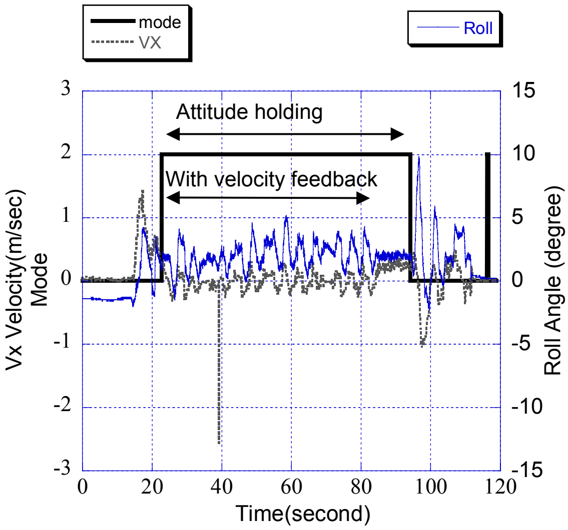


Figure 13: Roll angle and Vx while velocity feedback control

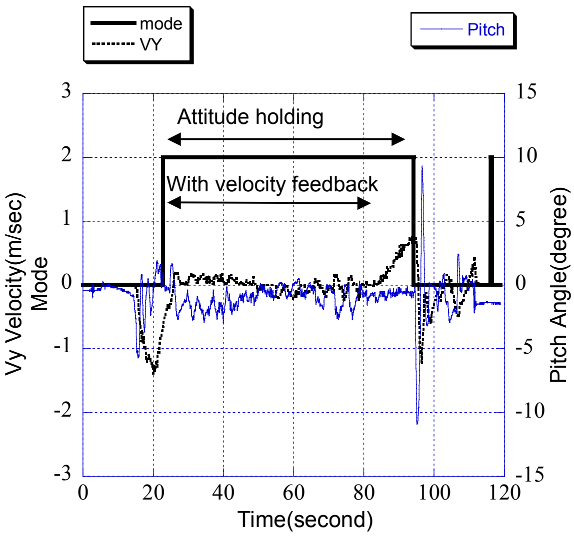


Figure 14: Pitch angle and Vy while velocity feedback control

is on, else manual control. As the results, height holding system kept the airframe within 1m errors.

Next, Figure 11 shows the results of automatic landing. Target height values were changed while height holding control. The test R/C helicopter succeeded automatic landing. In both experiments, maximum wind speed was 3 [m/s].

3.5 Position holding system

VTG command data of GPS unit was used as horizontal velocity for velocity feedback of position holding system. Feedback velocity parameters are treated as targeted inclination angle of the airframe that will cancel the velocity. Figure 12 shows the situation of position holding system. Horizontal migrations by wind disturbance were restrained. Also autonomous hovering maneuver and position holding were succeeded as shown in Figure 12.

Figure 13 and 14 show the variations of attitude angle of roll / pitch axis and airframe velocity components of x (lengthwise of the helicopter) / y (lateral direction of the helicopter) directions during autonomous hovering maneuver with / without velocity feedback. In both experiments, maximum wind speed was 3m/s, and wind speed component of y direction was larger than that of x direction.

In Figure 13, corrective steering by attitude holding system with velocity feedback was done, and its frequency was about 5 [Hz], also roll angle of the airframe was changing synchronously. As the result, x direction velocity of the airframe was suppressed within 0.3 [m/sec]. When velocity feed back was released and only attitude holding system was activated, target roll angle was 2 [degree] and the airframe was titled as about 2 [degree], and x direction velocity was constantly about 0.3 [m/sec].

In Figure 14, corrective steering frequency was about 5 [Hz], pitch angle of the airframe was changing synchronously, similarly. Y direction velocity of the airframe was almost constant, and suppressed within 0.2 [m/sec]. When velocity feed back was released and only attitude holding system was activated, target pitch angle was –0.5 [degree] and the airframe was titled as about –0.5 [degree], but y direction velocity increased to about 0.9 [m/sec].

These results show that attitude hold system with velocity feedback is effective to hold the airframe position.

4 CONCLUSION

In this study, we propose remote control support system (RCSS) for R/C helicopter. RCSS consists of attitude holding system, height holding system, and position holding system. It holds horizontal level, nose direction, altitude, and position automatically when the sticks are released. When the operator tilts the control sticks, tilting angles are treated as targeted inclination angle of the airframe. Attitude holding system was evaluated by hovering maneuver and evaluation test by trial subjects. As the results of assessments, RCSS is effective for keep the horizontal level. Also height holding system and position holding system were developed and tested, and attained good results. We will evaluate our system in trial aerial photographing in future.

5 ACKNOWLEDGMENTS

This study is partly supported by Yashima Dengyo Co. Ltd., and Aero service T.K.Factory. Also Visuatool Co.Ltd., lent us the laser distance meter.

We express an acknowledgment for them.

6 REFERENCES

[1] K.Hazawa , J.Shin , D.Fujiwara , K.Igarashi , D.Fernando , K.Nonami : Autonomous Flight Control of Unmanned Small Hobby-Class Helicopter Report2 : Modeling Based on Experimental Identification and Autonomous Flight Control Experiments, Journal of Robotics and Mechatronics Vol.15 No.5,2003 pp.546-554

[2] K. Nakata, M. Miwa, K. Tokuda : Remote Control System for R/C Helicopter, Proceedings of the 2006 JSME Conference on Robotics and Mechatronics （2006）2P2-C14

[3] K. Kato, Y. Imanaga : Introduction to helicopter, University of Tokyo Press (1985)

Finite-Time Settling Control restraining Amplitude of Control Input by Linear Programming

Junso Kimura[1]

[1] Dept. of Mechanical System Engineering, Faculty of Engineering, Fukuyama University, Fukuyama, Japan

Abstract

In this paper, a finite-time settling control restraining amplitude of control input response and disturbance response is proposed especially. In the proposed control, FIR parametrization is introduced and some control responses are analyzed. As a result, each response is described depending on affine transformation form. And a control system can be designed using optimization by linear programming.

Keywords:

Finite-Time Settling Control; Control Input; Disturbance; Linear Programming

1 INTRODUCTION

Being different from convergence property of continuous time control, in finite-time settling control of discrete time control, rapid response is regarded as important and controlled variable is made to completely track to reference signal in finite time [1]. In some recent researches, the response between sample points is considered. In finite-time settling control, excessive control input and response to vibrate intensely often occur and problems are caused as well known conventionally because fast response is realized.

Considering the square sum of control error of transient response or the maximum control error using l_1 norm, the finite-time settling control restraining the amplitude is researched [2], [3], [4]. Those efficiency is shown but the control system and design method are complicated.

In real control, for the purpose of considering constraint with plant or protecting plant, the control response amplitude such as control input or state is often limited. It is important to design a control system considering the control response amplitude consequently.

In this paper, a finite-time settling control restraining amplitude of control input response and disturbance response is proposed. In the proposed control, FIR parametrization is introduced and control responses are analyzed easily and in detail. And a method to restrain control input response and disturbance response is researched. The analysis result is used effectively, a control system can be designed by very simple linear programming.

2 PLANT AND FINITE-TIME SETTLING CONTROL

2.1 Plant and control system

A Plant described by (1) is considered.

$$y(k) = \frac{z^{-l}B(z^{-1})}{A(z^{-1})}(u(k)+d(k)) \tag{1}$$

$$A(z^{-1}) = 1+\sum_{i=1}^{n} a_i z^{-i} \qquad B(z^{-1}) = \sum_{i=0}^{m} b_i z^{-i}$$

Where, polynomials $A(z^{-1})$ and $B(z^{-1})$ are coprime each other, l is time delay, $y(k)$, $u(k)$, $d(k)$ are output, control input, disturbance respectively. It is supposed that the reference signal $r(k)$ is step signal and $r(k)=r$. For the plant (1), a next controller is used for control.

$$u(k) = \frac{S(z^{-1})}{C_d(z^{-1})R(z^{-1})}(r(k)-y(k)) \tag{2}$$

$R(z^{-1})$, $S(z^{-1})$ in this controller are as follows.

$$R(z^{-1}) = R'(z^{-1}) + z^{-l}B(z^{-1})Q(z^{-1}) \tag{3}$$

$$S(z^{-1}) = S'(z^{-1}) - C_d(z^{-1})A(z^{-1})Q(z^{-1}) \tag{4}$$

$$R'(z^{-1}) = 1+\sum_{i=1}^{m+l-1} r'_i z^{-i} \tag{5}$$

$$S'(z^{-1}) = \sum_{i=0}^{n+n_d-1} s'_i z^{-i} \tag{6}$$

$$Q(z^{-1}) = \sum_{i=0}^{n_q} q_i z^{-i} \tag{7}$$

$Q(z^{-1})$ is FIR parametrization. And the coefficient vector q of $Q(z^{-1})$ is defined according to the next (8).

$$q = \begin{bmatrix} q_0 & q_1 & \cdots & q_{n_q} \end{bmatrix}^T \tag{8}$$

A control system is designed to achieve some purpose using the degrees of freedom of this FIR parametrization $Q(z^{-1})$. $1/C_d(z^{-1})$ in (2) is a disturbance compensator. $C_d(z^{-1})$ is determined by (9) and then integral compensation is performed.

$$C_d(z^{-1}) = 1 - z^{-1} \tag{9}$$

The block diagram of above control system is shown in Fig.1.

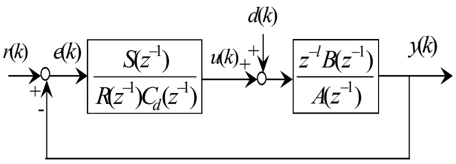


Figure 1: Block diagram of proposed finite-time settling control system.

By the calculation, closed loop transfer functions of this control system are the next (10) and (11).

$$y(k) = T_{yr}(z^{-1})r(k) + T_{yd}(z^{-1})d(k) \tag{10}$$

$$u(k) = T_{ur}(z^{-1})r(k) + T_{ud}(z^{-1})d(k) \tag{11}$$

$$T_{yr}(z^{-1}) = \frac{z^{-l}B(z^{-1})S(z^{-1})}{C_d(z^{-1})A(z^{-1})R'(z^{-1}) + z^{-l}B(z^{-1})S'(z^{-1})}$$

$$T_{yd}(z^{-1}) = \frac{z^{-l}C_d(z^{-1})R(z^{-1})B(z^{-1})}{C_d(z^{-1})A(z^{-1})R'(z^{-1}) + z^{-l}B(z^{-1})S'(z^{-1})}$$

$$T_{ur}(z^{-1}) = \frac{A(z^{-1})S(z^{-1})}{C_d(z^{-1})A(z^{-1})R'(z^{-1}) + z^{-l}B(z^{-1})S'(z^{-1})}$$

$$T_{ud}(z^{-1}) = -\frac{z^{-l}B(z^{-1})S(z^{-1})}{C_d(z^{-1})A(z^{-1})R'(z^{-1}) + z^{-l}B(z^{-1})S'(z^{-1})}$$

Where, it is necessary to be careful to the next condition.

$$\begin{aligned} &C_d(z^{-1})A(z^{-1})R(z^{-1}) + z^{-l}B(z^{-1})S(z^{-1}) \\ &= C_d(z^{-1})A(z^{-1})R'(z^{-1}) + z^{-l}B(z^{-1})S'(z^{-1}) \end{aligned} \tag{12}$$

2.2 Finite-time settling control

In the transfer functions of closed loop control system expressed by (10) and (11), $R'(z^{-1})$ and $S'(z^{-1})$ are decided satisfying the condition (13) in order to realize finite-time settling control.

$$C_d(z^{-1})A(z^{-1})R'(z^{-1}) + z^{-l}B(z^{-1})S'(z^{-1}) = 1 \tag{13}$$

All closed loop poles are assigned in the origin of z plane by this condition and finite-time settling control is achieved. In other words, the output response of closed loop control system for a reference signal is described by (14), and finite time settling is realized in the case of constant step reference signal.

$$y(k) = \frac{z^{-l}B(z^{-1})S(z^{-1})}{1} r(k) \tag{14}$$

And the control input response is described by (15) similarly, and finite time settling is also realized.

$$u(k) = \frac{A(z^{-1})S(z^{-1})}{1} r(k) \tag{15}$$

3 FINITE-TIME SETTLING CONTROL RESTRAINING CONTROL INPUT

In this chapter, a new method to easily restrain the amplitude of control response applying the well known linear programming is proposed especially.

3.1 Restraint of control input response

In this paper, for the purpose of restraining the amplitude of control input, the response of above control system is analyzed based on (15) and the desired control system is designed. (4) is substituted for (15) and the following (16) is got.

$$\begin{aligned} u(k) &= \frac{A(z^{-1})S(z^{-1})}{1} r(k) \\ &= \frac{A(z^{-1})\left\{S'(z^{-1}) - C_d(z^{-1})A(z^{-1})Q(z^{-1})\right\}}{1} r(k) \\ &= \frac{A(z^{-1})S'(z^{-1}) - C_d(z^{-1})A(z^{-1})A(z^{-1})Q(z^{-1})}{1} r(k) \end{aligned} \tag{16}$$

It is clear that the pulse transfer function in (16) becomes FIR polynomial. While noticing this fact, in the case of step reference signal, it is possible to calculate the control input response of this finite-time settling control system. By the above analysis, about the proposed finite-time settling control and its control system restraining control input, the next theorem is given.

[Theorem 1]

In the finite-time settling control system from (2) to (7) with FIR parametrization, the control system which restrains maximum amplitude by minimizing the upper bound of control input response can be designed using q, which can be found solving the linear programming problem described by next constraint condition (17) and performance index (18). Then, the value of minimized $u_{\max}$ is the upper bound of control input response.

$$\begin{bmatrix} -M_1M_3 & -\frac{1}{r} \\ & -\frac{1}{r} \\ & \vdots \\ & -\frac{1}{r} \\ M_1M_3 & -\frac{1}{r} \end{bmatrix} \begin{bmatrix} q \\ u_{\max} \end{bmatrix} \le \begin{bmatrix} -M_1M_2 \\ M_1M_2 \end{bmatrix} \tag{17}$$

$$J = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} q \\ u_{\max} \end{bmatrix} \to min \tag{18}$$

Where, M_1 , M_2 , M_3 in (17) are as follows.

$$M_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

$$M_2 = \begin{bmatrix} s'_0 \\ s'_1 + s'_0\, a_1 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$M_3 = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 2a_1 - 1 & 1 & \vdots & \cdots & \vdots \\ (a_1{}^2 + 2a_2) - 2a_1 & 2a_1 - 1 & 1 & \cdots & \vdots \\ \vdots & (a_1{}^2 + 2a_2) - 2a_1 & 2a_1 - 1 & \cdots & \vdots \\ \vdots & \vdots & \cdots & \cdots & a_n{}^2 - 2a_{n-1}a_n \end{bmatrix}$$

Proof

At first, time sequence of control input response for step reference signal is defined according to the next (19).

$$u = [u(0)\ u(1)\ \cdots\ u(2n + n_q + n_d)]^T \tag{19}$$

This is clarified in a form depending on FIR parametrization $Q(z^{-1})$, q . For analysis, the pulse transfer function described by FIR polynomial in (16) is noticed. In the above finite-time settling control system, using the coefficient vector q of FIR parametrization, it is possible that the time sequence (19) of control input response is described as follows.

$$u = \begin{bmatrix} u(0) \\ u(1) \\ \vdots \\ u(2n + n_q + n_d) \end{bmatrix} = M_1(M_2 - M_3 q)r \tag{20}$$

If (20) is examined, the time sequence of control input response is described by and depending on an affine transformation form of FIR parametrization. This notation is a method that author already have proposed [5],[6]. By using this technique, because not only control input response but also various control responses of control system can be easily evaluated by matrix calculation, it is very convenient.

Next, for the purpose of restraining the amplitude of control input response, the value of the response in each sample point is suppressed by a certain upper and lower bound.

$$|u(k)| \le u_{\max} \quad for\ \forall k \tag{21}$$

(21) is rewritten by the next inequality.

$$-u_{\max} \le u(k) \le u_{\max} \quad for\ \forall k \tag{22}$$

Furthermore, it is rewritten by the next inequalities.

$$u(k) \le u_{\max}\ ,\ \ u(k) \ge -u_{\max} \quad for\ \forall k \tag{23}$$

Where, when (23) is applied to (20), the following conditions are obtained.

$$M_1(M_2 - M_3 q)r \le \begin{bmatrix} u_{\max} \\ u_{\max} \\ \vdots \\ u_{\max} \end{bmatrix},\ \ M_1(M_2 - M_3 q)r \ge -\begin{bmatrix} u_{\max} \\ u_{\max} \\ \vdots \\ u_{\max} \end{bmatrix} \tag{24}$$

(24) is transformed as follows.

$$-M_1 M_3 q - \frac{1}{r}\begin{bmatrix} u_{\max} \\ u_{\max} \\ \vdots \\ u_{\max} \end{bmatrix} \le -M_1 M_2\ ,\ \ -M_1 M_3 q + \frac{1}{r}\begin{bmatrix} u_{\max} \\ u_{\max} \\ \vdots \\ u_{\max} \end{bmatrix} \ge -M_1 M_2$$

$$-M_1 M_3 q + \begin{bmatrix} -\frac{1}{r} \\ -\frac{1}{r} \\ \vdots \\ -\frac{1}{r} \end{bmatrix} u_{\max} \le -M_1 M_2\ ,\ \ M_1 M_3 q + \begin{bmatrix} -\frac{1}{r} \\ -\frac{1}{r} \\ \vdots \\ -\frac{1}{r} \end{bmatrix} u_{\max} \le M_1 M_2$$

$$\begin{bmatrix} & -\frac{1}{r} \\ -M_1 M_3 & -\frac{1}{r} \\ & \vdots \\ & -\frac{1}{r} \end{bmatrix} \begin{bmatrix} q \\ u_{\max} \end{bmatrix} \le -M_1 M_2\ ,\ \ \begin{bmatrix} & -\frac{1}{r} \\ M_1 M_3 & -\frac{1}{r} \\ & \vdots \\ & -\frac{1}{r} \end{bmatrix} \begin{bmatrix} q \\ u_{\max} \end{bmatrix} \le M_1 M_2 \tag{25}$$

Two inequalities in (25) are composed and previous constraint condition (17) is got. Here, suppose that the upper and lower bound $u_{\max}$ limiting the amplitude of response is not fixed, and it is regarded as a variable value same as the coefficient vector q of FIR parametrization. Furthermore, the coefficient vector q of FIR parametrization is adjusted to restrain the amplitude of control input response, and this $u_{\max}$ is reduced namely minimized. Consequently the performance index based on objective optimization becomes the above (18). Through such an analysis, in this paper, the design method of proposed finite-time settling control restraining the amplitude of control input response is regarded as a well known linear convex optimization problem, which minimize $u_{\max}$ by adjustment of q under constraint condition (17) and performance index (18) .

The proof ends.

The linear programming problem can be very easily solved with finite step using various techniques. In the proposed method, above affine transformation notation and optimization by linear programming are used. And a control system is designed reducing maximum amplitude of control input response. The design method is summarized in theorem 1.

3.2 Restraint of disturbance response

If only the amplitude of control input response is noticed too much, other control ability would deteriorate. In finite-time settling control, the disturbance robustness problem often occurs. Therefore a disturbance response is also described similarly by an affine transformation form in order to consider robust property for disturbance and the amplitude of disturbance response is restrained by similar technique.

Being based on the pulse transfer function $T_{yd}(z^{-1})$ in (10), FIR parametrization restraining the amplitude of response for step disturbance $d(k)=d$ can be obtained by a procedure same as the above control input. The method is summarized in the next theorem described as a linear programming problem.

[Theorem 2]

In the finite-time settling control system from (2) to (7) with FIR parametrization, the control system which restrains maximum amplitude by minimizing the upper bound of response for step disturbance can be designed using q, which can be found solving the linear programming problem described by next constraint condition (26) and performance index (27). Then, the value of minimized $y_{\max}$ is the upper bound of response for step disturbance.

$$\begin{bmatrix} -L_1L_3 & \begin{matrix} -\frac{1}{d} \\ -\frac{1}{d} \\ \vdots \end{matrix} \\ L_1L_3 & \begin{matrix} -\frac{1}{d} \\ -\frac{1}{d} \end{matrix} \end{bmatrix} \begin{bmatrix} q \\ y_{\max} \end{bmatrix} \le \begin{bmatrix} -L_1L_2 \\ L_1L_2 \end{bmatrix} \tag{26}$$

$$J = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} q \\ y_{\max} \end{bmatrix} \to min \tag{27}$$

Where, same as theorem 1, L_1, L_2, L_3 in (26) are matrices to describe the time sequence of response for step disturbance by an affine transformation. Elements in each matrix are omitted.

The proof is omitted

3.3 Control system design method

In the both case of control input response for step reference signal and response for step disturbance, each evaluation method considering maximum amplitude is introduced as above. But both evaluations cannot be optimized at the same time. Consequently weighting is introduced between two evaluations. The balance between two evaluations is changed using this weighting and optimization is adjusted.

[Theorem 3]

In the finite-time settling control system from (2) to (7) with FIR parametrization, the control system which restrains maximum amplitude of those responses by minimizing the linear combination of the upper bound of control input response and the upper bound of response for step disturbance can be designed using q, which can be found solving the linear programming problem described by next constraint condition (28) and performance index (29). Where, ρ in (29) is coefficient of weighting. Then, the value of minimized $u_{\max}$ is the upper bound of control input response and the value of minimized $y_{\max}$ is the upper bound of response for step disturbance.

$$\begin{bmatrix} -M_1M_3 & -\frac{1}{r} & 0 \\ & \vdots & \vdots \\ M_1M_3 & -\frac{1}{r} & 0 \\ -L_1L_3 & 0 & -\frac{1}{d} \\ & \vdots & \vdots \\ L_1L_3 & 0 & -\frac{1}{d} \end{bmatrix} \begin{bmatrix} q \\ u_{\max} \\ y_{\max} \end{bmatrix} \le \begin{bmatrix} -M_1M_2 \\ M_1M_2 \\ -L_1L_2 \\ L_1L_2 \end{bmatrix} \tag{28}$$

$$J = \begin{bmatrix} 0 & 1 & \rho \end{bmatrix} \begin{bmatrix} q \\ u_{\max} \\ y_{\max} \end{bmatrix} \to min \tag{29}$$

The proof is omitted

4 NUMERICAL EXAMPLE

Some numerical simulations are executed to confirm the efficiency of the proposed control.

A plant is supposed to be the next (30).

$$G(s) = \frac{1}{(1.5s+1)(4s+1)} \tag{30}$$

Step reference signal is supposed to be $r=1$. In the case of zero order holder, the next pulse transfer function is obtained by the discretization with sampling period of 1(sec).

$$G(z^{-1}) = \frac{0.0620 + 0.0457z^{-1}}{1 - 1.2922z^{-1} + 0.3998z^{-2}} z^{-1} \tag{31}$$

For the pulse transfer function of (31), proposed finite-time settling control system is designed based on theorem 3 in this paper. Then, weighting coefficient ρ in (29) is changed according to $\rho = 0.0$, $\rho = 10.0$, $\rho = 15.0$, $\rho = 20.0$. Here, design in case of $\rho = 0.0$ corresponds to design using theorem 1. Control simulations are executed by these designed finite-time settling control system.

Those results are shown in figures from Fig.2 to Fig.5. Both responses of output and control input are shown in each figure from Fig.2 to Fig.5. Some simulation results of response for step disturbance are shown in Fig.6. Dashed line stands for response in case of $\rho = 0.0$, solid line stands for response in case of $\rho = 15.0$, chain line stands for response in case of $\rho = 20.0$. The disturbance response in case of $\rho = 10.0$ is omitted because it closely is resembled with the disturbance response in case of $\rho = 0.0$. And, for comparison, control simulation of conventional finite-time settling control without FIR parametrization is executed, too. This is shown in Fig.7.

At first if the control result of proposed method is compared with that of finite-time settling control without FIR parametrization, the control input amplitude is restrained enough and finite time settling can be achieved as it is desired at first. But, due to the use of FIR parametrization, settling time is delayed. And the evaluation on disturbance response is added, the performance index considering restraint of both control input response and disturbance response is introduced, balance is adjusted changing the coefficient of weighting, as a result, good robust property for disturbance is obtained. Judging from these results of numerical simulations, the utility of proposed control method is clear.

5 CONCLUSION

In the finite-time settling control system, the control input response and disturbance response are calculated and evaluated by simple original method in linear. Being based on this analysis, the control method which restrains the amplitude of control input response and disturbance response is proposed and considered. And numerical simulations are executed, and the effectiveness of proposed method is clarified and confirmed. According to the technique proposed in this paper, it is possible to easily design the controller with FIR parametrization reducing the amplitude of control input and disturbance response.

6 REFERENCES

[1] Jury, E. I., 1958, Sampled-Data Control Systems

[2] Urikura, S., Nagata, A., 1987, Deadbeat Control of Discrete-Time Systems Considered the Transient Response (in Japanese), Tran. of the Society of Instrument and Control Engineers, 23/4:353-357.

[3] Sugie, T., Hata, Y., 1992, Design of Deadbeat Control Systems Considering the Maximum Error Bound (in Japanese), Tran. of the Society of Instrument and Control Engineers, 28/8:931-938.

[4] Urikura, S., 1992, Deadbeat Control in Consideration of the Continuous-Time Transient Response (in Japanese), Tran. of The Institute of System,Contorl and Information Engineers, 5/6:229-234.

[5] Kimura, J., Nishimura, Y., Yonezawa, Y., 1998, A Robust Finite Settling Time Control System by Simultaneous Design for Constrain of Control Input Amplitude, Proceedings of The 3rd International Conference on Advanced Mechatronics (ICAM'98), 843-848.

[6] Kimura, J., Saeki, M., Nishimura, Y., 1999, Robust Finite Settling Time Control System considering Constrains of Control Input and State Variable Amplitude, Proceedings of Pioneering International Symposium on Motion and Vibration Control in Mechatronics (P-MOVIC in Mechatronics'99), 290-295.

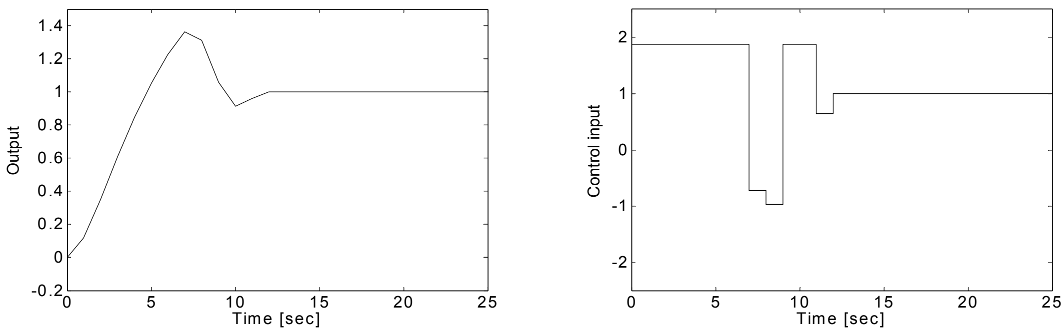


Figure 2: Simulation result ($\rho = 0.0$, $u_{\max} = 1.8724$, $y_{\max} = 174.3790$)

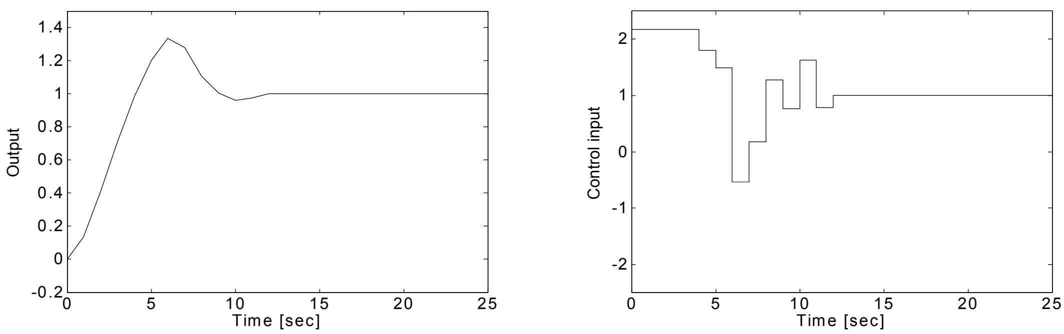


Figure 3: Simulation result ($\rho = 10.0$, $u_{\max} = 2.1785$, $y_{\max} = 0.3391$)

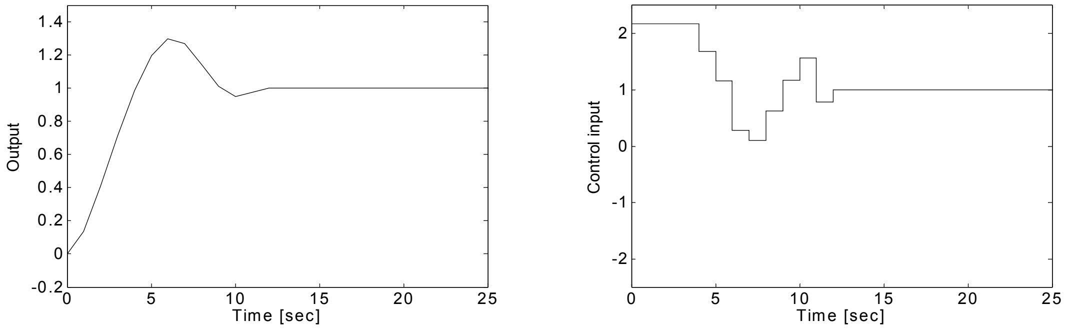


Figure 4: Simulation result ($\rho = 15.0$, $u_{\max} = 2.1785$, $y_{\max} = 0.3391$)

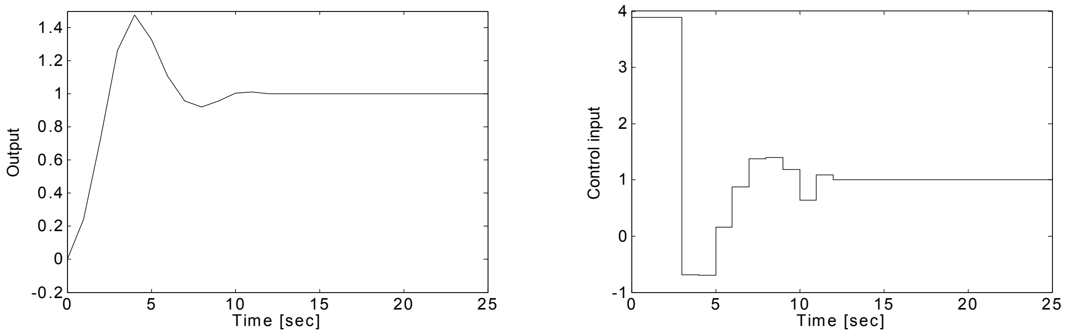


Figure 5: Simulation result ($\rho = 20.0$, $u_{\max} = 3.8831$, $y_{\max} = 0.2500$)

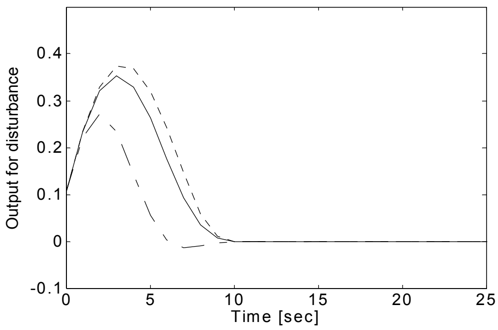


Figure 6: Response for step disturbance (dashed $\rho = 0.0$, solid $\rho = 15.0$, chain $\rho = 20.0$)

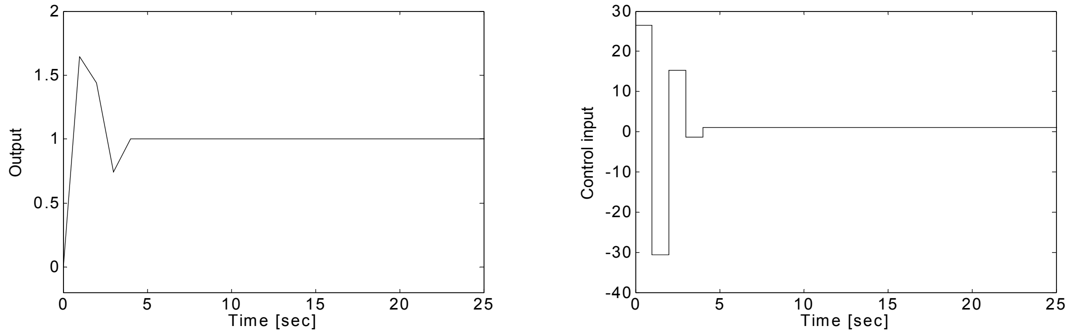


Figure 7: Simulation result without parametorization

Robust Control of Mechanical Systems with Nonlinear Friction Considering Compliance of Transmission Mechanism

Koji Funamoto, Naoki Uchiyama, Masaya Hattori, Shigenori Sano, Shoji Takagi
Dept. of Mechanical Engineering, Toyohashi University of Technology, Aichi, Japan

Abstract

Mechanical systems generally consist of three components; an actuator, load and power transmission mechanism such as a lead screw and gear coupling. Although there have been many researches on the controller design for mechanical systems with nonlinear friction, the compliance of the transmission mechanism has been little considered. Because the effect of nonlinear friction is significant when the transmission mechanism has low stiffness, its compliance should be considered in mechanical controller design for achieving the superior control performance. This paper presents a new approach to the controller design for mechanical systems with nonlinear friction considering the compliance of transmission mechanism. Because the effect of friction is generally unknown in advance, we design a robust controller based on the variable structure control that requires only the upper bound of friction magnitude. The effectiveness of proposed controller is demonstrated by comparative experimental results with a conventional state feedback controller.

Keywords:
Mechanical Control, Robust Control, Nonlinear Friction, Variable Structure Control

1 INTRODUCTION

To achieve accurate motion performance in mechanical systems, frictional effects should be considered in controller designs. Because the friction exhibits nonlinear and complicated properties such as Stribeck effect and hysteresis, many studies have been done to model the frictional characteristics and design sophisticated controllers thus far [1][2]. One of the difficulties in controlling the machines with friction is that the frictional property is generally unknown and depends on the operating conditions such as machine temperature.

Recent studies on friction compensation are categorized as follows: The first one is based on the analysis of frictional properties such as limit cycles induced by the controller. The analytical results are applied to the industrial PI or PID controllers [3][4]. The second one is to design the controller based on the estimation result of frictional parameters by preliminary experiments [5][6]. The disadvantage of these two approaches is that they are not effective for mechanical systems operated in varying environmental conditions.

To solve this problem, controller designs based on adaptive techniques are developed for real-time estimation of frictional effects [7]-[10]. These approaches require complicated estimation algorithms of frictional parameters, since frictional effects generally have nonlinear characteristics. Because the real-time estimation is not required in the robust control approach, it provides a simple structure of the controller. The variable structure control is known to be effective for nonlinear systems [11].

Mechanical systems generally consist of three components; an actuator, load and power transmission mechanism. For example, lead screw drive systems, which are used in many industrial machines such as machine tools and gantry robot systems, consist of the actuator, table and transmission mechanism such the lead screw and gear coupling as shown in Fig.1. Electric motor systems also generally consist of three components; a motor itself, load and gear coupling. For precision control, the compliance of transmission mechanism should be considered, because the frictional effect is increased if the transmission mechanism does not have enough stiffness. However, the compliance of transmission mechanism is little considered in most of the existing studies on the friction compensation except [10] and [11].

The controllers developed in [10] and [11] consider a sandwich dynamic system with friction, which corresponds to the system dealt with in this study, although only simulation results are shown to demonstrate the effectiveness. This paper presents a simpler design of the variable structure controller for the mechanical systems and demonstrates its effectiveness by experiments.

2 DYNAMICS OF MECHANICAL SYSTEMS WITH NONLINEAR FRICITON

Mechanical systems generally consist of three components; an actuator, load and power transmission mechanism. In this study, we consider a general model of mechanical systems as shown in Fig.2, where the transmission mechanism is modeled by a flexible torsion bar. Frictional torques exist in both the motor and the load disk sliding surfaces. The effect of friction at the load disk sliding surface becomes significant when the torsion bar has low stiffness.

The mechanical model in Fig. 2 has the following dynamics:

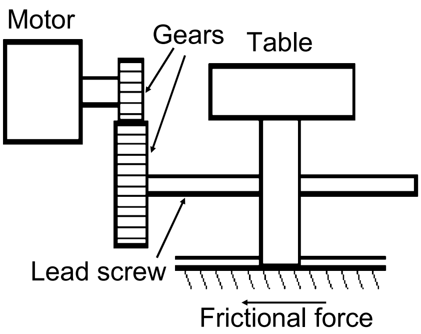


Figure 1: Example of mechanical system

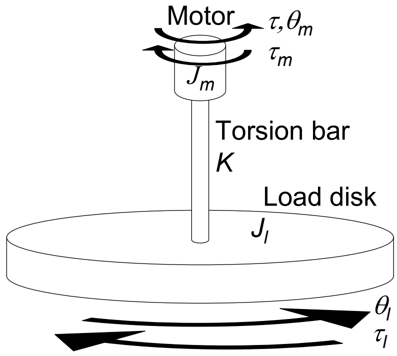


Figure 2: Model of mechanical system

$$J_m \ddot{\theta}_m + K(\theta_m - \theta_l) = \tau - \tau_m \,, \tag{1}$$

$$J_l \ddot{\theta}_l + K(\theta_l - \theta_m) = -\tau_l \,. \tag{2}$$

where θ_l is the load disk angle, θ_m is the motor angle, J_l is the load disk inertia, J_m is the motor inertia, K is the spring constant of torsion bar, τ_l is the frictional torque applied to the load disk, τ_m is the frictional torque applied to the motor, and τ is the control input torque. In Eqs.(1) and (2), τ_m can be compensated directly by the input torque τ, whereas it is not possible for τ_l. In other words, τ_l does not satisfy the matching condition. In addition, magnitudes of both τ_m and τ_l are generally unknown and variable. Hence, we propose a new design of the variable structure control system to compensate for effects of τ_m and τ_l by assuming that upper bounds of their magnitudes are known in advance.

3 CONTROLLER DESIGN

Tracking errors of the motor and the disk, e_m and e_l, are defined as follows:

$$e_m = \theta_{mr} - \theta_m \,, \tag{3}$$

$$e_l = \theta_{lr} - \theta_l \,, \tag{4}$$

where θ_{mr} and θ_{lr} are reference signals for the motor and the disk, respectivey. Substituting Eq.(3) into (2), we have

$$J_l \ddot{\theta}_l + K(\theta_l + e_m - \theta_{mr}) = -\tau_l \,. \tag{5}$$

We consider the following reference signal for the motor and the control input torque.

$$\theta_{mr} = \frac{J_l}{K}\ddot{\theta}_{lr} + \theta_{lr} + \frac{w_l}{K} \,, \tag{6}$$

$$\tau = J_m \ddot{\theta}_{mr} + K(\theta_{mr} - \theta_{lr}) + u + w_m \,, \tag{7}$$

where w_m and w_l are additional inputs for compensating for nonlinear frictions of the motor and disk, respectively, the symbol u is employed for the pole-placement control. Applying Eqs.(6) and (7) to Eqs.(1) and (5), we have

$$J_m \ddot{e}_m + K(e_m - e_l) = \tau_m - u - w_m \,, \tag{8}$$

$$J_l \ddot{e}_l + K(e_l - e_m) = \tau_l - w_l \,. \tag{9}$$

It should be noted that the nonlinear friction τ_l can be directly controlled by the input w_l in Eq.(9). Eqs.(8) and (9) can be rewritten as the following state-space equation:

$$\dot{e} = Ae - bu + d \,, \tag{10}$$

where

$$e = \begin{bmatrix} e_m \\ \dot{e}_m \\ e_l \\ \dot{e}_l \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -\dfrac{K}{J_m} & 0 & \dfrac{K}{J_m} & 0 \\ 0 & 0 & 0 & 1 \\ \dfrac{K}{J_l} & 0 & -\dfrac{K}{J_l} & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad d = \begin{bmatrix} 0 \\ \tau_m - w_m \\ 0 \\ \tau_l - w_l \end{bmatrix}$$

To design control inputs u, w_m and w_l, we consider the following Lyapunov function candidate:

$$V = e^T P e \,, \tag{11}$$

where P is the positive definite matrix. Differentiating Eq.(11) with respect to time and substituting Eq.(10) into it, we have

$$\dot{V} = e^T P\left(Ae - bu + d\right) + \left(e^T A^T - b^T u + d^T\right)Pe \,. \tag{12}$$

We consider the following state feedback control:

$$u = f^T e \,, \tag{13}$$

where f is the feedback gain vector. Applying Eq.(13) to (12), we have

$$\dot{V} = e^T \left\{P\left(A - bf^T\right) + \left(A - bf^T\right)^T P\right\} e + 2e^T P d \,. \tag{14}$$

It is known that there exist positive definite matrices P and Q satisfying the following equation if $(A\text{-}bf^T)$ is a stable matrix:

$$P\left(A - bf^T\right) + \left(A - bf^T\right)P = -Q \,. \tag{15}$$

where Q is positive definite. Substituting Eq.(15) into (14) leads to

$$\dot{V} = -e^T Q e + 2e^T P d \,. \tag{16}$$

A new vector e_p is defined as follows:

$$e_p^T = e^T P = \begin{bmatrix} e_{p1} & e_{p2} & e_{p3} & e_{p4} \end{bmatrix}. \tag{17}$$

Using some elements of this vector, we propose to design friction compensation inputs w_m and w_l as follows:

$$w_m = \mathrm{sgn}(e_{p2}) D_m \,, \tag{18}$$

$$w_l = \mathrm{sgn}(e_{p4}) D_l \,, \tag{19}$$

where $\mathrm{sgn}(x)$ denotes the sign function of x, and D_m and D_l are constants that satisfy the following inequalities.

$$\begin{aligned} \max\left(\left|\tau_m\right|\right) &\le D_m, \\ \max\left(\left|\tau_l\right|\right) &\le D_l. \end{aligned} \tag{20}$$

The symbol of $\max(x)$ denotes a maximum value of x.

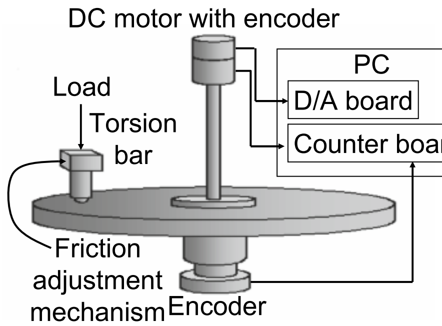


Figure 3: Experimental system

The resulting controller is given as follows:

$$\theta_{mr} = \frac{J_l}{K}\ddot{\theta}_{lr} + \theta_{lr} + \frac{\mathrm{sgn}(e_{p4})D_l}{K}, \tag{21}$$

$$\tau = J_m\ddot{\theta}_{mr} + K(\theta_{mr} - \theta_{lr}) + f^T e + \mathrm{sgn}(e_{p2})D_m . \tag{22}$$

This controller leads to the following relation from Eq.(14):

$$\begin{aligned}\dot{V} &= -e^T Qe + 2e_{p2}\{\tau_m - \mathrm{sgn}(e_{p2})D_m\} \\ &\quad + 2e_{p4}\{\tau_l - \mathrm{sgn}(e_{p4})D_l\}\end{aligned} \tag{23}$$

Because $\dot{V}$ is negative definite, the asymptotic stability is achieved. To avoid the chattering phenomena, which may give some damages to mechanical systems, we use the following function instead of $\mathrm{sgn}(e_{pi})$, (i=2,4):

$$f_s(e_{pi}) = \frac{e_{pi}}{|e_{pi}| + \varepsilon}, \tag{24}$$

where ε is a small positive constant.

Because the vector e_p includes w_l, the value of w_l can not be easily obtained. Thus, we define a new error vector that uses a past error vector obtained at the one sample past instant e_{past} as follows:

$$e_p^{*T} = e_{past}^T P = \begin{bmatrix} e_{p1}^* & e_{p2}^* & e_{p3}^* & e_{p4}^* \end{bmatrix}. \tag{25}$$

Using some elements of this vector, we generate the inputs w_l and w_m as follows instead of Eqs.(18) and (19):

$$w_m = f_s(e_{p2})D_m , \tag{26}$$

$$w_l = f_s(e_{p4}^*)D_l . \tag{27}$$

The resulting controller is given as follows:

$$\theta_{rm} = \frac{J_l}{K}\ddot{\theta}_{lr} + \theta_{lr} + \frac{f_s(e_{p4}^*)D_l}{K}, \tag{28}$$

$$\tau = J_m\ddot{\theta}_{mr} + K(\theta_{mr} - \theta_{lr}) + f^T e + f_s(e_{p2})D_m . \tag{29}$$

4 EXPERIMENTS

4.1 Experimental system

The experimental system is shown in Fig.3. The experimental machine consists of the motor, torsion bar, load disk and friction adjustment mechanism. A DC servo motor with a harmonic drive gear is employed for the actuator. The motor power and gear ratio are 20[W] and 1/50, respectively. The inertia of geared motor at the output shaft is J_m =0.021[kgm^2].

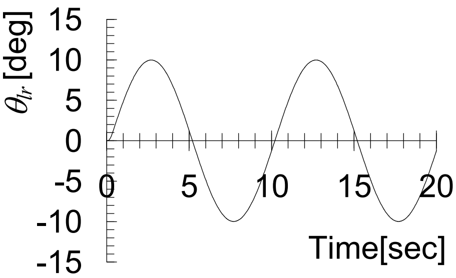


Figure 4: Reference signal of load disk

Torque control system of the motor is installed using a power amplifier electric circuit.

A piano steel is employed for the material of the torsion bar, whose spring constant is K = 1.2[Nm/rad]. The inertia of load disk is J_l = 0.0185[kgm^2]. The angles of motor and load disk are measured by rotary encoders whose resolutions are 1.8×10^{-3} [deg] and 2.2×10^{-4} [deg], respectively. Their velocities are numerically calculated by the backward difference operation.

4.2 Experimental results

Comparative experiments are performed with the conventional state feedback controller (SFC) which is obtained by setting both w_m and w_l in Eqs.(6) and (7) to zero.

The reference signal of load disk is given as follows:

$$\theta_{lr} = F(s)\times a\sin(\omega t) , \tag{30}$$

$$F(s) = \frac{1}{(Ts+1)^4}, \tag{31}$$

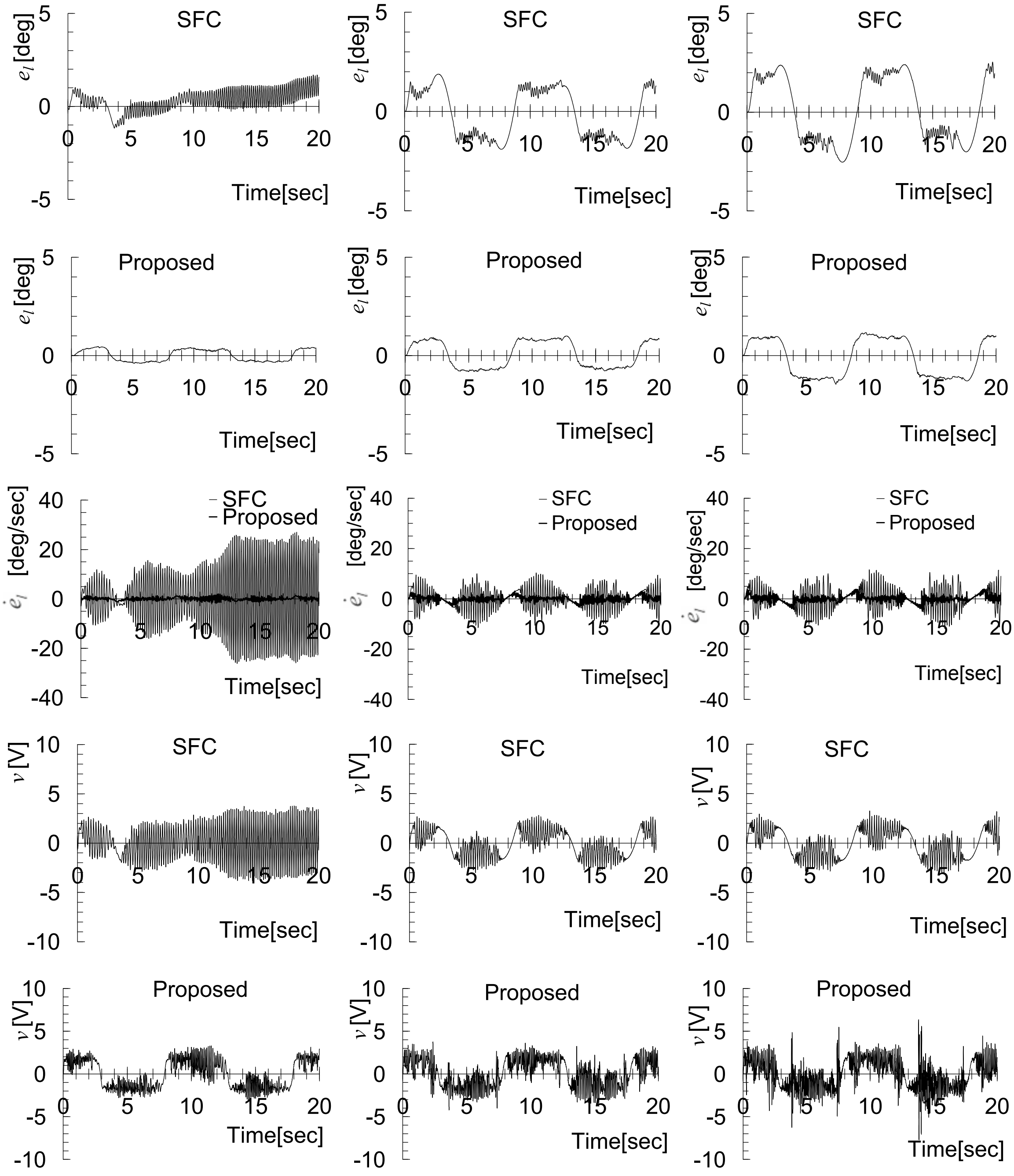
where a and ω are the amplitude and angular frequency of the desired sinusoidal curve, respectively. They are set as a=10[deg] and $\omega = 2\pi/10$[rad/sec]. The symbol s denotes the variable of the Laplace transformation, and T is the time constant of the low-pass filter $F(s)$ (T = 0.05[sec/rad]). The filter $F(s)$ is included for avoiding that the initial values of derivatives of θ_{lr} become infinity. Figure 4 shows the reference signal in Eqs.(30) and (31). The feedback gain vector f in Eq.(13) is given as follows:

$$f^T = [68.3 \quad 2.40 \quad 5.65 \quad 12.4] , \tag{32}$$

which places the desired control system poles to -20 [1/sec] as the repeated poles. The matrix Q in Eq.(15) is set to the identity matrix. The parameters in Eqs.(20) and (24) are set as ε=1.0, D_m=0.771[Nm] and D_l=0.32[Nm], respectively.

The load applied to the friction adjustment mechanism is varied as m=0 (Exp.1), 0.55 (Exp.2) and 0.9 (Exp.3) [kg]. Frictional torque τ_l exists even in the first case (Exp.1).

Experimental results are shown in Fig.5, where v is the control input voltage that is proportional to the input torque τ (The proportional gain is 0.70[Nm/V]). When the frictional torque applied to the load disk τ_l is small (Exp.1), oscillating motion of the load disk occurs especially in the last half period in the SFC result. Because the high-gain feedback controller is required for precision control in the conventional method, the control system is easily destabilized by disturbances. On the other hand, it is confirmed that the



(a) Exp.1 (m=0[kg]) (b) Exp.2 (m=0.55[kg]) (c) Exp.3 (m=0.9[kg])

Figure 5: Expreimental results

proposed controller can reduce the disturbance effect. When the magnitude of τ_l is increased (Exp.2), the oscillating motion in the SFC result is reduced. This is because the friction often increases the system stability (i.e., energy dispassion). However, tracking error magnitude $|e_l|$ increases especially when the load disk changes the rotational direction. It also can be seen from the graph of $\dot{e}_l$ that the oscillating motion is still significant compared to the proposed method. The proposed controller provides smaller magnitude of tracking error. When the magnitude of τ_l is further increased (Exp.3), larger tracking error is seen in the SFC result, whereas the error is less increased in the proposed control result.

The average and maximum of $|e_l|$, and the standard deviation of e_l are summarized in Figs. 6-8. The advantage of the proposed method is clearly seen.

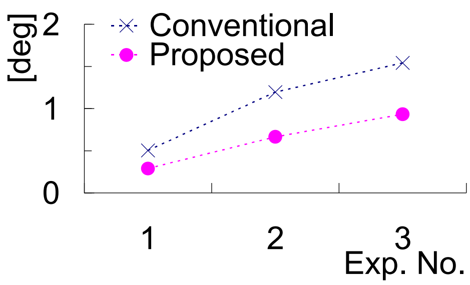


Figure 6: Average of $|e_l|$

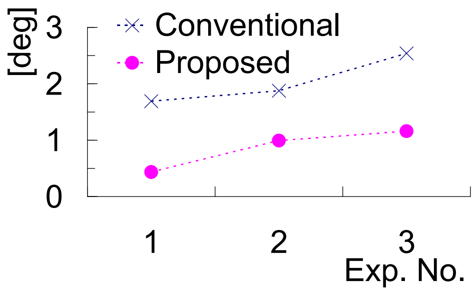


Figure 7: Maximum of $|e_l|$

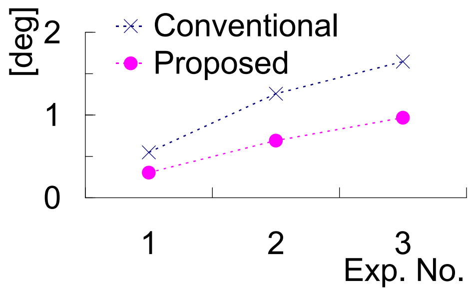


Figure 8: Standard deviation of e_l

5 CONCLUSIONS

This paper presents a new robust controller design for mechanical systems with nonlinear friction. A general model of mechanical systems, which consist of an actuator, load and power transmission mechanism, is given, and the compliance of the transmission mechanism that increases the frictional effect is considered. Because the friction occurred at the load sliding surfaces generally dose not satisfy the matching condition, we have proposed a new method to employ the variable structure control, which can achieve the robust performance to the unknown frictional effect. The proposed design has a simpler structure compared to the existing method by transforming the system dynamics appropriately. Comparative experiments have been conducted to verify the effectiveness of the proposed method. Controller gains are required to be increased for improving the control performance in the conventional feedback controllers, although high gain controller reduces the system stability and often brings the oscillating motion. The proposed robust controller has not only exhibited the better control performance (i.e., smaller magnitude of tracking error) but also reduced the oscillating motion compared to the conventional approach.

6 ACKNOWLEDGMENTS

The authors would like to sincerely thank the Naito Science & Engineering Foundation, Japan for financial support.

7 REFERENCES

[1] Armstrong-H´elouvry, B., Dupont, P., Canudas de Wit, C., 1994, A Survey of Models, Analysis Tools and Compensation Methods for the Control of Machines with Friction, Automatica, 30-7:1083-1138.

[2] Olsson, H., ˚Astr¨om, K. J., Canudas de Wit, C., G¨afvert, M., Lischinsky, P., 1998, Friction Models and Friction Compensation, European Journal of Control, 4:176-195.

[3] Putra, D., Nijmeijer, H., van de Wouw, N., 2007, Analysis of Undercompensation and Overcompensation of Friction in 1DOF Mechanical Systems, Automatica, 43:1387-1394.

[4] M´arton, L., 2008, On Analysis of Limit Cycles in Positioning Systems near Striebeck Velocities, Mechatronics, 18:46-52.

[5] Ro, P. I., Shim, W., Jeong, S., 2000, Robust Friction Compensation for Submicrometer Positioning and Tracking for a Ball-Screw-Driven Slide System, Precision Engineering, 24:160-173.

[6] Tjahjowidodo, T., Al-Bender, F., Van Brussel, V. H., Symens, W., 2007, Friction Characterization and Compensation in Electro-Mechanical Systems, Journal of Sound and Vibration, 308:632-646.

[7] Garagi´c, D., Srinivasan, K., 2004, Adaptive Friction Compensation for Precision Machine Tool Drive, Control Engineering Practice, 12:1451-1464.

[8] Choi, J. J., Han, S. I., Kim, J. S., 2006, Development of a Novel Dynamic Friction Model and Precise Tracking Control using Adaptive Back-Stepping Sliding Mode Controller, Mechatronics, 16:97-104.

[9] Nouri, K., Dhaouadi, R., Braiek, N. B., 2008, Adaptive Control of a Nonlinear DC Motor Drive Using Recurrent Neural Networks, Applied Soft Computing 8:371-382.

[10] Taware, A., Tao, G., Pradhan, N., Teolis, C., 2003, Friction Compensation for a Sandwich Dynamic System, Automatica, 39:481-488.

[11] Kwatny H. G., Teolis C., Mattice, M., 2002, Variable Structure Control of Systems with Uncertain Nonlinear Friction, Automatica, 38:1251-1256.

Size Reduction and Performance Improvement of Automatic Discharge Gap Controller for Curved Hole Electrical Discharge Machining

Masahiko Kita[1], Tohru Ishida[1], Koji Teramoto[2], Yoshimi Takeuchi[1]

[1] Dept. of Mechanical Engineering, Graduate School of Engineering, Osaka University, Osaka, Japan

[2] Dept. of Mechanical Systems Engineering, Faculty of Engineering, Muroran Institute of Technology, Muroran, Japan

Abstract

The study deals with size reduction and performance improvement of automatic discharge gap controller (ADGC) for a microrobot which can fabricate a long curved hole by electrical discharge machining (EDM). It is strongly required that curved hole machining method is developed so that the pipelines can be made which have arbitrary shape and are located in arbitrary position. However, such pipelines cannot be fabricated since holes are generally formed by drilling. To solve the problem, the authors have proposed the curved hole machining method by means of a microrobot with an EDM function. The microrobot has to have the function of performing stable EDM in the limit space such as the bottom of a long curved hole, in other words, the function of always keeping the equipped electrode in the appropriate position automatically and autonomously so as to maintain stable EDM. To realize the function, ADGC has been devised. Actually, the prototype of ADGC was developed and it had been proved that the prototype had the function. However, the prototype was too large to install on a microrobot and did not achieve the full performance designed. Therefore, in the study, the improvements of ADGC are done so that its size is reduced and its full performance is achieved. From the experimental results, it is found that the improved ADGC, whose size is reduced so that the installation to a microrobot is enabled, realizes the designed full performance.

Keywords:

Electrical Discharge Machining; Discharge Gap; Shape Memory Alloy

1 INTRODUCTION

A hole is the shape employed to achieve a variety of aims in mechanical engineering. For example, a hole makes a object pass, makes a object cool or warm, makes a object fixed or positioned, makes a object become light weight, makes a object be measured, etc. From such actual facts, it is known that a hole is a very fundamental shape used very frequently and plays very important roles. However, a hole is usually machined by a linear tool as represented by a twist drill. Consequently, the shape of machinable holes inevitably becomes straight line or polygonal line, that is, the realizable hole shape is limited. This concept has already been the common sense in the field of machining.

Despite the common sense, the limitation of the realizable hole shape can cause a problem in making holes achieve their functions to the maximum. Typical example of the problem is brought about in the design and fabrication of water channels, pipelines built in molds. Regulating flow rate and temperature of coolant flowing through water channels enables the thermal controls of molds, namely, control temperature distribution in mold surfaces and heat flow in molding, which leads to obtaining products without defects occurring in the several stages of molding process. Accordingly, the shape and position of water channels in molds are a very important factor in mold design. Although computer simulation gets to find the optimal shape and position of water channels, it is usual that their optimal shape is curved and that their optimal position is located in severe areas in molds [1]. Consequently, it is impossible to freely control the shape and position of water channels since drilling is the sole practical method to form water channels. That is to say, it is completely impossible to realize the optimal shape and position of water channels by conventional drilling methods. As a result, mold designers have compromised on polygonal-line-shaped water channels located in the area where twist drills can reach in spite of having found the optimal shape and position of water channels.

To solve the problems occurring in fabricating pipelines as represented by water channels, it is strongly required to establish a curved hole machining method and to put the method in practical use. To meet the requirement, some methods for machining a curved hole have been developed and can actually fabricate curved holes [2-5]. The respective methods have developed the original devices for machining curved holes and the devices have following common structure and function. The common structure is that a certain tool for machining is located in the bottom of a curved hole in the middle of fabricating and is connected to a certain control equipment and power supply, which are placed in the outside of the workpiece, through a certain mechanism and cable. The common function is that certain energy for machining is provided from the power supply to the tool through the cable and that the control of the position and posture of the tool, determining the moving locus of the tool, i.e., the shape of the machined curved hole, is done by transmitting a certain displacement generated by the control equipment to the tool through the mechanism. Accordingly, these curve hole machining methods have common problem in which the accurate transmission of the displacement for controlling the position and posture of the electrode becomes more difficult as the length of the machining curved hole becomes longer. This is because the mechanism between the tool and the control equipment is needed whose length is same as the

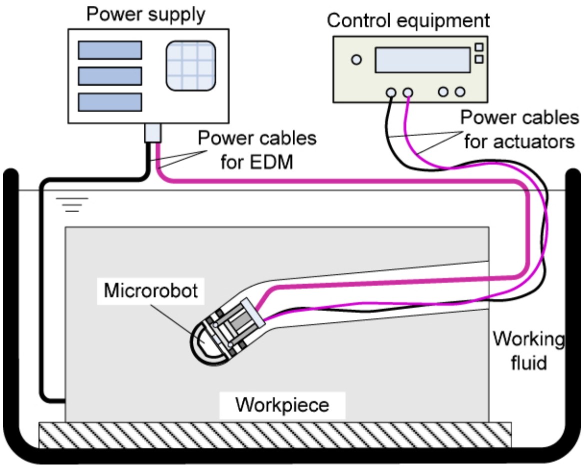


Figure 1: Conceptual view of curved hole machining by means of a microrobot with EDM function.

machining curved hole length, causing the stiffness of the mechanism to get lower.

While the authors have devised the curved hole machining methods of the same style as the above-mentioned methods [6, 7], they have also proposed a new method for the purpose of overcoming the fatal weakness of the methods. The new one is the method in which the control of the position and posture of the electrode does not depend on the displacement generated by the control equipment, in other words, in which the mechanism transmitting the displacement does not be needed.

Figure 1 illustrates a conceptual view of the new method which employs a microrobot with EDM function [8]. The microrobot has the ability of moving along an arbitrary trajectory in a workpiece simultaneously with performing EDM. This results in fabrication of long and complicated curved holes which is identical with the moving locus of the microrobot. To realize the method, the authors plan to respectively develop three units constituting the microrobot, namely, autonomous EDM unit, machining direction determining unit and in-pipe self-mobile unit. The paper deals with the autonomous EDM unit.

The autonomous EDM unit is required to automatically and autonomously control discharge gap so that stable EDM is performed in the almost isolated space such as the bottom of a curved hole. To embody the unit, ADGC has been devised and its prototype and tester have been developed. From the results of the experiments using them, the discharge gap control ability of ADGC has been confirmed [8] and proven quantitatively [9]. However, the prototype of ADGC had the problems that its size was so large that a microrobot could not be equipped with the prototype and that its full performance expected in the design stage could not be achieved. To solve the problems of the prototype, in the study, a new ADGC has been developed by extensively and drastically improving the design of its components. This yields so significant reduction of the size of the improved ADGC, compared with the prototypical ADGC, that the improved ADGC can be installed on a microrobot. From the results of the experiments employing the improved ADGC, it is proven that the improved ADGC achieves its full performance designed.

2 AUTOMATIC DISCHARGE GAP CONTROLLER

Figure 2 illustrates a basic structure of ADGC. ADGC consists of a bidirectional actuator by use of shape memory alloy (SMA) and an electrode and power supply for EDM. The bidirectional actuator has a shaft, on which the electrode is mounted. The power supply is connected to the bidirectional actuator and a workpiece. Discharge current occurring between the electrode and the workpiece in performing EDM is designed to pass through the SMA in the bidirectional actuator.

The bidirectional actuator is basically composed of a spring made of SMA, a bias spring made of spring steel, a shaft, and some mechanical parts. The SMA spring and the bias spring are assembled so as to apply load to each other through the some parts, namely, the shaft, the case, the disk and the electrode. Accordingly, the force equilibrium between the springs determines the shaft position, i.e., the electrode position. SMA is known as an actuator driven by the change in temperature. Concretely, SMA changes from a deformed shape to the memorized shape when its temperature rises more than the phase transformation temperature of the SMA. Consequently, the temperature of the SMA spring changes the force balancing position of the springs, which results in the change in the shaft position, i.e., the electrode position.

The temperature of SMA can be controlled by electric current passing through SMA since electric resistance of SMA is as comparatively high as SMA can be heated by electric current. Accordingly, the discharge current occurring between the electrode and the workpiece in performing EDM can control the SMA spring temperature, that is to say, the shaft position,

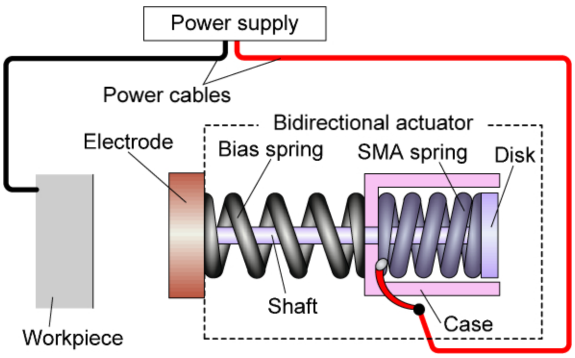
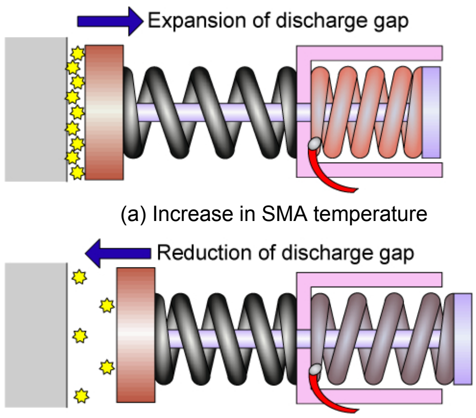


Figure 2: Basic structure of ADGC.



(a) Increase in SMA temperature

(b) Decrease in SMA temperature

Figure 3: Discharge gap control process of ADGC.

i.e., the electrode position. Besides, discharge current generally depends on a discharge gap. Concretely, a narrow gap makes discharge current increase, and to the contrary, a wide one makes it decrease.

As illustrated in Figure 3, consequently, ADGC can automatically and autonomously change the electrode position so as to make the discharge gap wide when the discharge current is high and so as to make the gap narrow when the current is low. As a result, ADGC allows the discharge gap to be adjusted to continue stable EDM by controlling the electrode position according to the discharge current.

3 PROTOTYPICAL ADGC AND IMPROVED ADGC

3.1 Prototypical ADGC and its problems

Figure 4 illustrates the structure of the prototypical ADGC [8]. An upper case and a lower case are assembled with screws and insulated spacers. In the space among the cases and the spacers, a bidirectional actuator and a slider are incorporated so that the position of the slider can be controlled; besides, a guide tube and the lower case are attached so that the motion of the slider can be perpendicularly guided by the tube and the case and so that the range of the slider motion can be determined by the case playing another role of a mechanical stopper for the slider motion. Since the slider is equipped with an electrode for EDM, the position and motion of the electrode and the range of the electrode motion can be done as well as the slider.

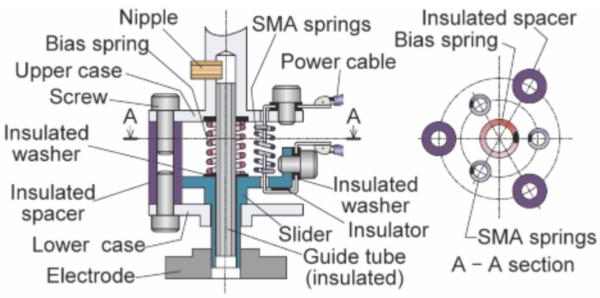


Figure 4: Structure of prototypical ADGC.

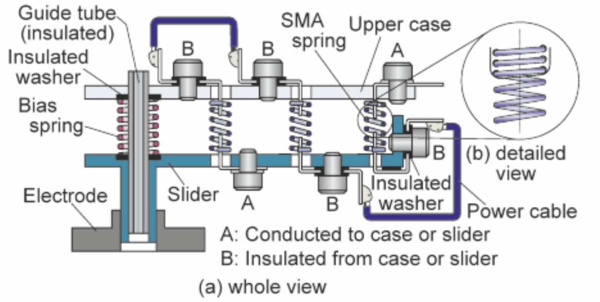


Figure 5: Fastening manner of SMA springs.

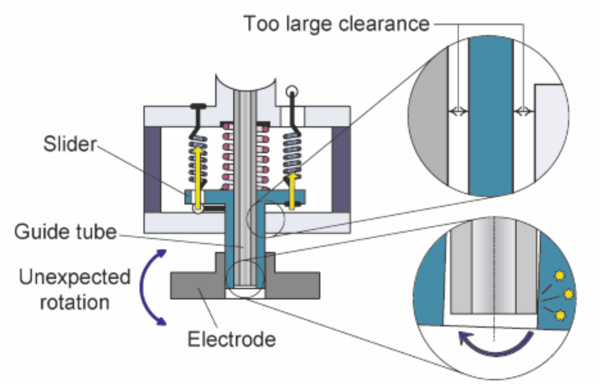


Figure 6: Failure of electrode motion by occurrence of unexpected rotation of slider.

The bidirectional actuator actually employed in the prototypical ADGC consists of the bias spring made as a helical compression spring and the SMA spring made as three helical extension springs. The three SMA springs are respectively positioned at the equal angle around the bias spring. In other words, the bias spring and the three SMA springs are joined to the upper case and the slider mechanically in parallel. Simultaneously, the three SMA springs are connected each other electrically in series so that the discharge current passing through each SMA spring is identical, namely, so that the heat quantity generated in each SMA spring is same due to equal electric resistance of each SMA spring, which results in the almost same motion of the three SMA springs theoretically. To realize the parallel mechanical join and the serial electrical connection concurrently, each SMA spring is fastened to the upper case and the slider in the manner as illustrated in Figure 5. This causes obstacles to size reduction, motion reliability enhancement and full performance achievement of the prototypical ADGC.

The size of the prototypical ADGC cannot but be large due to a lot of components and its complicated structure. This particularly results from the mechanical structure, the electrical conducting or insulating structures and wiring of the springs constituting the bidirectional actuator for embodying their parallel mechanical join and serial electrical connection, as shown in Figures 4 and 5(a). The electrode used in the prototypical ADGC is a cylindrical-shaped one of 22.0mm in diameter, made of oxygen-free copper. However, the diameter of the entire prototypical ADGC including the wiring is much larger than the electrode diameter.

The motion reliability of the prototypical ADGC is not high. This is caused by the fastening manner of the SMA springs, dispersion in the characteristics as SMA of the SMA springs and the low guiding ability of the guide tube and the lower case. The SMA springs are fastened by making their respective ends go through tiny holes of thin metal sheets, as shown in Figure 5(b), since they are helical extension springs. The metal sheets are attached to the upper case and the slider through insulators or conductors and screws, as shown in Figure 5(a). The fastening manner has the possibility of causing a poor electrical connection between the metal sheet and the SMA spring since their physical contact is uncertain and contact area is small. In addition, it is difficult to make the SMA characteristics of the respective SMA springs completely identical. Therefore, it seems that the respective SMA springs do not have a perfectly identical motion even if they are connected in series. Besides, it appears that the guide tube and the lower case do not well work as a guide for the slider motion since the clearance among the slider, the tube and the case is too large. As a consequence, the dispersion and the low guiding ability result in unexpected rotation of the slider, causing the failure of the electrode motion, as shown in Figure 6.

The low motion reliability above-mentioned prevents the prototypical ADGC from achieving its performance to the full. Concretely, the prototypical ADGC could not use at the maximum only 5% of the slider motion range, i.e., the electrode motion range, which had been set to 1000μm in the design stage of the prototypical ADGC.

3.2 Improved ADGC

To overcome the defects of the prototypical ADGC, the improved ADGC is developed. Figure 7 illustrates the structure of the improved ADGC. An insulated disk is mounted on a case. An SMA spring, made as a helical compression spring, is fastened to lower and upper sleeves and the lower sleeve is installed on the insulated disk. To the upper sleeve, a T-shaped shaft is attached so as to penetrate the SMA spring, the sleeves and the insulated disk. On the end of the shaft, an electrode for EDM is mounted. Between the insulated disk and the electrode, a bias spring, made as a helical compression spring, is incorporated.

That is to say, the improved ADGC has a new bidirectional actuator whose bias spring and SMA spring are respectively one helical compression spring. Besides, the case and the lower sleeve are connected by a power cable, so the discharge current passes through the case, the power cable, the lower sleeve, the SMA spring, the upper sleeve, the shaft and the electrode. All other paths are insulated. Moreover, the

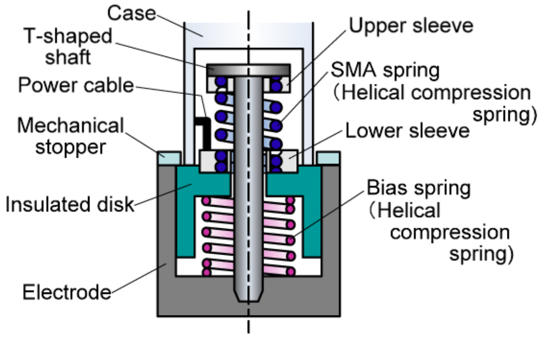


Figure 7: Structure of improved ADGC.

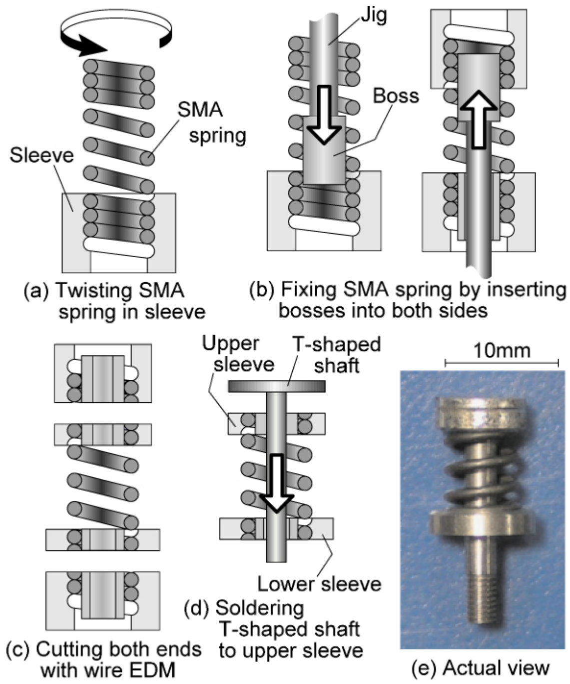


Figure 8: Assenbling procedure of SMA spring and its periphral parts and actual view of them after assembling.

SMA spring and the sleeves are joined in the manner as illustrated in Figures 8(a) ~ 8(d). First, the end of the SMA spring is twisted in the sleeve. On the inside wall of the sleeve, the circle-shaped thread is beforehand machined whose shape is identical with the outside shape of the SMA spring. Next, a boss is inserted into the SMA spring so as to fix the SMA spring to the sleeve and a through hole is machined to the boss. These procedures are done to the other end of the SMA spring. After that, both the ends are cut with wire EDM. Finally, the T-shaped shaft is inserted into the through holes and is soldered to the upper sleeve in the state of no contact with the lower sleeve. Figure 8(e) depicts an actual view of the SMA spring and its peripheral parts after assembling.

Furthermore, the insulated disk plays a role of perpendicularly guiding the electrode motion generated by the new bidirectional actuator. Additionally, it also plays the other role of determining the electrode motion range with a mechanical stopper attached to the electrode.

Figure 9 depicts actual whole views of the prototypical ADGC and the improved ADGC by the same scale. From the comparison between Figures 9(a) and 9(b), it is found that the size of the improved ADGC is drastically reduced compared with the prototypical ADGC. Actually employed electrode in the improved ADGC is made of oxygen-free copper, whose outer shape is a cylinder of 20.0mm in diameter and of 20.0mm in height. As seen from Figure 9(b), the diameter of the parts except the electrode is smaller than the electrode diameter. This means that the improved ADGC can be installed on the microrobot, in other words, that the microrobot equipped with the improved ADGC can at least go through the hole machined by the improved ADGC. This size reduction is realized particularly by the improvement of the new bidirectional actuator in which bias spring and SMA spring are respectively one helical compression spring. The improvement yields simplifications of the mechanical structure, the electrical conducting and insulating structures and wiring of the improved ADGC.

In the improved ADGC, its motion reliability is also enhanced by the following improvements. First, the fastening manner of the SMA spring is improved so as to keep a sure electrical connection between the SMA spring and the sleeves, as shown in Figure 8. Second, the number of the SMA spring is changed from three to one, which means that the dispersion of the SMA characteristics does not influence the electrode motion. Finally, the guiding ability of the electrode motion is made higher, which gives a smooth motion of the electrode.

The new bidirectional actuator and the insulated disk and mechanical stopper are designed so that the electrode movable range of the improved ADGC is 850μm.

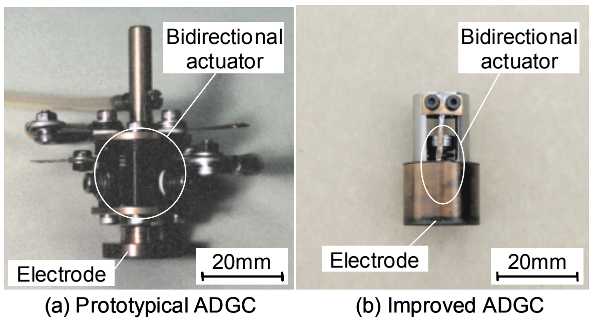


Figure 9: Actual whole views of prototypical ADGC and improved ADGC.

4 VERIFICATION EXPERIMENTS

4.1 Fundamental motion experiment

Motion experiment is conducted by the following manner to confirm the electrode movable range of the improved ADGC. After setting the improved ADGC in water whose temperature can be controlled, the position of the electrode is measured in the process of warming and cooling the water temperature sufficiently slowly. This yields the relationship between the temperature of the SMA spring in the improved ADGC and the moving distance of the electrode since the water temperature is equal to the SMA spring temperature.

Figure 10 shows the relationship between the SMA spring temperature and the electrode moving distance actually obtained in the experiment. From the figure, it is shown that the electrode motion has a hysteresis to the SMA spring temperature in the centre of about 60°C, the hysteresis of displacement to temperature is a typical feature of SMA actuators; besides, it is seen that the improved ADGC clearly has the feature of bidirectional actuator in which the motions in the opposite directions are realized according to the heating or cooling process of its SMA spring; moreover, it is found that the improved ADGC has the electrode movable range of about 850μm, which is almost same as the designed value.

4.2 Machining experiment

Machining experiment is carried out to verify the discharge gap control ability of the improved ADGC. Figure 11 illustrates the experimental manner. Figure 11(a) shows the initial stage

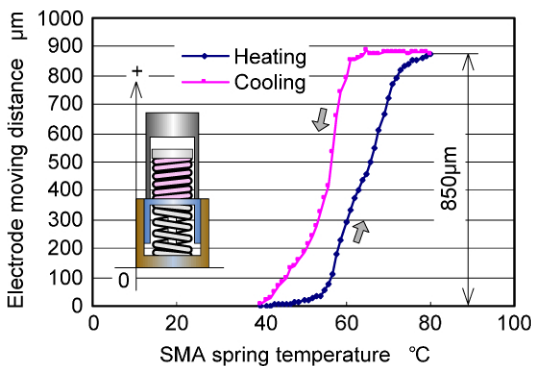


Figure 10: Relationship between SMA spring temperature and electrode moving distance.

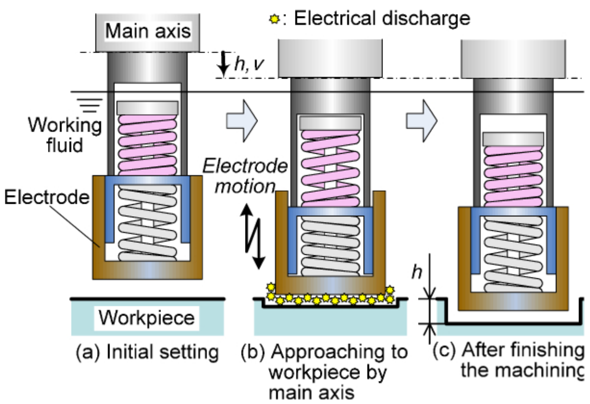


Figure 11: Experimental manner for verifying discharge gap control ability of AGDC.

of the experiment. In the initial stage, the improved ADGC is mounted on the main axis of a die-sinking electrical discharge machine as well as a normal electrode and is positioned so that the gap between its electrode and a workpiece becomes the shortest distance in the state of no occurrence of an electrical discharge. Concurrently, the electrical discharge machine is set to the situations that its own discharge gap control function is turned off and that the voltage and current for EDM is in activation so that EDM can be performed under a certain EDM condition. Next, when the improved ADGC is moved to the workpiece a certain distance at a certain speed and is stopped there by means of the main axis, it starts and continues stable EDM by controlling the discharge gap automatically and autonomously, as illustrated in Figure 11(b). Finally, the EDM ends when the improved ADGC machines the hole of the depth identical with the distance in which the main axis is moved, as shown in Figure 11(c). The distance and the speed in moving the improved ADGC to the workpiece by the main axis are respectively called ‘approach distance’ and ‘approach speed’, which are respectively indicated as h and v, as shown in Figures 11(b) and 11(c).

To make the situation of the initial stage as shown in Figure 11(a), the EDM by means of the improved ADGC as shown in Figures 11(b) and 11(c) is preliminarily performed. That is to say, the gap of the shortest distance without occurrence of an electrical discharge is the gap generated when the preliminary EDM by means of the improved ADGC ends. In addition, the preliminary EDM is performed until the depth of the hole machined by it becomes 0.5mm so that the EDM in the machining experiment is performed using all the area of the bottom of the cylindrical electrode. Actually employed electrical discharge machine in the study is a numerically controlled die-sinking electrical discharge machine (EDNC65) made by Makino Milling Machine Co., Ltd.

Letting the average approach speed be 1.2mm/min, actual machining experiments are done under the EDM condition

Table 1: EDM condition

Electrode	Oxygen-free copper
Workpiece	Aluminum alloy (A5052)
Work fluid	Oil
Discharge current	8 A
Pulse direction	135 μs
Duty factor	80 %
Polarity	Electrode(+) / Workpiece(-)
Jump	Not used
Orbital motion	Not used

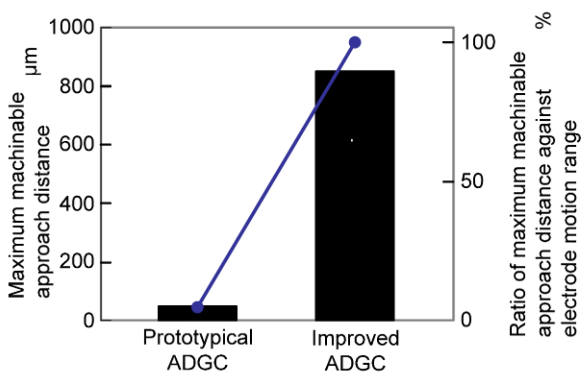


Figure 12: Comparison of performance between prototypical ADGC and improved ADGC.

listed in Table 1. From the results of the experiments, it is confirmed that the improved ADGC can maintain stable EDM by controlling the discharge gap and can finish fabricating the hole of the depth identical with the approach distance without short circuit or unexpected stop of EDM, even when the approach distance is set to 850μm which is almost equal to the electrode movable range as shown in Figure 10. This means that the improved ADGC can achieve its performance to the full.

Figure 12 shows comparisons of the maximum machinable approach distances and the ratios of maximum machinable approach distance against the electrode motion range between the prototypical ADGC and the improved ADGC. In case of the prototypical ADGC, the former is 50μm and the latter is 5%. In case of the improved ADGC, on the other hand, the former is 850μm and the latter is 100%. From the comparison, it is found that the former is increased 17 times and the latter is raised to the maximum. This proves the effectiveness of the improvement and means that the minimum positioning resolution required to the microrobot to be developed has increased 17 times.

5 CONCLUSIONS

Aiming at establishment of a new curved hole machining method by means of the microrobot with EDM function, ADGC has been devised which is one of the three units constituting the microrobot and can automatically and autonomously control discharge gap to realize stable EDM in the almost isolated space such as the bottom of a curved hole. The experiments using the prototype of ADGC have proven that ADGC has the automatic and autonomous discharge gap control ability. However, the prototypical ADGC had the problems that it could not be installed on the microrobot since its diameter was larger than the diameter of the electrode mounted on it, i.e., the diameter of a hole machined by it, and that it could control discharge gap only in the range of 50μm at the maximum although the electrode motion range was set to 1000μm in the design stage. To solve the problems, in the study, the improvement of ADGC has been given. By the improvement and from the results of the experiments employing the improved ADGC, the obtained conclusions are summarized as follows:

1. The improvement can make the size of ADGC drastically reduced by thoroughly redesigning the components of ADGC so that ADGC can be installed on the microrobot.
2. The improvement can make the discharge gap control ability of ADGC enhanced so that ADGC can control discharge gap even in the range of 850μm, which is 17 times of the ability before the improvement and is 100% of the electrode motion range after the improvement due to setting it to 850μm, meaning that the improvement can make ADGC achieve its full performance.

6 ACKNOWLEDGMENTS

The study is partly supported by the Ministry of Education, Culture, Sports, Science and Technology, Grant-in-Aid for Exploratory Research 17656051 and Grant-in-Aid for Young Scientists (B) 18760099, and MAZAK Foundation. The authors would like to express their sincere gratitude to them for their support.

7 REFERENCES

[1] Koresawa, H., Suzuki, H., 1999, Autonomous Arrangement of Cooling Channels Layout in Injection Molding, Proceedings of Annual Technical Conference '99 (ANTEC99), 1073-1077.

[2] Fukui, M., Kinoshita, N., 1989, Developing a 'Mole' Electric Discharge Digging Machining, Annuals of the CIRP, 38/1:203-206.

[3] Ichiyasu, S., Takeuchi, A., Watanabe, K., Goto, A., Magara, T., 1997, Machining Curved Tunnels for Coolant with Mole EDM, Proceedings of 4th International Conference on Die & Mould Technology, 224-230.

[4] Goto, A., Watanabe, K., Takeuchi, A., 2002, A Method to Machine a Curved Tunnel with EDM, International Journal of Electrical Machining, 7:43-46.

[5] Uchiyama, M., Shibazaki, T., 2004, Development of an Electromachining Method for Machining Curved Holes, Journal of Materials Processing Technology, 149/1:453-459.

[6] Ishida, T., Nakajima, Y., Miyake, Y., Teramoto, K., Takeuchi, Y., 2006, Development of Electrode Motion Control Device for Electrical Discharge Curved Hole Machining, Proceedings of 39th CIRP International Seminar on Manufacturing Systems, 307-312.

[7] Ishida, T., Kogure, S., Miyake, Y., Takeuchi, Y., 2004, Creation of Long Curved Hole by Means of Electrical Discharge Machining Using an In-Pipe Movable mechanism, Journal of Materials Processing Technology, 149/1:157-164.

[8] Ishida, T., Takeuchi, Y., 2001, Development of Automatic Discharge Gap Controller for Curved Hole machining, Proceedings of 13th International Symposium for Electromachining, 1:63-73.

[9] Ishida, T., Hosoda, T., Fukuda, T., Takeuchi, Y., 2003, In-Process Measurement of Electrode Motion in EDM by Means of Automatic Discharge Gap Controller, Proceedings of 2nd International Conference on Leading Edge Manufacturing in 21st Century (LEM21), 965-970.

Design of a Motion Platform for a Mobile Machine Simulator by Utilizing 6-D Measurements and Inverse Dynamics Analysis

Joni Sallinen, Tero Eskola, Heikki Handroos

Institute of Mechatronics and Virtual Engineering, Lappeenranta University of Technology, Finland

Abstract

This paper describes the design process of a motion platform for a mobile machine simulator. Design process was started by measuring dynamics of an underground loader. Based on the measurements and operational analysis of the machine behaviour the basic structure of the platform was selected. A simulation model that consists of measurement data, signal filtering, inverse kinematics and inverse dynamics was used in dimensioning the actuators and the mechanical structure of the platform.

The aim of the work was to develop a motion platform that creates realistic feeling of motion. Main requirement for the mechanical structure was low height. Movements were planned to be carried out by using electrical actuators.

As a result motion platform with three degrees of freedom and servomotors as actuators was designed. The physical prototype of the motion platform is going to be built based on the design and it’s going to be connected to a real-time simulator of the underground loader.

Keywords:

ICMA2008; Motion platform; mobile machine; simulator; measurements; signal filtering; inverse dynamics.

1 INTRODUCTION

Motion platforms are used in training simulators to create feeling of motion, which is very important especially in vehicle and mobile machine simulators [1], [3]. The signal filtering between the simulator and motion platform is an important issue, because the platform has a limited work space [1], [3], [4]. Major part of the publication discuses about the filtering. In [2] design issues apparent in a simplified driving simulator are discussed. The present study introduces a method by means of which the motion platform can be designed by employing motion measurements from the real vehicle and simulation of inverse and direct dynamics of the platform.

By measuring the dynamics of machine, whose simulator is going to be constructed, it is possible to define required degrees of freedom and kinematic structure of platform. By utilizing the measured data and inverse dynamics analysis of the mechanical structure of the platform it is possible to define the performance requirements of actuators. After determination of the performance requirements through inverse dynamic analysis the final performance which is strongly dominated by the filtering and motion driving algorithms can be estimated through simplified direct dynamic analysis.

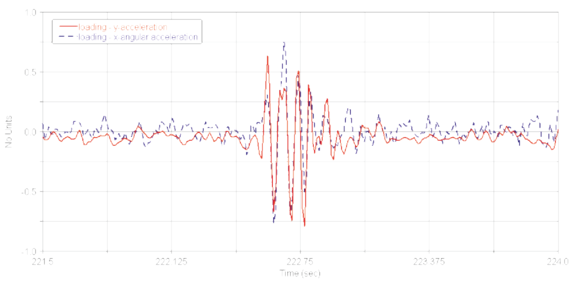
The present paper describes a design process of a motion platform for an underground loader. Measurements, data analysis, basic structure design, signal filtering and inverse dynamics analysis are described. Finally results of direct dynamic analysis with approximative dynamic model for the platform and the selected filtering algorithm are shown. The simulated and measured accelerations and angular velocities are compared with those originally measured from the real vehicle.

2 MEASUREMENTS

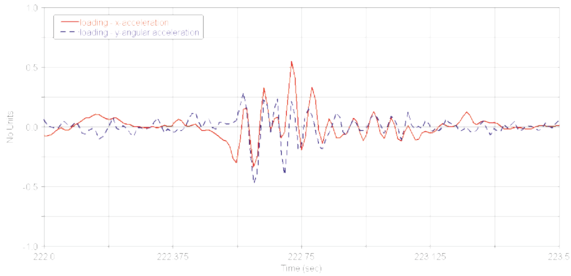
The design process was started by measuring an underground loaders dynamics by using a 6-D sensor (Microstrain, Inertia-Link). The sensor consists of three accelerometers and three gyros and can measure at 100 Hz frequency and its measuring range is ±10 g and ±1200˚/s. The sensor was rigidly mounted with the cabin floor under the driver’s seat to measure the accelerations that driver feels in during the basic work cycles of the machine. Work cycle consists of loading, hauling and unloading.

Measured data was analyzed by using FFT- analysis and by searching maximum translational accelerations and angular velocities in each work cycle situations. Maximum accelerations and angular velocities took place during loading at frequencies 2-3 Hz and 10-15 Hz. Another clearly distinguishable point was during hauling when driving at bumpy surface.

Linkage between translational accelerations and angular velocity in horizontal axes was found in measured data. When machine tilts around *x*-axis forms *y*-acceleration and when it tilts around *x*-axis forms *y*-acceleration. This linkage is shown in figure 1. Angular accelerations were derived from angular velocities and scaled to size of translational accelerations.



(a)



(b)

Figure 1: Linkage between translational and angular accelerations.

3 KINEMATIC STRUCTURE OF THE MOTION PLATFORM

Measured data, operational analysis of the machine and feeling of motion during operation were used to define the kinetic structure of the platform.

Measured data showed that accelerations and angular velocities occur to all directions during operation. It also showed linkage between translational accelerations and angular velocity in horizontal axes. Hence horizontal accelerations are mimicked by tangential components of the angular accelerations. Rotation around vertical axis can be replaced by using visualization.

3-DOF (roll, pitch and heave) kinetic structure of the platform was selected. The structure is shown in figure 2. It consist of upper and lower frame, support beam between frames and three actuators. Actuators can be linear movement actuators or rotating actuators connected to crank mechanism. Support beam is rigidly mounted to upper frame and with universal joint to lower frame. Beam also enables translational movement. By placing universal joint to lower part of the beam it is possible to get larger radius to form tangential accelerations along horizontal axes.

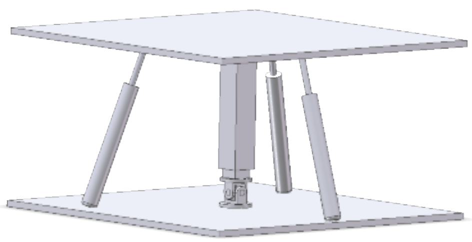


Figure 2: Selected kinetic structure of the platform.

Coordinate system of the platform is defined with the *xy*-plane along the upper frame and *z*-axis is pointing upwards. Actuators are connected to the frames at the corners of an equilateral triangle shown in figure 3 to get symmetrical operation. [2]

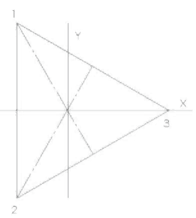


Figure 3: Layout of actuator's joint places.

4 REQUIRED PERFORMANCES OF THE ACTUATORS

After determining the basic structure of the platform the required performances of actuators were defined by the procedure presented in figure 4.

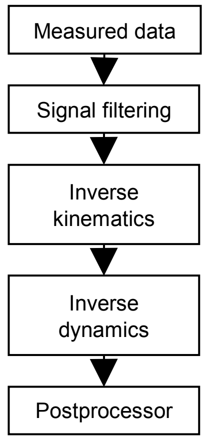


Figure 4: Determination of design requirements by inverse dynamic simulation.

Firstly, the measured accelerations were converted into forms in which they can be used in control of the movements of the platform. This was carried out by applying appropriate filters and double integrators. Because the work space of the platform is limited, the low frequency accelerations are described by tilting the platform. Secondly, the references into the servo-actuators are generated form the desired trajectories of the platform through inverse kinematics. The same filtering and inverse kinematics can later be used in actual control of the motion platform and they were calculated in Matlab/Simulink.

The reference values of the actuator velocities are taken into the inverse dynamics model of the platform as the actual velocities of the actuators. The inverse dynamic simulation model which is made in Adams software calculates the forces and moments acting against the actuators. Also the forces affecting in the joints can be checked. From postprocessor the following requirements for the actuators can be obtained: the force/moment with maximum speed, speed with maximum force/moment, maximum power (speed times force/moment) and continuous force/moment with continuous speed. Using these values actuators can be dimensioned and selected.

4.1 Signal filtering

Diagram of the signal filtering is shown in figure 5. Only the measured accelerations are used.

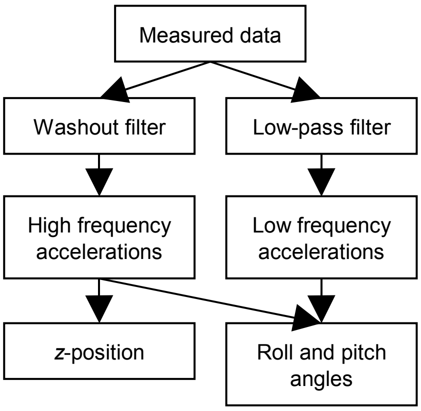


Figure 5: Diagram of the signal filtering.

Classical washout filter consist of band-pass filter, high-pass filter and a double integrator [1]. Parameters chosen to filters:

- band-pass filter: filter order 2 and frequency 1…30 Hz
- high-pass filter: filter order 1 and frequency 0.5 Hz
- low-pass filter: filter order 2 and frequency 0.1 Hz

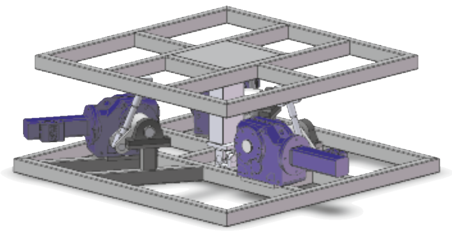
After filtering *z-position* is formed directly from washout filtered *z*-acceleration and roll and pitch angles (θ_1 and θ_2) are combined from washout and low-pass filtered *x*- and *y*-accelerations.

High frequency angles are formed by dividing washout filtered accelerations by length from universal joint to control point in the upper frame. Low frequency angles are formed by utilizing gravitation. By tilting the platform it is possible to describe angles in road and velocity changes of the machine.

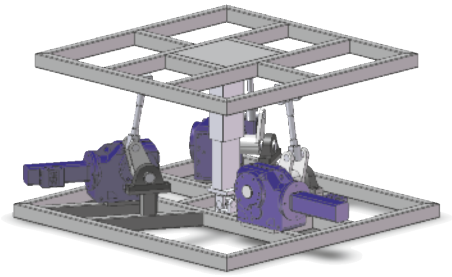
5 RESULTS

5.1 Final design

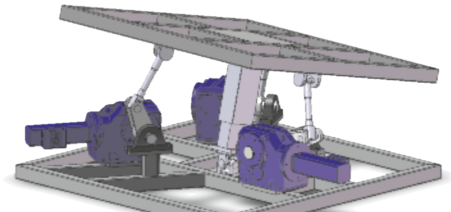
As a result motion platform with three degrees of freedom (roll, pitch and heave) and servomotors as actuators was designed. The structure is shown in figure 6.



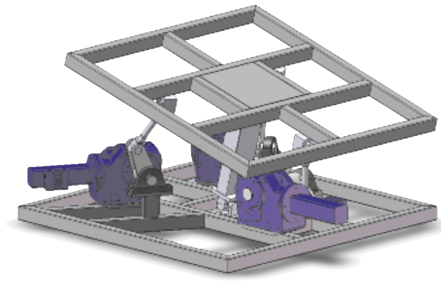
(a)



(b)



(c)



(d)

Figure 6: 3-DOF motion platform. (a) Lowest position. (b) Highest position. (c) Roll. (d) Pitch.

Work space of the motion platform:

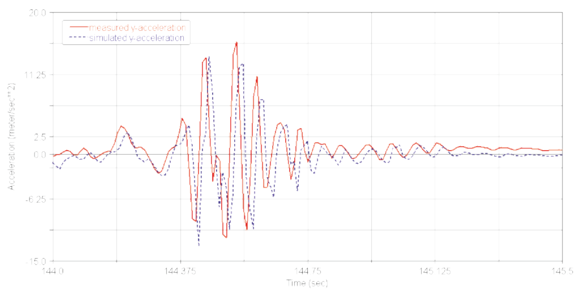
- *z-translation* ±100 mm.
- Roll ±14.5°.
- Pitch ±15°.

Chosen actuators:

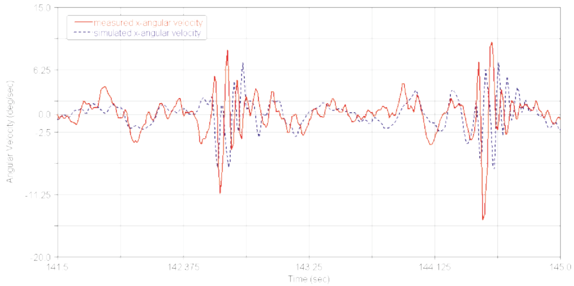
- Synchronous geared servomotor with helical-bevel gear unit K67DS56H, M=4 Nm, n_M=4500 rpm, i=123,54, n_A=36.
- MOVIDRIVE B frequency inverter MDX61B0030.

5.2 Comparison of simulated and measured dynamics

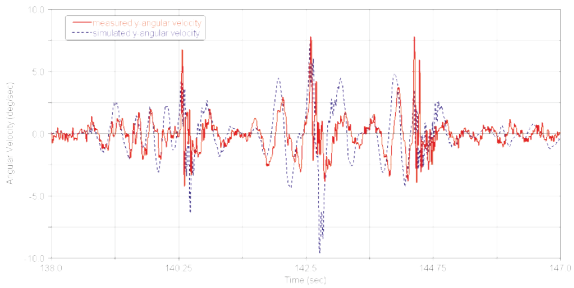
Figure 7 shows how well designed motion platform mimics measured accelerations and angular velocity of the real machine. Platforms simulated values don't involve delays that come from control circuit and servomotors.



(a)



(b)



(c)

Figure 7: Comparison of simulated and measured dynamics. (a) *y*-acceleration. (b) *x*-angular velocity. (c) *y*-angular velocity.

Figure 7 (a) shows that curves are corresponding hence signal filtering and control of the motion base is correct. Delay comes from signal filtering and there is offset because sensor was not exactly horizontal plane at measurements. *x* and *z*-accelerations behave similar way as demonstrated *y*-acceleration.

Figure 7 (b) and (c) shows that platform mimics quite well real machine's behavior. Simulated angular velocities are close to measured values though only acceleration was used in control.

6 SUMMARY

A design process of a motion platform for an underground loader simulator was presented. Firstly, the dynamics of a real machine was measured. Secondly, based on the measurements and operational analysis of the machine behavior, the kinematic structure of the platform was selected. After that the required performances of actuators were defined by using simulation tools. As a result motion platform with three degrees of freedom (roll, pitch and heave) and servomotors as actuators was designed.

Comparison of measured and simulated values showed that motion platform mimics well the real machines behavior. Horizontal accelerations can be mimicked by tangential components of the angular accelerations and rotation around vertical axis can be replaced by using visualization.

Based on this work it is possible to utilize dynamic measurements and simulation tools in simulators motion platform design. Although success of the design can be fully confirmed by building a prototype of the platform, because it is hard to estimate with simulation what is the feeling of motion of the platform.

Before building a prototype of the platform the design needs to be focused on the areas of mechanical joints, control circuit, safety and strength calculations.

7 REFERENCES

[1] Nehaoua L., Amouri A. and Arioui H. Classic and Adaptive Washout Comparison for a Low-Cost Driving Simulator. Proceedings of the 13th Mediterranean Conference on Control and Automation, held at Limassol, Cyprus, 27-29 June, 2005. pp. 586-591.

[2] Webjorn R. Design and Application of a Motion Platform for a High-Speed Craft Simulator. IEEE International Conference on Mechatronics, ICM, 2006.

[3] Nehaoua L., Arioui H., Espie, S., and Mohellebi, H. Motion Cueing Algorithm for Small Driving Simulator. Proceedings of the 2006 IEEE International Conference of Robotcs and Automation, held at Orlando, Florida, May 2006. pp. 3189-3194

[4] Nahon Meyer A. and Reid Lloyd D. Simulator Motion-Drive Algorithm: A Designer's Perspective. Journal of Guidance, 1990. Vol. 13:2. pp. 356-362

A Position Measurement Method for a Miniature Mobile Robot Using Three Moving Landmarks

Tomohiro Tanahashi[1], Akihiro Torii[2], Masaaki Banno[1], Akiteru Ueda[2], Kae Doki[2]

[1] Graduate School of Engineering, Aichi Institute of Technology, Toyota, Japan

[2] Faculty of Engineering, Aichi Institute of Technology, Toyota, Japan

Abstract

We studied a position measurement system for a small autonomous mobile robot. The system uses three line laser projectors above the robot, and the line laser projectors move on three linear stages. Three position sensitive detectors (PSDs) arranged in a triangle are carried on the robot. The line laser beams are irradiated on the active surface of the PSDs, and the robot's position is obtained by the lasers' positions and PSDs' outputs. In our previous method, the position of the robot is measured by the positions of movable lasers and PSD outputs. Since the system required that the laser beams be irradiated on the center the PSDs, a real-time measurement is difficult. In this paper, the lasers' positions are obtained by the elapsed time of the position measurement. By measuring the time when the PSD output becomes zero, the position of the robot is calculated with the PSDs' outputs and the timer output. The preliminary experimental results are discussed. In our experimental conditions, the measurement accuracy is smaller than 0.01 mm.

Keywords:

Miniature Robot; Position Sensitive Detector; PSD; Position Measurement System

1 INTRODUCTION

Many industrial robots are used in the production of electrical and mechanical components. The size of the products varies from some millimeters to some meters. We have some problems in the small components production by the use of large industrial robots. The conventional production system using the industrial robots, which are usually large, wastes of energy and space.

Recently, the concept of precision manufacturing system which is organized by several miniature robots was introduced [1]. The robot is incorporated with microsensors for monitoring, micromanipulators for precise operation, and small actuators for precision moving. The miniature robots are organized to work for manufacturing in cooperation with each other and conventional machines.

When the miniature mobile robot cooperates with other robots or machines, it must know its position. In our laboratory, we use two position measurement systems. One is a camera vision system, and the other is a precise displacement sensor. A camera vision system is used in a wide measurement area which is about 10 cm x 10 cm. The resolution of the position measurement is determined by the number of pixels of the camera. A precise displacement sensor is used in a small measurement area. Although the resolution of the sensor is about 0.01 mm or higher, the maximum working range is several millimeters.

We previously proposed the position measurement system using position sensitive detectors (PSDs) and lasers [2]. Three lasers are arranged around miniature robot and irradiated on the PSDs which outputted the laser spot position. The robot's position was obtained by the laser spot position on the PSD surface. The working area is restricted by the length of the effective area of the PSDs.

We introduced moving lasers so that the measurement area is equal to the working length of the moving lasers [3]. We irradiated the moving lasers on the center of the PSDs, and obtained the robot position. The lasers, however, kept chasing the center of the PSDs.

In this paper, we describe a position measurement method using a timer and three moving landmarks, which are line laser projectors carried by three linear stages. In the following sections, the structure of the miniature robot with three PSDs, the measurement system using three linear stages with line laser projectors, the principle of the position measurement, and preliminary experimental results are described.

2 MINIATURE ROBOT

A lot of miniature robots were developed by many research groups. They have some functions corresponding to their applications. The autonomous miniature robot requires a precise motion and accurate positioning. Some robots detect their position by using on-board position sensors. Some robots use communication devices which are used to exchange position information with other robots. The position measurement system for the miniature robot should be the system which is suitable for the function of the robot.

Figure 1 shows the miniature robot we developed. The bottom of the robot is a moving mechanism. The deformation of piezoelectric elements (piezos) moves the robot. The size of the piezo mainly decides the dimensions of the robot. Two principles of operation are usually used. One is rapid deformation of the piezo. The robot moves by the repetition of quick and slow deformation of the piezos. The other is an inchworm which consists of thrust mechanisms and clamping mechanisms. We used the piezos as the thrust mechanism and electromagnets as the clamping mechanisms. The size of the piezo mainly decides the dimensions of the robot. The detail of the miniature robot is described in Reference [4].

The PSD is an optoelectronic position sensor utilizing photodiode surface resistance. It provides continuous position

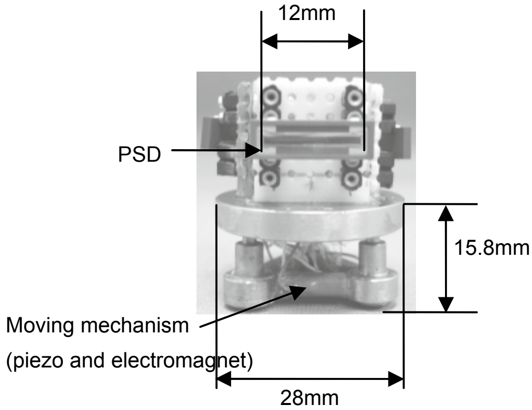


Figure 1: Miniature robot. Piezo and electromagnet move the robot. PSDs on the robot are used for position measurement.

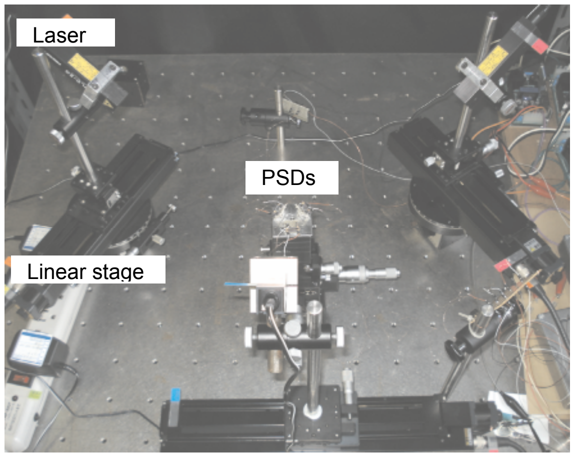


Figure 2: Measurement system. Linear stages in a triangle form carry line laser projectors. Laser beams apply to PSDs.

data and features high position resolution, and is suitable for the position measurement of the miniature robot. The top of the robot shown in Figure 1 is a one-dimensional PSD with belt-shaped active area designed to detect the longitudinal positions. The resolution of the PSD is determined by the sensitivity of the PSD and an electronic noise of a measurement system.

3 MEASUREMENT SYSTEM

In some conventional position measurement systems, a polygon scanner is used. A laser beam is scanned by the polygon scanner and the position of the scanner is used as a reference position of the position measurement. The position of a moving robot is determined by the polar coordinate system, and is obtained by the product of a distance and an angle. The distance measurement error and the angle measurement error cause the position measurement error. The angle error is enlarged by the distance, when we measure the position away from the reference. We, therefore, use linear stages which has an equal uncertainty along the moving length, instead of the polygon scanner.

Figure 2 shows a measurement system for a small mobile robot. Three linear stages are arranged in an equilateral triangle around the robot with three PSDs, which are arranged in an equilateral triangle on the robot. Three line laser projectors, which draw three straight lines on a measurement floor, move with the linear stages. The lasers are above the miniature robot so that the laser beam is not interfered with an obstacle on the measurement floor. The laser lines are perpendicular to the linear stages. The measurement area is determined by the length of the linear stages.

Figure 3 shows the coordinate of the position measurement. Capital letters (X_1, X_2, X_3) express the lasers and their positions and small letters (x_1, x_2, x_3) express the laser spot position on the PSDs, respectively. The center of gravity of the triangle formed by the linear stages is defined as the origin of the xy coordinate of the position measurement.

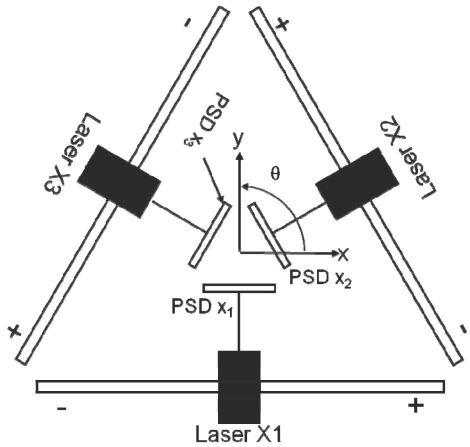


Figure 3: Coordinates. Laser X1, X2, X3 apply to PSD x1, x2, x3, respectively. Angle θ is the rotation in the counter clockwise direction.

When the linear stage is at the center of their travel range, X_1, X_2 and X_3 are zero and the laser beams intersect at one point where the origin of the position measurement of the robot. The rotational angle θ is defined by the right-handed screw rule.

4 MEASUREMENT PRINCIPLE

4.1 Conventional measurement system

In the measurement system we firstly proposed, the position of a miniature robot was measured by using three lasers and three PSDs [2]. The lasers are fixed on an operating surface, and the robot's position is measured by the PSDs' outputs. The disadvantage of this method is small range of measurement. The measurement area is limited by the length of the active area of the PSD. The longitudinal length of active measurement area of the PSD determines the range of measurement. Since the maximum length of the PSD is several centimeters, the robot cannot move on a desktop manufacturing stage.

We then introduced a position measurement method using moving lasers [3]. The moving lasers were carried by linear stages and they chased the center of the PSDs continuously. In order to irradiate laser beams continuously, line laser projectors were also used. When we used Gaussian lasers, the laser beams must be in parallel with the surface where the robot moves. The advantage of the method is that the length of linear stages determines the measurement area. On the other hand, the disadvantage of the method is that the moving lasers kept chasing the center of the PSDs continuously. The parallel laser beam was interfered with an obstacle between the laser source and the PSD on the surface. Therefore, the moving lasers are above the miniature robot. The line laser projectors are also carried by linear stages, and irradiated to the active area of the PSDs easily. The robot's position was measured by the position of the linear stages, and the PSD outputs. This method, however, have to obtain the position of the moving lasers, and the laser position have to be synchronized by the PSD output.

4.2 Proposed principle

The system we propose in this paper uses linear stages which move at constant speed. Assume that the speed of a miniature robot is slower than that of the linear stage. We use three linear stages, and all the linear stages move at constant and equal speed. The position of the laser projector is obtained by the constant *k* which denotes the speed of the linear stage, and the time *t* which denotes the time elapsed.

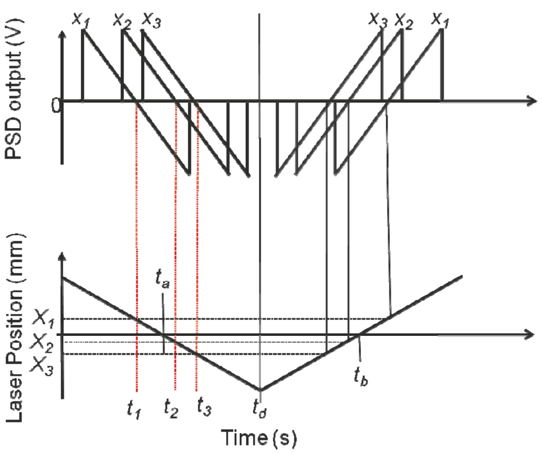


Figure 4: Schematic principle of the laser position detection. (top) PSD outputs change according to the laser position X1, X2, X3. (bottom) Laser positions X1, X2, X3 are determined from zero crossing of PSD outputs x1, x2, x3.

The three lasers start at the same time move from one end toward the other end of the linear stages. They move at constant and equal speed on the linear stages. When they reach the other end of the linear stages, they change the moving direction and return to the start position. The linear stages change their directions when they arrive at the end of the linear stages. Since the linear stages move with the equal speed, the position of the linear stage is expressed by the function of the time elapsed.

The PSD outputs and laser positions are illustrated in Figure 4 while lasers move from one end to the other end and returns. The PSD outputs and laser positions change as a function of time elapsed. For example, the laser position Xi is expressed as

$$\begin{cases} X_i = -k(t - t_a) & (0 < t < t_d) \\ X_i = k(t - t_b) & (t_d < t) \end{cases} \tag{1}$$

where subscript i denotes the stage number shown in Figure 3. The positions of the lasers are synchronized with each other so that the lasers positions X_1, X_2, X_3 are equal along with the elapsed time. The PSDs output the position where a laser beam irradiates. The lasers move at constant speed so that the PSD output changes linearly. Since the length of the PSD is shorter than the length of the linear stage, the PSD output is nearly zero while the laser beams do not apply to the active area of the PSD. The laser position X_i and the PSD output x_i are symmetry with respect to the passage of time, because the linear stage moves at constant speed in the reverse direction.

The position and posture of the miniature robot are expressed by the following [3].

$$x = \frac{2X_1 - X_2 - X_3}{3} - \frac{2x_1 - x_2 - x_3}{3}\cos\theta \tag{2}$$

$$y = \frac{X_2 - X_3}{\sqrt{3}} - \frac{x_2 - x_3}{\sqrt{3}}\cos\theta \tag{3}$$

$$\theta = \sin^{-1}\frac{(X_1 + X_2 + X_3)/(3h)}{\sqrt{1 + ((x_1 + x_2 + x_3)/(3h))^2}} - \tan^{-1}\frac{x_1 + x_2 + x_3}{3h} \tag{4}$$

Capital letters denote positions of the linear stages and small letters denote PSD outputs. The constant h is a distance between the centre and PSD. The capital letters X_1, X_2, X_3 are replaced by Equation (1), then we obtain the position at a general time t_1, t_2, t_3.

The following equations are obtained.

$$x = \frac{2kt_1 - kt_2 - kt_3}{3} - \frac{2x_1 - x_2 - x_3}{3}\cos\theta \tag{5}$$

$$y = \frac{kt_2 - kt_3}{\sqrt{3}} - \frac{x_2 - x_3}{\sqrt{3}}\cos\theta \tag{6}$$

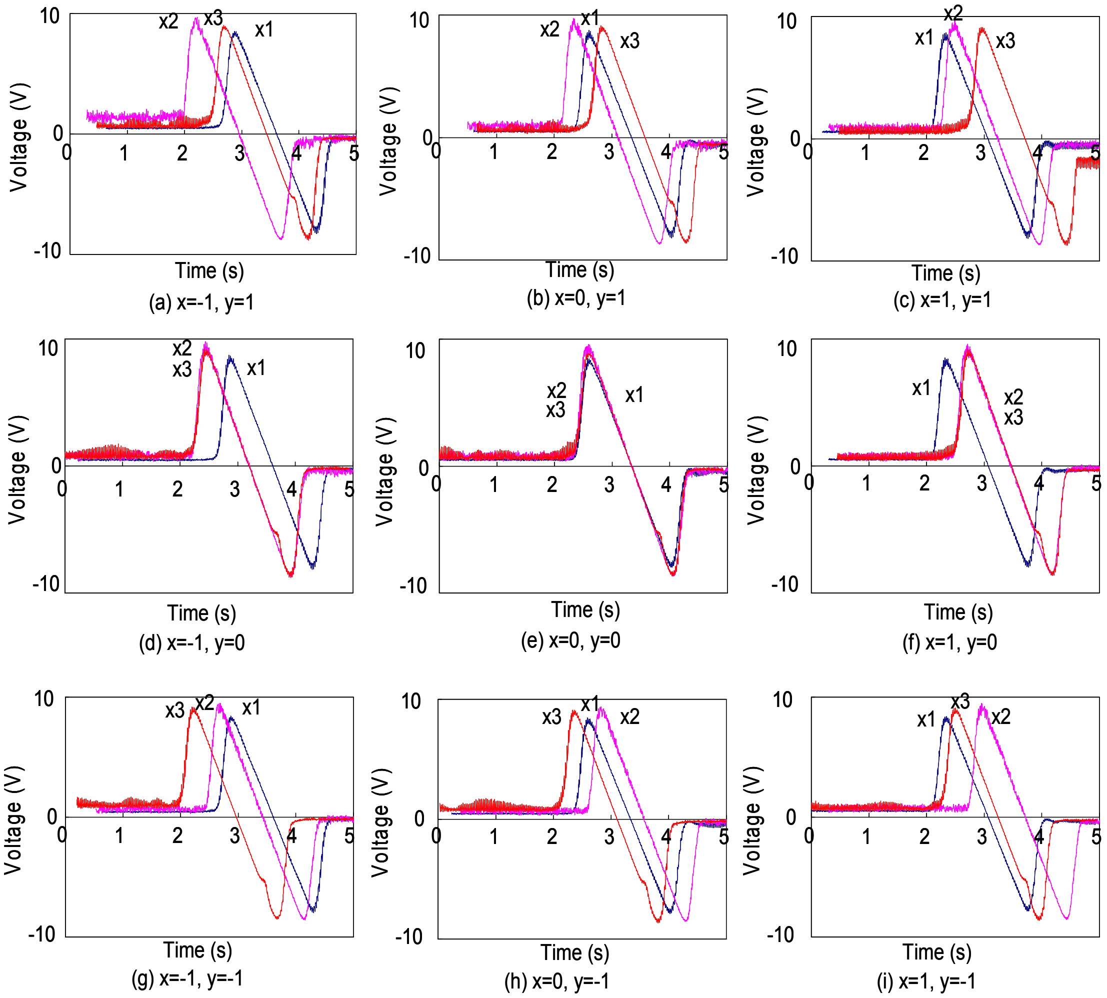


Figure 5: Nine measurement results at x=-1, 0, 1, and y=-1, 0, 1. The nine reference positions are indicated.

$$\theta = \sin^{-1} \frac{(kt_1 + kt_2 + kt_3 - 3kt_a)/(3h)}{\sqrt{1 + ((x_1 + x_2 + x_3)/(3h))^2}} - \tan^{-1} \frac{x_1 + x_2 + x_3}{3h} \tag{7}$$

Suppose that the robot moves in the linear direction. The rotational angle θ of the robot is zero and the PSD surface is parallel with the linear stages. The robot position is simplified as follows.

$$x = \frac{2kt_1 - kt_2 - kt_3}{3} - \frac{2x_1 - x_2 - x_3}{3} \tag{8}$$

$$y = \frac{kt_2 - kt_3}{\sqrt{3}} - \frac{x_2 - x_3}{\sqrt{3}} \tag{9}$$

The time t_1, t_2, t_3 are obtained when the PSD outputs x_1, x_2, x_3 are zero, as shown in Figure 4.

In our system, the lasers' positions X_1, X_2 and X_3 are directly proportional to the elapsed time, and the lasers move at constant speed k. The robot position x and y, therefore, is simply expressed by

$$x = \frac{(2t_1 - t_2 - t_3)k}{3} \tag{10}$$

$$y = \frac{(t_2 - t_3)k}{\sqrt{3}} \tag{11}$$

The position of the robot is calculated by the time elapsed when PSDs' outputs are zero.

Time measurement error causes the position measurement error in x and y direction. The errors are expressed by the followings.

$$dx = (2|dt_1| + |dt_2| + |dt_3|)\frac{k}{3} \tag{12}$$

$$dy = (|dt_2| + |dt_3|)\frac{k}{\sqrt{3}} \tag{13}$$

5 EXPERIMENTAL

In the experiment, an xy stage simulates a miniature robot, and three PSDs (Hamamatsu Photonics, S3932) arranged in an equilateral triangle are installed on the xy stage. The position of the xy stage is used as the reference of the position measurement. The reference x stage moves from -1 mm to 1 mm while the reference y stage also moves from -1 mm to 1mm. The length of active surface of the PSD is 12 mm. Three linear stages (Suruga Seiki, KS162 Series) carry the line laser projectors (Newopto, OPLD-20). The length of the linear stage is 200 mm and its speed k is 6.00 mm/s. The constant ta in Figure 4 is 3.33 s. Since the stage starts from X=10 mm at t=0, the laser position is expressed as follows.

$$X_i = -6.00(t - 3.33) \tag{14}$$

The wavelength, power, and projection angle of the line projector is 670 nm, 5 mW, and 45 degrees. The position of the linear stage is directly proportional to the time elapsed. The PSDs' outputs are monitored.

6 RESULTS AND DISCUSSION

Figure 5 shows experimental results, which are obtained at nine reference positions. The horizontal axis denotes the time elapsed and the vertical axis denotes the PSD outputs. Although the PSD usually outputs the light spot position, the vertical axis indicates the output voltage which is in proportional to the laser spot position. The PSD outputs zero when the light spot is at the center of the PSD surface. From the zero crossing of the PSD output, the time t_1, t_2, and t_3 are obtained, and the position x and y are calculated by Equations (10) and (11), where the speed constant is k=-6.00 mm/s in the experimental conditions.

Figure 6 explains one example, which shows the magnified zero crossing in Figure 5 (b). The intersections of the PSD outputs and the horizontal axis are t_1=3.34 s, t_2 = 3.11 s, and t_3 = 3.58 s, and then the position is obtained as x=0.012 mm and y=0.999 mm, which agree well with the reference position x=0.0 mm and y=1.0 mm by Equations (10) and (11). The rotational angle θ is calculated by Equation (7) where constant h is 15 mm, and is obtained as 4.0 mrad, which agreed the experimental conditions.

Error is estimated by the noise of the PSD output. In Figure 6, the PSD output x_2 contains a lot of noise, although the outputs x_1 and x_3 include small noise. The maximum error of zero cross reading included in x_2 is about 0.010 s. The errors in x_1 and x_3, however, are about 0.005 s. The error dx and dy are 0.05 mm and 0.02 mm by the use of Equations (12) and (13). Eliminating electronic noise is important in order to reduce the position measurement error.

Figure 7 shows summarized results shown in Figure 5. All nine results agree with the reference positions as shown in Figure 7(a). Figure 7(b) shows the measurement errors

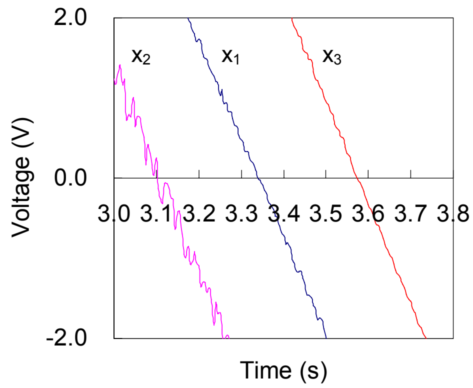


Figure 6: Magnified zero crossing of PSD outputs. The reference position is x=0 and y=1.

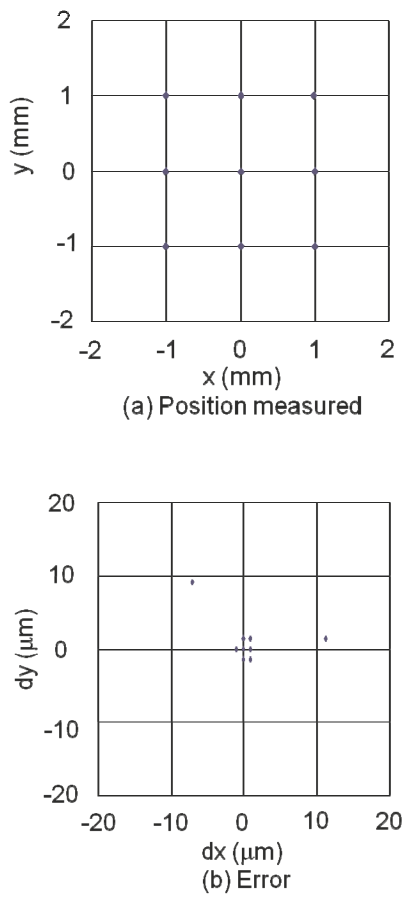


Figure 7: Results and errors in position measurement.

compared with the reference positions. Most errors obtained by the proposed position measurement system is smaller than 0.01 mm.

7 SUMMARY

This paper describes a position measurement system for a small autonomous mobile robot. The system used three line laser projectors above the robot, and three linear stages which carry the lasers. The line projectors and linear stages are used as landmarks.

The movement of three liner projectors are synchronized with each other and the position of the lasers are in proportional to the measurement time. Since we used the liner laser projectors the laser beams irradiate the PSD surface easily.

Three PSDs are installed on a mobile miniature robot, although we used a position reference stage in the preliminary experiment. The PSD outputs a laser spot position on its active area.

The position of the robot, which is simulated by the use of a reference xy stage, is obtained by the zero crossing time and the constant *k* which is the speed of the linear stages.

In the preliminary experiment, we monitored the PSD outputs, detected the zero crossing, and obtained the robot position. The error include in x_2 is the maximum, but is random noise. We, therefore, obtain the average of the zero cross time and neglect the effect of the random noise. The accuracy of the position measurement was smaller than 0.01 mm.

We will extend the measurement area and evaluate the dynamics of the proposed measurement system in the future.

8 ACKNOWLEDGMENTS

This research was partially supported Grant-in-Aid for Scientific Research (C), 20560245, 2008.

9 REFERENCES

[1] Aoyama, H., Nakano, H., Hayakawa, N., Sasaki, A., Shimokohbe, A., 1993, Miniature Robots for Ultra Precision Measurement and Machining, Proc. of the 7th Int. Precision Engineering Seminar, 275 - 289.

[2] Torii, A., Kawane, Y., Doki, K., Ueda, A., 2003, Position and Angle Measurement of a Miniature Robot by using Position Sensitive Detectors, IEEJ Trans. IES, 123/12, 2152 - 2159 (in Japanese)

[3] Torii, A., Inoue, T., Doki, K., Ueda, A., 2005, A Macro/Micro Position Measurement System for a Miniature Robot, IEEJ Trans. EIS, 125/12, 1830 - 1836 (in Japanese)

[4] Torii, A., Kato, H., Ueda, A., 2000, A Miniature Actuator with Electromagnetic Elements, T. IEE Japan, 120C/1, 40 - 45 (in Japanese)

Transport Equipment Positioning System Using Accelerometer and PHS

Yasuhiro Kawahara, Hiroshi Yoshida and Hiroshi Hosaka

Dept. of Human and Engineered Environment Studies, Graduate School of Frontier Sciences, the University of Tokyo, Tokyo, Japan

Abstract

An algorithm to distinguish moving/stationary status of distribution equipment based upon vibrations is developed to realize a low-cost and power-saving positioning system, using PHS. It gets position data measured by PHS once or twice a day, measures vibrations continuously, updates maximum and minimum acceleration for days when the equipment is moving or is stationary, and sets the threshold of acceleration to their average value. With this algorithm, it becomes possible to distinguish whether equipment is moving irrespective of the measurement error of the PHS, vibration characteristics of the distribution machinery, or the installation position of the accelerometer.

Keywords:

Logistics; PHS; Positioning; Vibration

1 INTRODUCTION

Positioning systems using the personal handyphone system (PHS) are used in logistics, since the system is easily installed. The users can track PHS terminals attached to transport equipment inside buildings. Also, the terminals require less electricity [1]. Today, logistics applications use more than 4000 PHS terminals with PHS positioning systems. Figure 1 shows an outline of this positioning system. Since the main cost in using this system are communication charges, it is important to reduce the number of required position searches. Logistics applications require a battery life of more than half a year, thus, reducing the search time is important. Typically, users usually search for positions once or twice a month, however, some transport equipment, such as pallets, containers, and chassis, which do not move for long periods of the year, need not to be searched during these periods.

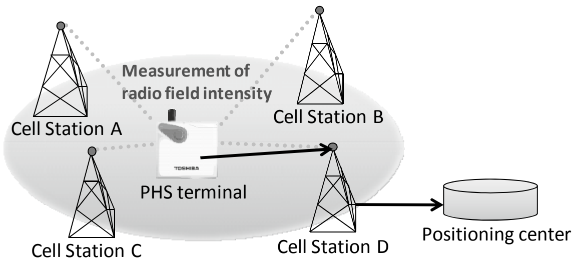


Figure 1: Positioning system using PHS.

This paper discusses a method for detecting the vibrations of transport equipment using an attached accelerometer as accelerometers consume less electricity. The continuous measurement of the level of acceleration of the transport equipment determines whether it is moving. The measurement of the acceleration of transport equipment is described, as well as a method for setting a threshold for determining whether the equipment is moving—the validity of this method is verified experimentally. The threshold setting method contains a learning process using position data located by PHS, as well as the maximum and minimum values of acceleration recorded every day. This method can determine whether transport equipment is moving, regardless of its vibration characteristics and how the accelerometer is attached.

2 ACCELERATION MEASUREMENT ON A CAR

Speed and acceleration were measured using an accelerometer and a GPS set on a car seat to evaluate whether it is possible to use an accelerometer to determine whether the transport equipment is in motion. For this measurement, an acceleration logger was developed (Figure 2). The logger records two axis accelerations at a frequency of 640 Hz within a range of $-9.8\ \mathrm{m/s^2}$ to $9.8\ \mathrm{m/s^2}$ for up to a week. The recording media used was a compact flash memory card.

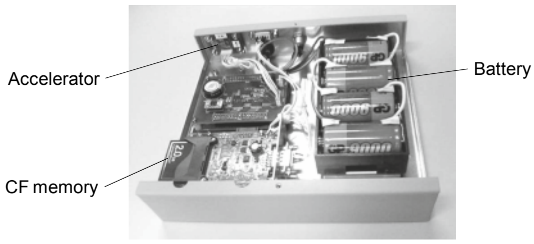


Figure 2: Acceleration logger.

Figure 3 shows the acceleration of a car in its traveling direction. The acceleration level of the moving car (a) was higher than that of the stationary car (b). Thus, an accelerometer is capable of determining whether the transport equipment is moving.

3 ALGORITHM FOR JUDGING MOVING STATUS OF TRANSPORT EQUIPMENT

In practical measurements, it is difficult to set a threshold for determining whether the equipment is in motion. The

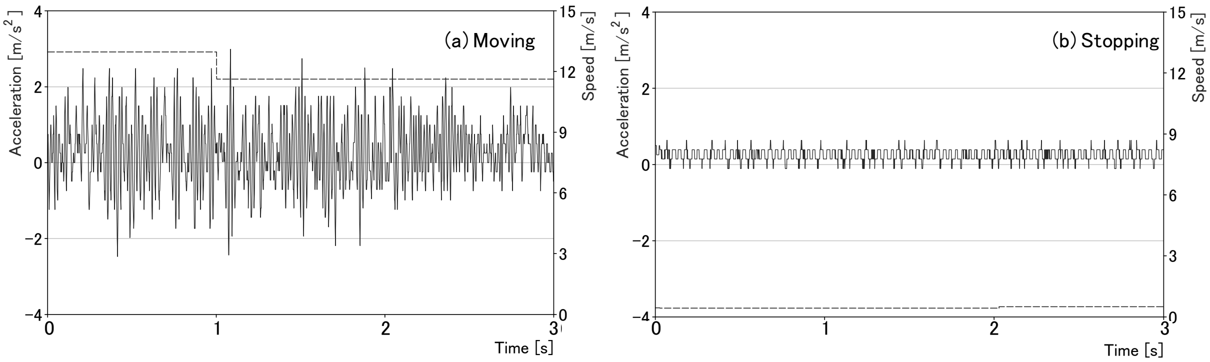


Figure 3: Acceleration of a car (direction of travel).

acceleration level of moving equipment depends on the equipment type and the direction of motion; the acceleration level of moving equipment and the noise level of stationary equipment depends on the stiffness of the material and the location of the accelerometer. Therefore, a threshold needs to be set after observing the status of the transport equipment in motion. Large distance movement of the transport equipment is observed using PHS positioning, but as the accuracy of this method is about 100 m, it cannot detect small movements as well as the time the equipment moves, because in practice PHS positioning is done only once or twice a day. In this research, a method to extract the approximate minimum acceleration value of moving equipment and maximum value of stationary equipment was devised. The method is explained using the flowchart shown in Figure 4.

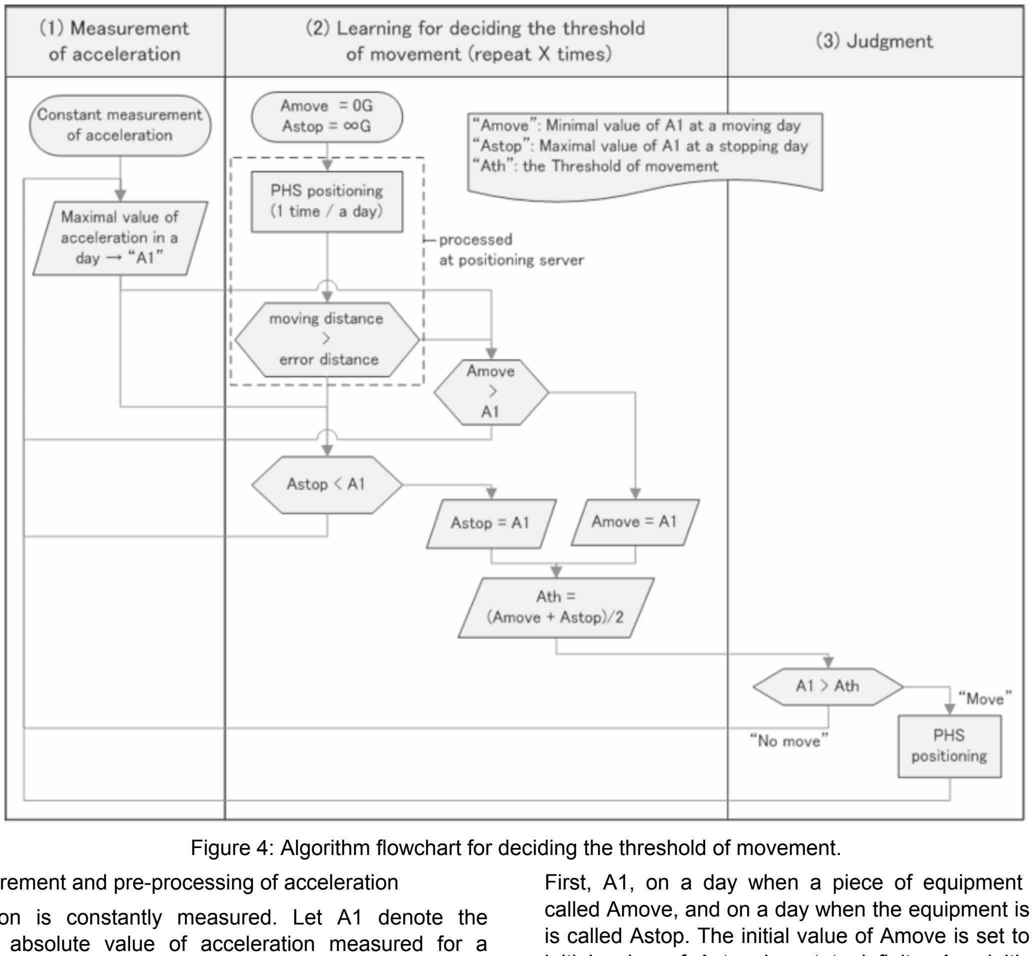


Figure 4: Algorithm flowchart for deciding the threshold of movement.

(1) Measurement and pre-processing of acceleration

Acceleration is constantly measured. Let A1 denote the maximum absolute value of acceleration measured for a period from one PHS positioning to the next PHS positioning.

(2) Method for setting threshold

First, A1, on a day when a piece of equipment moves, is called Amove, and on a day when the equipment is stationary is called Astop. The initial value of Amove is set to 0 and the initial value of Astop is set to infinity. Any initial value of Amove is valid, as long as it is higher than the estimated value of A1; any initial value of Astop is valid as long as it is lower than the estimated value of A1. Next, the position of the transport equipment is located by PHS once a day; its status is judged in terms of whether it has been moved within that day or not. The status is judged "moved" if the distance between two consecutive positions measured by PHS is greater than the positioning error distance of the PHS positioning method; it is judged as "stationary" if the distance between two consecutive measured positions is shorter than the error distance of the PHS positioning method. The PHS positioning error distance is calculated with a curve describing the error distance in relation to the distance between PHS cell stations (a curve including 99% data). The curve was drawn based on actual measured data (Fig. 5). Next, Amove and Astop are refreshed. Amove is set to the minimum value of A1 for days when the transport equipment moves; thus, Amove is set to judge whether the transport equipment moves if the value of A1 is higher than that of Amove. Specifically, the value of Amove is refreshed to that of new A1 when the value of the new A1 is higher than the recorded value of Amove, and it is not refreshed when the value of the new A1 is not higher. Astop is set to the maximum value of A1 on days when the transport equipment is stationary; Astop is set to judge that the equipment is stationary if the value of A1 is lower than that of Astop. Astop is refreshed to the value of A1 when the value of the new A1 is lower than the recorded value of Astop, and it is not refreshed when the value of the new A1 is not lower. Finally, the threshold of the acceleration level, Ath, for judging the motion status of the transport equipment is set. The threshold is set to the average of the values of Amove and Astop.

Astop exactly expresses the maximum value of the acceleration level when the transport equipment is stationary, and Amove expresses the minimum value among maximum acceleration levels for days when the equipment is moved. This algorithm is designed in such a manner, because the transport equipment is always stationary during the stationary period, but during the moving period, it is both in motion as well as stationary. This algorithm judges the motion status of transport equipment only at daily intervals, but the interval is adequate for practical use in logistics.

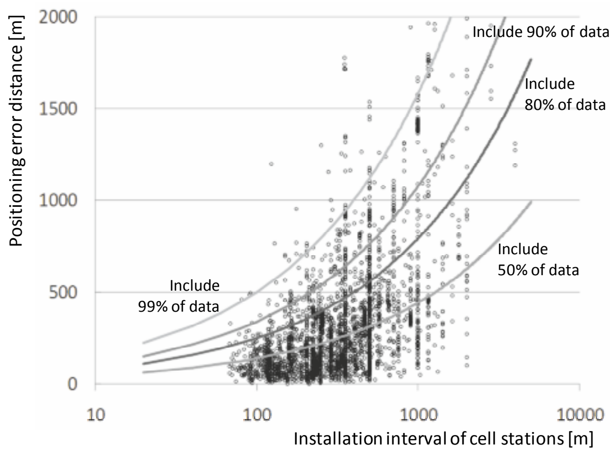


Figure 5: Relation of PHS positioning error distance to cell stations' interval.

(3) Judgment of transport equipment using the threshold of acceleration level.

After setting the threshold, the motion status of the equipment can be determined using only the acceleration level. The transport equipment is judged as moving if A1 is larger than Ath and stationary if it is lower. If judged as moving, a PHS terminal is powered on, the position of the transport equipment is located by PHS, and the located position is displayed. If judged as stationary, positioning by PHS is not carried out.

Figure 6 shows an example of setting the threshold. The acceleration data are not actual measured data, but assumed data. On day 1, the transport equipment is stationary and the maximum absolute value of acceleration is A1(1). Since the initial value of Astop is infinity, the condition Astop > A1(1) applies, and the value of Astop is refreshed to that of A1(1). On day 2, the transport equipment moves and the maximum absolute value of acceleration is A1(2). Since the initial value of Amove is 0, the condition Amove < A1(2) applies, and the value of Amove is refreshed to that of A1(2). On day 3, the equipment is stationary and the maximum acceleration absolute value is A1(3). Because of the relationship A1(1) < A1(3), Astop < A1(3) applies, and the value of Astop is again refreshed to that of A1(3). On day 4, the equipment is stationary and the maximum absolute value of acceleration is A1(4). Because of the relationship A1(2) > A1(4), Amove > A1(4) applies, and the value of Amove is again refreshed to that of A1(4). The values of Amove and Astop are refreshed in this way and the threshold Ath is expressed as the average of these two values. The above sequence is repeated N times where N is the number of elapsed days during the time when equipment-moving days are counted.

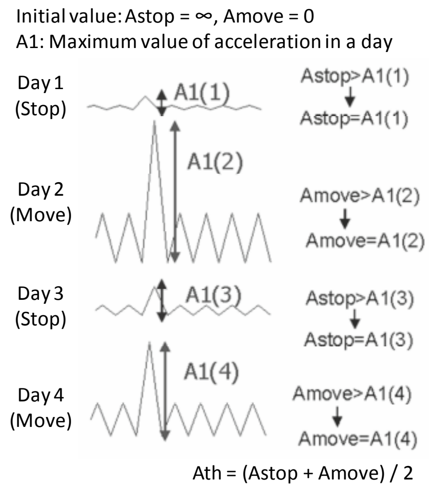


Figure 6: Conceptual scheme of threshold setting.

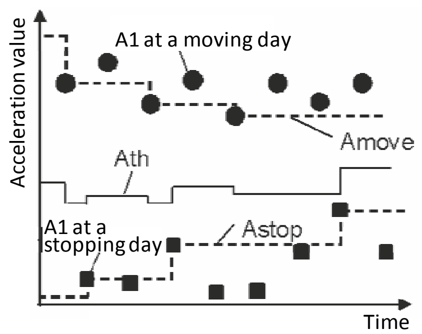


Figure 7: Expected threshold refreshing.

Figure 7 shows a conceptual diagram of the threshold setting. The value of Astop is refreshed every time A1 is higher than Astop on a day when the transport equipment is stationary. The value of Amove is refreshed every time the value of A1 is lower than that of Amove on a day when the equipment moves. As a result, the value of Astop increases monotonously and the value of Amove decreases monotonously; thus, the two values approximate gradually. The value of Ath is the average of the two values (Astop and Amove) and its magnitude shows small oscillations after several days.

4 APPLICATION OF THE ALGORITHM

4.1 Sampling of acceleration

Transport equipment was moved and stopped, based on the assumptions about its practical use. Its acceleration was measured by the attached accelerometer under several conditions of attachment. The output voltage of the dual-axis accelerometer (ADXL202E from Analog Devices, Inc.) was recorded with a data logger (ZE-DMR10 from Omron Corporation) at a sampling interval of 1 kHz. The following four cases represent differing combinations of attaching the accelerometer to transport equipment. In each case, accelerations along the travelling direction and the vertical direction of the equipment are measured. The four cases are: (1) the accelerometer is attached directly to the car floor, (2) the accelerometer is on a cushion attached to the car floor, (3) the accelerometer is attached directly to a dolly and, (4) the accelerometer is on a cushion attached to a dolly.

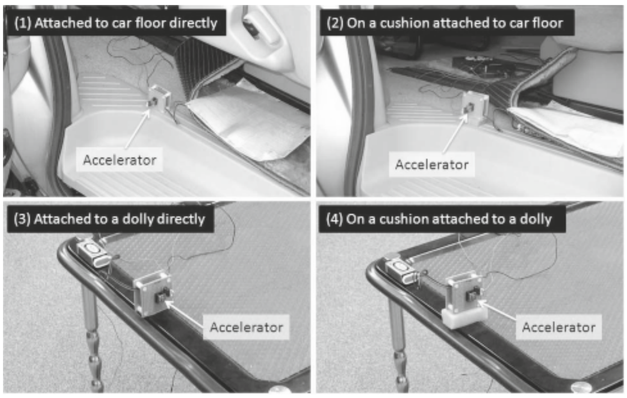


Figure 8: Conditions of accelerator attachment.

Figure 8 shows pictures of these four conditions.

The results of acceleration analysis by fast Fourier transform for transport equipment motion for cases (1) and (3) are shown in Figure 9.

The maximum acceleration of moving equipment measured in this experiment was 10 m/s^2 and its average acceleration was 5 m/s^2. Considering the maximum acceleration on the floor of a container being carried by rail is 80 m/s^2 and its average is 15 m/s^2 (for example) [2], the measurements in this experiment are taken under the condition that vibration level of the moving equipment is less than in an actual environment. It is assumed that the samples in this experiment are measured under conditions where the status of the equipment is more difficult to detect.

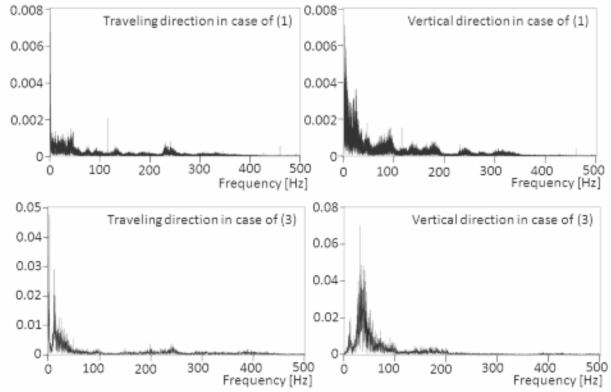


Figure 9: Analysis by fast Fourier transform.

4.2 APPLICATION OF THE ALGORITHM

The threshold setting process using measured samples from the previous section was confirmed.

Transport equipments with attached accelerometers were moved intermittently for 30 minutes; the accelerations along their traveling direction and vertical direction were recorded, and the absolute values of acceleration were calculated. This experiment was carried out on the assumption that the PHS positioning interval is 3 minutes (one day in actual use) and the values of Astop and Amove are refreshed every 3 minutes. The days when equipment moved were distinguished manually from the days when they did not move. Figure 10a, 10b shows the process of setting the threshold for each case.

Figure 10a shows the absolute value of acceleration along the travelling direction for case (1). The measurement time is divided into consecutive three minute spans; the shaded parts represent the times when the equipment was stationary and the other parts were the times when it moved. The initial values of Astop and Amove were set to 0 m/s^2 and 20 m/s^2, respectively; the initial value of Ath was 10 m/s^2. The car moved in the first three minutes; the maximum absolute value of acceleration A1 in this period was 2.24 m/s^2. The circles in Figure10 indicate points when A1 is generated. The value of Amove was refreshed to 2.24 m/s^2 and Ath was refreshed to 1.13 m/s^2 after 3 minutes. In the period between 3 and 6 minutes, the car was stationary and the value of A1 was 0.13 m/s^2; thus, Astop was refreshed when 6 minutes elapsed. In the period between 6 and 9 minutes, the car moved and A1 was 2.50 m/s^2; thus, Amove was not refreshed, since the value of A1 was higher than that of Amove. In the period between 9 and 12 minutes, the car was stationary and Astop was not refreshed, since the value of A1 was lower than that of Astop. In the period between 12 and 15 minutes, the car moved and Amove was refreshed, since the value of A1 (2.05 m/s^2) was lower than that of Amove (2.24 m/s^2). In the periods between 15 and 18 , 21 and 24 , 27 and 30 , and between 33 and 36 minutes, the car moved, and A1 was 2.35, 2.21, 1.80, and 2.17 m/s^2, respectively; thus, Amove was refreshed by A1 between 27 and 30 minutes and when 30 minutes had elapsed. In the periods between 18 and 21, 24 and 27, and 30 and 33 minutes, the car was stationary, and A1 was 0.14, 0.09, and 0.09 m/s^2, respectively; thus, Astop was refreshed by A1 in the period between 18 and 21 minutes and when 21 minutes had elapsed. Ath changed until 30 minutes elapsed, but changed little after 6 minutes elapsed; thus, the effective repeat count N mentioned above was 1.

Figure 10b shows the acceleration level along the vertical direction for case (1). In the period between 0 and 3 minutes, the car moved and Amove was refreshed, since the value of A1 (5.08 m/s^2) was less than that of Amove; thus, Ath was refreshed to 2.54 m/s^2 when 3 minutes had elapsed. In the period between 3 and 6 minutes, the car was stationary and Astop was refreshed, since the value of A1 (0.18 m/s^2) was higher than that of Astop; Ath was refreshed to 2.63 m/s^2 after 6 minutes had elapsed. In the periods between 6 and 9, 15 and 18, 21 and 24, 27 and 30, and 33 and 36 minutes, the car moved and A1 was 2.73, 2.77, 2.07, 2.88, and 1.41 m/s^2, respectively; thus, Amove was refreshed after 9, 24, and 36 minutes. In the periods between 9 and 12, 18 and 21, 24 and 27, and 30 and 33 minutes, the car did not move and A1 was 0.09, 0.23, 0.10, and 0.10 m/s^2, respectively; thus, Astop was refreshed after 21 minutes elapsed. Ath's change lasted until the final period, but Ath changed little after 12 minutes; thus, the effective repeat count N was 2.

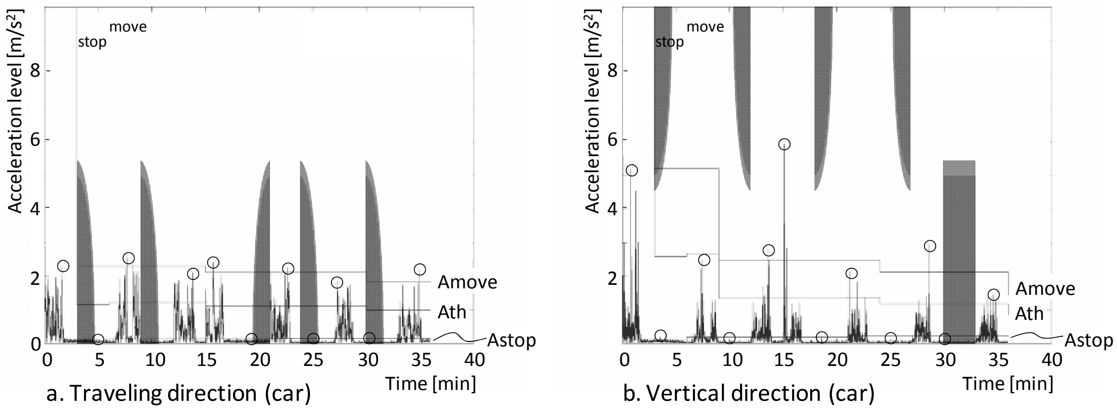


Figure 10: The result of the algorithm application.

On the acceleration level along the travelling direction for case (2), Ath was not refreshed after 12 minutes; thus, the repeat count N was 2. On the acceleration level along the vertical direction for case (2), Ath was refreshed after 21 minutes, but the value of Ath afterwards was sufficiently lower than the value of Amove before the time that the repeat count N was 2.

On the acceleration level along the travelling direction for case (3), Ath was not refreshed after 15 minutes elapsed; thus, the repeat count N was 3. On the acceleration level along the vertical direction for case (3), Ath changed little after 9 minutes; thus, the repeat count N was 1.

On the acceleration level along the travelling direction for case (4), Ath changed little after 12 minutes had elapsed; thus, the repeat count N was 2. On the acceleration level along the vertical direction for case (4), Ath changed little after 9 minutes elapsed; thus, the repeat count N was 1.

As described above, the threshold becomes stable within 3 days of the equipment moving. When a piece of equipment moves every other day, the learning process with this algorithm finishes in 6 days. After the threshold is set, the time when equipment starts to move can be detected—PHS positioning is carried out at that time, since the status is judged constantly. In this experiment, the thresholds were set with two accelerometer directions, with two attachment conditions, and with two kinds of transport equipment. The experiment was carried out under limited conditions, but transport equipments with low vibration levels were used and the algorithm was applied under conditions where equipment movement was more difficult to detect. As a result, the threshold setting is thought to be similar to that which would be required for other transport equipment.

5 DEVELOPMENT OF A TERMINAL FOR JUDGING EQUIPMENT MOVEMENT STATUS

5.1 Configuration of the terminal

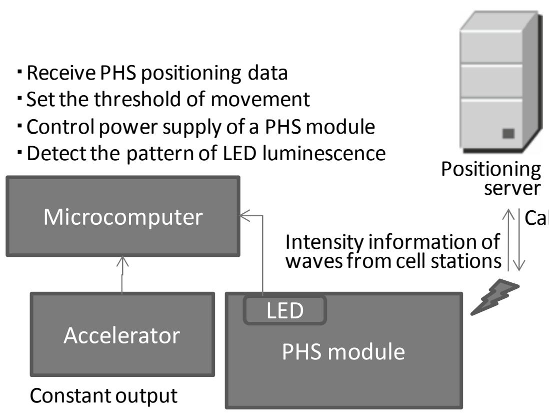


Figure 11: The composition of a terminal for judging the movement status of equipment.

A PHS terminal, including an accelerometer, was developed to set a threshold for judging whether equipment is moving, using the algorithm explained in the previous section. The terminal locates the equipment position in conjunction with the PHS positioning server when the judgment result is "moving". Figure 11 shows the terminal configuration.

The terminal consists of a dual-axis accelerometer (ADXL202 from Analog Devices, Inc.), a microcomputer (PIC-16F873 from Microchip Technology Inc.), a PHS module (LC102 from Toshiba Corporation) and four AA alkaline batteries.

The PHS module receives a positioning request from the remote positioning server, sends measured strength of the wave from PHS cell stations surrounding it to the positioning server, and receives its status of movement, determined by the PHS positioning. Its LED lights show a specific pattern when it receives the status information “moving”. The microcomputer detects the LED light pattern, thus, the microcomputer knows the movement status during the learning process for the threshold setting algorithm.

The microcomputer receives the output voltage of the accelerometer, calculates the threshold setting for the equipment movement status, judges the moving status accordingly, and controls the power of the PHS module. After a day when a piece of equipment moves three times, it completes the learning process in the algorithm and sets the threshold for judging the equipment’s movement status. After the learning process is finished, the equipment is judged to be “moving” when the absolute value of the acceleration, continuously sampled and processed by the microcomputer, becomes higher than the threshold. The PHS power is then turned on, and the terminal starts to locate the position in conjunction with the PHS positioning server.

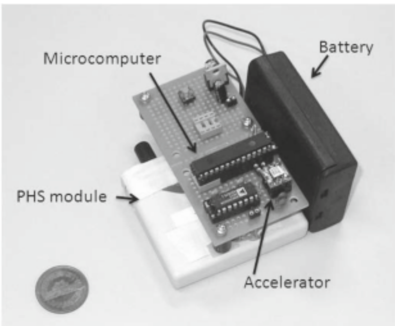


Figure 12: A picture of the movement judgment terminal.

Figure 12 shows a picture of the terminal. It is 85×90 ×48 mm and weighs 110g.

5.2 Operation of the terminal

The developed terminal was set on the floor of a car and its operations were checked. In this experiment, the repeat count in the algorithm N was set to 2 and, during the algorithm learning process, daily PHS positioning was triggered manually via the remote positioning server. The car moved every other day. After the fifth day, the PHS module was powered on and PHS positioning was carried out on the remote positioning server at the time when the car moved.

5.3 Applications for management of transfer equipment

The terminal developed in this study connects to the remote positioning server and the server calculates the terminal’s position when the absolute value of measured acceleration exceeds the threshold set in the algorithm’s learning process. To reduce communication costs, the timing of communication to the positioning server needs to be determined to accommodate the purposes of the specific logistical management application.

For management applications for confirming the positions within depots where transport equipment, such as containers and chassis stay, they should be positioned by PHS within half a day after they start to move because they are expected to stop within half a day. For position management applications for transport equipment that does do not move for several months, the communication cost should be reduced drastically.

Transport equipment, such as construction equipment and heavy machinery, are often stolen [3]. Through constant tracking and providing alerts when they are judged to be “moving”, their paths can be traced. .

6 SUMMARY

An algorithm for judging whether transport equipment is moving, by detecting their vibrations, was devised to reduce communication costs and terminal power consumption. The algorithm judges whether a transport equipment is moving using position information provided by the PHS positioning server daily—it refreshes the maximum absolute acceleration value (when equipment is stationary) or the minimum absolute acceleration value (when moving), and then sets a threshold to the average of these two values. This study shows that learning acceleration levels of moving equipment for three days provides sufficient data for judging movement status (moving or stationary) for several kinds of transport equipment, without depending on the vibrational characteristics of the equipment and conditions of the attached accelerometer.

7 ACKNOWLEDGMENTS

This research is supported by JST/CREST.

8 REFERENCES

[1] Kawahara, Y., Sawa, Y., Matsubara, R., Sakata, K., Hirota, T., 2005, Pallet Location Finding System for Physical Distribution Using PHS Terminal, Micromechanics, 49/192:12-21 (in Japanese).

[2] Nakajima, S., 2007, The Influence of Changes in the Transport Environment on Cargos, Material Handling Journal, 250:10-14 (in Japanese).

[3] Salzano, J., Hartman, S., 1997, Cargo Crime, Transnational Organized Crime, 3/1:39-49.

A Four-legged Robot Connected to Ubiquitous Devices in a Home

Xiaofeng Wu[1], Florent Servillat[1], Shinji Yamazaki[1], Tsuyoshi Taki[1], Ryohei Nakatsu[1, 4]
Masafumi Okajima[2], Tadashi Enomoto[2],
Satoshi Kagami[3], Simon Thompson[3], Yoshifumi Nishida[3]

[1] Nirvana Technology, Kyoto, Japan
[2] Kansai Electric Power Co., Hyogo, Japan
[3] Digital Human Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan
[4] Kwansei Gakuin University, Hyogo, Japan

Abstract
This paper describes basic research conducted of a robotic system that can support people in general domestic surroundings. The integration of robot with facilities for personal interaction are essentially important for robots. However it is technically and physically difficult to achieve those functions by using standalone robot. Proposed here is a way of accomplishing this using a system that combines ubiquitous microphone arrays with ultrasonic locators installed in the room ceiling. This paper describes a ubiquitous device, an autonomous locomotive function via cooperation with a robot, and the experimental results. The efficacy of the system is then clarified.

Keywords: Service robot, Security, Sensing systems

1. INTRODUCTION

It is expected that service robots designed to support our everyday lives will be developed and introduced into society. For this purpose, basic research was begun here to develop a robot system that can provide this support.

A system was first proposed and developed consisting of two subsystems: a robot with wheel-based locomotion capability and a ubiquitous sensor network composed of ubiquitous ceiling ultrasonic locators and microphone arrays [1]. A ubiquitous sensor network mounted within a house can provide information to facilitate human–robot interaction. The system's main feature is that the robot's processing load is reduced by distributing its intelligent processing to both the robot and the environment.

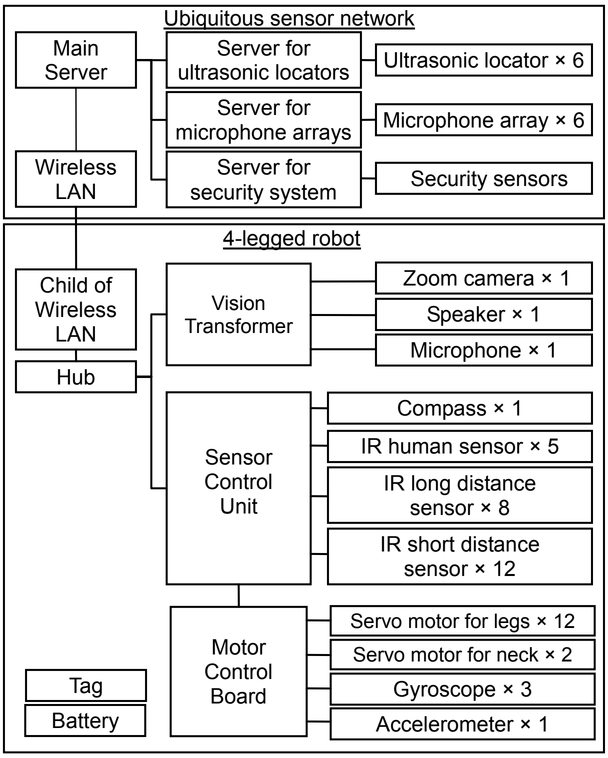


Figure 1: System diagram

After evaluating this first prototype, a second prototype was developed (Figure 1). Because the perception and planning parts of the intelligent processing could be shifted to the environment side, it was expected that robot hardware components would become smaller and lighter. Low power consumption was also expected. Maximizing these merits, a four-legged robot was introduced that could function as a moving intelligent sensor in a complete robot-environment system.

A four-legged robot was introduced primarily because of its friendly appearance; people would not likely feel great difficulty in interacting with it. Another reason is the four-legged robot's capability of traversing imperfectly flat floor in a house. It was expected that the robot could achieve the following scenarios.

(1)If a family member calls the robot, it moves to the room where he/she is and identifies him/her. Then the interaction begins.

(2)The robot can collaborate with a home security system and achieve highly reliable home security system. For example, once the security system sends an alarm to the robot due to something like a unexpected door opening, the robot hurries to the location of the alarm, tries to capture an image, and then sends the image to family members by e-mails.

To realize these scenarios, it is indispensable that the robot can move robustly to the target location. This paper describes a ubiquitous device, the autonomous locomotive function featured in a four-legged robot, and its experimental results.

2. Ubiquitous Devices in the Home

2.1 Ubiquitous ceiling ultrasonic locators

An ultrasonic tag system was developed in order to measure the three-dimensional tag locations. The system consists of the following three parts: a synchronous unit, which receives radio signals synchronously from tags mounted on the robot; a receiving line installed in the ceiling; and a tag(Figure 2).

A synchronous unit transmits a tag's ID via 315 MHz carrier wave, and the tag sends a 40 kHz ultrasound if its unique ID is received. A receiving line measures the time the ultrasound takes to reach the tag, and then transmits the time to the

server PC via RS485 bus.
The server PC computes the three-dimensional location of the tags statistically from the arrival time to each receive unit.

Figure 2: (Left) Ultrasonic tag. (Right, outer ring) 32-channel ceiling microphone array, and (right, inner cone) 15-channel ultrasonic receivers

2.2 Ubiquitous ceiling microphone array

This 16-bit system is laid out horizontally on the ceiling using an array of disks each 540 mm in diameter. A 16 kHz sound-source acquires data through 360-degree sampling, with a wave angle oriented perpendicularly 0 degrees horizontally to the array side and performed across a range of 0–90 degrees. This hemispherical surface is split into 160 cells. The sound pressure of each cell orientation is determined, and four cells are detected in order to calculate the optimal direction.
Using the DSBF (delayed-sum beam forming) and FBS (frequency band selection) methods, the sound-source location was searched and the determined sound separated.

After sound sequences from sources with augmented sound pressure ratio, the system transfer those sound sequences to a sound recognition system based on a Julian. Sound recognition system contains about 30 words to command destination location, user and motion of the robot.

2.3 Experimental house "Holone"

A ceiling sensor setup comprised an ultrasonic receiving line of the above-mentioned 15 channels in a core, with a microphone array in a side part. The unit was installed in six locations in the experimental house "Holone" (bedroom, living room, dining room, kitchen, a door, and study room, totaling approx. 106 m^2) (Figure 3).

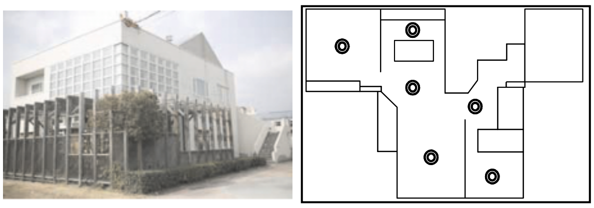


Figure 3: Map of Experimental House "Holone" and sensor locations

3. Four-legged Robot

3.1 Structure

The four-legged robot in Figure 4 was developed for the experiment. To aim a future-commercial use, a robot was to be a mounted cheap and lightweight general-purpose devices.

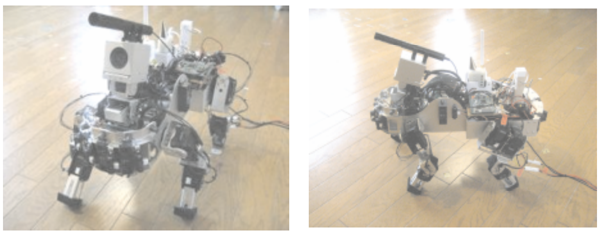


Figure 4: Two-directional views of the four-legged robot

Three control units are used for the control of sensors and motors. The block diagram of these controllers is illustrated in Figure 5. Also the specification of the whole robot hardware is shown in Table 1.

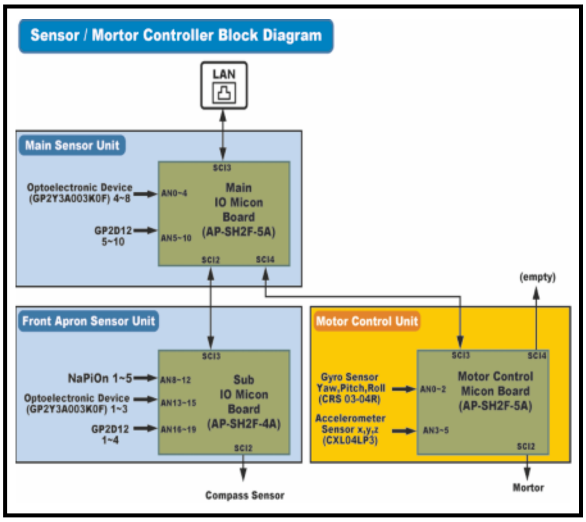


Figure 5: Block diagram of sensor/motor control units

Table 1: Specification of the robot

Item	Specification	Number
Controller 1 (Sensor and locomotion control)	AP-SH2F-５A SH7047 49MHz 85×60 mm	2
Controller 2 (Sensor control)	AP-SH2F-４A SH7046 49MHz 70x50mm	1
Zoom camera	SONY: FCB-IX11A	1
IR human sensor	Matsushita: NaPion	5
IR long distance sensor	Sharp: GP2Y3A003K0F	8
IR short distance sensor	Sharp: GP2D12	12
Compass	Hitachi: HM55B	1
Actuator	ROBOTIS: RX-64 ROBOTIS: DX-117	12 2
Gyroscope	Silicon sensing system: CRS03-04	3
Accelerometer	CXL04LP3	1

3.2 Localization

Here the aim is to develop an algorithm of detecting the position of the robot using the output of several sensors. As the basic algorithm for the detection of robot position taking sensor data and robot locomotion into consideration, the Monte-Carlo Localization method was adopted [5]. The algorithm has been developed in order to manage the different types of sensors efficiently. In our first algorithm, the generic part of the particle filter was in charge to decide whether or not it accepts or rejects the sensor as an incorrect data, based only on the posterior distribution of the robot pose. If the value was too low, the algorithm decided to reject the update. Afterward, given the number of rejected sensors, the robot was making the decision to continue to use this distribution or decided that it was lost. The algorithm improvement consisted in moving this decision from the generic part of the algorithm

to the sensor model. As a consequence, the algorithm updates the distribution of the robot pose update of the particle weight for each sensor and doesn't reject any sensor. When the sum of the particle weight after the update is close to zero, the robot decides that it is lost.
The Monte-Carlo Localization is a recursive Bayes filter that estimates the posterior distribution of the robot pose variable, called X given the sensor data [5]. The distribution of robot pose X at the time t, is called the belief and is related to the current data.

Bel(Xt) = P(Xt | d0…t)

When applied to the localization of a mobile robot, the data can be divided into 2 types.

Actions: the movement the robot performed $a_0 \ldots a_n$
Observations: measurements provided by the sensors $o_0 \ldots o_n$

Assuming that actions and observations are successive events, that the data are independents from the previous status, the belief at time t is desctibed as:

Bel(Xt) = n * P(Ot|Xt)* ∫P(Xt|Xt-1, at-1) * Bel(Xt-1) * dXt-1

Where n = $P(Ot|\ at\text{-}1,\ldots o0)^{-1}$ is the normalization constant

Since the model is time invariant and actions are described in discrete form (a walking step unit), we can express the update by this formula:

Bel(Xt) = n * P(O|X)* ∫P(X'|X, a) * Bel(X) * dX

In this formula, P(O|X) is the perceptual model which represents the probability of the observation given the pose. We have several observations for each localization, thus, the expression of P(O|X) will be:

$P(O|X) = \prod_i P(O_i|X)$, where $P(O_i|X)$ is the probability for one observation.

The part P(X'|X, a) is the motion model which describes the probability of the next state density given current position and action.

The algorithm updates its position in two steps. First, the belief of current pose is updated to the next pose given the action performed by the robot. Then, the probability distribution function is updated using the sensor data. To express the continuous probability density function for the robot pose is topical issue in mobile robotics. The representation will use a sampling-based approach, called particle filter.The continuous distribution is replaced by a set of particles, each particle being a copy of the robot pose with a weight (X, W). The weight represents the importance of the particle inside the set. Using a large number of particles, the set is a good approximation of the pdf.
To estimate the pose given the particles, we will use the weighted mean:

$X_{Est} = \sum_i(w_i * X_i)$

In the four-legged robot project, the robot pose X is expressed using 3 components $[x, y, \theta]^t$

x, y : the robot position in the world coordinates
θ: the robot orientation

The observations are performed using several distance sensors, an ultrasonic locator and a compass. The model is provided in the following part. The movement the robot performed is provided by the robot status, and since the robot is only able to move of 1 step unit, the movement is described as an action type and the number of steps. The robot uses 2 movement patterns: a rotation on its center and a curved walk movement.

This algorithm will cause the particle filter to decide that it is lost, if the noise on sensors is too large. We must filter the data before the localization to prevent the localization from failing. Furthermore, for each sensor type we will define the most suitable model.

3.3 Sensor mode optimizations

The evaluation of the posterior distribution of the robot pose is based on the evaluation of the deviation between the expected sensor value, given the position for each particle, and the current sensor value. If the particle position is a good estimate of the current robot pose, the error should be small. However, if there is a large deviation between expected sensor value and the current value, we assume that the particle pose is not a good estimate of the posterior distribution, and thus the weight of the particle will decrease. The evaluation of P(Obs | X) will vary given the sensor type.

Example: Distance sensor

To represent the probability P(Obs|X) for the distance sensor, we took into account several parameters:
The valid range of the sensor: Outside the valid range of the sensor, the value is not used, we return the same probability for each evaluation.
The noise on the sensor value: We assume that we have a white noise, independent from the distance measurement. Since the error is a white noise, the error is modelled using a Gaussian centred on expected value which standard deviation (σs) is fixed. The evaluation is performed by this formula:

P = 1/(sqrt(2*pi) *σs) * exp(-(expectedVal-sensorDistance)^2 /2σs^2)

The sensor can intersect obstacles (the evaluation of the noise will fail): In that case, the previous probability evaluation will be close to 0. To prevent the sensor probability to come close to 0, we represent the obstacle as a uniform probability distribution (PLim) which is fixed. When the previous evaluation is lower than the fixed obstacle probability Plim, we will use this one instead of the probability P.

3.4 Movement model

The movement model is used to update the state density after an action. The robot platform provides the current number of steps the robot performed during the action. As a consequence, this variable is discrete and let us knows only when the step has been performed. There are 2 types of movement available in the localization as is shown in Figure 6.

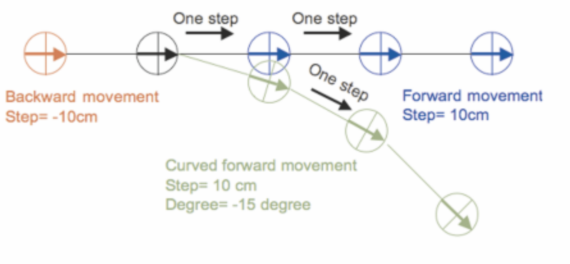


Figure 6: The translation model

The translation model consists in a forward or backward movement along the robot YR axis. The distance parameter describes the distance the robot moved within one step. The additional drift parameter describes the number of degree the orientation will be modified during one step. Using the drift parameter, the robot can approximate curved trajectory.
The robot is acting in the real world. As a consequence, the robot movement will be modified by several parameters such as the slippery on the floor or the motor errors. Thus, the robot won't move exactly the same distance from one step to the other and won't keep the same orientation . It is necessary to represent the error on distance and drift in the model.
Both errors will be assumed as a white noise centred on the average value. Thus, we will represent the translation using the average distance the robot will move during one step and the standard deviation on this distance. In the same way, the average for the drift will be the number of degrees the orientation varies during one step.

translationMean, translationStDev
driftMean, driftStDev

We succeed to represent both forward and curved movement with the same description. Using this description we defined 4 different actions the robot can use:

Straight forward movement
Backward movement
Curved forward movement on the left
Curved forward movement on the right

3.5 The rotation model

The rotation model consists in a modification of the robot orientation without modification of robot position. However, the robot is unable to perform a perfect rotation on its axis, due to the slippery of the floor and motor errors. As a consequence, the deviation on the position must be inserted in robot model. We will represent this error as a white noise on the position.
The parameters for the rotation will be:
rotationMean, rotationStDev
distStDev

Given the estimated pose $[x_k, Y_k, \theta_k]$ at time k, the update of robot pose will be:
$\theta_{k+1} = \theta_k$ + randomNormal(rotationMean, rotationStDev)
distDev = randomNormal(0, distStDev)
orientDev = randomUniform(0, PI)
$x_{k+1} = x_k + \cos(\text{orientDev}) * \text{distDev}$
$y_{k+1} = y_k + \sin(\text{orientDev}) * \text{distDev}$

The formula will generate a noise centered on robot position with a random orientation.
Using this description, we defined 4 rotate actions:

Small step rotation on the right
Large step rotation on the right
Small step rotation on the left
Large step rotation on the left

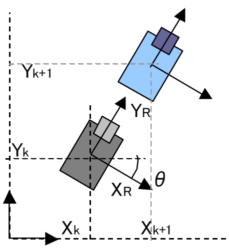


Figure 7: The rotation model

Given the estimated pose $[x_k, Y_k, \theta_k]$ at time k, the update of robot pose will be (Figure7):

$\theta_{k+1} = \theta_k$ + randomNormal (driftMean, driftStDev)
dist = randomNormal (translationMean, translationStDev)
$x_{k+1} = x_k - \sin(\theta_{k+1}) * \text{dist}$
$y_{k+1} = y_k + \cos(\theta_{k+1}) * \text{dist}$
(The translation is along the Y axis of robot coordinates.)

3.6 Navigation

To improve the execution speed, we concentrated our effort on improving the distance evaluation in the map. In order to check the distance between the estimated position of a sensor and the wall, we need to project the segment, defined by its origin, orientation and length, in the map. We start from the origin, and find the exit point for the first cell. Then, we perform the same operations for the next cell recursively, until we find a wall or reach the end of the segment.
To improve the execution speed, we concentrated our effort on improving the function that evaluates the distance between a point and the wall in the map. Since the localization is performed on the static map (the map of the walls), and won't be modified during the execution, we can optimize the structure of the map in order to decrease the number of intersection between a ray and the map.
When the robot controller starts, the map of the wall is converted into a quadtree map. A quadtree map is a tree data structure, in which each node may contain 0 or 4 nodes. Each node represents a region of the map. A quadtree manages a 2 dimensional area of 2^N per 2^N cells (where n is the maximum length of a branch). If all the cells within an area have the same value, the node is a leaf of the tree. Otherwise, the area is divided into 4 regions and the same evaluation is performed on these cells. If the map is larger than the maximum size of a quadtree region (2^N cells), then we split it into several trees and store them in a table.
For each node, we can classify them into 3 types:
A free node: all the cells inside this node are free
An obstacle node: none of the cells inside this node are free.
A mixed node: the node is compose of free cells and obstacles, and thus it include 4 child nodes.
We decided to choose a quadtree structure, because it reduces the necessary information when representing uniform areas. Figure 8shows an example of a quadtee map representing a house where we carried out the evaluation. As we can see in Figure 8, the smallest node represents the case that the node is composed of one cell. The density for these nodes is really small. The only possibility to have a high-density of small cells would happen in the case that we had a succession of occupied and empty areas. The original map is composed of 144*128 cells. We generated the quadtree map using 6 levels of depth for each tree, which means that each side of the quadtree has $2^{6-1} = 32$ cells size. As a consequence, the map is split into a table of 8 per 9 quadtrees.

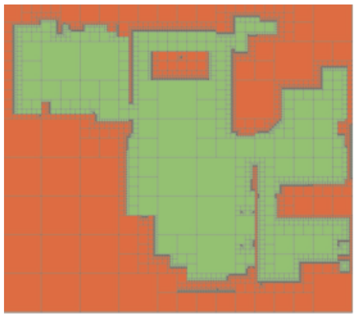


Figure 8: A map of a house represented as quadtree.

4. Experimental Results

4.1 Experimental results of Localization

The effectiveness of the self-positioning estimation was experienced. Manual movement of the four-legged robot was conducted inside the experimental house, and the results of self-positioning estimation using a particle filter inputting ultrasonic positioning data, distance data, and bearing compass data, were compared with ultrasonic positioning data to obtain a true localization values. The results are shown in Table 2 and Figure 9. Error evaluation was performed using the Euclidean distance of the actual locations of each real coordinate and the presumed location.

At corners and near entrances to rooms, the ultrasonically determined error was large, but as Table 2 and Figure 9 show, this can be corrected by a particle filter. Repeatable results are the goal. In these experiments, the mean ultrasonic positioning error was approximately 27.7 cm and average self-positioning estimation error was 19.9 cm.

Table 2: Experimental results of Localization

	real			localization				US device + compass			
time	x	y	Θ	x	y	Θ	deviation	x	y	Θ	deviation
589	540	525	47	537	522	49	4.243	560	551	70	32.80
644	534	527	89	533	526	95	1.414	551	565	105	41.63
1093	534	527	89	526	529	93	8.246	542	557	106	31.05
1181	363	544	81	365	530	87	14.14	NG	NG	89	NG
1627	363	544	81	390	545	83	27.02	NG	NG	86	NG
1707	198	589	75	198	576	80	13.00	224	584	82	26.48
2129	198	589	75	202	575	81	14.56	223	590	82	25.02
2188	198	590	30	203	576	39	14.87	217	595	38	19.65
2458	198	590	30	224	578	40	28.64	223	585	44	25.50
2547	121	738	27	127	733	32	7.810	120	727	37	11.05
2880	121	738	27	135	736	31	14.14	106	737	34	15.03
2949	120	741	74	134	738	78	14.32	87	765	80	40.80
3295	120	741	74	121	740	81	1.414	79	773	80	52.01
3359	14	777	68	8	777	74	6.000	30	773	64	16.49

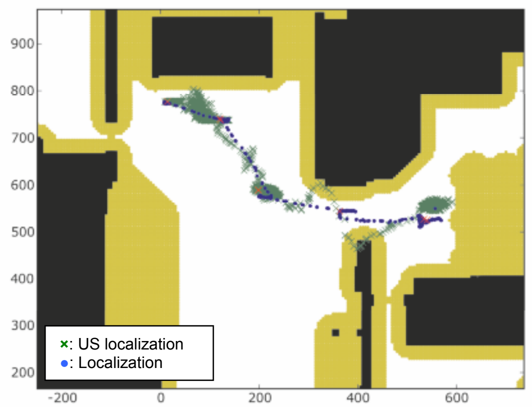


Figure 9: Experimental results of Localization

4.2 Experimental results of Navigation

To check the navigation's effectiveness, verification tests were conducted. A test was set up for the robot to follow a line between the bedroom and the study within 180 s, and performed ten times. The duration times and position errors are shown in Table 3. The mean arrival time was 171 s, for a success rate of 70%.

Table 3: Experimental results of Navigation

	Direction	Duration	Deviation to goal position	Result
1	Bedroom → Study	2 min 21.6 s	9 cm	OK
2	Bedroom → Study	2 min 32.1 s	13 cm	OK
3	Study → Bedroom	2 min 59.6 s	26 cm	OK
4	Bedroom → Study	2 min 59.9 s	9 cm	OK
5	Study → Bedroom	2 min 42.0 s	27 cm	OK
6	Bedroom → Study	2 min 50.5 s	11 cm	OK
7	Study → Bedroom	3 min 26.4 s	12 cm	NG
8	Bedroom → Study	2 min 21.6 s	9 cm	OK
9	Study → Bedroom	3 min 12.1 s	15 cm	NG
10	Study → Bedroom	3 min 5.1 s	28 cm	NG

The successful results of Test 2 are shown in Table 4 and Figure 10. The results show that a self-positioning estimation was performed, reducing error in the ultrasonic positioning from figures, an arc operation was performed, and the robot continues walking smoothly.

Table 4: Experimental results of Navigation

	Start point	End point
Goal	—	600, 160
Real position	−298, 895	587, 160
Error at goal position	—	13 cm
Localization result	−318, 892	589, 152
Error for localization	20 cm	8 cm
Time	13:38:16.5	13:40:58.6
Duration	—	2 min 32.1 s

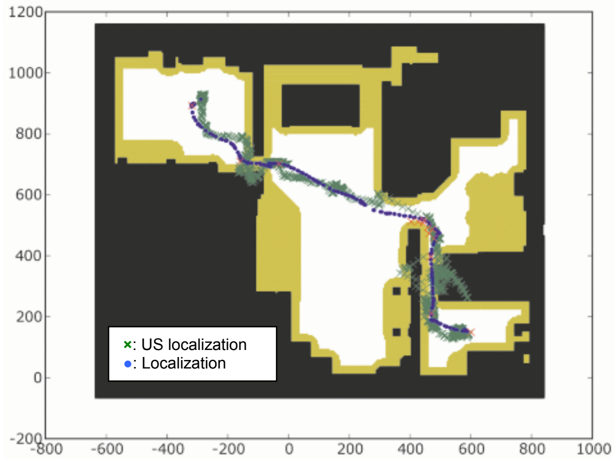


Figure 10: Experimental results of Localization

The successful results in Table 5 and Figure 11 are derived from Test 9, which was not able to perform a movement within the target time. Narrow areas such as the entrance to a room required more time, preventing smooth movement from being performed. In narrow parts, sensors date from ultrasonic locators and distance sensors are not obtained precisely. Therefore, unable to perform a self-positioning estimation correctly but able to acquire a bearing, repeating a turn could thus be set and folded and the robot able to recover. This may await some improvement. However, when location data are not obtained from the ultrasonic positioning system, only distance information is selected to perform continuous self-positioning estimation. In general, the proposed navigation practice for leg-type robots shown here is considered to be effective.

Table 5: Experimental results of Navigation

	Start point	End point
Goal	—	−300, 850
Real position	600, 160	−301, 835
Error at goal position	—	15 cm
Localization result	604, 159	−295, 843
Localization error	4 cm	10 cm
Time	16:15:10.5	16:18:22.6
Duration	—	3 min 12.1 s

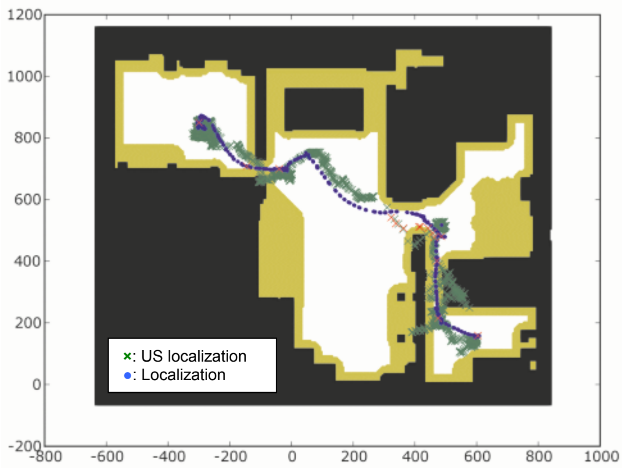
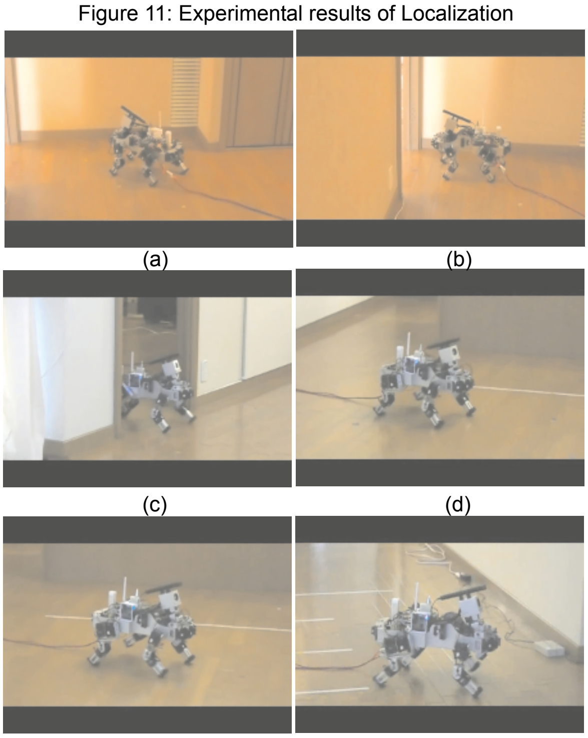


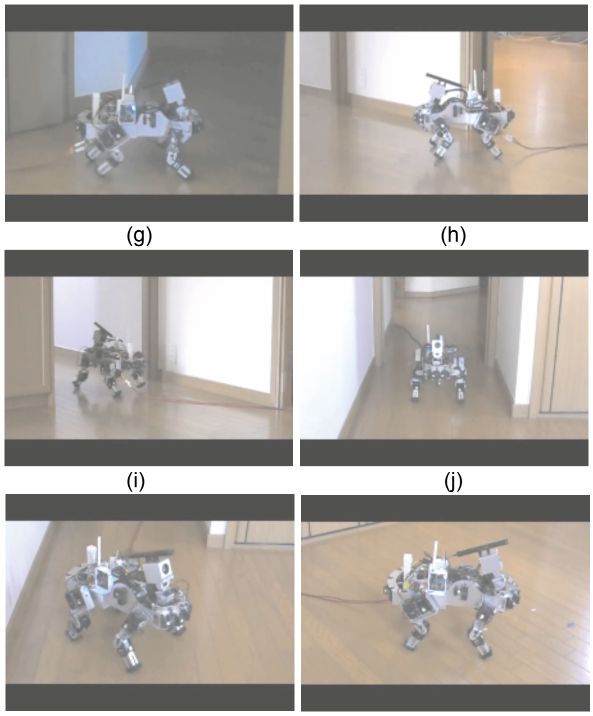
Figure 11: Experimental results of Localization



(a) (b)

(c) (d)

(e) (f)



(g) (h)

(i) (j)

(k) (l)

Figure 12: Snapshots of Navigation

5. SUMMARY

This paper described an autonomous locomotive function of a four-legged robot that utilized cooperative actions between ultrasonic locators and microphone arrays installed indoors. The results illustrate the efficacy of such a system, thus showing the objectives of this scenario to be realizable.

<References>

[1] Satoshi Kagami, Simon Thompson, Yoko Sasaki, Yoshifumi Nishida, Hiroshi Mizoguchi, and Tadashi Enomoto,
"Calling from the Other Room by Ceiling Ultrasonic Locators and Microphone Array," in Proceedings of the IEEE/RSJ 2007 International Conference on Intelligent Robots and Systems (IROS2007), pp. 2028–2034, San Diego, California, USA, October 2007.

[2] Satoshi Kagami, Simon Thompson, Yoshifumi Nishida, Tadashi Enomoto, and Toshihiro Matsui,
"Home Robot Service by Ceiling Ultrasonic Locators and Microphone Array," in Proceedings of the 2006 IEEE International Conference on Robotics and Automation (ICRA2006), Orlando, Florida, May 2006, pp. 3171–3176.

[3] Golubovic, D., Li, B., and Hu, H., 2003, A Hybrid Software Platform for Sony AIBO Robots. RoboCup 2003: Robot Soccer World Cup VII, 478-486.

[4] Kerepesi, A., Kubinyi, E., Jonsson, G. K., Magnusson, M. S., and Kiklosi, A., 2006, Behavioural Comparison of Human-Animal (Dog) and Human-Robot (AIBO) Interactions, Behavioural Processes, 73/1, 92--99.

[5] Thrun, S., Fox, D., Burgard. W., and Dellaert F., 2001, Robust Monte Carlo Localization for Mobile Robots, Artificial Intelligence, 128/1-2, 99-141.

Development of Invisible Mark and Its Application to a Home Robot

Sho Komai[1], Tomomi Kuroda[1], Masaharu Takano[1], Seiji Aoyagi[1], and Eiji Fukui[2]
[1] Dept. of Mechanical Engineering, Kansai University, Osaka, Japan
[2] R&D Headquarters, OG Corp., Tokyo, Japan

Abstract
To develop a practical multipurpose robot, the authors have previously proposed the RECS (Robot-Environment Compromise System) concept, which means the technology to compromise a robot and its environment in order to increase the robot's performance. As the modification in the environment side, the author gave a mark made from a commercial retro-reflective sheet to the tableware: however, this mark was slightly visible to a human. Moreover this mark was not resistant to the dishwashing. To overcome the above-mentioned problem of the previous mark, the present paper reports the development of a completely invisible mark.

Keywords:
Invisible barcode mark; Invisible landmark; Photochromic material; Ultraviolet light; Image processing; Clearing the table; Navigation

1 INTRODUCTION

In the near future a multipurpose robot performing domestic tasks should be indispensable for social needs, and this type of robot requires high robot technologies. A humanoid robot is not really practical at present to the extent that it could be actually employed in a home or a hospital, considering its reliability and cost. To develop a practical multipurpose robot, the authors have previously proposed the RECS (Robot-Environment Compromise System) concept, which means the technology to compromise a robot and its environment in order to increase the robot's performance [5]. This concept aims to share the technical difficulties with the robot and the environment in order that robot tasks would be possible and easy. As the modification in the environment side, the author gave a mark made from a commercial retro-reflective sheet to the tableware [2]: however, this mark was slightly visible to a human, and the surface of tableware was made rugged, giving the discomfort to a human during eating the meal. Moreover this mark was not resistant to the dishwashing.

To overcome the above-mentioned problem of the previous mark, the present paper reports the development of a completely invisible mark. As the modification in the environment side based on the RECS, this mark is given to the tableware and the ceiling of the room. The former makes the home robot task of clearing the table easy [4], and the latter makes the localization of a mobile home robot in the room easy [3]. This mark is invisible at usual state and only colored illuminated by an ultraviolet light when the robot recognition process is necessary, so the human is not aware of it while doing his/her daily home work. The emerging color is extracted from the raw image data captured by a CCD camera, and it is binarized by setting an appropriate threshold. Eventually, almost only the mark exists in the image, the extracting of which is much easier than the usual image recognition dealing with raw data, i.e., the robot task of visual recognition is simplified by applying the RECS.

First, as the material of the invisible mark, photochromic material (OG Corp., "stealth pigment", type INOR-1 and INOG-1, turning rose and green, respectively) is employed, which is usually colorless and transparent, and is only colored when illuminated by ultraviolet rays. For the task of clearing the table, first, the mark turning rose is added to the edge of the tableware. The marked tableware is coated by polymer Parylene film. This polymer is biocompatible, i.e., safe for the human. The film-coated tableware is confirmed to be washable, and be available for the microwave cooking. The shape of tableware is assumed to be circular. As the image recognition, the rose color is extracted from the image, followed by excluding the apparently too large or small areas. Then, the procedure of fitting ellipse to the image is carried out. The four intersection points between the elliptic line and the major/minor axes are extracted. Using their data, the position and orientation of the tableware with reference to the camera coordinate system can be calculated, provided that the radius of the tableware is known in advance.

Second, by using two colors, i.e., rose and green, a barcode mark is achieved, which includes the radius information necessary to calculate the position and orientation of the tableware as mentioned above. Other information effective for handling task of the tableware could be included in this barcode mark.

Third, an invisible barcode landmark is also developed, many of which are attached to the ceiling. They are effective for localization of a mobile home robot in the room. Finally the effectiveness of the proposed invisible mark was experimentally confirmed, in which a mobile robot equipped with a 7-DOF manipulator was navigated to the table position using the landmark image, and the tableware was handled by a robot hand using the tableware image.

2 EXPERIMENTAL SYSTEM

2.1 Mobile robot

A mobile robot is developed, which has two independently driving wheels in the front, and passive two casters in the rear, as shown in **Figure 1**. As a vision sensor, a USB camera of 300,000 pixels is employed. The performance of a board computer (PC) for controlling the mobile robot is as follows: CPU PentiumMobile 1.3 MHz, memory: 512 MB, slots: PCIx8, OS: Linux FedoraCore6.0. A counter board is inserted to the PCI slot, for receiving pulses from encoders attached to the driving shafts, which are used for odometry

Figure 1: Mobile robot.

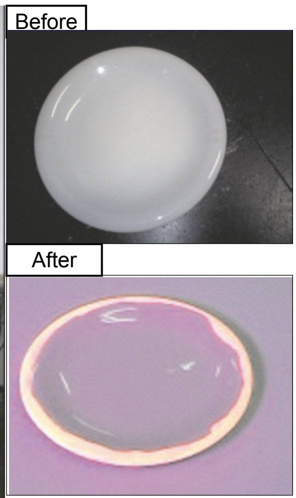


Figure 2: Ultraviolet rays is irradiated. Edge turns rose color.

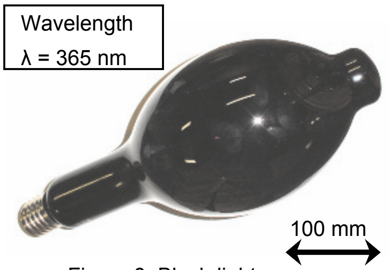


Figure 3: Black-light.

(dead reckoning). A D/A board is also inserted to the PCI slot, for sending control voltage to servo-amplifiers of the driving motors. The board computer also controls an image processing board. As the image processing software, HALCON (Linx Crop.) is employed, which is used to extract marks from a captured image.

2.2 Invisible mark

In this research, the effect of photochromic material is investigated, which is usually transparent and colorless and it is only colored when illuminated by ultraviolet rays.

As the result of testing samples made by various companies, photochromic material (OG Corp., “stealth pigment”, type INOR-1) is employed. **Figure 2** shows the results of turning color of the photochromic material.

This coating material is colored to rose with fast response time of approximately two seconds after ultraviolet rays are irradiated. As an ultraviolet source, a black-light shown in **Figure 3** (TOSHIBA Corp, type: H400BL-L, glass tube diameter: 150mm, its length: 370 mm, 400W) is employed.

Although it is not declared that ultraviolet rays are completely safe to the human body, there is a possibility of practical use of the black-light, considering two reasons: one is that the wavelength of it is within near ultraviolet rays area (365 nm), which is safety area to the human body: another is that the robot irradiates the rays in a short duration only when the robot needs recognition of the environment.

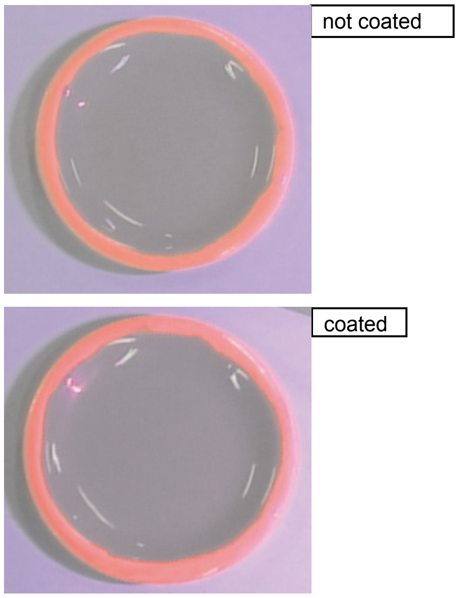


Figure 4: Effect of Parylene coating.

3 INVISIBLE BARCODE MARK ON THE TABLEWARE

3.1 Coating using polymer membrane (Parylene)

The ingredient of photochromic material has not been announced. Therefore, it cannot be declared that it is harmless to the human body, although it is generally used for sun glasses, etc. Considering these circumstances, we coat polymer Parylene film on the tableware which is given circular invisible mark of photochromic material [1].

Parylene is colorless, transparent, biocompatible, heat resistive, and corrosion resistive. Parylene has obtained authorization as a coating material of the medical equipment in the United States. In addition, CVD (Chemical Vapor Deposition) can realize a conformal deposition (that is, the deposition is performed not only on the top surface of a target object but also on the back/side surface of it).

Ultraviolet rays are irradiated to the tableware, which is coated by Parylene (film thickness 5 μm) using a CVD apparatus. Then, it is observed that the photochromic mark is surely recognized without any degradation on visibility compared with non-coated tableware. **Figure 4** shows the effect of Parylene coating on visibility of mark.

To investigate the practicability of the developed invisible mark, we tried using the tableware in a microwave oven, and washing the tableware with the detergent. As a result, problem was not confirmed at all.

3.2 Four point mark method

In this study, the photochromic material is coated on the whole circular edge of the tableware. So, assuming that the tableware is circle, the image of the invisible mark is ellipse. HALCON of Image processing software is used to fit ellipse to the image. Then, the major/minor axes of the ellipse are obtained. Next, the four intersection points between the elliptic line and the major/minor axes are extracted.

As shown in **Figure 5**, it is assumed that the coordinates (X, Y, Z) is based on the tableware coordinate system Σ_T, and it is assumed that they are projected to (x, y) in the camera

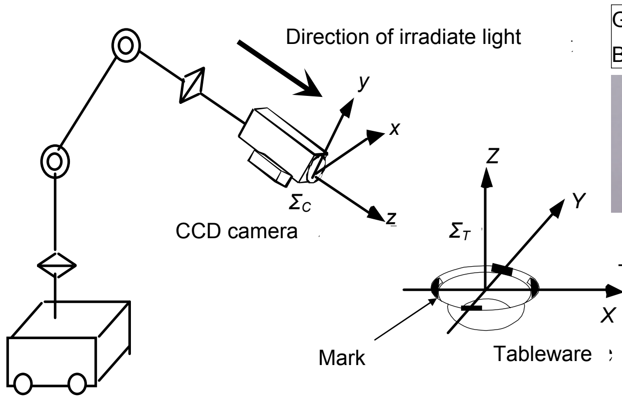


Figure 5: Relationship between camera coordinate system and tableware coordinate system.

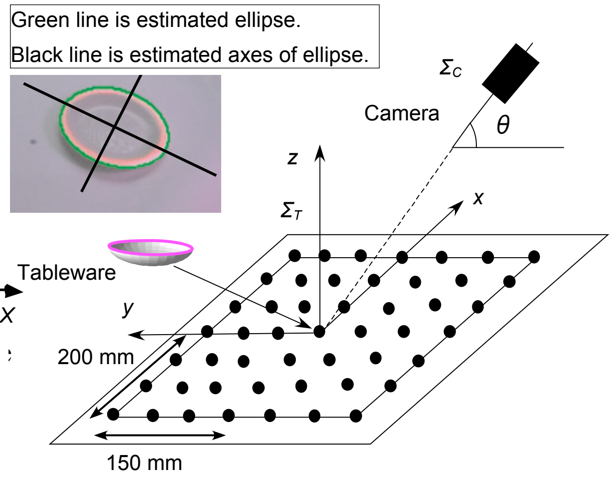


Figure 6: Experimental condition.

coordinate system Σ_C. Then the relationship between them is expressed as follows [6]:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \dfrac{f}{C_{31}X + C_{32}Y + C_{33}Z + n}(C_{11}X + C_{12}Y + C_{13}Z + m) \\ \dfrac{f}{C_{31}X + C_{32}Y + C_{33}Z + n}(C_{21}X + C_{22}Y + C_{23}Z + m) \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix} = \begin{bmatrix} C_\theta C_\phi C_\psi - S_\theta S_\psi & -C_\theta C_\phi S_\psi - S_\theta C_\psi & C_\theta S_\phi \\ S_\theta C_\phi C_\psi + C_\theta S_\psi & -S_\theta C_\phi S_\psi + C_\theta C_\psi & S_\theta S_\phi \\ -S_\phi C_\psi & S_\phi S_\psi & C_\phi \end{bmatrix} \quad (2)$$

where (l, m, n) is the coordinate origin of Σ_T, based on Σ_C, and (θ, ϕ, φ) is the Euler angle which expresses the relative orientation of Σ_T to Σ_C, f is focal length of the camera (assumed to be known in advance).

We set the coordinate origin of Σ_T to the center position of the tableware, then, the four mark positions based on this coordinate system are (X, Y, Z) = (r, 0, 0), (0, r, 0), ($-r$, 0, 0), (0, $-r$, 0), assuming the radius of the tableware is r. By measuring the four 2-D positions in the image corresponding to their 3-D position, 2×4=8 equations are obtained according to Eq. (1). Then, the parameters of (l, m, n, θ, ϕ, φ) are calculated so as to satisfy the 8 equations most accurately by the nonlinear least square method. As the result, the center of the tableware based on Σ_C can be acquired as (l, m, n), and the orientation of the tableware based on Σ_C can be acquired as (θ, ϕ, φ).

The feature of this four point mark method is that the position and orientation of the object (tableware in this case) can be obtained by using one camera, whereas two cameras must be required for stereovision.

3.3 Estimation of accuracy in position measurement of the tableware

Under the condition shown in **Figure 6**, the accuracy in position measurement of the tableware based on the proposed method was estimated.

The tableware is positioned at the grid points on the table one by one, where the position of the tableware was calculated by using CCD camera based on the proposed four point mark method, which is already explained in the previous Section 3.2.

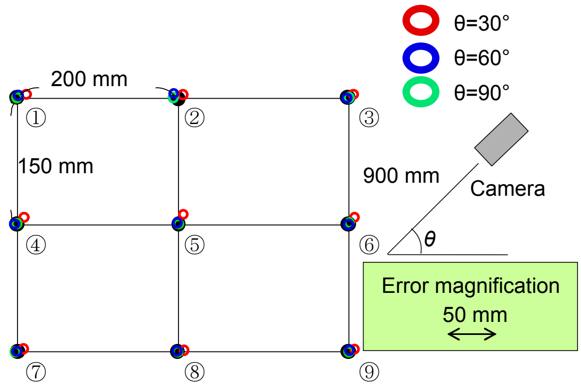


Figure 7: Measuremet error for 400×300 mm area.

Table 1: Measurement error for each point.

Position	30°[mm]	60°[mm]	90°[mm]
①(-200, 150)	(9.1, 7.5)	(2.0, 1.6)	(0.1, 0.1)
②(0, 150)	(5.3, 6.6)	(-0.8, 1.3)	(-1.1, 0.2)
③(200, 150)	(4.1, 8.7)	(-0.2, 1.0)	(0.3, 1.5)
④(-200, 0)	(6.5, 7.9)	(-1.6, 1.1)	(-1.3, 0.0)
⑤(0, 0)	(4.2, 14.1)	(-1.0, 1.0)	(-0.2, -0.4)
⑥(200, 0)	(8.8, 11.3)	(0.1, 0.3)	(0.4, 0.4)
⑦(-200, -150)	(5.7, 4.0)	(-1.8, 0.3)	(-1.0, -0.1)
⑧(0, -150)	(8.4, 3.5)	(-0.6, 0.1)	(-0.2, 0.0)
⑨(200, -150)	(5.5, 4.6)	(1.0, 1.3)	(-0.0, 0.0)

The results are shown in **Figure 7** and **Table 1**. Looking at this figure, considerable large error of 10 mm at the maximum arises in case that the angle between the camera and the table is 30 deg. However, in case that the angle is 60 or 90 deg, the measurement can be performed with the error less than 1 mm, which is finely accurate. Adding to say, this error is absorbed later in the handling task, since the mechanism of the robot hand can accept such a small error.

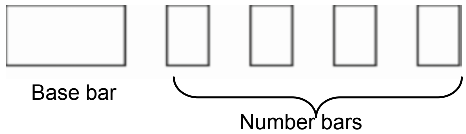


Figure 8: Schematic of barcode mark.

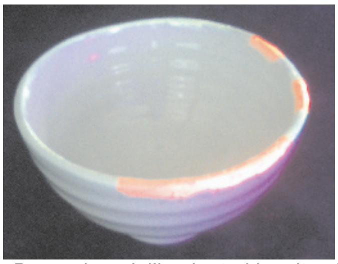


Figure 9: Barcord mark illuminated by ultraviolet rays.

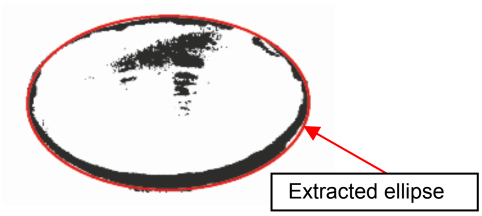


Figure 10: Extracted barcode of tableware (rose and green colors are used).

3.4 Barcode mark using two colors

In the four point mark method, the radius of tableware should be known in advance. In this research, the invisible barcode mark is developed, and the information of the radius is embedded to this barcode.

If the barcode mark is provided on the edge by using only one color, the color is lost at the area of “0” code. For example, if the barcode is 4 bit and it is “0000”, then, almost the mark disappears: thus the fitting of ellipse to the tableware image is almost impossible.

To address this problem, in this study, besides the photochromic material turning rose color (OG Corp., INOR-1), that turning green color (OG Corp., INOG-1) is also used.

First of all, a green coloring material is applied to all circumference on the edge of the tableware. Next, the rose coloring material is applied to the number bars corresponding to “1” bit, and it is also applied to the base bar, which indicates that the number bar exists in the right-hand side of it. The number bar and the base bar are schematically shown in **Figure 8**.

Figure 9 shows the barcode mark (number bar and the base bar) on the tableware, when the ultraviolet rays are irradiated.

The procedures of image processing are as follows: the captured image is binarized by using the color information of both rose and green. Then, the all circumference of the edge is extracted. The ellipse is fitted to the extracted image, as shown in **Figure 10**. Next, the extracted image is further binarized by using the color information of only rose. Then, the barcode is extracted, as shown in **Figure 11**.


Figure 11: Extracted barcode (only rose color is used).


Figure 12: Robot hand grips a dish.

Table 2**:** Example Information embedded in RF tag.

ID	Tableware	Height	Approach	Hand	Radius
1	dish	low	side	grip	75 mm
2	cup	high	edge	scoop	45 mm

In addition, for example, it is possible to embed such information as shown in **Table 2** to the barcode. By reading the barcode, the robot can easily acquire information on how and where to grasp the tableware. Examples of the information are as follows: if the tableware is a tea cup, the side surface is easily grasped by a gripping type hand. If the tableware is a plate, the edge area is easily grasped by a scooping type hand. **Figure 12** shows an example situation that the robot grasp a dish plate by using a scooping hand.

4 INVISIBLE BARCODE LANDMARK ATTACHED TO CEILING

4.1 Localization of mobile robot by landmark

Caused by the slip or/and friction between the wheel and the floor surface, odometry is not sufficient for localizing a mobile in indoor environment. Here, localization means to get information of mobile robot’s position accurately based on the coordinate system, which is set in the environment. To realize accurate localization, employing landmarks, absolute coordinate of which are known in advance, is very effective.

4.2 Barcode landmark

In this research, invisible barcode landmark is developed, which consists of recognition bar, base bar, and number bars of 7 bits, as shown in **Figure 13**. Photochromic material (OG Corp., type INOR-1, turning rose) is painted on each bar.

Information on absolute coordinate of the mark in the room is embedded to the barcode. Many invisible barcode marks are set to the ceiling. They are obtrusive, i.e., colorless and transparent, at usual state. When the robot irradiates ultraviolet rays, their colors turn to rose. By image processing these landmarks, localization of the robot would be possible.

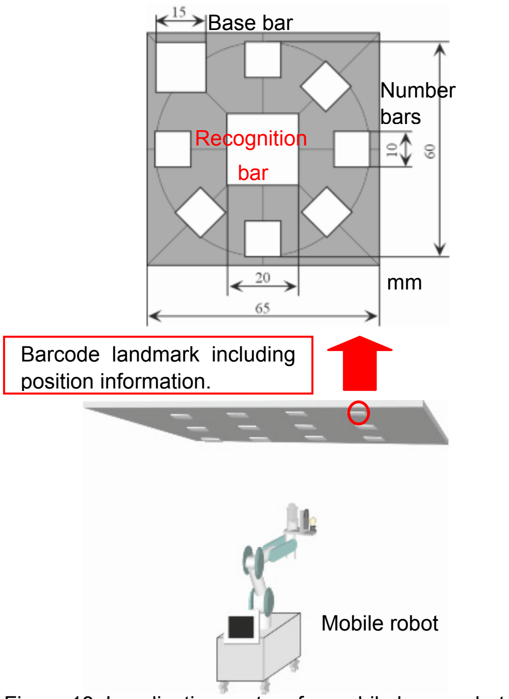


Figure 13: Localization system for mobile home robot.

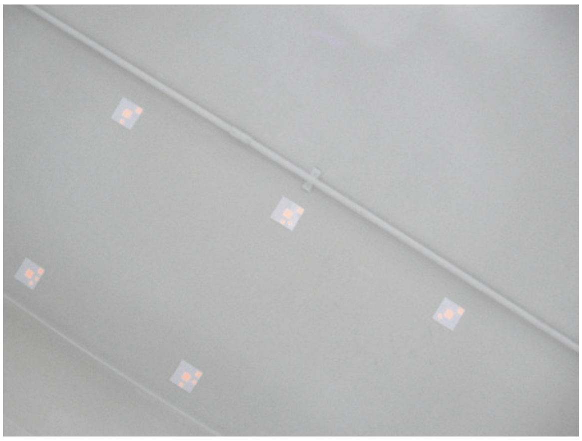


Figure 14: Bar-cord landmarks attached to the ceiling


Figure 15: Bar-cord landmarks attached to the ceiling

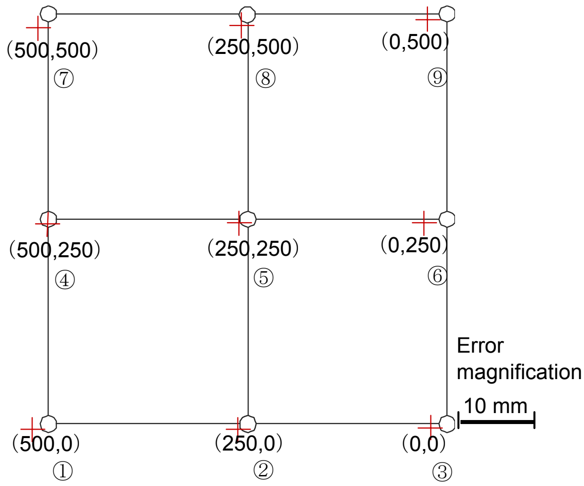


Figure 16: Error distribution in localization using ceiling landmarks..

Table 3: Results of localization using ceiling landmarks.

Position	Landmark[mm]	Localization[mm]	Error[mm]
①	(500, 0)	(533.2, -18.0)	(33.2, -18.0)
②	(250, 0)	(271.4, -21.0)	(21.4, -21.0)
③	(0, 0)	(47.3, -14.2)	(47.3, -14.2)
④	(500, 250)	(503.6, 229.6)	(3.6, -20.4)
⑤	(250, 250)	(270.7, 231.4)	(20.7, -18.6)
⑥	(0, 250)	(53.7, 240.8)	(53.7, -9.2)
⑦	(500, 500)	(528.1, 463.1)	(28.1, -36.9)
⑧	(250, 500)	(278.2, 474.4)	(28.2, -25.6)
⑨	(0, 500)	(51.5, 481.0)	(51.5, -19.0)

Figure 14 shows barcode landmarks attached to the ceiling when irradiated by an ultraviolet light. **Figure 15** shows extracted barcodes by image processing, i.e., binarization using rose color, excluding the apparently too large or small areas, etc.

How to read the barcode is described herein. The role of each bar is judged by using the area information of each bar. Namely, the bar having the biggest area is judged to be the recognition bar. The bar having the second biggest area is judged to be the base bar. The number bars are read clock-wise from the right-hand side of the base bar.

4.3 Estimation of accuracy in position measurement of the mobile robot

If at least the two barcode landmark images are included in the captured image, the robot absolute position in the room can be calculated by using the information of barcode landmarks' positions, which can be obtained by reading barcodes.

Figure 16 shows the experimental data in 500x500 mm area, in which the accuracy of the proposed system using barcode landmarks is estimated. **Table 3** also shows the same results. From these results, the error of approximately 50 mm at the maximum arises.

The reason for somewhat large error is supposedly caused by the inclination of the robot, since the calculation of the mobile robot's position is assuming the parallelism between the floor surface and the ceiling surface, and assuming the verticality of CCD camera. However, these assumptions are not strictly satisfied in the real environmental condition. By employing an inclinometer, compensating the inclination of the mobile robot is ongoing work.

5 CONCLUSIONS

A completely invisible mark is developed by using photochromic material. As the modification in the environment side based on the RECS (Robot-Environment Compromise System), this mark is given to the tableware and the ceiling of the room. This mark is invisible at usual state and only colored illuminated by an ultraviolet light when the robot recognition process is necessary, so the human is not aware of it while doing his/her daily home work.

The emerging color is extracted from the raw image data, and it is binarized. Eventually, almost only the mark exists in the image, the extracting of which is much easier than the usual image recognition dealing with raw data.

First, as the material of the invisible mark, photochromic material is employed. The position and orientation of the tableware with reference to the camera coordinate system can be calculated.

Second, by using two colors, a barcode invisible mark is achieved, which includes the radius information necessary to calculate the position and orientation of the tableware. Other information effective for handling task of the tableware could be included in this barcode mark.

Third, an invisible barcode landmark is also developed, many of which are attached to the ceiling. They are effective for localization of a mobile home robot in the room.

ACKNOWLEDGMENTS

This work was supported by 2008 Collaborative Development Innovative Seeds, Japan Science Technology Agency (JST): title is “Development of Invisible Mark and Its Application to Intellectualization of Environment”.

REFERENCES

[1] S. Aoyagi, etc, 2006, Proposal of Recognition System of Tableware utilizing Invisible Mark and RF tags for Home Robot Application, JSPE Autumn Conference, 2006:471-472.

[2] Y. Nakazato, M. Kanbara, N, Yokoya, 2004.07, A Localization System Using Retro-reflective Markers and an IR Camera, Proc. MIRU2004, 2004:119-120.

[3] S. Aoyagi, Y.Kiguchi, K. Tsunemine, M. Takano, 2001.02.01, Position and Orientation Measurement of a Mobile Robot by Image Recognition of Simple Barcode Landmarks and Compensation of Inclinations, IEEJ Transactions on Electronics, 121-C/2:375-384.

[4] S. Aoyagi, H. Kinomoto, S. Ieuji, M. Takano, 2000, Edition on Robotics for Social Application. Recognition and Handling of Tableware by Using a Robot Based on RECS Concept, IEEJ Transactions on Electronics, 120-C, 5:615-624.

[5] M. Takano, T. Yoshimi, K. Sasaki, H. Seki, 1996, Development of Indoor Mobile Robot System Based on RECS Concept, Journal of JSPE, 62/6:1334-1338.

[6] H. Takahashi, F. Tomita, 1992, Camera Calibration for Robot Vision, Journal of RSJ, 10/2:177-184.

Adaptive Polyhedral Subdivision for Image-based Lighting

Shinji Yamazaki[1], Yoshihiro Yasumuro[2] and Masahiko Fuyuki[2]

[1] Graduate School of Engineering, Kansai University, Osaka, Japan

[2] Faculty of Environmental and Urban Eng., Kansai University, Osaka, Japan

Abstract

Augmented Reality (AR) is a technique that synthesizes and displays virtual objects such as CG in images of real environments. It is important to harmonize real environment with CG and to render the harmonized scene. In this research, we propose a method for rendering virtual objects by using Image-based Lighting (IBL) that uses polyhedral subdivision in AR. We confirmed that hierarchical light source map constructed by polyhedral subdivision reduces the computation cost and maintains the effect of IBL.

Keywords:

Image-Based Lighting; Polyhedral Subdivision; Augmented Reality

1 INTRODUCTION

Augmented Reality (AR) is a visual integration technique capable of fusing virtual and real scenes by overlaying computer-generated information onto real scenery images. Computer graphics (CG) is often used for representing digital contents such as additional information related to the physical object in front of the user. The AR technique is gathering attention and has come into use in various fields. In a factory, for instance, virtual products are superimposed on an existing production line for simulating the working situations and extracting the possible problems beforehand [1]. A marker-based image fusion technique allows on-line process and geometrically accurate image overlaying, that is used in TV shows such as weather forecasts [2]. The marker-based AR especially enables an interactive application with a single USB camera instantly. The ARToolKit [3] is a good example for implementation of a realtime AR/MR application software.

Visual fusion of CG and a real scene is known as mixed reality (MR) as well [4,11,12]. In terms of “natural” appearance of the synthesized object, as if it really exists in front of you, a key issue is managing consistency between virtual and physical space, including geometry, illumination and time conditions. As for illumination conditions, image-based lighting (IBL) is an effective approach for constructing lighting conditions, taking existing lighting situations into account [5]. IBL constructs a light source distribution from skylight images. Every single pixel on the image can be a reference of a single light source, which illuminates the CG object. The intensity or the colour of the pixel can be the one of the light sources. The pixel position indicates the direction of the incoming light. Since coverage of the light source directions are supposed to be preferably wide, the skylight images are captured with a wide-angle or an omni-directional camera.

Since each pixel of the image is used as a light source, IBL requires costly computation even applying a local lighting model that uses only straight lines for the light paths from the sources to the object. To determine a colour of a single point on the object needs summing up all the effects of the light sources visible from the point. Needless to say in the case of global lighting model, that takes into account multiple light paths reaching the object from the light source through reflections and/or refractions.

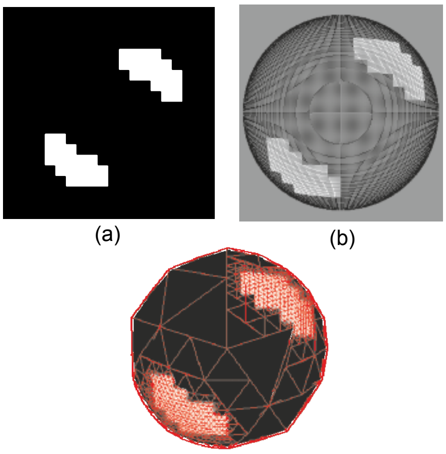
Several methods have been proposed for decreasing the computational cost, and the down-sampling of pixels instead of using all pixels. The K-means method can be applied for dividing the light source area by clustering the pixels of light source [6]. The number of light sources is decreased by each area to a single light source. However, since the whole area is evenly divided, decreasing the number of light sources directly affects the rendering results. Importance sampling techniques are applied hierarchically, which divide the light source area, clustering with an index named an importance metric, which is most likely the brightness of the pixels [7,10]. Consequently, since the division number of the area is different according to the importance metric in the area, an efficient sampling light source can be done.

In this research, we propose a geometrically hierarchical approach for constructing light source distribution. The aim is to decrease the number of light sources and to automatically suggest an appropriate level of detail of the light sources based on rendering results.

2 PROPOSED APPROACH

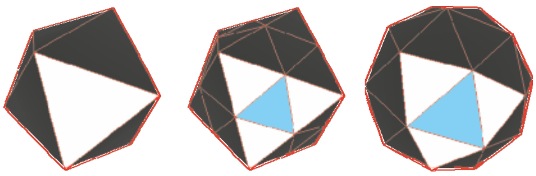
2.1 Overview

In our approach, the lighting environment map is approximated by the polyhedral subdivision of the sky dome, and each face of the polyhedron is treated as a source of light. The mechanism to adaptively divide faces is based on the hierarchical structure of the polyhedron. The subdivision of a certain face creates detailed light source distribution within the face. The detailed lighting improves rendering accuracy, reflecting actual illumination conditions captured by the input image. Considering that intense light sources drastically effect the rendering results, faces containing brighter light sources must be subdivided independently. The subdivision process can be monitored in terms of contribution



(a) (b) (c)

Figure 1 : Light source map by polyhedron subdivision



(a) (b) (c)

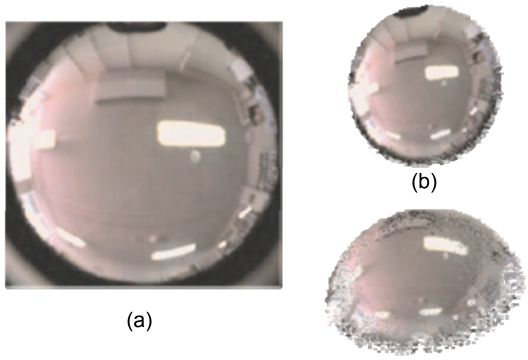
Figure 2 : Division procedure in each face. A face of the polyhedron(a) is divided into four area(b). New vertices are moved to the surface of the sphere (c).

for the rendering quality. Taking the computation cost for the rendering and processing frame rate into account, the system performance is adjustable. The following describes the overview of the proposed method that applies a polyhedron subdivision to IBL.

First, the light source distribution of the surroundings is obtained from an omni-directional image as shown in Figure 1 (a). An initial light source map is shown in Figure 1 (b). The initial light source map is made by projecting the light source image onto to a hemisphere that shows the directional distribution. The light source map is re-sampled, according to subdivided polyhedron faces as shown in Figure 1 (c), in which each face can be referred as a source of light. Finally, virtual objects are rendered based on the light source information obtained above.

2.2 Light Source Mapping to Polyhedron

A light source map is a distribution chart of light sources mapped on a hemisphere whose direction is the same as the physical environment in AR application. The map is constructed by projecting each pixel on a captured image, or an image probe [5] onto a 3D hemisphere. An omni-directional image is taken with a fisheye camera that has 180 degrees of view angle. Conversion from the image plane to the 3D polar coordinate system is specified in each lens type. This paper uses equi-solid angle representation. The position on the environmental map of the point with the image in this case is computable as follows:



(a) (b) (c)

Figure 3 : Light source mapping from a photo : photo Image of a ceiling(a), mapped pixels onto a dome((b), (c))

$$e_x = \frac{x - x_c}{x_c},\; e_y = \frac{y - y_c}{y_c},\; e_z = \sqrt{1 - \left(e_x^2 + e_y^2\right)} \tag{1}$$

where, *P(x, y)* is a single point on the image plane, vector ***e***(*ex, ey, ez*) is a mapped direction on a unit hemisphere. Each light source is supposed to be settled at infinite distance on the direction e. This research applies the local lightning model with the point light sources referring to this map. The vector ***e*** intersects one of the faces in the geodesic dome. Finding the intersecting face specifies the pixel and the direction where the incoming light comes from. Interction between a triangle face ABC and a vector $\boldsymbol{e}_i$ can be detected as follows:

$$\alpha(x - a_x) + \beta(y - a_y) + \gamma(z - a_z) = 0 \tag{2}$$

$$\alpha = \{(b_y - a_y)(c_z - a_z) - (c_y - a_y)(b_z - a_z)\}$$
$$\beta = \{(b_z - a_z)(c_x - a_x) - (c_z - a_z)(b_x - a_x)\}$$
$$\gamma = \{(b_x - a_x)(c_y - a_y) - (c_x - a_x)(b_y - a_y)\}$$

where, equation (2) reparents a plane which contains three vertex A(a_x, a_y, a_z), B(b_x, b_y, b_z), C(c_x, c_y, c_z). The intersecting point P of the plane and e P is expressed as equation (3) with a paremeter t shown in equation (4).

$$P = (e_x t, e_y t, e_z t) \tag{3}$$

$$t = \frac{\alpha\, a_x + \beta\, a_y + \gamma\, a_z}{\alpha\, e_x + \alpha\, e_y + \alpha\, e_z} \tag{4}$$

where α, β, γ are constant coefficients. If three normal vectors, N_{ABP}, N_{ABP}, N_{ABP} of the triangle ABP, APC and PBC respectively are identical, intersectio P is located withing the triangle ABC.

$$N_{ABP} = N_{APC} = N_{PBC} \tag{5}$$

2.3 Hierarchical Lighting Map with Geodesic Dome

The geodesic dome is used for structural construction such as a domed roof. In general, it indicates the polyhedron structure composed of triangles inscribed to a sphere. The geodesic dome is constructed by approximating sphere shape by starting from a regular icosahedron and then each initial face is equally subdivided into new faces until reaching a required approximation criterion. Since the face area is constant anywhere, a geodesic dome can be a base grid for evenly sampling light energy from every direction. In this research, we utilize the base grid and the approximation process of the geodesic dome for hierarchically re-sampling the light sources. The faces are divided according to the following procedure. The initial dome is an icosahedron as mentioned above. Each face is a regular triangle, whose sides are evenly divided by n equally-spaced points. New sides pass through the points parallel to the original sides. New sides form 2n new triangles, which are projected onto the hemisphere surface from the center. Finally, projected vertices are connected to form a new triangle mesh. By doing the processing above over the dome, faces are divided into n^2 sub faces.

2.4 Hierarchy Control based on Light Source Brightness

Since all the new faces are included in the existing face before division, an inclusion relation can be expressed as a hierarchical structure. We call the number of division m the number of the hierarchy. An optimal number of hierarchies can be determined considering the processing speed of hardware, required rendering quality and complexity of light source image. The area including a higher luminance source strongly influences the rendering result. Therefore, subdividing the faces containing higher luminance points enhance the rendering accuracy, since the direction of the light source is settled more accurately after the division. In the procedure, light source distribution for each face is examined, then the highest luminance source included in each face is picked up. By thresholding, faces with higher brightness are divided into deep hierarchies. On the other hand, faces with only dark sources are not used for the following processing because of less influence in the total luminance calculation. The procedure above can be express in a pseudo-code as:

```
Repeat the following
For each light source {
    For each polygon {
        Calculate t so that t*(light source vector) is on
        the polygon plane
        If ( t <=0 ) {
            search in the next face.
            }
        P = t*(light source vector)
        If ( P is within the polygon ){
            Add the light source in the polygon
            Break
        }
    }
}
For each polygon {
    If (maximum intensity in the face > threshold )
        &&(division number < max hierarchy number ){
        divide the polygon
    }else{
        Exit
    }
}
```

2.5 Rendering

Using every single pixel on the lighting map as a light source, a CG object is illuminated and rendered. We use local illumination that uses rays directly reached from the light source. Every ray is used for computing specular and diffuse reflection, and for each vertices on the CG object, the color to be rendered is determined as in the equation below.

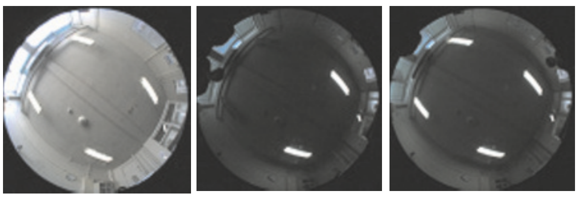
$$\begin{aligned} I &= (\textit{specular}) + (\textit{diffuse}) \\ &= K_d \sum_m^M I_d (N L_m) + K_s \sum_m^M I_s \left[N \frac{V + L_m}{|V + L_m|} \right]^{\mu} \end{aligned} \tag{6}$$

where, I is the colour of the vertex, K represents the reflectance coefficient, and N denotes the normal vector on the vertex. V is a vector from the vertex to the viewpoint. μ denotes a sharpness coefficient for the specular. Face m on the dome emits light of intensity L_m. The total number of light source is M. Suffix s and d denotes the component of specular and diffuse reflection, respectively.

3 IMPLEMANTATION AND DISCUSSION

We implemented the process mentioned above using C++ with graphics libraries, OpenGL and GLSL. A CG object

(a)



(b) (c) (d)

Figure 4 : Rendering materials ; CG object, 'happy budha'[9](a), Light source images of ceiling with daylight (b), with 2 lights (c) and 4 lights (d)

(40,000 vertices) is rendered using fisheye pictures as light source images (120x120 pixels) (Figure 4). The division numbers and rendering results are shown in Tables 1, 2 and 3. Figure 7, 8 and 9 show the lighting map for each hierarchy and its rendering results. The higher hierarchy shows more intense highlights and the specular reflection varies in its position. The increase in bright light sources lifts the total image brightness. Table 5 shows absolute average of pixel value differences between rendered images of adjacent hierarchies. Table 6 shows pixel value differences between the proposed method and the original light map. The depth of the hierarchy clearly changes the rendering quality. Systematic subdivision may automatically show the optimized balance between process time and output quality, according to the application demands.

4 SUMMERY

This paper proposed a hierarchical subdivision approach for re-sampling image pixels that can be systematically used for IBL. Experimental results showed that the depth of the hierarchy might control the LOD of the light source distribution. Our future work will be focused on optimization criteria for re-sampling, considering the user's intention to control the dynamic range of the camera to render CG objects and demonstrate real-time AR/MR applications.

5 REFERENCES

[1] http://www.metaio.com/products/planner/.

[2] Advanced Virtual Studio -Seamless image composition using actual and virtual studio sets-, 2003, NHK Broadcast Technology no.13, p.22.

[3] P. Milgram and A. F. Kishino, Taxonomy of Mixed Reality Visual Displays, 1994, IEICE Transactions on Information and Systems, E77-D(12): 1321-1329.

[4] Kato, H. and Billinghurst, M., 1999, Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System. Proceedings of the 2nd International Workshop on Augmented Reality, 85-94.

[5] Paul Debevec,1998, Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography, Proc. 25th annual conference on Computer graphics and interactive techniques,189-198.

[6] LLOYD, S. , 1983, An optimization approach to relaxation labeling algorithms. Image and Vision Computing 1, 2, 85-91.

[7] OSTROMOUKHOV, V., DONOHUE, C., AND JODOIN, P.-M., 2004, Fast hierarchical importance sampling with blue noise properties. ACM Transactions on Graphics 23/3: 488-495.

[8] Large Geometric Models Archive, Georgia Institute of Technology, http://www-static.cc.gatech.edu/projects/large_models/

[9] The Stanford 3D Scanning Repository, http://www-graphics.stanford.edu/data/3Dscanrep/

[10] Sameer Agawal, Ravi Ramamoorthi, Serge Belongie and Henrik Wann Jensen, 2003, Structured Importance Sampling of Environment Maps, ACM Transactions on Graphics 22/3, 605-612

[11] Yoshihiro Yasumuro, Yu Ishikawa, Koichi Minami, Masataka Imura, Yoshitsugu Manabe, Kunihiro Chihara, Consistent Presentation of Interactive Virtual Objects in Real Space with 3D Markers, 2003, Journal of the Institute of Image Information and Television Engineers VOL.57;NO.10, 1307-1313

[12] Paul E. Debevec and Jitendra Malik, Recovering High Dynamic Range Radiance Maps from Photographs,1997, International Conference on Computer Graphics and Interactive Techniques, 369-378

Table 1: Processing time and division number for daylight image (from Figure 4(b))

Hierarchy Number	1	2	3	5
processing time (Sec/ frame)	2.797	3.703	4.188	6.031
creation time(Sec)	2.641	3.422	3.891	4.828
division number	65	131	281	1742

Table 2: processing time and division number for 2 lights image (from Figure 4(c))

Hierarchy Number	1	2	3	5
processing time (Sec/ frame)	0.157	0.250	0.281	0.671
creation time(Sec)	0.328	0.406	0.484	0.375
division number	47	80	119	281

Table 3: processing time and division number for 4 lights image (from Figure 4(d))

Hierarchy Number	1	2	3	5
processing time (Sec/ frame)	0.250	0.313	0.391	0.547
creation time(Sec)	0.406	0.500	0.609	0.906
division number	50	86	143	398

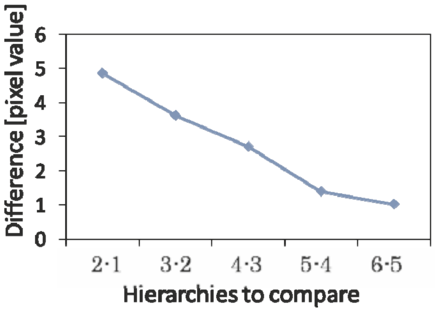


Figure 5 : Rendering difference between adjacent hierarchies (dark room)

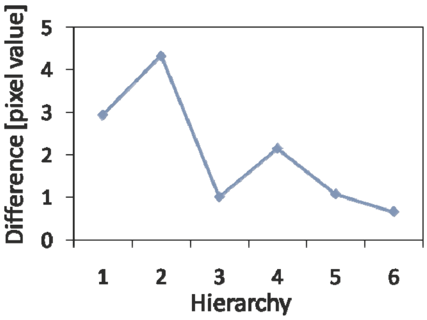


Figure 6 :Rendering difference between subsampled and original light map (dark room)

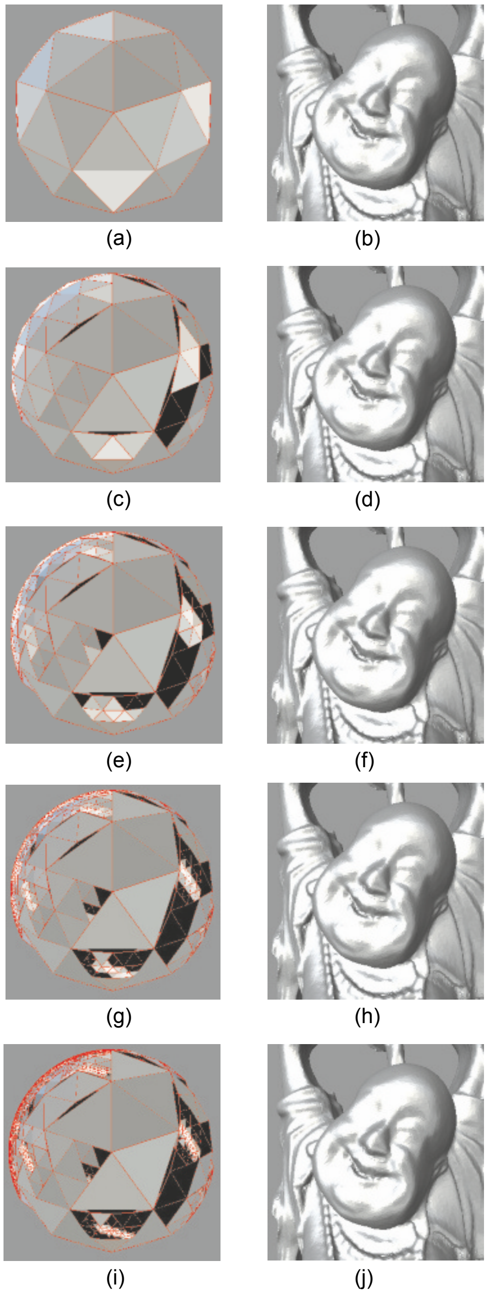


Figure 7 : light map for daylight image (left column) and rendering results (right column)

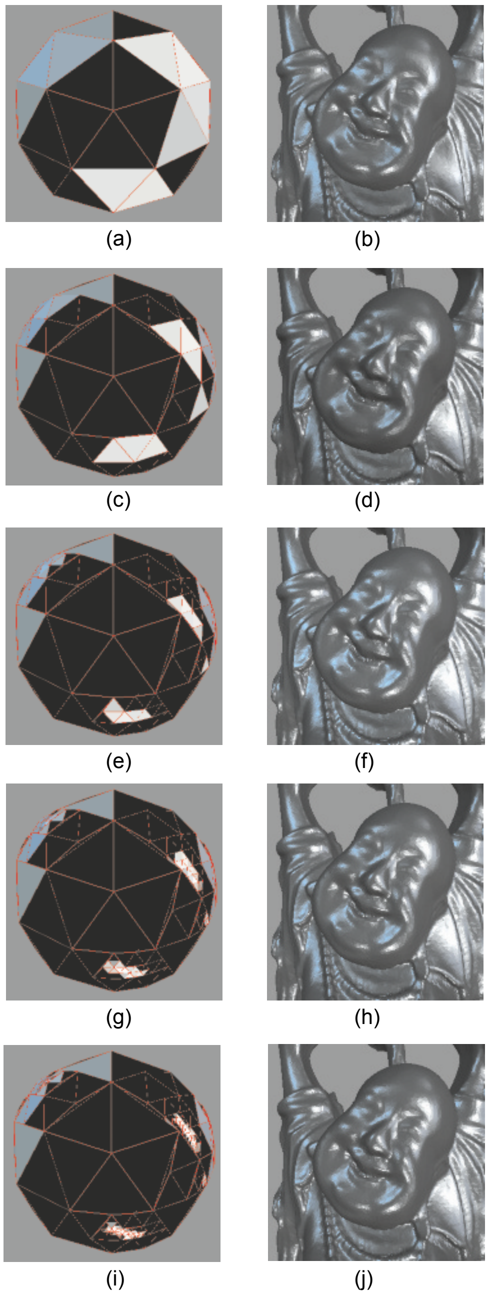


Figure 8 : light map for 2 lights image (left column) and rendering results (right column)

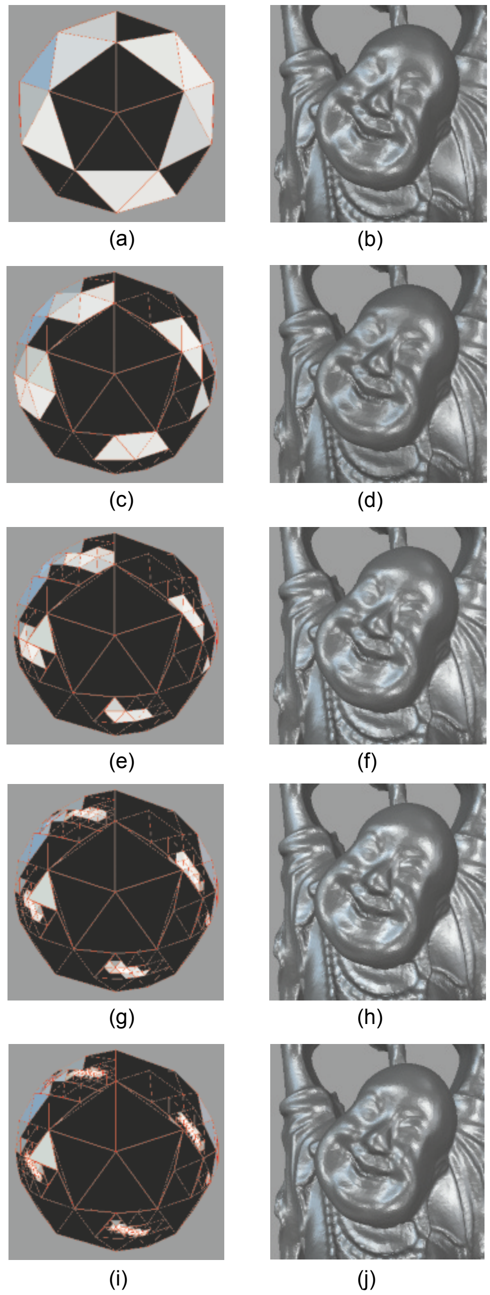


Figure 9 : light map for 4 lihgts image (left column) and rendering results (right column)

Comparison of Driving Performance of Piezoelectric Actuator
- Current Pulse Drive and Voltage Linear Drive -

Katsushi Furutani[1], Atsushi Furuta[2]

[1] Dept. of Advanced Science and Technology, Toyota Technological Institute, Nagoya, Japan

[2] Nano Control, Tokyo, Japan

Abstract

The authors have proposed a driving method of a piezoelectric actuator by using current pulses for the high resolution. It was compared with the voltage drive with a linear amplifier. For a step input, the linearity of the steady-state values by the current pulse drive was better than that by the voltage drive because the hysteresis of the displacement to the applied voltage was small. The direction of the creep was in reverse each other. The leak current also affects the stability of the displacement.

Keywords:

Piezoelectric Actuator; Positioning; Hysteresis; Creep; Linearity

1 INTRODUCTION

Piezoelectric actuators extend several tens of micrometers in maximum, and they have a potential of a nanometer-order or finer resolution. The deformation of the piezoelectric actuator is generally controlled by adjusting applied voltage [1]. In a simple feed-forward control, the hysteresis and creep of the displacement to the applied voltage to the piezoelectric actuator is observed. The displacement is sometimes fed back for the precision positioning to avoid the hysteresis [2]-[4]. A linear current drive circuit is usually complicated and neither the terminals of a piezoelectric actuator are generally connected with the ground [5].

On the other hand, a charge control by supplying continuous current or detecting induced charge on the electrode attached at the ends of a piezoelectric actuator has been proposed in which the hysteresis of the displacement is not observed [6][7]. This method also has the same problem as the voltage source because it uses a digital to analogue converter.

The Δ or $\Sigma\Delta$ conversion and the pulse density modulation (PDM) of the applied voltage to the actuator as input and output methods have ensured simple configuration of control system and high resolution [8]. It is expected that a series of current pulses may allow better performance because the piezoelectric actuator is a capacitive load. The authors have proposed a driving method by using a series of current pulses [9]. A piezoelectric actuator can be regarded as a linear device by driving this method because it is a kind of charge control.

The performance of the current pulse drive has been compared with that of the voltage drive through a simulation with a Martin's model [10]. This model does not represent the hysteresis [11]. Some models that can represent the hysteresis by using Preisach or generalized Maxwell model [12][13]. Their parameters are experimentally identified and adjusted for the voltage drive. Therefore, experimental evaluation is easier than theoretical one for the hysteresis and creep performance

In this paper, a driving performance of a piezoelectric actuator by using the current pulse is compared with a voltage linear drive through experiments. At first, the principle and results by the driving method by using the current pulse is introduced. Then the comparisons between them are described.

2 DRIVING PRINCIPLE

2.1 Configuration of Driving Circuit

Figure 1 shows a driving circuit of the piezoelectric actuator by current pulse [9]. It consists of n current souses, $I_{p1} > I_{p2} > \ldots > I_{pn}$, and m current sinks, $I_{m1} > I_{m2} > \ldots > I_{mm}$, connected by switches S_{p1}, S_{p2}, …, S_{pn}, S_{m1}, S_{m2}, …, S_{mm} in parallel where m, n=1, 2, … Each switching duration is T_{p1}, T_{p2}, …, T_{pn}, T_{m1}, T_{m2}, …, T_{mm}. The output current pulse from the

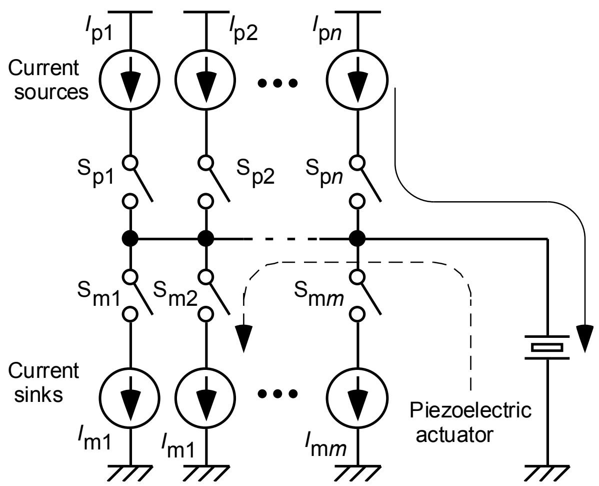


Figure 1: Configuration of driving circuit.

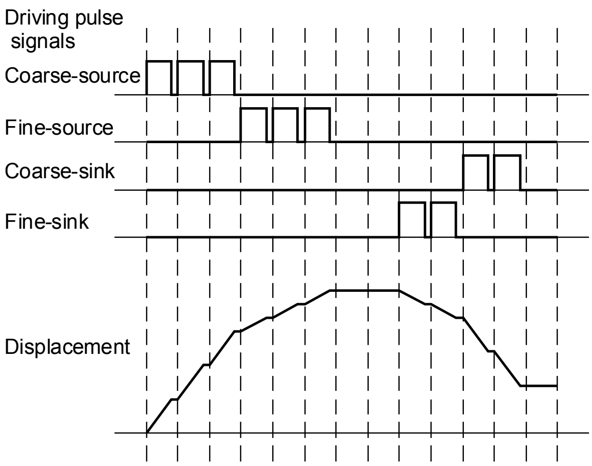


Figure 2: Driving signals and displacement of piezoelectric actuator.

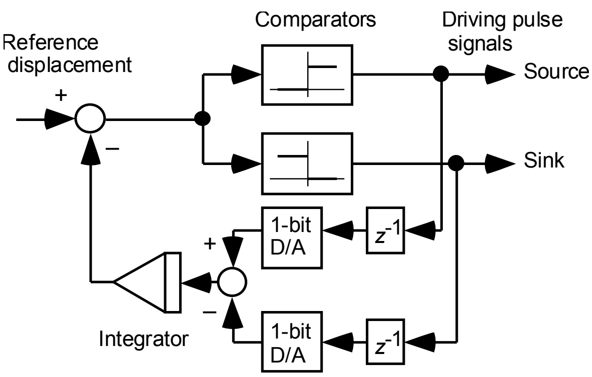


Figure 3: Δ converter.

driving circuit follows the driving signals to select one of the current sources or sinks provided by the controller.

2.2 Driving Method

Figure 2 shows a schematic displacement of the piezoelectric actuator [9]. If the error against a reference value is large, the driving circuit provides I_{p1} or I_{m1}, for a quick deformation. Then the corresponding current with the error is provided. If the error is smaller than the deformation driven by I_{pn} or I_{mm}, the driving circuit is disconnected from the piezoelectric actuator with a high output impedance and does not provide any current pulse.

This method is a kind of the charge control because the charge amount is represented by the product of the pulse duration by the current value. Consequently, the displacement with little hysteresis in any range of the applied voltage is promised. The displacement can be predicted by counting the given pulses without the displacement feedback.

The ΣΔ conversion is generally more preferable than the Δ one [14] for the generation of the driving signals. Because the piezoelectric actuator performs as an integrator of the current pulse, the driving signal was generated based on the Δ conversion for pulse density modulation (PDM) shown in Figure 3 [9], implemented as the software. The driving pulse signals were separately provided to each source and sink.

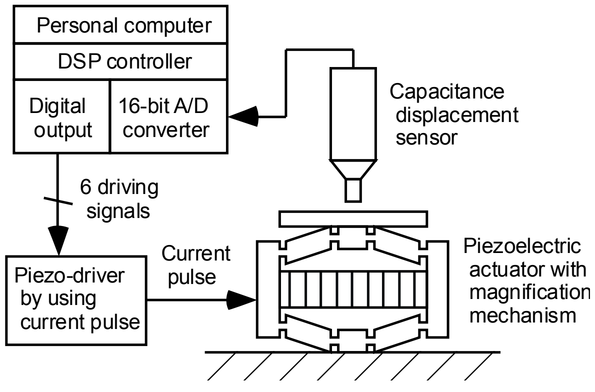


Figure 4: Experimental setup.

Table 1: Peak and duration of current pulse.

Resolution	Current mA	Pulse duration μs	Step nm	Threshold nm
Coarse	1000	16	3000	4500
Medium	20	20	112	224
Fine	0.5	50	7	14

3 EXPERIMENTAL SETUP

Figure 4 illustrates an experimental setup. The driving circuit consists of three sets of the current sources and sinks: coarse- (1 A), medium- (20 mA) and fine-step ones (0.5 mA) as shown in Table 1. The thresholds are used in the feedback control to switch the current pulse values.

Figure 5 shows a brief circuit diagram of a set of a current source and sink. A set of the current source and sink were mounted on a board and three sets were connected in parallel. N-channel MOS-FETs were used for switching devices. The output current I is regulated to:

$$I = \frac{V_{\mathrm{ref}}}{R_{\mathrm{f}}} \tag{1}$$

where V_{ref} and R_{f} are a reference voltage and feedback resistor. The high-side gate circuit is isolated with a photo coupler and DC-DC converter. The measured minimum pulse duration of the current sources and sinks was 16 μs.

The displacement of a stacked piezoelectric actuator with a capacitance of 1.4 μF (AE0505D16 by NEC-Tokin) is magnified by a flexure hinge mechanism made of aluminium. It deforms 70 μm at an applied voltage of 100 V with a hysteresis of 14 %. The first and second natural frequencies were 770 and 1800 Hz, and the anti-natural frequency was 1200 Hz in the case of attached a small steel plate as a sensor target. The space around the piezoelectric actuator was filled with gel to passively suppress the residual vibration. The damping ratio measured by the logarithmic decrement method [15] was 6.25×10^{-2}.

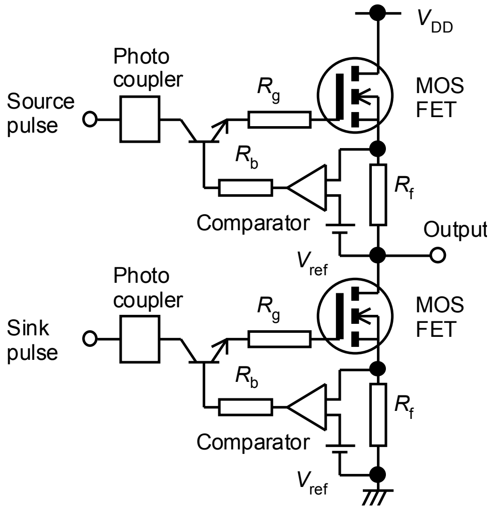


Figure 5: Set of current source and sink.

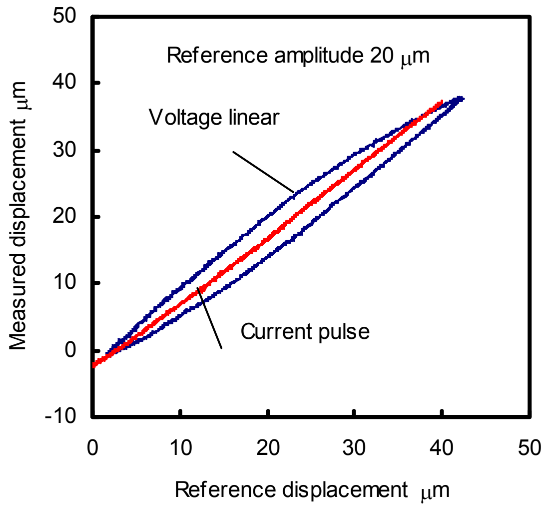


Figure 6: Hysteresis curves of measured displacement to hysteresis.

A digital signal processor (DSP) was used as a controller. The control program was written in C language. A desired number of the current pulses for a reference were sequentially given in the current pulse drive.

Because the current sources and sinks have a high output impedance, the displacement was kept unless additional current pulse was given. Therefore, the piezoelectric actuator was initialized by discharging through a resistor of 100 kΩ after each experiment.

For the voltage linear drive, a reference was provided from a 14-bit D/A converter and then was magnified 20 times with a power amplifier with a band width of 500 kHz and the maximum peak current of 5.7 Ap-p. Because the maximum current of the voltage linear amplifier is larger than the current pulse drive, its slew rate is larger than that of the current pulse drive.

The displacement was measured with a capacitance displacement sensor with a measurement range of 50 μm, a resolution of 10 nm and a bandwidth of 20 kHz. It was used for the evaluation of the performance in the feed-forward control and for the displacement measurement in the feedback control. The signals were recorded with a 10-bit digital oscilloscope (its input impedance: 10 MΩ).

4 EXPERIMENTAL RESULTS

4.1 Example of displacement control

Figure 6 shows a comparison of displacement by the current pulse drive with that by a voltage linear drive. The reference displacement was set to the sinusoidal wave with a frequency of 0.5 Hz and an amplitude of 20.0 μm. The cycle time of the every current pulse was set to 100 μs. The hysteresis was negligible in the current pulse drive. The average step a pulse was 6 nm/pulse, which is equivalent to a 13-bit DAC.

The frequency elements of the noise in the Δ conversion are squeezed out of the low frequency range in principle. In addition, the current pulse is integrated by the piezoelectric actuator. Therefore, the displacement mainly contained a frequency element of 0.5 Hz though the driving signals contained higher ones [9].

The dielectric loss by the current pulse drive is equal to that by the voltage linear drive and smaller than that by the voltage pulse drive [16].

4.2 Linearity

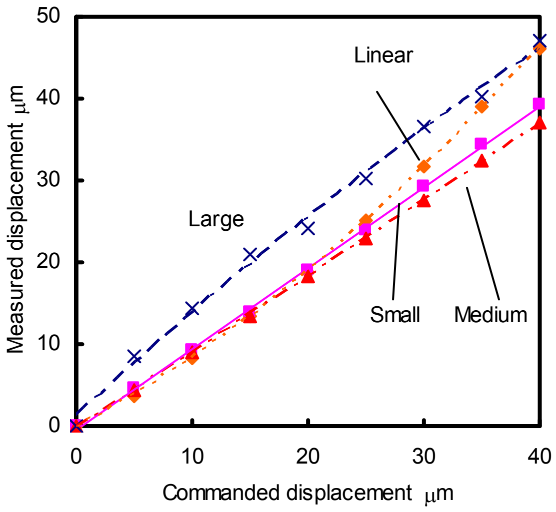
Because the step heights varied a little depending on the ambient conditions, the step heights were adjusted in each experiment at an amplitude of 20 μm both in the current pulse drive and voltage linear drive as shown in Table 1. All cycle times were set to 100 μs by adjusting the intervals.

The reference was changed from 5 to 40 μm. Figure 7 (a) shows displacements by step inputs. The displacement was linearly increased by the current pulse drive with the small and medium pulses. The resolution of the displacement by the large pulse is too coarse to drive the piezoelectric actuator precisely. The displacement by the voltage linear drive followed the hysteresis curve. Figure 7 (b) shows linearity errors. The errors were almost within the steps by the current pulse drive. The capacitance of the piezoelectric actuator is decreased with an increase of the terminal voltage or displacement. This also affected the linearity.

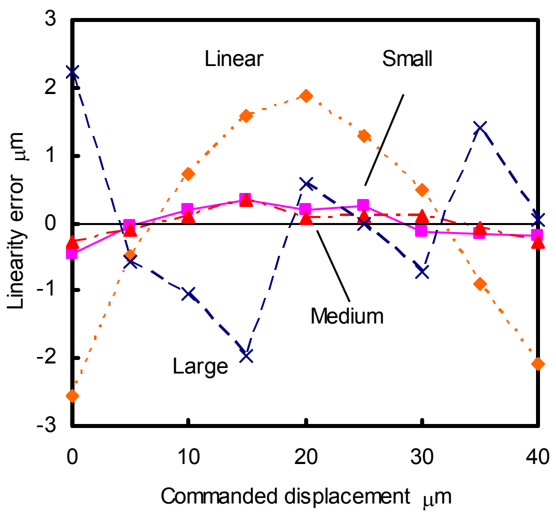
4.3 Creep

Figures 8 and 9 show the creep of the piezoelectric actuator with a rising and falling step input, respectively. To equalize the displacement each other, a step of 30 V was applied by the voltage linear drive and small pulses for 280 ms were provided by the current pulse drive.

The creep by the voltage linear drive was large at the beginning but became smaller after 30 s as shown in each Figure (a). The directions of the creep by the current pulse drive were in reverse as shown in each Figure (b). The creep in the initial 10 s was smaller than that by the voltage linear drive. In the case that the voltage amplifier was disconnected just after the rising or falling of the step voltage, the direction of the creep was similar to that by the current pulse drive as shown in each Figure (c). The decrease of the displacement



(a) Displacement.



(b) Linearity error.

Figure 7: Displacement by step inputs.

in each Figure (b) was larger than that in each Figure (c). The leak current in the driving circuit affected the stability of the displacement even though its output impedance is very high. The leak current to cancel the spontaneous polarization flows inside the piezoelectric actuator so that the displacement was decreased as shown in Figures (b) and (c).

The delay of the polarization causes the creep, and the reverse piezoelectric effect by the deformation of the piezoelectric actuator generates spontaneous charges. Because the voltage amplifier always provides the charge after the change of the applied voltage is ended, the spontaneous charges are cancelled. On the contrary, the charges are not cancelled by the current pulse drive because its output impedance is very high. Because the combination of the phenomena mentioned above with the leak current caused the displacement shown in Figures (b) and (c), the change was not monotonic.

5 SUMMARY

In this paper, a driving performance of a piezoelectric actuator by using the current pulse was compared with a voltage linear drive through experiments. Conclusions can be drawn as follows.

1. For a step input, the linearity of the steady-state values by the current pulse drive was better than that by the voltage drive because the hysteresis of the displacement to the applied voltage was small.
2. The direction of the creep was in reverse each other.
3. The leak current also affects the stability of the displacement.

6 ACKNOWLEDGMENTS

The authors would like to thank Mr. Katsuhiko Iida and Mr. Koichi Mori of Nano Control for their helpful cooperation.

This work was financially supported by Grant-in-Aid for Scientific Research on Priority Areas, “Actuator” (16078214), 2004-2008 by Ministry of Education, Culture, Sports, Science and Technology, Japan.

7 REFERENCES

[1] Xu, W. L., Han, L., 1999, Piezoelectric actuator based active error compensation of precision machining, Meas. Sci. Technol., 10/2, 106-111.

[2] Cuttino, J. F., Miller, Jr., A. C., Schinstock, D. E., 1999, Performance Optimization of a Fast Tool Servo for Single-Point Diamond Turning Machines, IEEE/ASME Trans. Mechatronics, 4/2, 169-179.

[3] Furuta, A., Munekata, M., Higuchi, T., 2002, Precise Positioning Stage Driven by Multilayer Piezo Actuator using Strain Gauge, Jpn. J. Appl. Phys., 41/1/10, 6283-6286.

[4] Furutani, K., Suzuki, M., Kudoh, R., 2004, Nanometer-cutting Machine Using Stewart-platform Parallel Mechanism, Meas. Sci. Technol., 15/2, 467-474.

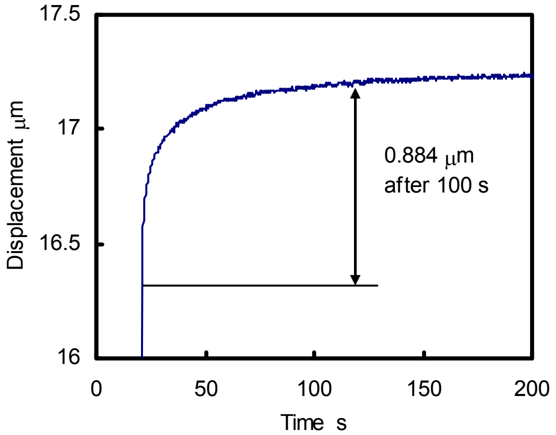
[5] Newcomb, C. V., Flinn, I., 1982, Improving the Linearity of Piezoelectric Ceramic Actuators, Electron. Let., 18, 442-444.

[6] Furutani, K., Urushibata, M., Mohri, N., 1998, Displacement Control of Piezoelectric Element by Feedback of Induced Charge, Nanotechnology, 9/2, 93-98.

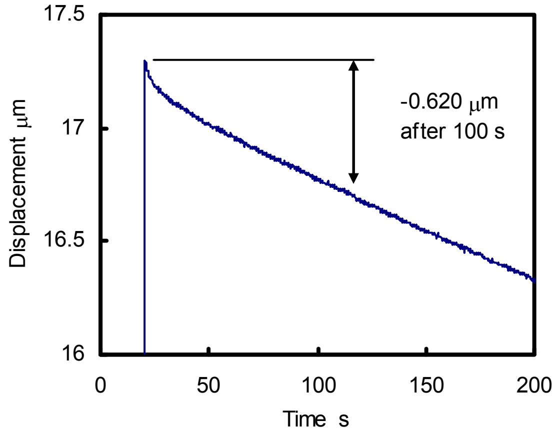
[7] Yi, K. A., Veillette, R. J., 2005, A Charge Controller for Linear Operation of a Piezoelectric Stack Actuator, IEEE Trans. Control Syst. Technol., 13/4, 517-526.

[8] Fukui, T., Segawa, M., Kurosawa, M., Oka, K., Higuchi, T., 1996, A control System with single-bit digital processing, Proc. Int. Conf. Control, Autom. Rob. Vision, Singapore, 1992-1996.

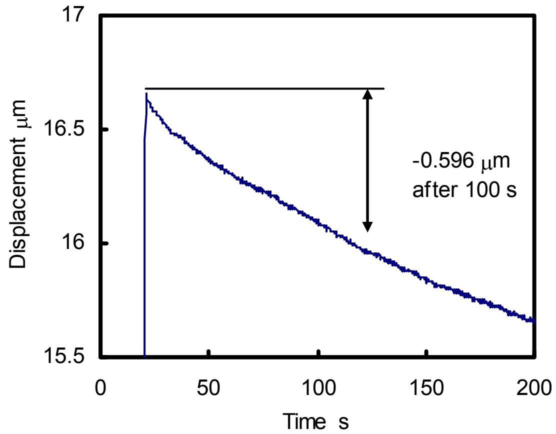
[9] Furutani, K., Iida, K., 2006, Driving method of piezoelectric actuator by using current pulse, Meas. Sci. Technol., 17/9, 2387-2394.



(a) Voltage linear drive.

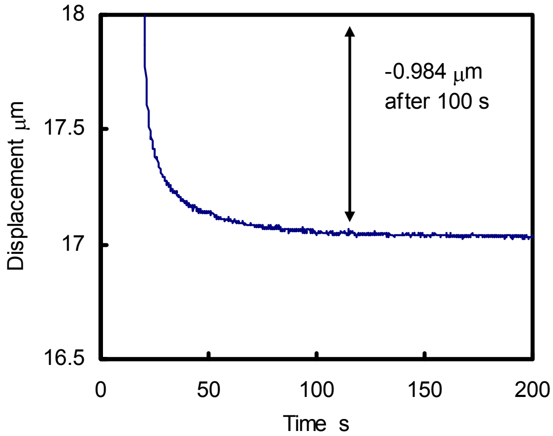


(b) Current pulse drive.

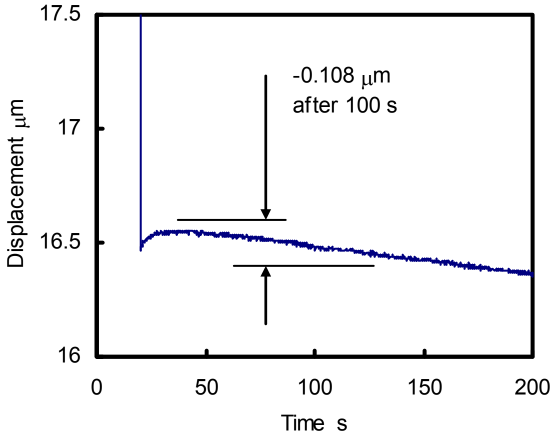


(c) Disconnection just after rising (voltage linear drive).

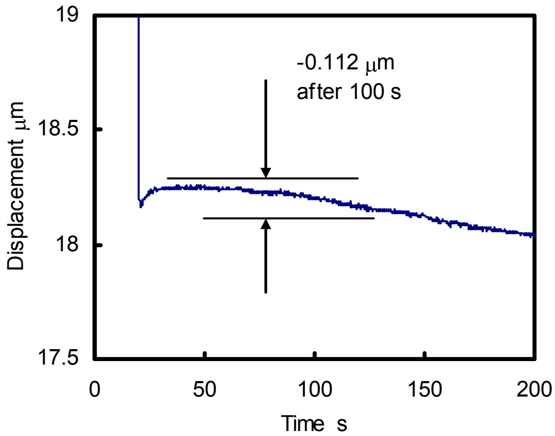
Figure 8: Creep with rising step input.



(a) Voltage linear drive.



(b) Current pulse drive.



(c) Disconnection just after falling (voltage linear drive).

Figure 9: Creep with falling step input.

[10] Furutani, K., Furuta, A., 2008, Evaluation of Driving Performance of Piezoelectric Actuator with Current Pulse, Proc. 10th Int. Workshop Adv. Motion Cont., Trento, Italy, 387-392.

[11] Martin, G. E., 1964, On the Theory of Segmented Electromechanical Systems, J. Acoust. Soc. Am., 37/7, 1366-1370.

[12] Ge, P., Jouaneh, M., 1995, Modelling Hysteresis in Piezoceramic Actuators, Precis. Eng., 17/3, 211-221.

[13] Goldfarb, M., Celanovic, N., 1997, A Lumped-Parameter Electromechanical Model for Describing the Nonlinear Behavior of Piezoelectric Actuators, ASME J. Dyn. Syst., Meas. Cont., 119/3, 478–485.

[14] Aziz, P. M., Sorensen, H. V., Van der Spiegel, J., 1996, An overview of Sigma-Delta Converters, IEEE Signal Process. Mag., 13/1, 61-84.

[15] Franklin, G. F., Powell, J. D., Emami-Naeini, A., 2006, Feedback Control of Dynamic Systems 5th Ed., Pearson Prentice Hall, Upper Saddle River, NJ, 159.

[16] Ikebe, Y., Nakada, T., 1974, On a Piezoelectric Flapper Type Servovalves Operated by a Pulse Width Operated Signal, Trans. Am. Soc. Mech. Engr., G, J. Dyn. Syst. Meas. Control, 96/1, 88-94.

Steady State Analysis of Gyroscopic power Generator

Jun Iwasaki, Tomohiro Ishii, Satoru Yoshikawa, Hiroshi Hosaka, Ken Sasaki
Dept. of Human and Engineered Environment, Graduate School of Frontier Sciences,
The University of Tokyo, Kashiwa, Japan

Abstract

Supplying electric power to information and communication devices is a critical issue for both current mobile networks and coming ubiquitous systems. To achieve this, we focus on dynamic energy conversion systems that utilize human vibrations in daily life. The gyroscopic power generator has a rotor, which moves three dimensionally and spins at a high speed from low frequency vibrations, such as human movements. In this paper, simple equations that indicate the relationships among input vibration, rotor movement, impedance of the output circuit, and the critical conditions for stable rotation are derived. Next, by measuring the phase difference between input vibration and rotor precession angle, and critical conditions for stability using prototype generators, validity of the theory is verified.

Keywords:
Gyroscopic power generator; Phase difference; Stability analysis

1 INTRODUCTION

To solve energy supply problems for mobile and ubiquitous equipment, many power generation systems utilizing energy in nature have been studied [1]. Generators using vibration of moving bodies is a prospective technology because low-level vibrations occur in many environments and they can generate electric power at various locations.

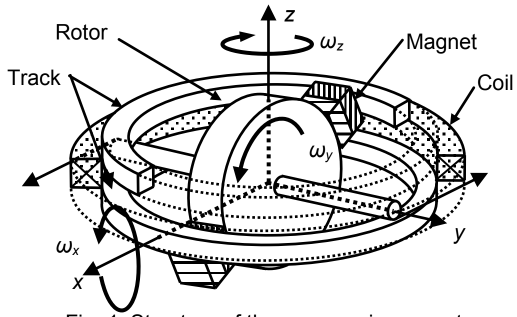
In general, their potential power is proportional to their oscillatory mass and the square of the oscillating velocity. Thus, it is difficult to produce generators that are small in size and can generate high power at a low vibration level.

To solve this problem, we focus on a gyroscopic power generator utilizing precession and friction of a rotating cylinder [2]. This generator can generate a large amount of power (>1 W) but has difficulty in achieving stable rotation. Thus, it is necessary to study its rotating state both theoretically and experimentally to clarify the stability design conditions and develop a practical device.

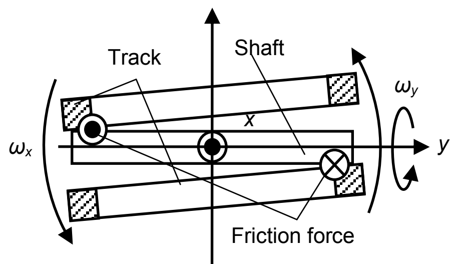
In this paper, the structure and power generation principle of the generator is explained. Next, a dynamic model of the device is presented and the relationships among input vibration, rotor movement, damping force to the rotor, and the critical conditions for stable rotation are derived. Finally, by measuring the phase difference between input vibration and rotor precession angle and critical conditions for stability using a prototype generator, validity of the theory is verified.

2 CONFIGURATION

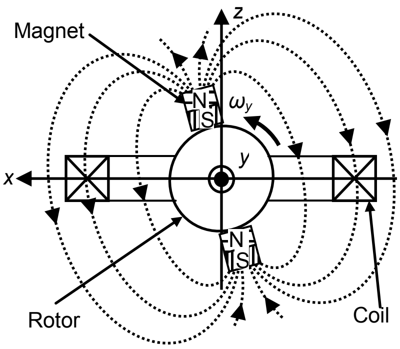
In this section, we describe the principle of the gyroscopic power generator (Figure.1). There is a rotor with magnets that rotates about the y-axis at ω_y and about the z-axis at ω_z. The shaft of the rotor is supported by tracks convolved with the coil. The track gap is slightly larger than the diameter of the shaft. When the track is rotated about the x-axis at ω_x, torque is produced in the rotor in the direction of ω_x, and the rotor starts precession movement at ω_z by the law of angular momentum. Then, the frictional force from the track is applied to the shaft. Figure 2 shows the generator as seen from the x-axis. The shaft contacts only one of the track sides because of the space between tracks. The frictional force generates torque in the direction where rotation ω_y of the rotor is increased. As a result, the rotor spinning speed increases as the track is rotated and vibrated. This principle of operation is used in the toy called Dynabee [3][4]. Figure 3 shows the generator as seen from the y-axis. Since the rotor shaft changes direction by precession, the coil and the rotor are arranged perpendicularly. By rotating the magnets, the magnetic field in the coil surface is changed and an alternating voltage is generated in the coil.



Fig, 1: Structure of the gyroscopic generator.



Fig, 2: Principle for increasing the spinning velocity



Fig, 3: Magnetic field distribution

3 THEORY

To analyze the intricate rotation, we define some coordinate systems (Figure.4). The right-handed orthogonal coordinate system *E* is an inertial coordinate system that consists of vectors $\boldsymbol{E_1}$, $\boldsymbol{E_2}$, and $\boldsymbol{E_3}$. The center of *E* is *O*, which is fixed at the center of the tracks. $\boldsymbol{E_1}$ and $\boldsymbol{E_2}$ are aligned in a horizontal plane and $\boldsymbol{E_3}$ is vertical to them. The track swing rotation by angle θ around $\boldsymbol{E_2}$ is expressed as

$$\theta = \theta_0 \sin \tau t \tag{1}$$

where θ_0 is the amplitude, and τ is the frequency of the input vibration. We defined vectors $\boldsymbol{e_{t1}}$, $\boldsymbol{e_{t2}}$, and $\boldsymbol{e_{t3}}$ from a right-handed orthogonal coordinate system e_t, which corotates with the track. Vector $\boldsymbol{e_{t2}}$ is matched to $\boldsymbol{E_2}$, and $\boldsymbol{e_{t3}}$ is aligned normal to the track's plane.

The rotor precesses around $\boldsymbol{e_{t3}}$. We defined vectors $\boldsymbol{e_1}$, $\boldsymbol{e_2}$, and $\boldsymbol{e_3}$ from a right-handed orthogonal coordinate system e, which corotates with the rotor precession. Vector $\boldsymbol{e_3}$ is matched with $\boldsymbol{e_{t3}}$, and $\boldsymbol{e_2}$ is located on the shaft. The top of the rotor axis contacts with an upper track at point P (Fig.5). The e_π system is rotated *e* by $-\beta$—the angle between OP and $\boldsymbol{e_1}$ around $\boldsymbol{e_2}$. $\boldsymbol{e_{\pi 2}}$ is matched to $\boldsymbol{e_2}$, and $\boldsymbol{e_{\pi 1}}$ is matched with OP. We define ξ as

$$\xi = \frac{R_t}{R_a} = \frac{1}{\tan \beta} \tag{2}$$

where R_a and R_t are the rotor shaft and the track radii, respectively. If it is assumed that the shaft rolls the track without sliding, the relationship between γ and α is

$$\dot{\gamma} = \xi \dot{\alpha} \tag{3}$$

where γ is the spin angle, and α is the precession angle. The angular velocity of the rotor $\boldsymbol{\omega_r}$ is given by the sum of the three rotation vectors; rotation of the e_t-system in the *E*-system, rotation of the *e*-system in the e_t-system and rotation of the rotor in the *e*-system. By transforming the result into *e*-system, $\boldsymbol{\omega_r}$ is given as follows.

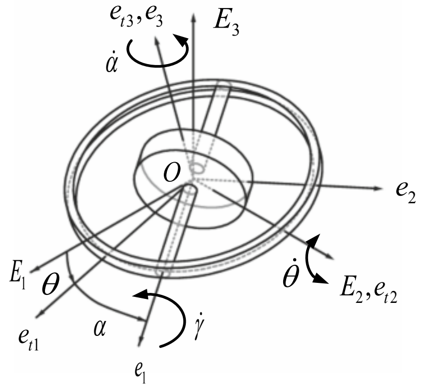
$$\omega_r = (\dot{\gamma} + \dot{\theta} \sin \alpha) e_1 + (\dot{\theta} + \cos \alpha) e_2 + \dot{\alpha} e_3 \tag{4}$$

The angular velocity $\boldsymbol{\Omega}$ of the *e*-system to *E*-system is

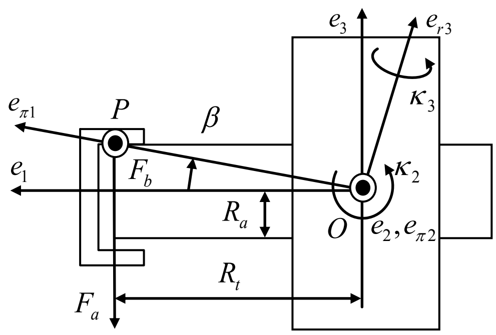
$$\Omega = (\dot{\theta} \sin \alpha) e_1 + (\dot{\theta} \cos \alpha) e_2 + \dot{\alpha} e_3 \tag{5}$$

Because the *e*-system corresponds to the principal axes of inertia of the rotor, the rotor angular momentum $\boldsymbol{L}$ is

$$L = I_1 (\dot{\gamma} + \dot{\theta} \sin \alpha) e_1 + I_2 (\dot{\theta} \cos \alpha) e_2 + I_2 \dot{\alpha} e_3 \tag{6}$$



Fig, 4: Track and rotor coordinates



Fig, 5: Definition of the $\boldsymbol{e_\pi}$ coordinate system

where $\boldsymbol{I_1}$ and $\boldsymbol{I_2}$ are the principal moment of inertia about the rotor principal axis and the orthogonal axis. The torque applied to the rotor from the track is given by

$$\kappa_2 = 2 F_a R_t \tag{7}$$

$$\kappa_3 = \frac{2 F_b R_t}{\cos \beta} \tag{8}$$

where F_a is the normal component of reaction, and F_b is the frictional force(Fig.5). The torque is represented by Eq. (9) using the moment against the rotation,

$$M = \kappa_2 e_{\pi 2} + \kappa_3 e_{\pi 3} - \sigma \dot{\gamma} e_1 \tag{9}$$

where σ is a damping coefficient for rotor spin caused by the electromagnetic force and mechanical resistance. Substituting Eqs. (3), (5), (6), and (9) into Euler's equation Eq. (10) gives Eq. (11).

$$M = \frac{\partial L}{\partial t} + \Omega \times L \tag{10}$$

$$\begin{aligned}
&\ddot{\alpha}\left(\xi^2 I_1 + I_2\right) + \xi I_1 \ddot{\theta} \sin\alpha \\
&+ \left(I_2 - I_1\right)\dot{\theta}^2 \sin\alpha \cos\alpha + \xi^2 \sigma \dot{\alpha} = 0 \\
&I_2 \ddot{\theta} \cos\alpha + \left(I_2 - 2I_1\right)\dot{\theta}^2 \dot{\alpha} \sin\alpha + I_1 \xi \dot{\alpha}^2 = 0 \\
&\ddot{\alpha}\xi\left(I_2 - I_1\right) - I_1 \ddot{\theta} \sin\alpha - \dot{\alpha}\dot{\theta}\left(1 + \xi\right)\cos\alpha \\
&+ \xi (I_2 - I_1)\theta \cos^2\alpha - \xi\sigma\dot{\alpha}\frac{\kappa_3}{\sin\beta} = 0
\end{aligned} \tag{11}$$

This is a nonlinear differential equation; although, an analytical solution is not obtained, an approximate solution in the steady state is obtained. In the steady state, the track vibration synchronizes with the rotor precession movement. If it is assumed that the rotation speed is constant, then

$$\alpha = \varphi + \tau t \tag{12}$$

is obtained, where ϕ is the phase difference between α and τt and a constant value. Substituting Eq. (12) and the relationship that the first and second differentials of ϕ are 0 into the first equation of Eq. (11) and integrating it in a vibration cycle, the following equation is obtained.

$$\cos\varphi = \frac{2\xi\sigma}{I_1 \tau \theta_0} \tag{13}$$

Other unknowns, κ_2 and κ_3, are determined from other equations obtained from of Eq. (9), which give the condition for non-slip rotation. This condition, however, stands automatically when the static friction coefficient is sufficiently large. Thus, we consider only Eq. (13) from now on. The next expression is necessary for Eq. (13) to have a solution, as $\cos\phi \le 1$.

$$\sigma \le \frac{I_1 \theta_0 \tau}{2\xi} \tag{14}$$

This expression gives the condition that the stationary rotation exists. Since the generated electric power is proportional to the damping coefficient σ when mechanical loss is neglected, the maximum σ given by Eq. (14) decides the optimum σ for the maximum power generation.

4 EXPERIMENTAL APPARATUS

To verify the theory in the preceding section, an experimental apparatus that generates sinusoidal vibration and measures precession angle of the rotor is constructed. A servomotor is used to give the generator a sinusoidal input vibration (Fig.6).

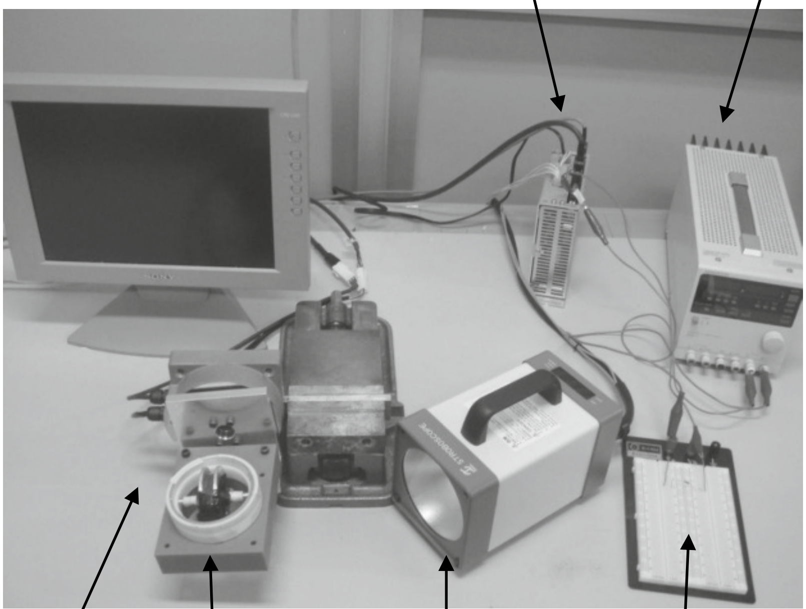
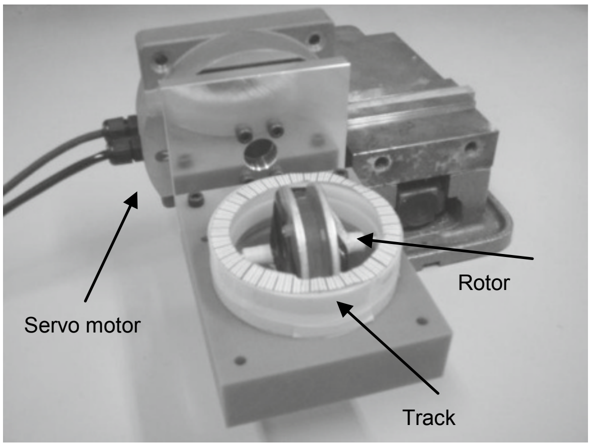
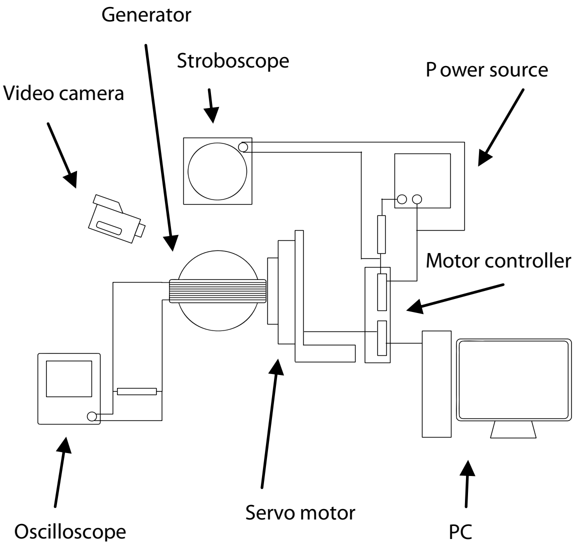


Figure 6 : Photo of the measurement system



Fig, 7: Sinuisoidal input oscillation device

To capture an instant position of gyroscopic precession, we developed stroboscope light from a servomotor pulse. If the rotor movement is in a steady state and the strobe flashes synchronized with the input vibration, we can see the rotor as if it were in a stationary state at a particular vibration phase. Using this apparatus (Figs.7 and 8), we set various parameters for rotation and study steady state operation.

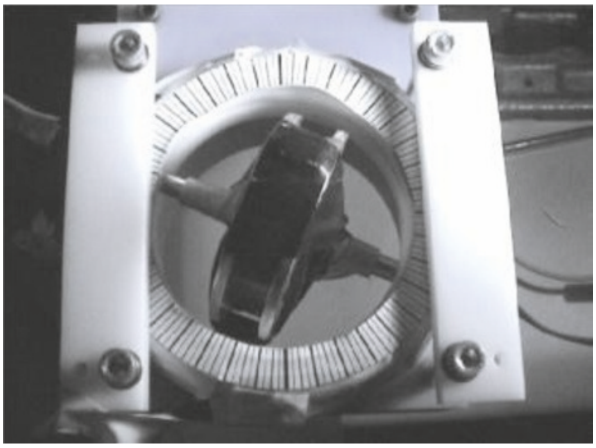


Fig, 8: Outline of the measurement system .

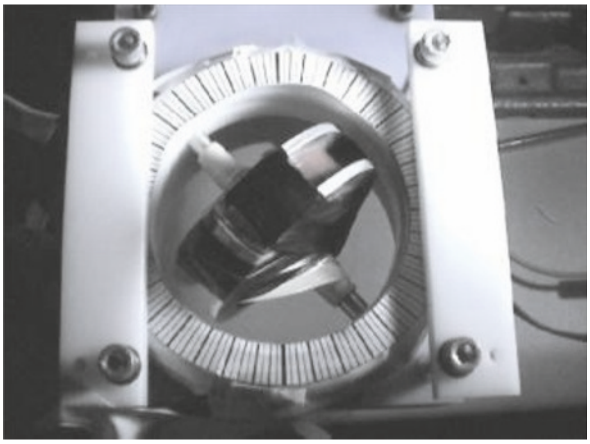
5 EXPERIMENTAL AND CALCULATED RESULTS

5.1 Phase difference

To verify Eq. (13), we measured the phase difference ϕ in the apparatus of Figs. 7 and 8 by changing the external resistance, number of coil turns, and servomotor vibration amplitude and frequency. The resistance and coil turns affect σ in Eq. (13). When each parameter meets the requirements of Eq. (14), the rotor can reach the steady state. Figures 9 and 10 show photographs taken at a steady state and when the input phase was zero (τt = $2n\pi$), where θ_0 was 35 degrees, τ was 2.5 Hz, and the external resistance R_o was 10 Ω. There are 150 coil turns N in Fig.9 and 25 in Fig.10.

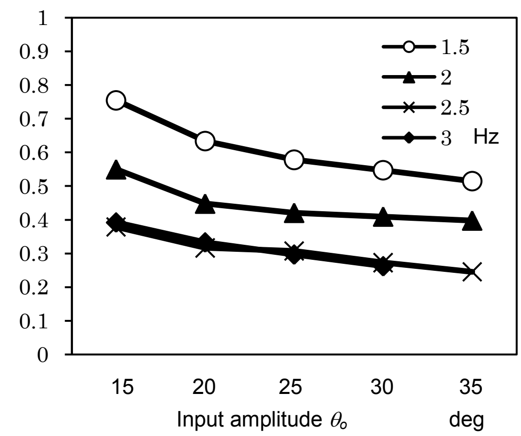


Fig,9: Photo at θ = 0 for N = 150.



Fig, 10: Photo at θ = 0 for N = 25.

Figure.11 shows a graph in which the measured phase difference is plotted when the input amplitude and frequency are changed. In this graph, the plot number is different in each parameter grouping. We can assume that the gyroscopic generator exceeds the stability limit in such cases. In the curves connecting measured points, cosϕ decreases as the input amplitude and the frequency increase. These results qualitatively agree with the theoretical result shown by Eq. (13).



Fig, 11: Relationships between Input amplitude and phase difference for various input frequency.

Next, to compare the experimental and the theoretical results quantitatively, we calculate σ using the voltage wave pattern of the generator (Fig,12). When T is the period of the rotor spin cycle and R_i is the internal resistance, the averaged electrical damping σ_e in a spin cycle is given by Eq. (15). Furthermore, mechanical damping σ_m is calculated as in Eq. (16) by using the measured phase difference ϕ_o at the open circuit condition. Then the total damping σ is given by Eq. (17).

$$\sigma_e = \frac{1}{(R_i + R_o)\xi^2\tau^2 T}\int_0^T V^2(t)dt \tag{15}$$

$$\sigma_m = \frac{I_1 \tau \theta_0}{2\xi}\cos\varphi_0 \tag{16}$$

$$\sigma = \sigma_e + \sigma_m \tag{17}$$

Figure 13 shows the measured and theoretical values of cosϕ. In the experiment, the number of coil turns N and the outer resistance R_o are changed to change electrical damping σ_e. σ_e increases as N increases and R_o decreases. In the calculation, σ was given by Eqs. (15) to (17) and the results are shown for R_o = 10Ω. In the experimental results, cosϕ increases as N increases and R_o decreases, that is as σ_e increases. This agrees qualitatively with Eq.(13). Furthermore, at R_o =10 Ω, the calculated results agree quantitatively with the experimental ones.

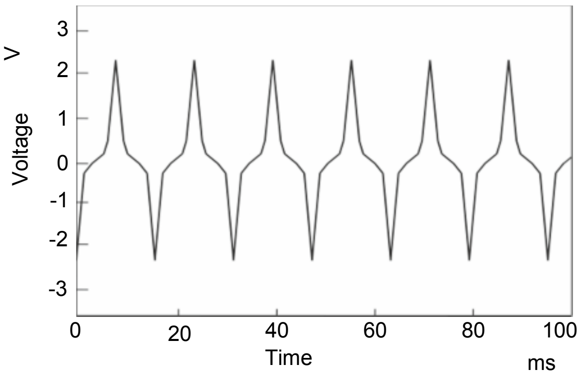


Figure 12: Generator output valtage .

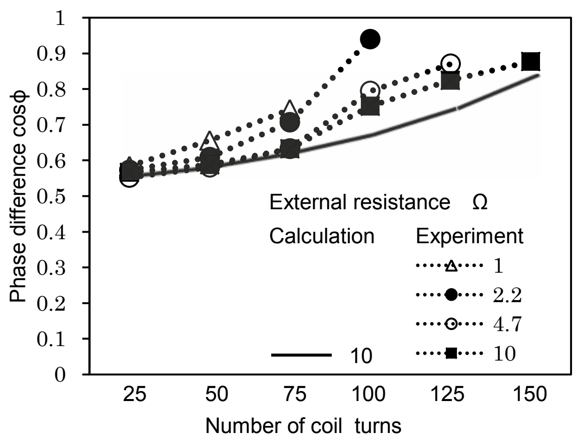


Figure 13: Relationchips between electrical Impedance and phase difference.

5.2 Stability limit

Finally, we verify the stability condition. From the equality condition of Eq. (14), Eq. (18) is obtained as a critical condition for stability.

$$\sigma = \frac{I_1 \tau \theta_{0\min}}{2\xi} \tag{18}$$

Since σ is obtained under both dynamic and electrical conditions, we verify the validity of Eq. (18) by comparing σ obtained by the two methods. σ is calculated dynamically by measuring I_1, θ_0, τ and ξ at the critical conditions. First, we let the servomotor produce vibrations large enough to provide stable rotation. Next, τ is maintained constant, and θ_0 is decreased gradually until stability is lost. At that moment, θ_o gives $\theta_0 min$, and σ is calculated from Eq. (18). Also σ is obtained electrically by Eqs.(15)-(17).

Figure 14 shows the damping constant σ at the critical conditions for various output resistances R_o at τ = 2 Hz. In both results, σ decreases as R_o increases and they coincide well.

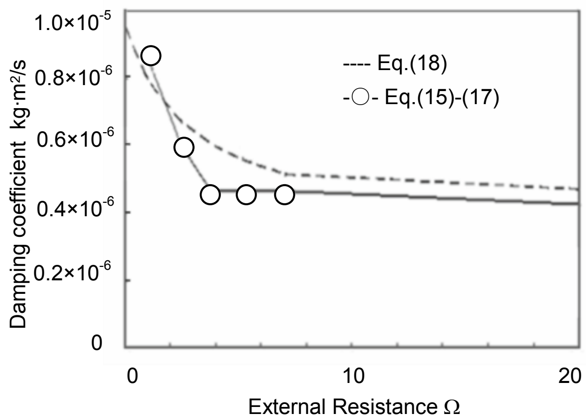


Figure 14: Relationships between damping constant and external resistance at critical condition

6 CONCLUSION

To realize a compact and high-power electric generator for mobile and ubiquitous equipment, we studied the dynamic characteristics of a gyroscopic generator. First, simple equations indicating the relationships among input vibration, rotor movement, damping constant, and the critical condition for stable rotation were derived. Next, by measuring the phase difference between the input vibration and rotor precession angle and the critical conditions for stability using a prototype generator, validity of the theory is verified.

These results show that the phase difference changes from 90 degrees to 0 as the dynamic conditions, such as the input vibration, and mechanical and electrical damping, come close to the stability limit.

The output power of the generator increases as the electrical damping increases. Thus, the electrical damping should be increased as much as possible and set to the stability limit. However, the critical damping is not constant and cannot be known beforehand, as the total damping includes mechanical damping, which changes with the driving conditions.

This study implies that the phase difference indicates the stability margin, and electrical damping should be set such that phase difference is slightly larger than zero. By measuring the phase difference, electrical damping is set to the optimum value without knowing the mechanical damping and other parameters. The electrical damping of the generator can be easily changed by changing the number of coil turns [5]. This characteristic is useful for driving the generators efficiently even when environmental and input conditions vary.

7 REFERENCES

[1] S. Roundy, P. Kenneth, J. M. Rabaey : Energy Scavenging for Wireless Sensor Networks with Special Focus on Vibrations, Kluwer Academic Publishers, Nowell, 2004.

[2] T. Ishii, Y. Goto, T. Ogawa and H. Hosaka : Study on gyroscopic power generator, J. of Japan Soc. Precision Eng., 74. 7, pp.764-768, 2007 (in Japanse).

[3] A. L. Mishler, Gyroscopic Device, United States Patent 3726146, Apr. 10, 1973

[4] D.W. Gulick, O.M.O'Reilly : On the Dynamics of the Dynabee, J. of Applied Mechanics, Vol.67, pp. 321-325, 2000

[5] K. Sasaki, Y. Osaki, J. Okazaki, H. Hosaka, K. Itao : Vibration-based automatic power-generation system, Microsystem Technologies, Vol.11, No.8-10, pp. 965-969, 2005

Characterization of Shape Memory Piezoelectric Actuator and Investigation of the Origin of the Imprint Electrical Field

Yoichi Kadota[1], Hiroshi Hosaka[1] and Takeshi Morita[1]

[1] Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Japan

Abstract

Shape memory piezoelectric actuator had been fabricated by inducing the imprint electrical field by applying dc field of 3.5 kV/mm in 150 ˚C environment for several hours. The actuator was operated by pulse shaped voltages of 100 ms width and ±130 V amplitude alternately. The actuator kept more than 150 μm shape memory even after 10^4 cycling. Moreover, a detailed investigation about the origin of this asymmetric phenomenon was carried out by measuring the strain hysteresis and polarization hysteresis simultaneously. These observations suggest that the irreversible polarization takes a major role for the asymmetry and imprint electrica field.

Keywords:

Piezoelectric Actuator; Shape Memory; Polarization Reversal

1 INTRODUCTION

Ferroelectric materials are widely utilized for various applications due to their multifunctional properties, such as ferroelectric random access memories (FeRAMs), piezoelectric actuators, sensors, and so on [1-4]. These applications are fall into two general categories in terms of switching of its polarization. One is FeRAM application that uses switching of their remnant polarization as 1-bit information. The other application is the application that never utilizes their polarization switching after the poling process. These applications use quite different principles of ferroelectric materials. In the latter case, for example, piezoelectric actuators use converse piezoelectric effect and are applied for precision positioning systems, fuel injection systems, ultrasonic actuators, micro actuators and etc [5-7]. In recent years, piezoelectric actuators are expected to be potential candidates for micro-electro-mechanical-system (MEMS) actuators because of their simple structure and high energy density [8-9]. However, there are some problems of piezoelectric actuator for application to MEMS device. One is that they need high operating voltages. Another one is that they need continuous driving voltage to keep a certain piezoelectric displacement.

To solve these problems, we have proposed the shape memory piezoelectric actuator which is operated by pulse shaped voltage reversing its polarization. This actuator is able to keep a certain piezoelectric displacement without any operating voltages. The pulse shaped voltage can be easily fabricated with one capacitor and transducer from low voltage source; then this actuator realizes low voltage operation and low energy consumption.

In the previous paper [10-11], the detailed characterization of the actuator and the exact origin of this phenomenon are not discussed yet. Therefore there is no design consideration to fabricate the shape memory piezoelectric actuator. In this report, the origin of the phenomenon was investigated and discussed using the results about the strain hysteresis curve measured with polarization hysteresis curve simultaneously.

2 PRINCIPLE

2.1 Memory effect with imprint electrical field

The principle of the shape memory piezoelectric actuator is based on the memory effect caused by imprint electrical field in ferroelectric materials. The principle of the memory effect is shown in Fig.1. Field dependence of the ferroelectric properties, polarization and strain, are shown with or without imprint electrical field. In the case of conventional ferroelectrics, spontaneous polarization is used for memory effect. On the other hand, their strain does not, because their strain hysteresis curve (butterfly curve) is symmetric and has only one stable strain condition regardless of the direction of its polarization. On the other hand, if the imprint electrical field exists in the ferroelectric materials, this property changes. The imprint electrical field is well known phenomenon in the field of ferroelectric thin film [12-14] as an offset inner field. This phenomenon causes serious failure of FeRAMs. Thus, the researches to prevent imprint behaviour have been intensively studied. However its origin is unclear yet. With imprint electrical field, the D-E hysteresis of the ferroelectric shifts to the direction of field axis and become asymmetric. At the same time, not only their D-E hysteresis, but also their butterfly curve shits to the direction of field axis and become asymmetric. With this asymmetricity, the ferroelectric materials become to have two distinct stable values of their strain at the point of zero fields according to their polarization state; this means memory effect of strain. Thus it can be the shape memory piezoelectric actuator by reversing its polarization.

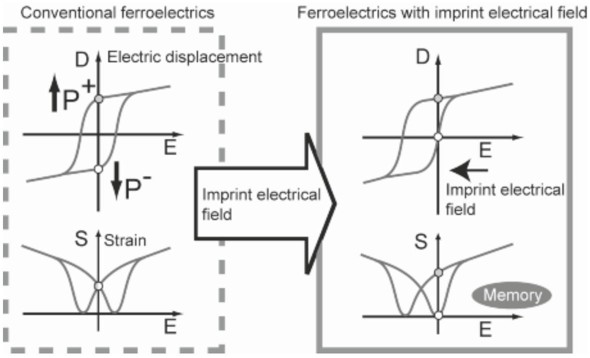


Figure 1. Principle of the memory effect with imprint electrical field

2.2 Pulse driving of the actuator

The operating sequence of the shape memory piezoelectric actuator by pulse shaped voltage is shown in Fig.2. In this illustration, a bending type piezoelectric actuator is shown. First, if the conventional ferroelectric actuator, which has symmetric butterfly curve, is operated by the pulse shaped voltage plus and minus alternately (the amplitude should be enough large to switch its polarization), the displacements at the point of zero voltage are same regardless of the polarization. Thus, the actuator doesn't have memory effect. On the other hand, if the actuator with imprint electrical field, which has asymmetric butterfly curve, is operated in the same fashion, the actuator has two distinct stable displacements at the point of zero voltage according to its polarization state. As a result, the actuator realizes the shape memory for the difference of two stable displacements.

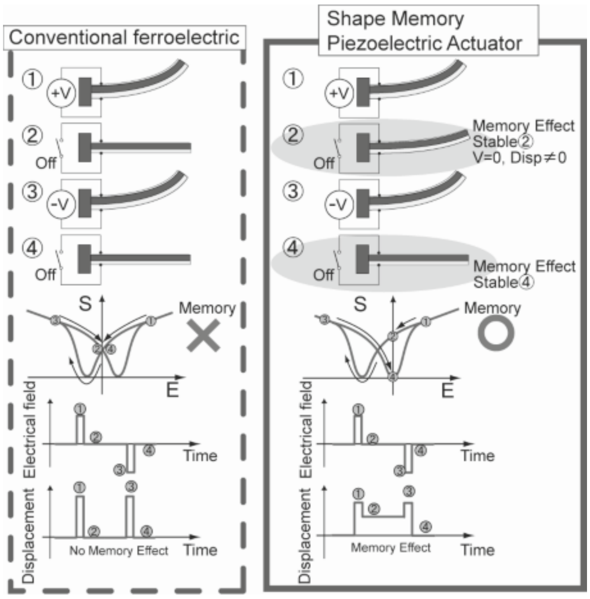


Figure 2. Pulse operation of shape memory piezoelectric actuator.

3 EXPERIMENTAL SET UP

3.1 Experimental instruments

Commercially available soft type PZT unimorph actuators (Nihon Ceratec, LPD3713X) were used for the experiments. The actuator is composed of a metal pate (0.2 mm thick, 37 mm long, 13.4 mm wide) and soft type PZT (0.2 mm thick, 28 mm long, 13.4 mm wide) on one side of the metal plate. To measure the butterfly curve of the actuator, one side of metal plate was clamped and was operated. The displacement of the actuator was measured by laser displacement sensor (Keyence, LC2450). Simultaneously, its polarization was measured by ferroelectric tester (Radiant, Precision LC). The picture of experimental setup is shown in Fig.3.

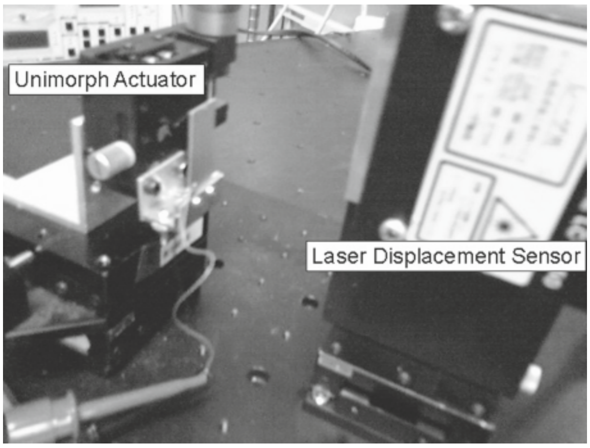


Figure 3. The picture of experimental setup

3.2 Imprint treatment

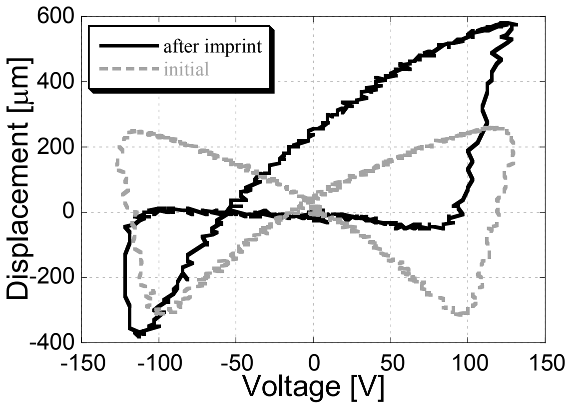
To induce the imprint electrical field to the actuator, high dc voltage of 3.5 kV/mm was applied for several hours in high temperature oven set at 150˚C. The actuator was put into the oven set at 150˚C and was electrically shorted to avoid the charges that were appeared by pyroelectric effect. Then, high voltage was applied, gradually increasing the voltage to +700 V. When the voltage became to +700 V, the time was started to be counted to set hours. After that, the oven was turned off and the actuator was cooled down to about 60˚C. It took 3 minutes for cooling. After cooling down, the applied voltage was gradually decreased to zero. It took about 1 minute. Then, the actuator was aged in room temperature before the measurement. For all experiments, the direction of the applied high field was from the top electrode to the bottom electrode.

4 CHARACTARIZATION OF ACTUATOR

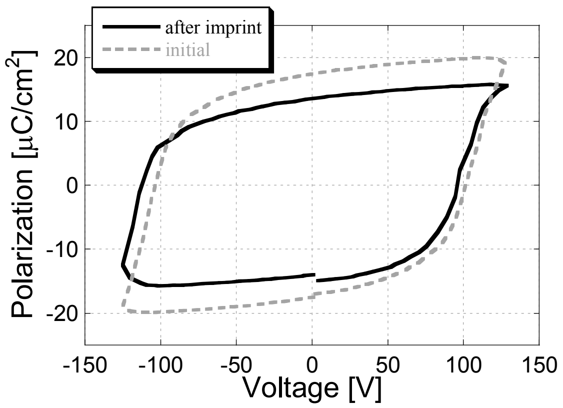
4.1 Change of the polarization hysteresis and butterfly hysteresis of the actuator by imprint treatment

A shape memroy piezoelectric actuator was fabricated by 4 hour imprint treatment. The actuator was operated by 260 V_{pp} ramp wave. The butterfly curve of the actuator just after the imprint was changeable by the applied field. Therefore, the history of the applied field is notable. To measure its butterfly curve and polarization hysteresis, the actuator was operated with two sequence of 260 V_{pp} ramp wave after several sets of pulse shaped voltages of 100 ms width and ±130 V amplitude had been applied. The results are shown in Fig.4 with the initial (unpoled) ones. The initial butterfly curve was symmetric and didn't have a memory effect of displacement. On the contrary, after the imprint treatment, its butterfly curve became asymmetric and had memory effect of piezoelectric displacement.

From the polarization hysteresis results, it was shown that the imprint electrical field had been induced by the high dc voltage treatment. However, it should be noted that the imprint treatment reduced the switchable polarizatoin of the actuator. This result indicates the origin of the asymmetric behavior is not only the imprint electrical field but also the change of switchable polarization. The origin of this phenomenon is discussed in the latter section.



(a)



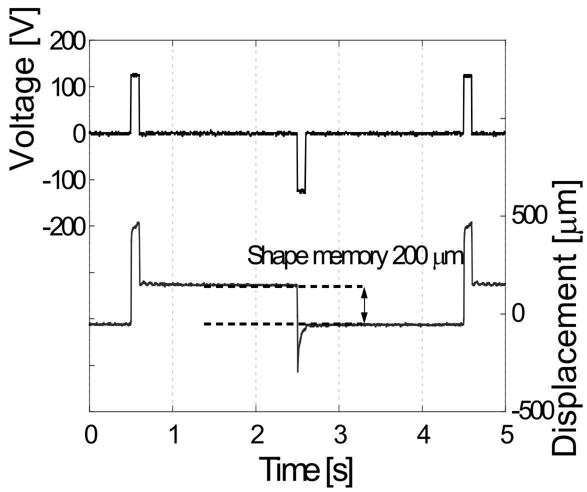
(b)

Figure 4. Butterfly curve of piezoelectric displacement(a) and polarization hysteresis (b) before and after the imprint treatment.

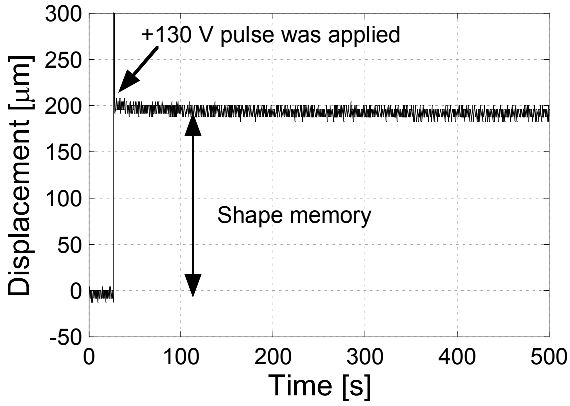
4.2 Pulse operation and time dependece

The fabricated shape memroy piezoelectric actuator was operated by the pulse shaped voltage of ±130V amplitude, 100 ms width. Fig.5 (a) shows the result of the pulse operation after some dozen times of ramp wave and pulse application. It is confirmed that the actuator had two stable displacements at the point of zero voltage and could keep the displacement without any applied voltages. This actuator realized the shape memory of 200 μm.

The time dependent change of the displacement is also shown in Fig.5 (b). After the application of plus pulse voltage of 130 V, the actuator could keep the displacement at least 500 second. The long range time dependence of the memory effect is on going work.



(a)



(b)

Figure 5. Pulse operation of the shape memory piezoelectric actuator.

4.3 Fatigue behavior by continuous cycling

That is already noted that the butterfly curve after the imprint treatment is changeable by the applied field. Therefore, it is reasonable that the shape memory gap changes during the continuous cycling of the polarization switching by the pulse shaped voltages. Then, the fatigue test was carried out with several hour imprint treated actuators. The actuators were operated by the continuous cycling of the pulse shaped voltage of ±130 V amplitude and 100 ms width. During the operation, the shape memory gap was measured. Fig.6 shows the results the dependency of the imprint treatment conditions. Almost all the sample showed similar fatigue tendency. The shape memory gap reduced in proportion to the logarithm of cycling number in the first stage of fatigue and after about several thousands cycling, the shape memory gap become to be stable at least about $5x10^4$ cycling.

Moreover, it was found that the longer impirnt time is important to reakuse stable shape memory gap.

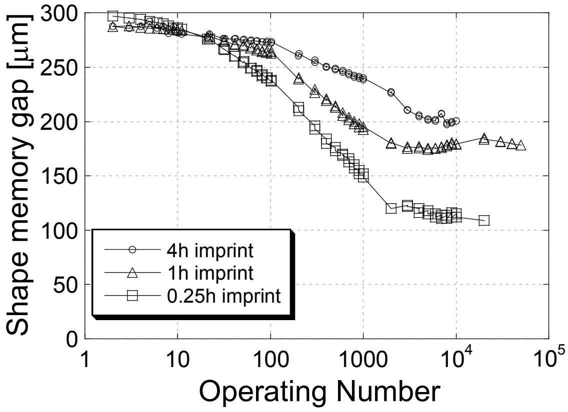


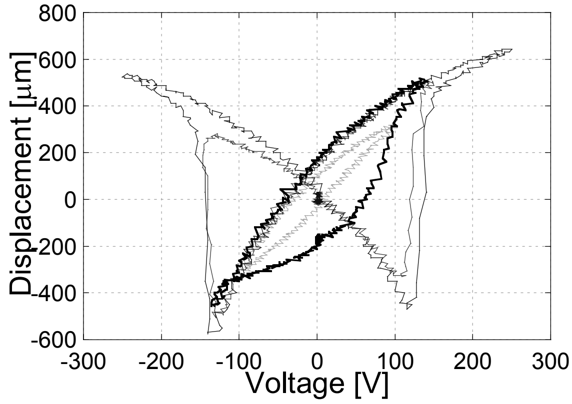
Figure 6. Fatigue propetis of shape memory gap by continous cycling of the polarization switching.

5 INVESTIGATION OF THE ORIGIN OF THE ASYMMETRIC BEHAVIOUR OF THE ACTUATOR

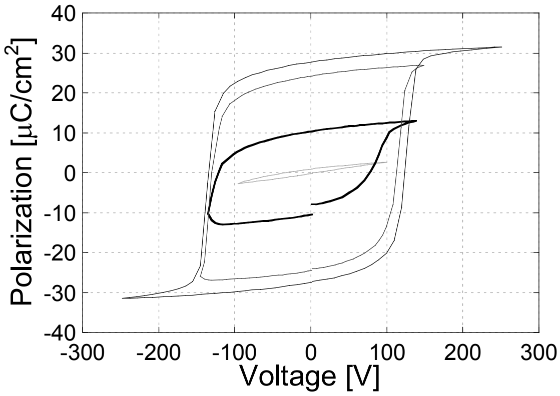
5.1 Imprint electrical field characteristics due to the applied field

In the above chapter, the shape memory piezoelectric actuator was characterized and the fatigue behaviour showed on the dependency of the imprint conditions or operating conditions. Therefore, there should be the optimum condition for realizing the high performance, such as large shape memory gap, high tolerance to fatigue, low voltage operation. To design the shape memory piezoelectric actuator, the detailed understanding of the imprint electrical field phenomenon is absolutely imperative. Thus, in this chapter, the detailed investigation of the asymmetric behaviour of the strain butterfly curve is discussed.

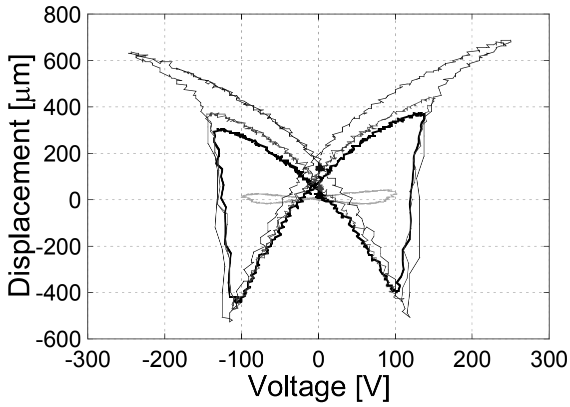
For the first step, the electrical field dependency of the piezoelectric strain and polarization were measured using the imprint treated sample and the virgin sample. For the whole measurement, the frequency of the applied ramp wave was fixed to 0.25 Hz and each wave was applied two times for each amplitude. The amplitude of applied voltages were from 200 V_{pp}, 260 V_{pp}, 270 V_{pp}, 280 V_{pp}, 290 V_{pp}, 300 V_{pp}, 310 V_{pp}, 320 V_{pp}, 340 V_{pp}, 400 V_{pp} and 500 V_{pp}. The results with applied voltage amplitude of 200 V_{pp}, 280 V_{pp}, 300 V_{pp} and 500 V_{pp} are only shown in Fig.7.



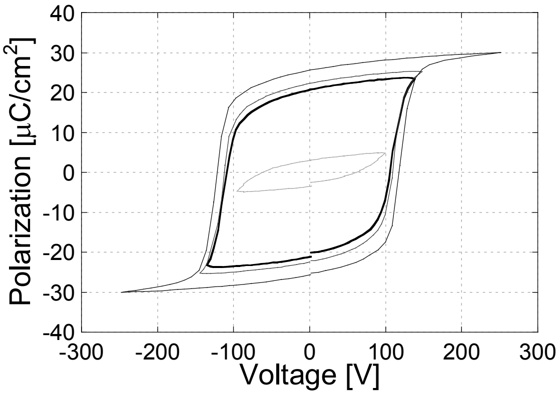
(a)



(b)



(c)



(d)

Figure 7. Piezoelectric displacement butterfly curves and polarization hysteresis of various applied voltage, (a), (b) after imprint treatment, (c), (d) initial condition.

5.2 Results and discussions about the imprint electrical field and polarization

Based on the displacement and polarization measurements of imprinted actuator shown in Fig.7, both of the hysteresis changed drastically in accordance with the incrementation of the applied voltage from 280 V_{pp} to 300 V_{pp}. When the applied voltage was 200Vpp, the displacement hysteresis didn't show polarization reversal. However, in the case of 280 V_{pp} their displacement hysteresis showed polarization reversal, which results in asymmetric butterfly curves. Moreover, the butterfly curve with application of 300 V_{pp} was remarkable because its butterfly curve suddenly returned to symmetric. Over 300 V_{pp} operation, all of the measured butterfly curves are nearly symmetric. Considering the measured polarization hysteresis, with over 300 V_{pp} application, the hysteresis were almost saturated. On the other hand, in the case of the virgin sample, almost all the measured butterfly curves were symmetric even from 280 V_{pp} to 300 V_{pp}. These results indicated that to realize the asymmetric butterfly curve, partial reversal of the polarization after imprint treatment plays an important role.

Next, the imprint electrical field was measured from the view of the polarization hysteresis. The imprint electrical field was defined as

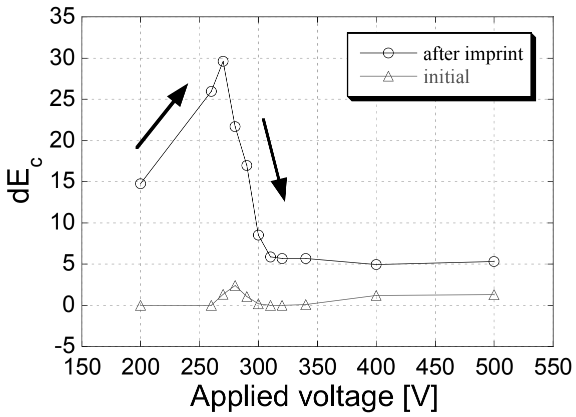
$$dE_c \equiv \left| \frac{E_c^+ + E_c^-}{2} \right|$$

where, E_c^+ is the positive coercive field and E_c^- is the negative coercive field, which were measured at the point of zero polarization in the polarization hysteresis curve. The relationship between the applied field and the imprint electrical field is shown in Fig.8 (a). It is apparent that the imprint electrical field didn't exist in the virgin sample but did in the imprint treated sample. Moreover, the imprint electrical field of the imprint treated sample had changed due to the amplitude of the applied voltages. Fig.8 (a) shows that the large imprint electrical field appeared when the applied voltage was near 270 V_{pp}. Furthermore, the imprint electrical field decreased drastically over 300 V_{pp}. This result related with the drastic change of the butterfly curve and polarization hysteresis. The difference of the imprint electrical field between the virgin sample and imprint treated one came from the existence of the offset polarization which didn't reverse under 300 V_{pp}. With the imprint treatment, the polarization of the sample was totally aligned. After that, the polarization partially switches with the application of about 270 V_{pp}. Then, the unswitchable polarization remained and it affected the behaviour of the imprint electrical field.

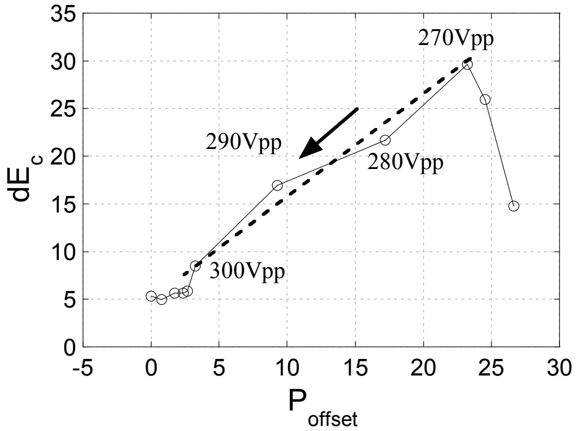
The offset polarization is defined as

$$P_{offset} = P_{\max} - P_r,$$

where, P_{max} is the switchable polarization with the application of 500 V_{pp}, P_r is the switchable polarization for each applied voltages. The relationship between the offset polarization and the imprint electrical field is shown in Fig.8 (b). The imprint electrical field indicated an approximately linear relationship with the offset polarization. The exceptions were the points with application voltage of 200 V_{pp} and 260 V_{pp}. In these cases, the switchable polarization was so small, then, the accurate determination of the imprint electrical field was difficult.



(a)



(b)

Figure 8. Relationships between the imprint electrical field and (a) applied, (b), offset polarization.

6 SUMMARY

In this paper, a detailed characterization of the shape memory piezoelectric actuator was performed. The shape memory piezoelectric actuator was fabricated by the imprint treatment and realized 200 μm shape memory operated by pulse shaped voltage of ±130 V amplitude and 100 ms width alternately. The shape memory effect continued for at least 500 seconds. The fatigue behaviour by continuous driving was investigated. In the first stage of fatigue, the shape memory gap of the actuator decreased proportional to the logarithm of cycling numbers. However, after 10^4 cycling, the change of the shape memory gap became to be stable and continued at least $5x10^4$ cycling.

The origin of the asymmetric butterfly curve and the imprint electrical filed was discussed using results of the butterfly hysteresis and the polarization hysteresis. It was suggested that there is offset polarization in the imprint treated sample and it takes major role for the imprint electrical field and the asymmetric butterfly curves.

7 ACKNOWLEDGMENTS

This research was partially supported by the Ministry of Education, Culture, Sports, Science, and Technology through a Grant-in-Aid for Scientific Research on Priority Areas, No. 438, “Next Generation Actuators Leading Breakthroughs”, as well as by the Murata Science Foundation, the Futaba Electrons Memorial Foundation and Japan Chemical Innovation.

8 REFERENCES

[1] T. Mikolajick Author, C. Dehm, W. Hartner, I. Kasko, M. J. Kastner, N. Nagel, M. Moert and C. Mazure, 2001, “FeRAM technology for high density applications”, *Microelectronics Reliability*, **41**, 947-950

[2] Paz de Araujo, C. McMillan, L. Joshi, V. Solayappan, N. Lim, M. Arita, K. Moriwaki, N. Hirano, H. Baba, T. Shimada, Y. Sumi, T. Fujii, E. Otsuki, T., “The future of ferroelectric memories”, 2000, *Solid-State Circuits Conference 2000*, 268-269

[3] A. H. Meitzeler and E. K. Sittig, “Characterization of Piezoelectric Transducers Used in Ultarasonic Devices Operating Above 0.1 GHz”, 1969, *Journal of applied physics*, **40**, 4341-4352

[4] F. Filhol, E. Defay, C. Divoux, C. Zinck, M. T. Delaye, “Resonant micro-mirror excited by a thin-film piezoelectric actuator for fast optical beam scanning”, 2005, *Sensors and Actuators A*, 123, 483-489

[5] T. Morita, "Miniature piezoelectric motor", *Sensors and Actuators*, 2000, **103**, .291-300

[6] Y. Lv and Y. Wei, “Study on Open-loop Precision Positioning Control of a Micropositioning Platform Using a Piezoelectric Actuator”, 2004, *Proceedings of the 5th World Congress on Intelligent Control and Automation*, 1255-1259

[7] T. Maeno, “Recent Progress of Ultrasonic Motors in Japan”, 2005, *The First International Workshop on Ultrasonic Motors and Actuators*

[8] H. C. Lee, J. Y. Park, and J. U. Bu, “Piezoelectrically actuated RF MEMS DC Contact Switches With Low Voltage Operation”, 2005, *IEEE Microwave and Wireless Components Letters*, **15**, 202-204

[9] J. Akedo, M. Lebedev and J. Park, “Aerosol Deposition and its Application to Micro Actuators”, 2006, *10th International Conference on New Actuators*, 70-73

[10] T. Morita, Y. Kadota and H. Hosaka, “Shape memory piezoelectric actuator”, 2007, *Applied Physics Letters*, **90**, 082909

[11] Y. Kadota, H. Hosaka and T. Morita, “Utilization of Permittivity Memory Effect for Position Detection of Shape Memory Piezoelectric Actuator”, 2008, *Japanese Journal of Applied Physics*, *Brief Communications*, **47**, 217-219

[12] T. Morita and Y. Cho, "Piezoelectric property of an epitaxial lead titanate thin film deposited by the hydrothermal method", 2006, *Applied Physics Letters*, **88**, 112908

[13] D. Nagasawa and H. Nozawa, “Imprint Model Based on Thermionic Electron Emission Under Local Fields in Ferroelectric Thin Film”, 1999, *Japanese Journal of Applied Physics*, **38**, 5406-5410

[14] G. L. Rhun, R. Bouregba and G. Poullain, “Polarization lop deformations of an oxygen deficient $Pb(Zr_{0.25},Ti_{0.75})$ ferroelectric thin film”, 2004, *Journal of Applied Physics*, **96**, 5712-5721

A Study on Real-time Scheduling for Holonic Manufacturing Systems - Application of Reinforcement Learning -

Koji Iwamura[1], Norihisa Mayumi[1], Yoshitaka Tanimizu[1], Nobuhiro Sugimura[1]

[1] Dept. of Mechanical Engineering, Graduate School of Engineering, Osaka Prefecture University, Osaka, Japan

Abstract

A real-time scheduling method based on utility values has been proposed and applied to the holonic manufacturing systems (HMS), in the previous paper. In the proposed method, all the job holons and the resource holons firstly evaluate the utility values for the cases where the holon selects all the candidate holons for the next machining operations. The coordination holon secondly determine a suitable combination of the resource holons and the job holons, based on the utility values. Reinforcement learning is newly proposed and implemented to the individual job holons and resource holons, in order to improve their capabilities for evaluating the utility values.

Keywords:

Holonic Manufacturing Systems; Real-time Scheduling; Reinforcement Learning; Coordination

1 INTRODUCTION

Recently, new architectures of manufacturing systems have been proposed to realize flexible control structures of manufacturing systems, which can cope with dynamic changes in volume and variety of products and also unforeseen disruptions, such as malfunctions of manufacturing resources and interruptions by high priority jobs. They are so called as ADMS (Autonomous Distributed Manufacturing Systems) [1], BMS (Biological Manufacturing Systems) [2], and HMS (Holonic Manufacturing Systems) [3] [4]. This paper deals with a real-time scheduling system of the HMS. In the previous paper, a real-time scheduling method based on utility values has been proposed and applied to the HMS [3]. The holons in the HMS are divided into three classes based on their roles in the manufacturing processes and the scheduling processes.

(a) Resource holons: They transform the job holons in the manufacturing process. In the scheduling process, they evaluate the utility values for the candidate job holons which are processed by the resource holons in the next time period.

(b) Job holons: They are transformed by the resource holons from the blank materials to the final products in the manufacturing process. In the scheduling process, they evaluate the utility values for the candidate resource holons which carry out the machining operations in the next time period.

(c) Coordination holon: It selects a most suitable combination of the resource holons and the job holons for the machining operations in the next time period, based on the utility values sent from the resource holons and the job holons.

Reinforcement learning is newly proposed and implemented to the individual job holons and resource holons, in order to improve their capabilities. A reinforcement learning method was proposed and applied to centralized scheduling problems for semiconductor manufacturing processes [5]. In the present research, a reinforcement learning is applied to the distributed scheduling processes for the manufacturing processes of machine products.

2 REAL-TIME SCHEDULING METHOD FOR HMS [3]

2.1 Information for real-time scheduling

It is assumed here that the individual job holons have the following technological information.

- M_{ik}: k-th machining operation of the job holon i. (i = 1,..,α), (k = 1,..,β).
- AC_{ik}: Required machining accuracy of machining operation M_{ik}. It is assumed that the machining accuracy is represented by the levels of accuracy indicated by 1, 2, and 3, which mean rough, medium high, and high accuracy, individually.
- R_{ikm}: m-th candidate of resource holon, which can carry out the machining operation M_{ik}. (m = 1,...,γ).
- W_i: Waiting time until the job holon i becomes idle if it is under machining status.

The individual resource holons have the following technological information.

- T_{ikm}: Machining time in the case where the resource holon R_{ikm} carries out the machining operation M_{ik}.
- MAC_{ikm}: Machining accuracy in the case where the resource holon R_{ikm} carries out the machining operation M_{ik}. MAC_{ikm} is also represented by the levels of 1, 2 and 3.
- MCO_{ikm}: Machining cost in the case where the resource holon R_{ikm} carries out the machining operation M_{ik}.
- W_{ikm}: Waiting time until resource holon R_{ikm} becomes idle if it is under machining status.

2.2 Real-time scheduling process based on utility values

A real-time scheduling process based on the utility values have been proposed, in the previous research [3], to select a suitable combination of the job holons and the resource holons which carries out the machining operation in the next time period.

At the time t, all the 'idling' holons have to select their machining schedules in the next time period. The following procedure is proposed for the individual holons to select their machining schedules.

(1) Retrieval of status data

The individual 'idling' holons firstly get the status data from the other holons which are 'operating' or 'idling'. The 'idling' holons can start the machining operation in the next time period.

(2) Selection of candidate holons

The individual 'idling' holons select all the candidate holons for the machining operations in the next time period. For instances, the job holon i selects the resource holons which can carry out the next machining operation M_{ik}. On the other hand, the resource holon j select all the candidate job holons which can be machined by the resource holon j.

(3) Determination of utility values

The individual 'idling' holons determine the utility values for the individual candidates selected in the second step. For instances, the job holon determines the utility values, based on its own decision criteria for all the candidate resource holons which can carry out the next machining operation. The utility values are given as follows.

$JUV_i(j)$ ($0 \leq JUV_i(j) \leq 1$): Utility value of the candidate resource j for the job holon i.

$RUV_j(i)$ ($0 \leq RUV_j(i) \leq 1$): Utility value of the candidate job i for the resource holon j.

(4) Coordination

All the 'idling' holons send the selected candidates and the utility values of the candidates to the coordination holon. The coordination holon determine a suitable combination of the job holons and the resource holons which carry out the machining operations in the next time period, based on the utility values. The decision criteria of the coordination holon is to maximize the total sum of the utility values of all the holons.

2.3 Real-time scheduling process based on utility values

The utility values are evaluated based on the decision criteria of the individual holons, and various decision criteria are considered for the holons. Therefore, it is assumed that the individual holons have one of the objective functions shown in Table 1 for evaluating the utility values.

The following procedures are provided for the resource holons to evaluate the utility values. Let us consider a resource holon j at a time t. It is assumed that $TT_{j \cdot t}$, $ME_{j \cdot t}$, and $MA_{j \cdot t}$ show the total time after the resource holon j starts its operations, the efficiency, and the evaluated value of machining accuracy of the resource holon j, respectively. If the resource holon j selects a candidate job holon i for carrying out the machining operation M_{ik}, the efficiency and the evaluated value of the machining accuracy are estimated by the following equations.

$$ME_{j \cdot t+1}(i) = (ME_{j \cdot t} \cdot TT_{j \cdot t} + T_{ikj})/(TT_{j \cdot t} + T_{ikj} + W_i) \quad (1)$$

Table 1: Objective functions of holons.

Objective functions		Objective function values
Resource holon	Efficiency	ΣMachining time / Total time
	Machining accuracy	Σ(Machining accuracy of resources – Required machining accuracy of jobs)
Job holon	Flow-time	Σ(Machining time + Waiting time)
	Machining cost	Σ(Machining cost of resources)

$$MA_{j \cdot t+1}(i) = MA_{j \cdot t} + (MAC_{ikj} - AC_{ik}) \quad (2)$$

where, the resource holon j can carry out the machining operation M_{ik} of job holon i ($j = R_{ikm}$).

As regards the job holons, the following equations are applied to evaluate the flow-time and the machining costs, for the case where a job holon i selects a candidate resource holon j ($= R_{ikm}$) for carrying out the machining operation M_{ik}. It is assumed that $JT_{i \cdot t}$ and $JC_{i \cdot t}$ give the total time after the job holon i is inputted to the HMS and the machining cost, respectively.

$$JT_{i \cdot t+1}(j) = JT_{i \cdot t} + T_{ikj} + W_{ikj} \quad (3)$$

$$JC_{i \cdot t+1}(j) = JC_{i \cdot t} + MCO_{ikj} \quad (4)$$

The objective functions mentioned above have different units. Some of them shall be maximized and others shall be minimized. Therefore, the utility values are normalized from 0 to 1.

3 APPLICATION OF REINFORCEMENT LEARNING

Reinforcement learning is newly proposed and implemented to the individual job holons and resource holons, in order to improve their capabilities. In the reinforcement learning method [6], an agent must be able to sense the status of the environment to some extent and must be able to take actions that affect the status. The agent also must have a goal or goals relating to the status of the environment.

Figure 1 summarizes the reinforcement learning procedure proposed here. The individual job holons and resource holons carry out the following four steps to obtain their suitable decision criteria for evaluation of the utility values by applying the reinforcement learning.

Step1 The individual job holos and resource holons carry out the real-time scheduling process described in section 2.2, when their previous machining operations are finished. The real-time scheduling process (1) and (3) are modified as following for implementation of reinforcement learning.

(1) Retrieval of status data

The individual 'idling' holons get the status data from the other holons which are 'operating' or 'idling', and observe the status s of the manufacturing systems.

(3) Determination of utility values

The Individual job holons and resource holons execute the action a based on the value $Q(s, a)$, to evaluate the utility values for all the candidate machining operations in the next time period. Where, s and a represent the status and actions in the reinforcement learning method.

Step2 The real-time scheduling process are repeated until all the machining operations of the job holons are finished by the resource holons in the HMS.

Step3 The individual job holons and resource holons obtain the reward r based on their own objective function values, and calculate the value $Q(s, a)$.

Step4 Step1 to Step3 are repeated for the new job holons to be manufactured in the manufacturing systems, in order to converge the value $Q(s, a)$ of the individual job holons and resource holons.

In these steps, the status s, the action a and the reward r are given as follows.

(1) Status s

The status s observed by the job holons and the resource holons is represented by the following equation, in the present research.

$$s = (s_1, s_2, s_3, s_4) \tag{5}$$

where, s_p (p =1, 2, 3, 4) are the number of 'idling' holons, that have the objective functions of efficiency, machining accuracy, flow-time, and machining cost, respectively. This means that the learning process of the individual holons are carried out based on these status.

(2) Action a

The individual job holons and resource holons determine the parameter n (= 1/5, 1/3, 1, 3, 5) in the following equation to evaluate the utility values.

$$UV' = (UV)^n \tag{6}$$

where, UV is the utility value calculated by the individual job holons and resource holons described in section 2.3.

ε-greedy method [6] is applied for the individual job holons and resource holons to determine the action a.

(3) Reward r

The individual job holons and resource holons obtain the reward r given by following equations.

$$r = (1/4) \sum_{p=1}^{4} [\sum_{h=1}^{\tau} r_h / \tau]_p \tag{7}$$

where, p and τ are the ID of objective functions, and the total number of holons with p-th objective functions of holons. r_h is calculated by following equations, based on its own objective functions.

(a) For the case that the objective function is efficiency

$$r_h = (a_h - b_h) / b_h \tag{8}$$

(b) For the case that the objective function is either machining accuracy, flow-time or machining cost

$$r_h = (b_h - a_h) / b_h \tag{9}$$

where, a_h and b_h are the objective function values for the cases where the individual holons h evaluate the utility values based on the proposed method with the reinforcement learning, and the method without the reinforcement learning.

The value $Q(s, a)$ is determined by applying the monte carlo method [6]. The individual job holons and resource holons save the n rules (s_t, a_t) (t = 0, 1, …, n-1) between the time when they obtain the reward r and the time when they obtain the new reward r. The rule (s, a) means the set of status s and action a. The value $Q(s, a)$ is calculated by the following equations.

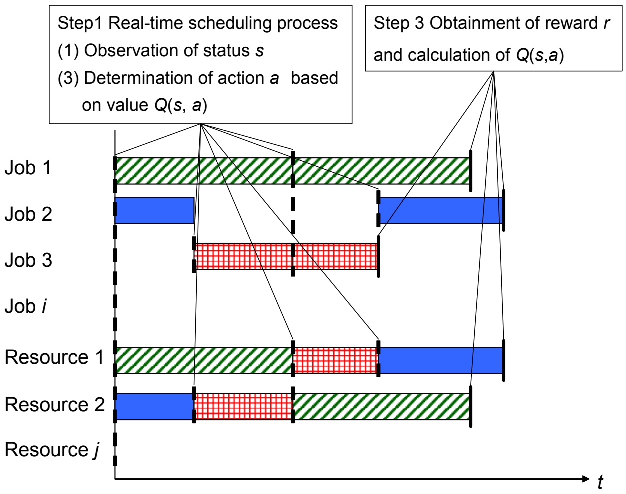


Figure 1: Application of reinforcement learning.

$$SumReward(s_t, a_t) \leftarrow SumReward(s_t, a_t) + r \tag{10}$$

$$Q(s, a) \leftarrow SumReward(s, a) / RewardCount \tag{11}$$

where, $SumReward(s, a)$ is the cumulative rewards in the case where the action a is carried out in the status s. $RewardCount$ is the total number in the case where the rule (s, a) obtain the reward r.

4 CASE STUDY

Some case studies have been carried out to verify the effectiveness of the proposed methods. The HMS model considered in the case studies has 10 resource holons. The individual resource holons have the different objective functions and the different machining capacities, such as the machining time T_{ikm}, the machining accuracy MAC_{ikm}, and the machining cost MCO_{ikm}.

As regards the job holons, 16 job holons are considered in the case study, which have the different objective functions and the machining sequences. It is assumed that the same job holons are inputted to the HMS after the resource holons finish all the manufacturing processes.

5 cases are considered, in the case study, by changing the machining capacities of the resource holons. ε is set to 0.2 for the ε-greedy method.

Figure 2 shows a typical example of the results. In the figure, the horizontal and vertical axes show the *episode* and the improvement ratio λ, respectively. The *episode* means here the number of repetitions of all the manufacturing processes of the inputted jobs. The improvement ratio λ means the ratio between the objective function values of all the holons obtained by the proposed method and the ones by the method without reinforcement learning. λ is calculated by following equation.

$$\lambda = \sum_{h=1}^{\nu} \mu_h / \upsilon \quad (12)$$

where, μ_h and υ are the improvement ratio of the objective function values of holon h and the total number of holons, respectively. The μ_h is calculated by the following equation based on the type of the objective functions.

(a) For the case that the objective function is efficiency

$$\mu_h = b_h / a_h \quad (13)$$

(b) For the case that the objective function is either machining accuracy, flow-time or machining cost

$$\mu_h = a_h / b_h \quad (14)$$

where, a_h and b_h are the objective function values of the individual holons h obtained by the proposed method and the previous method. As shown in the figure, the improvement ratio λ is converged until the *episode* reaches to 100.

Figure 3 shows the average improvement ratio λ*average* of the same case. Following equation gives the λ*average* which means the average of improvement ratio λ until the *episode* reaches to ω.

$$\lambda average = \sum_{epsiode=1}^{\omega} \lambda_{episode} / \omega \quad (15)$$

where, $\lambda_{episode}$ is the improvement ratio λ at the *episode*.

Table 2 summarizes the result of all the cases from the view point of the average improvement ratio. As shown in the table, the individual job holons and resource holons obtain the suitable decision criteria for evaluation of utility values and improve their objective function values.

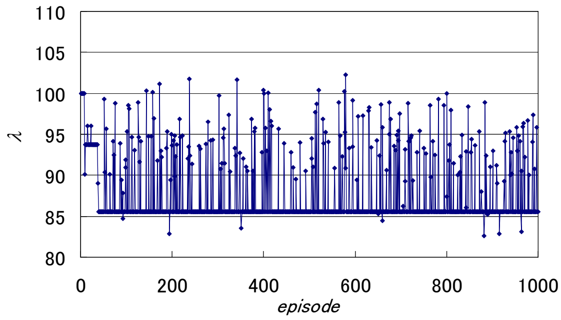


Figure 2: Improvement ratio.

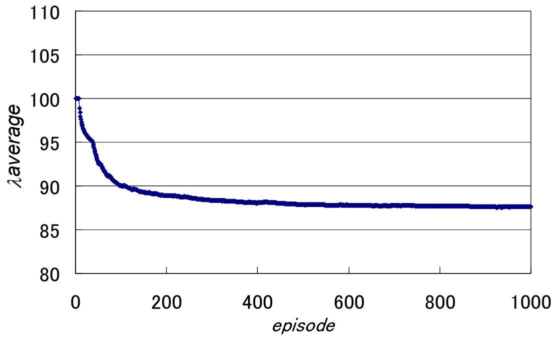


Figure 3: Average improvement ratio.

Table 2: Results of case study.

ε	50	100	150	200
Case 1	92.7	90.0	89.3	88.9
Case 2	97.8	97.5	97.6	97.5
Case 3	98.1	97.7	97.6	97.5
Case 4	98.7	97.6	97.3	97.1
Case 5	90.9	89.8	89.1	88.7

5 CONCLUSIONS

New systematic methods are proposed here to improve capabilities of the individual job holons and resource holons based on the reinforcement learning. The following remarks are concluded.

(1) The real-time scheduling process are modified for implementation of reinforcement learning in order to obtain the suitable decision criteria for evaluation of utility values.

(2) The status, the action and the reward are defined for the individual job holons and the resource holons to evaluate the suitable utility values based on the status of the HMS.

(3) Some case studies of the real-time scheduling have been carried out to verify the effectiveness of the proposed methods in comparison with the previous method. It was shown, through case studies, that the proposed methods are effective to improve the objective function values of the individual holons.

6 REFERENCES

[1] Moriwaki, T., and Sugimura, N., 1992, Object-oriented modeling of autonomous distributed manufacturing system and its application to real-time scheduling, Proc. of the ICOOMS '92, 207-212.

[2] Ueda, K., 1992, An approach to bionic manufacturing systems based on DNA-type information, Proc. of the ICOOMS '92, 303-308.

[3] Iwamura, K., Okubo, N., Tanimizu, Y. and Sugimura, N., 2006, Real-time scheduling for holonic manufacturing systems based on estimation of future status, International journal of production research, 44/18-19:3657–3675.

[4] Iwamura, K., Nakano, A., Tanimizu, Y., Sugimura, N., 2007, A study on real-time scheduling for holonic manufacturing systems -Simulation for estimation of future status by individual holons-, LNAI 4659, Holonic and multi-agent systems for manufacturing, Springer, 205–214.

[5] Fujii, N., Takasu, R., Kobayashi, M. and Ueda, K., 2005, Reinforcement learning based product dispatching scheduling in a semiconductor manufacturing system, Proc. of the 38th CIRP International seminar on manufacturing systems, CD-ROM.

[6] Sutton, R. and Barto, A., 1998, Reinforcement learning: An introduction, Bradford Books.

Flexible Production Systems developed and utilized in DENSO CORPORATION and their Evaluation

Shigeru Harashima[1], Katsuhisa Ohno[2]

[1] Production Engineering Department, DENSO CORPORATION, Kariya, Japan

[2] Faculty of Management and Information Science, Aichi Institute of Technology, Toyota, Japan

Abstract

As the global environment changes rapidly and the information technology develops quickly in recent years, requirements for the production become diverged widely and sophisticated highly. The flexible production systems developed in DENSO are evaluated by the performance function 'Life Cycle Cost' of the production volume and compared with conventional systems. The results show that the flexible production systems contribute to the management of the manufacturer.

Keywords:

Changeability, Reconfigurability, Life Cycle Cost

1 INTRODUCTION

The flexibility of the production has been extended through single manufacturing machines, manufacturing lines and factories. This is one of the key enablers for meeting the challenges of the global competition.

Researches on the flexibility of the production are roughly classified into (1) the concept (framework), (2) the design, (3) the case study and (4) the evaluation technique, but there seem to be a few researches based on the practical production systems in use.

Because of its complexity it needs some idea to express the concept of the flexibility including connection with the production system in particular.

Koste et al. [1] have described dimensions of the flexibility with a hierarchy structure and Wiendahl [2] has defined a changeability of a factory at plural levels.

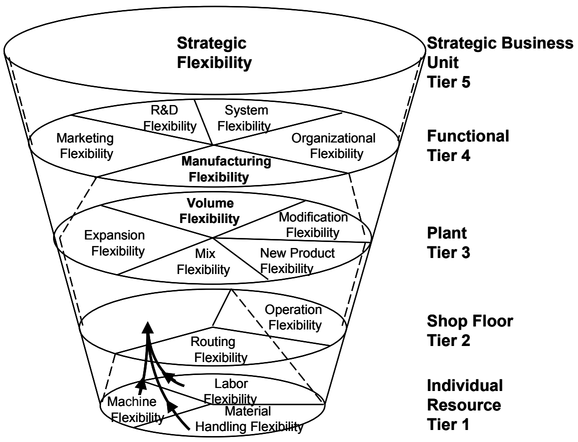


Figure 1: Hierarchy of flexibility dimensions (Koste et al.[1]).

In Figure 1 'Manufacturing Flexibility' is located in hierarchy 'Functional' right under 'Strategic Business Unit'. In Figure 2 'Reconfigurability' is described as the operative ability of a production system to switch with minimal effort and delay to a particular family of work pieces in the matrix of 'Factory Changeability'.

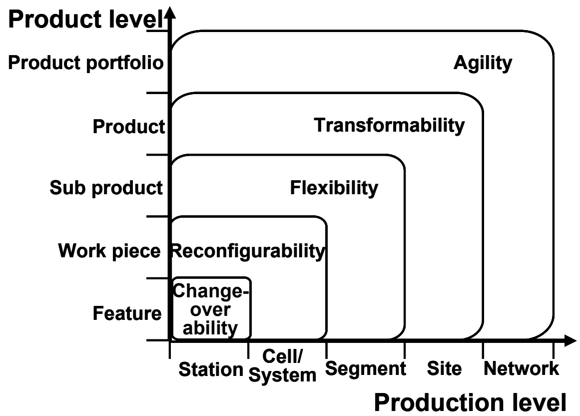


Figure 2: Classes of Factory Changeability (Wiendahl[2]).

In this paper several examples of the flexible production systems are demonstrated which the first author was really concerned with their development for practical use of many years. They are all 'the production systems corresponding to the quantitative fluctuation' in the extension line of 'the mixed production system'. According to the concepts of the flexibility, they are placed in 'Plant (Tier3)' of Figure 1 and in 'Reconfigurability' of Figure 2.

Those flexible production systems are evaluated by the performance function 'Life Cycle Cost (LCC)' of the production volume and compared with conventional systems such as a Manual line and a Transfer line. The production of an automobile part usually continues for around 10 years at least 5, although there is a risk of suffering a change of the business environment in the meantime. The function LCC is used to evaluate the overall cost over the long time horizon. In addition, in order to fairly compare those different production systems, a simulation technique is adopted under the assumption of several fluctuation patterns of the production volume. The results show that the flexible production systems contribute to minimize the risk of investment decision for manufacturing systems.

2 HISTORY OF THE PRODUCTION SYSTEMS IN DENSO

In the automotive parts manufacturer DENSO, not only the products but the production systems have been developed since the 1949 company establishment. As the automotive parts involve people's lives, extraordinary high quality and reliability are required and then human error should be removed when fabricating. The automation for quality assurance and productivity improvement made a significant progress before the '70s on the wave of the high economic growth. It realized the consecutive (from part processing to assembly) and automatic high speed TL (Transfer Line).

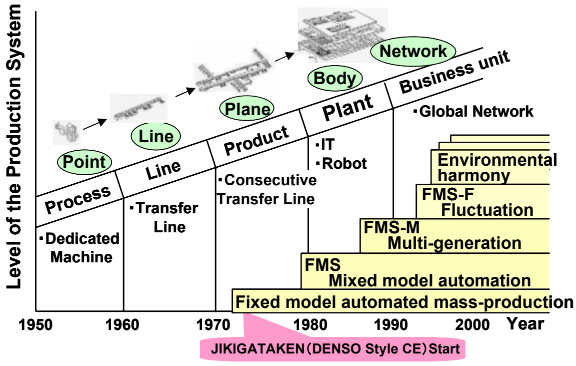


Figure 3: History of the production systems in DENSO.

It is from the '80s that the flexible production systems were developed. Expecting the diversification of the demands, the development of the production systems for various kinds of products commenced. It was carried out simultaneously by the production engineering department in cooperation with the product design department. Such a project activity has been called 'JIKIGATAKEN' in DENSO since 1974. It is just the same as CE (Concurrent Engineering) or SE (Simultaneous Engineering) called later. Many mixed model automated production systems were realized.

In the '90s cost competition made it fierce and foreseeing the sales and the product life became difficult. It is FMS-M (Flexible Manufacturing System for Multi generations) utilizing robots first to have appeared. Next is FMS-F (Flexible Manufacturing System corresponding to the Fluctuation of the production).

From the time when the 21st century approaches, the flexibility needed for a production system levels up and diversifies with global production correspondence, global environment correspondence and so on[3].

Before beginning the main subject "evaluation", three representative FMS-Fs of DENSO are introduced as follows.

3 THE MANUFACTURING SYSTEMS CORRESPONDING TO THE FLUCTUATION OF PRODUCTION VOLUME

3.1 CAC –Circular Assembly Cell-

The CAC (Circular Assembly Cell) [4] was developed for assembling the fuel injection pump of small diesel engines. This product group is essentially of small quantity with many kinds and also of the seasonal fluctuation in volume of production. Therefore the production system adopted was not automatic TL but manual line before the CAC was developed.

The CAC is one of the Assembly Cells which completes all assembling processes in one machine on the basis of automation and mix production technology developed so far. Main developed technologies are the work circulation system which the palettes of the equivalent number of a lot size make a round trip in the machine like a pendulum, the small and unequal pitch automatic warehouse and the high stiffness robot with four link arm.

Fortunately the volume of production lengthened smoothly, six CACs were made and are operating in one of the subsidiaries of DENSO even now.

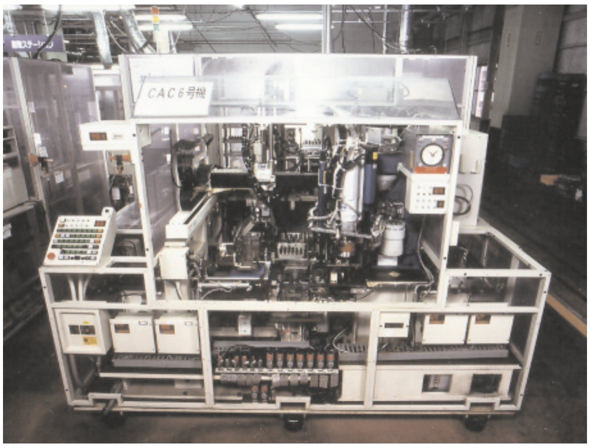


Figure 4: Appearance of CAC.

3.2 CTL –Clustered Transfer Lines-

The ABS (Antilock Brake System) is normally equipped with the car today, but it was optional when it appeared in the market. In addition, many manufacturers started to produce in advance of the future spread and fought for the share. Therefore, planning of the production amount was extremely difficult and there was particularly a risk in introducing the automatic TL.

On the other hand, the product structure varies with the grade of the car to be equipped with and it shows the difference of the number of the specific part built in. It results in the difference of the number of the specific machine even if for the same part at same production volume.

The production system is called the CTL (Clustered Transfer Lines), which are divided into several blocks of similar processes (machines group). Those blocks are connected with an automatic warehouse as a common dynamic buffer by conveyers [5]. The layout of CTL resembles a manifold, but the work flow is made rectification like the TL. Every block is FMS. If the volume increases, the bottleneck machine shall be increased.

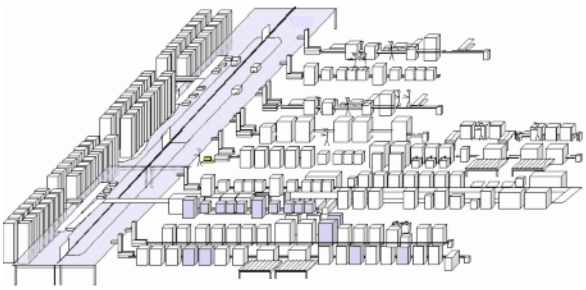


Figure 5: Layout of CTL.

3.3 PPS –The Protean Production Systems-

The automotive air-conditioner is the product which contributed to the growth of DENSO greatly.

Because the air-conditioner unit is comparatively big and has many assembling directions, adopting the robot is suitable. It also becomes the many kinds by a car model (body size, grade, a steering wheel position etc.). The production of the line for exclusive use of the car model may change largely even if the change of sales of the whole car is small.

The production system which has the concept such as the reduced drawing of the recycling society is the PPS (The Protean Production System) [6][7]. The PPS consists of the assembling robot cells which can be put on and taken off according as the volume of production. It is the AC at the state of one cell and is just the same as the TL when every cell connected performs only one process. The state between the AC and the TL is the FMS. The capacity of each line can be increased or decreased as expected. Surplus robot cells can be recycled at another line in the factory.

Main developed technologies are the lot circulation flow system, function divided modules and plug and play technique. Now ten systems are operating in an air-conditioner factory while recycling.

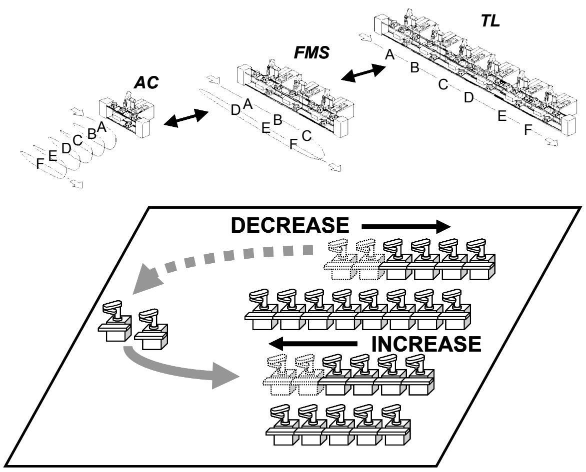


Figure 6: The Protean Production System.

4 RECONFIGURATION FOR THE FLUCTUATION

The classification of three FMS-Fs described above is shown in Figure 7. The horizontal axis is the capacity. The vertical axis is the process aggregation rate. It means how many processes are done in one machine. Assuming that the number of all processes is N and the number of machines is n, the aggregation rate A is equal to N/n. The aggregation rate of the AC taking charge of all processes with one machine is N, and that of the TL most suitable for the mass production becomes 1 (the minimum).

If the FMS-F corresponding to the fluctuation of production volume is recognized as the configuration change, it is illustrated as the domains surrounded by a dashed line in Figure 7. In other words, as for the CAC it is the change of the number of the AC, as for the CTL change of the FMS block and as for the PPS it is the change of the aggregation rate A. As well as these configuration changes, two more abilities should be considered to express the performance of FMS-F. One ability is the variable range of capacity R and the other is the number of segmentation S.

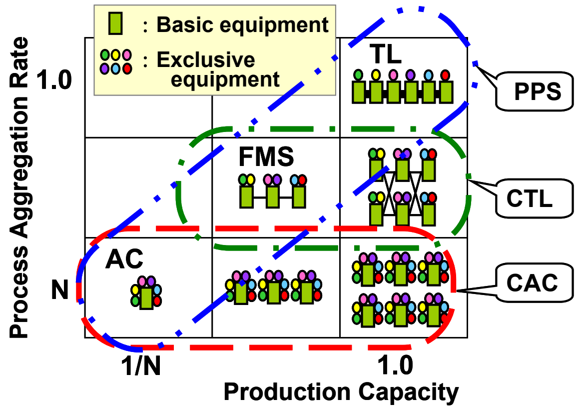


Figure 7: Classification of the FMS-F.

Table 1 shows those values of three FMS-Fs. The products assembled by each system are different. However, assuming that all three systems assemble the same product, the number of all processes is N (>1) and the capacity of TL is 1, the value of R can be decided by the value of A. From the value of R of three FMS-Fs, the CAC is suitable from small to middle production volume change, the CTL is from middle to large and the PPS is from small to large. About S, the real values are shown in parentheses. However, the value S of the CTL is equal to the number of the calking block simulated later and that of the PPS varies from 1 to 4 or 4 to 1.

Table 1: Changeability of FMS-F.

FMS-F	Category	A	R	S
CAC	AC	1≪~≦N	1/N≦~≪1	(6)
CTL	FMS	1≦~≪N	1/N≪~≦1	(12)
PPS	A-variable	1≦~≦N	1/N≦~≦1	(±4)

5 EVALUATION METHOD AND RESULTS

The production of an automobile part usually continues for around 10 years at least 5, although there is a risk of suffering a change of the business environment in the meantime. The LCC is used to evaluate the overall cost over the long time horizon. It will be the best idea to evaluate the efficacy of the FMS-Fs mentioned above. The Net Present Value (NPV) and Real Option Value are calculated with the production volume for the investment decision [8]. Verifying the effectiveness of the LCC, it is thought enough to calculate the LCC only by consideration of the NPV, because all FMS-Fs are in use and the actual data is available.

Tosatti [9] proposes the LCC calculated by (1).

$$LCC_T = \sum_{t=t_0}^{T+t_0-1} \frac{I_t + FC_t + \sum_{s=1}^{S_t} \left(p_{t,s} \cdot \min_{j \in AC_{t,s}} VC_{t,s}^{j} \right)}{(1+\rho)^t} \quad (1)$$

where I_t : portion of the investment paid at period t

FC_t : fixed cost at period *t*

$VC^j_{t,s}$: variable cost for configuration *j* in scenario *s* at period *t*

$p_{t,s}$: probability of the *s*th scenario at period *t*

$AC_{t,s}$: configuration available for the *s*th scenario at period *t*

ρ : the discount rate

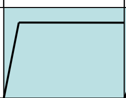
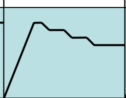
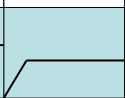
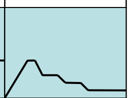
Because high automation is a premise for the FMS-F, the idea of step-by-step investment is included to reduce an investment risk. It changes the configuration of a production system. From the cost price point of view, the depreciation and amortization, normally treated as the fixed cost, may be treated as the variable cost by postponing the investment till the rate of operation rises.

The ability for the fluctuation of three FMS-Fs is compared with conventional systems using the LCC. In consideration of the life of automobile parts, the LCC is calculated every three months for 120 months (10 years). The depreciation period is 7 years and ρ is 0.1 per year.

Four scenarios of production volume fluctuation are shown in Table 2.

The scenario S1 reaching the 100% of the plan in one year is an almost favorable pattern and scenario S2, S3, S4 are undesirable patterns.

Table 2: Scenarios of production volume fluctuation.

	S1	S2	S3	S4
Pattern				
Lifetime Volume	1	0.74	0.47	0.21

The steps of evaluation are as follows:

(1) The processes and facilities of 'virtual TL and virtual Manual Line' assembling the same product assembled by each FMS-F are designed.

(2) The monthly capacity of TL is set as the total capacity of six cells for the CAC, set as the capacity of actually propagated calking block for the CTL and set as the capacity of the configuration of ten cells connected for the PPS.

(3) The production cost price **Cost** (machines depreciation and amortization + labor cost + vice-material and energy costs) except the direct materials expense of each system is normalized by dividing by **PtI** (the planed value of the virtual TL at full production) and the LCC which is integral calculus value of **Cost** is normalized by dividing by **ItI** (the investment of the virtual TL).

(4) The normalized LCC and **Cost** of every three months are illustrated as a graph and the performances of those systems are compared and analyzed.

5.1 CAC for scenario S3

Figure 8 illustrates LCC/**ItI** and **Cost/PtI** calculated for the CAC, the Manual line, the TL, respectively, for scenario S3.

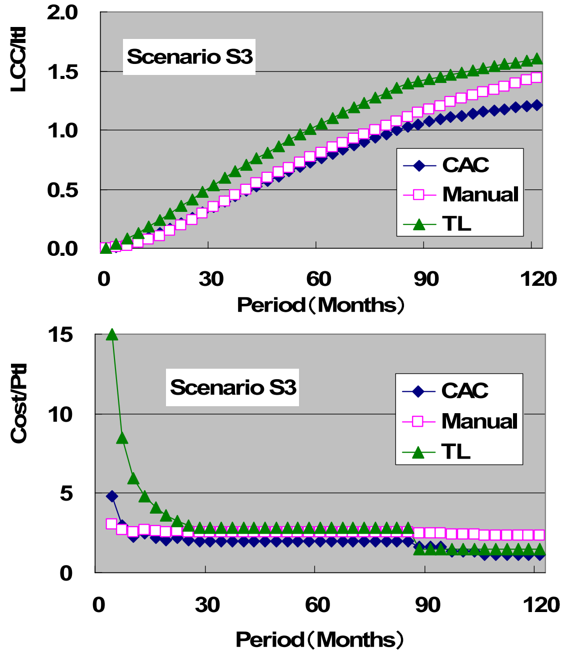


Figure 8: Results of the CAC for scenario S3.

In scenario S3 the amount of production reaches 50% after two years. Because the number of the CAC necessary is three, the smaller investment made both LCC/**ItI** and **Cost/PtI** of the CAC under those of the TL. Besides, by the automation effect of the CAC, LCC/**ItI** lessens than that of the Manual Line in three years shorter than the seven years of a depreciation period.

5.2 CTL for scenario S2

Figure 9 illustrates LCC/**ItI** and **Cost/PtI** calculated for the CTL, the Manual line, the TL, respectively, for scenario S2.

In scenario S2 the amount of production reaches 100% after two years at last but decreases afterward. By an automation effect, both LCC/**ItI** of the CTL and that of the TL lessen than that of the Manual Line in about two years. Furthermore, as the additional investment for the capacity of the CTL is postponed, LCC/**ItI** of CTL lessens than that of the TL. Because LCC/**ItI** is calculated using the NPV, the effect of delayed investment appears clearly, it seems that the validity of this evaluation method is proved.

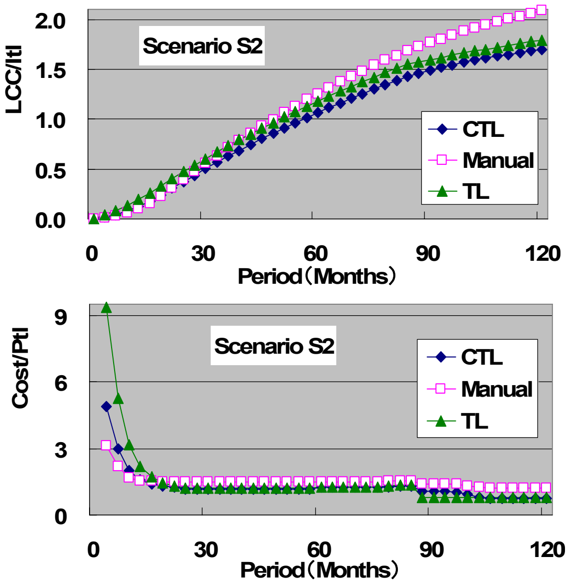


Figure 9: Results of CTL for scenario S2.

5.3 PPS for scenario S4

Figure 10 illustrates LCC/**Itl** and **Cost/Ptl** calculated for the PPS, the Manual line, the TL, respectively, for scenario S4.

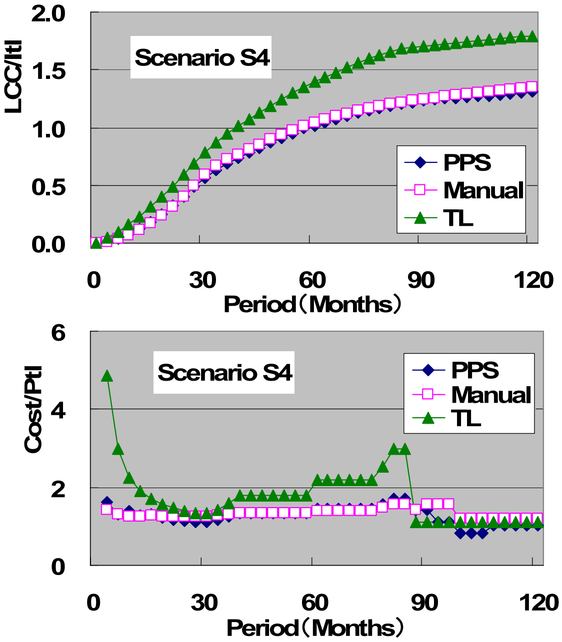


Figure 10: Results of PPS for scenario S4.

In scenario S4 the amount of production reaches 50% after two years at last but gradually decreases to 10% afterward. It is the considerably disappointing scenario which the lifetime volume ends up with 21% of scenario S1. As for LCC/**Itl** after 10 years, the Manual Line gives the smallest easily estimated by seeing the transition of **Cost**/**Ptl**, but even in such a case LCC/**Itl** of the PPS is obviously smaller than that of the TL. The choice of the PPS lowering an issue of quality outbreak risk by the automation seems to deserve examination.

5.4 CAC for combined scenario S1 and S4

Four typical patterns of production fluctuation scenarios (from S1 to S4) are assumed, however the real pattern to happen should be complicated. No matter how complicated the pattern is, it can be transformed to the combination of some typical patterns. By using the probability p of each typical pattern to happen, two combined scenarios Z1 and Z2 are made as shown in Table 3.

Table 3: Combined scenarios.

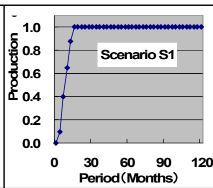
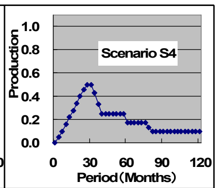
			
Combined scenario	Z1	p=0.7	p=0.3
	Z2	p=0.3	p=0.7

Figure 11 illustrates the simulation result of CAC for scenario Z1 and Z2.

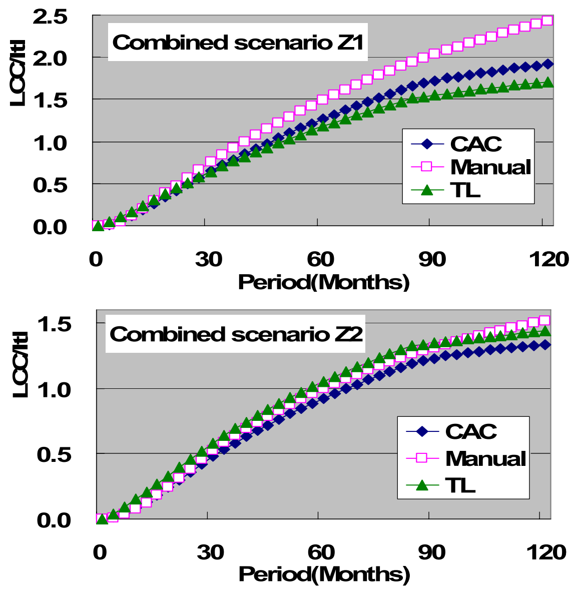


Figure 11: Results of CAC for combined scenario z1 and z2.

If there is the lifetime volume more than 70% of the scenario S1, the investment in the TL is possible even if the probability of undesirable scenario S4 is 30% from the result of scenario Z1. On the other hand, from a result of scenario Z2 in which the lifetime volume of undesirable scenario S4 is 70%, the Manual Line is passable but the CAC investing step by step while watching a production trend is more advantageous than the TL.

5.5 Comparison among three systems

Is it possible to compare the CAC, the CTL and the PPS one another?

As the preparations, the FMS-F and the Manual Line are examined how flexible they are compared to the TL for each fluctuation scenario. The LCC of the TL is denoted by **LCCtl** and the LCC of the FMS-F or the Manual Line by **LCCf**. Then the **FEF** (the Flexibility Efficiency for a Fluctuation) of the FMS-F or the Manual Line is defined by 1/ (**LCCf** / **LCCtl**).

Figure 12 illustrates the **FEF** of each FMS-F or Manual Line with respect to the Lifetime Volume. The smaller the Lifetime Volume becomes, the higher the **FEF** of the FMS-F or the Manual Line rises. When the Lifetime Volume is big enough, the **FEF** of the Manual Line is less than 1.0 that is to say inferior to the TL. On the contrary, when the Lifetime Volume becomes less than 20% the **FEF** of the Manual Line becomes superior to the FMS-F.

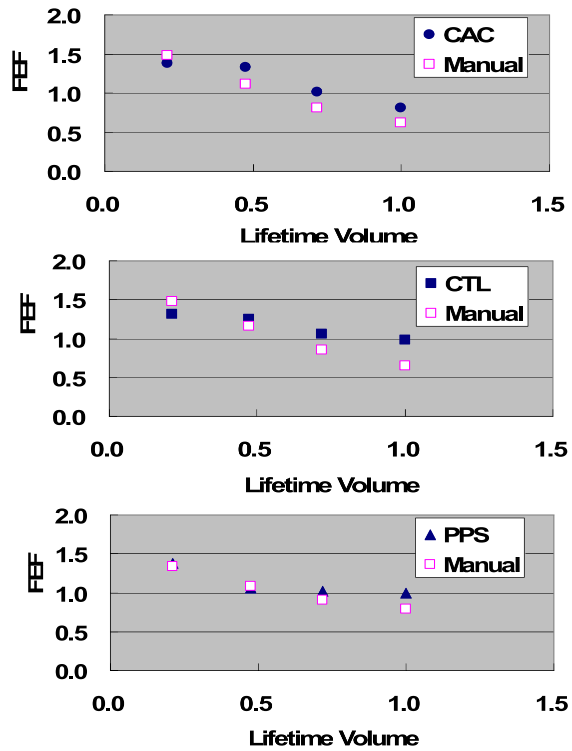


Figure 12: Efficiency of FMS-F and Manual Line.

If the Lifetime Volume is large enough, the **FEF** of the FMS-F is almost equal to that of TL. Therefore the **FEF** of each FMS-F can be compared one another when the Lifetime Volume decreases assuming all the **FEF** of the FMS-F in scenario S1 to be 1.0. This **FEF** normalized by **FEF** of the TL in scenario S1 expresses the essential performance for the fluctuation of the FMS-F changing its configuration. It may be defined as the Reconfigurability. Figure 13 illustrates the Reconfigurability of three FMS-Fs with the Lifetime Volume.

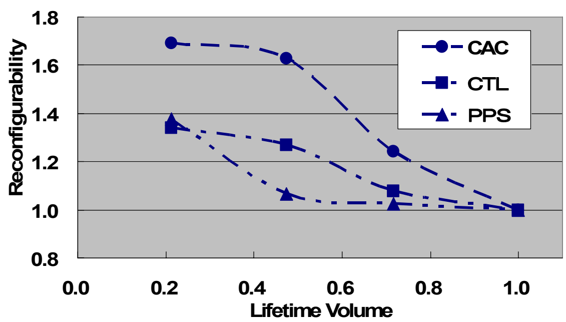


Figure 13: Reconfigurability comparison among FMS-Fs.

From the relative comparison of A and R in Table 1, the Lifetime Volume domain of high changeability seems to be from small to middle for the CAC, from middle to large for the CTL and from small to large for the PPS. In addition, the bigger the value of S is, the smoother the Reconfigurability curve will be. It is worth noting that Figure 13 effectively supports them.

6 CONCLUSION

The performance function LCC will be the best evaluation method for the production system corresponding to the fluctuation of the production volume. The FMS-Fs developed in DENSO change the configuration for the fluctuation. As the Reconfigurability newly defined in this paper expresses their performances well, the additional studies for reformation and application to the evaluation are suggested.

7 REFERENCES

[1] Koste, L., Malhotra, M., 1999, A Theoretical Framework for Analyzing the Dimension of Manufacturing Flexibility, Journal of Operation Management, Vol. 18, No.1, 75-93.

[2] Wiendahl, H.-P., 2002, Wandlungsfaehighkeit: Schluesselbegriff der zukunftsfaehigen Fabrik (Transformability: key concept of a future robust factory), wt Werkstattstechnik online, 92/4: 122-127.

[3] Ranky, P.G., 2007, Eighteen "monozukuri-focused" assembly line design and visual factory management principles with DENSO industrial examples, Assembly-Automation, Vol. 27, No.1, 12-16.

[4] Takiguchi, M., Suzuki, I., Tanaka, M., Sakai, M., Kominami, T., Ookuma, M., 2001, Mini Assembly Factory CAC -circular Assembly Cell-, Journal of Society of Automotive Engineers of Japan, Society of Automotive Engineers of Japan, Inc., Vol.55, No.12, 28-32.

[5] Yamazaki, Y., Yamanaka, T., Iwamatsu, R., Kitano, M., 2004, Production System Design Adapted to Production Volume and Product Variety Change in Market Demand, Denso Technical Review, Vol.9, No.1, 26-35.

[6] Sugito, K., Takata, S., Harashima, S., 2004, The Protean Production System as a Method for Improving Production System Responsiveness, 37th CIRP International Seminar on Manufacturing Systems, Proceedings of ISMS2004, 167-173.

[7] Sugito, K., Inoue, T., Kamijima, M., Takeda, S., Yokoi, T., 2004, The Protean Production System as a method for improving production system responsiveness, Journal of the Japan Society of Precision Engineering, Vol.70, No.6, 737-741.

[8] Abele, E., Liebeck, T., Woern, A., 2006, Measuring Flexibility in Investment Decisions for Manufacturing Systems, Optimization of Manufacturing Systems (O), Annals of the CIRP, Vol. 55, No.1,

[9] Tosatti, L.M., 2006, Life Cycle Cost Calculation for Investment Decision in Manufacturing, 13th CIRP International Conference on Life Cycle Engineering, Proceedings of LCE2006, 723-7.

Promotion Methods of Job Redesigning for Elderly Workers on the Production Line

Kiyohiko Hori[1], Tsuneo Kawano[2], Keiichi Shirase[3]
[1]Dept. of Mechanical Engineering, Graduate School of Engineering, Kobe University, Kobe, Japan
[2]Dept. of Industrial Engineering, Faculty of Engineering, Setsunan University, Osaka, Japan
[3]Dept. of Mechanical Engineering, Graduate School of Engineering, Kobe University, Kobe, Japan

Abstract

A shortage of labour force at production sites is a concern in ”an aging society with a fewer children”. In such a society, it is important to improve working environments in which workers are able to continue on the job up to the age of 65. To this end, it is necessary to effectively employ elderly workers systematically and continuously, to plan for the long term while paying attention to overall surroundings and not focusing merely on partial and temporary improvements, and to work for the empowerment of elderly workers. Therefore, job redesigning that takes into consideration the ergonomics of operations which middle-aged and elderly workers can perform on production lines easily, safely, comfortably, and without undue strain, is necessary. In light of the expansion of workplaces accommodating of elderly workers, in this study we propose four countermeasures, based upon specific evidence, with the objective of advancing the streamlining of job redesigning for middle-aged and elderly workers along the production line.

Keywords:
Elderly workers; Job redesign; Job evaluation criteria; Human-oriented manufacturing system

1 INTRODUCTION

In present-day Japan, not only is life expectancy increasing considerably, but also birth rates are falling to even lower levels, and ‘an aging society with a fewer children’ is now upon us. In 2005, Japan had both the most elderly population and the lowest birth rate of any country in the world. The percentage of children younger than age of 15 is 13.6% and the lowest in the world, while the percentage of elderly people aged 65 or older becomes 21.0%, the highest in the world. These are based upon the 2005 Census released on 30 June 2006 by the Ministry of Internal Affairs and Communications. Furthermore, in 2006, the National Institute of Population and Social Security Research estimated that by 2025, the current population of 127.76 million people would shrink by roughly 8.5 million [1]. The Ministry of Health, Labour and Welfare estimates that by 2030, the labour force of 66.57 million people over the age of 15 including employees and job seekers in 2006 will shrink by 10.7 million, by which time Japan will have taken the lead as being the country with the most rapidly shrinking society in the world.

In these circumstances, the expansion of elderly employment until age of 65 will become increasingly necessary. In order to employ elderly workers, it is necessary to adopt supplementary countermeasures (job redesigning for elderly workers) with ergonomics that take into consideration the decline in abilities that results as we age; it is also necessary to create jobs in which motivation is raised to higher levels, in other words, ‘humanization of work (Quality of Working Life)’ [2]. In the past, even though there was a tendency to sing the praises of a focus upon respects for workers, cases were encountered in which designers favoured improvements in the operational environment that placed greater emphasis upon increased efficiency. However, in the first part of the 1990s, the Ministry of Economy, Trade and Industry, and some Industries began to think about new factories [3] that took into consideration the human factor and the provision of amenities; the harmonious coexistence of productivity and humanistic values came to be a major policy in deciding the future direction of these factories. However, Japan was unfortunately faced with the collapse of a bubble economy, and it cannot be denied that the adoption of the measures outlined above was delayed.

In this study, we propose a kind of job in which elderly workers can work safely and comfortably, and in which elderly workers who so desire will be able to continue working until the age of 65. Our objective is the creation of streamlined policies for promoting job redesigning for elderly workers on the production line so that elderly workers may feel happy and satisfied. Figure 1 shows our four proposals: first, specific opinions on the expansion of job in which elderly workers are working; second, methods for extracting factors that impede elderly workers’ tasks, for implementing job redesigning; third, ways of measuring operational ability according to similar movements close to movements in actual working operations that make use of standard settings; and, fourth, methods of classification between counter-measures against aging on the job expansion for elderly workers and countermeasures for personnel-reduction through automation.

2 A MINIMUM AMOUMT OF JOB EXPANSION

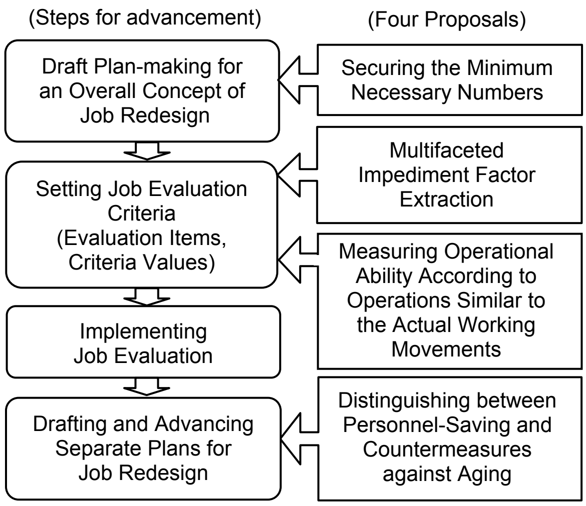


Figure 1 Steps and Four Proposals for Promoting Job Redesign

REQUIRED FOR ELDERLY WORKERS

It is necessary to pursue easy-to-work tasks in the production policies that place greatest emphasis on productivity with a sense of the value of the worker and respect for the individual, in order to create a working environment in which elderly workers can continue to perform their tasks along the production line as safely, comfortably, and free of strain as they always have done. Generally, the following two methods have been adopted in conventional measures for increasing workplaces accommodating of elderly workers (countermeasures against aging):

(1) Improving all jobs (tasks) so that elderly workers will also be able to perform any job.
(2) Preparing new jobs exclusively for elderly workers, without modifying the pre-existing production line.

However, method (1) involves reducing weights and operation postures in overall operations, and is not limited strictly to helping elderly workers. Furthermore, while method (2) is a test installation for fabricating an exclusive elderly worker production line, there are instances in which these kinds of installations have been discontinued due to a decline in motivation along production lines staffed only by elderly workers.

Therefore, in this study we propose the following method:

(3) Improving operations by the minimum necessary amount in accordance with the number of elderly workers on staff (also including middle-aged workers).

In other words, in order to put to the best use the valuable experience and technical know-how that elderly workers have acquired over the course of their careers, it is necessary to use machines to make up for elderly workers' declining abilities and various functions, to change job organisation in order to draw out elderly workers' rich talents, and also to think of how elderly workers feel and to treat them with respect, redesigning job settings in order to allow for tasks to be carried out more easily. Even if this might produce additional costs, affording elderly workers a sense of satisfaction will nevertheless tie in with the development of future corporate management, and will in no sense be wasted. In this sense as well, job redesign is a necessary corporate investment, for which purpose it is necessary to reduce the corporate burden as much as possible, and to base future thinking upon controlling minimum necessary levels. Decisions regarding the number of job redesign implementation plans are to be made by estimating the number of tasks matching the number of middle-aged and elderly workers in the future, investigating the 'Number of Tasks Suitable by Middle-Aged and Elderly Workers' based upon the job evaluation implementation to be discussed in the next chapter, and then subtracting the latter from the former.

3 TASK IMPEDIMENT FACTOR EXTRACTION METHODS FOR JOB REDESIGNS OF ELDERLY WORKERS

The greatest source of motivation for elderly workers is working in a familiar workplace where they are able to put their experience and technical know-how into practice to their hearts' content. However, an unfortunate consequence of aging is that middle-aged and elderly workers become slower to complete tasks they once did quickly, and also take longer to recover from fatigue. It is difficult to improve productivity by adopting policies in which job redesign calls for removing undue burdens upon elderly workers, and investment in such policies aims directly at workers' benefit for the elderly workers themselves similarly to the investment for environmental protection measures aimed at social benefit. Thus, corporations tend to adopt job redesign policies for elderly workers partially and superficially, inserting such policies only sporadically into new production preparations and newly-designed lines, with aspects in which future plans for elderly workers are indefinitely.

Table 1 Experiments in Three Kinds of Occupations

	Job Evaluation Item	Assessment	Action
1st Experiment	Weariness, 8 Items	Improper for Shortage of Items	Addition of Items
2nd Experiment	Weariness, 8 Items	Improper for Shortage of Items	Addition of Items
	Difficulty, 3 Items	Improper for Inadequacy of Items	Change in Items
3rd Experiment	Weariness, 11 Items Awkwardness, 3 Items	Proper for Considered Elaboration Items	—

Furthermore, there is a tendency for superficial adaptations to be insufficiently grasped such that they actually become a factor in impeding elderly workers from doing their jobs. Thus, through a multi-faced grasping of elderly workers' decline in physical functions, it becomes possible to remove even small details of unsuitable operations.

In order to advance carefully and overall job redesign, the overall, unified grasping of areas perhaps unrelated to the occupational category is necessary, which is a process similar to job evaluation. It is also necessary to consider proposals of a method for implementing short-term organisations that does not require specialised knowledge. Furthermore, it is necessary for the advancement of job redesign to cultivate a sense of fairness among the different types of occupations, for which purpose quantitative job evaluation observation is necessary, and is not preferable to subjective observation. Towards these ends, we have devised a method to fulfil the following three viewpoints:

(1) The evaluation method must be simple, and the evaluation results, must be impartial and appropriate
(2) Must be capable of clarifying operation improvement points from a variety of angles
(3) Evaluation must be carried out for items suitable for, and common to, jobs taking in all occupational categories

In this study, focusing our attention on physical functions [4], which decline as a worker ages, we propose job evaluation standards [5, 6, 7] for grasping a lot of impeding factors for elderly workers against tasks obtained from our experiments in three kinds of occupations, as shown in Table 1. Ultimately, we have chosen eleven and three evaluation items based upon the 'weariness' and 'awkwardness in doing tasks' that accompany aging, respectively.

For the first experiment, in order to choose evaluation items from the viewpoint of the decline in physical functions with aging, we referenced the 'Saitoh Calculation Chart' [4] proposed by the former director of the Institute for Science of Labour, Hajime Saitoh. Saitoh offers 36 different kinds of physical and mental functions; as he says, 'the pace at which age-related physiological and psychological functions decline differs according to the type of function'. Out of those 36, we chose 8 kinds of 'weariness', shown in Table 2, as being applicable to our present research. According to Saitoh, the functions that decline most markedly are memory and judgment, which fall off 47% and 23%, respectively, from the time when a worker is young. We found these figures to be

Table 2 Functional Decline and Evaluation Items (Reference) Saitoh Calculation Charts

Mental and Physical Function	Decline in Function (%)	Evaluation Item
Memory	47	Number of Different Parts to be Handled (Choices of Parts Differing by Model)
Judgment	23	
Leg Strength	37	Raising and Lowering (Total Up-and-Down Height)
Gripping Power	25	Tool Use Conditions (Number of Trigger Pulls Per Day)
Arm Strength	20	
Back Muscle Strength	25	Lifting of Heavy Weight (Maximum Weight Per Single Part, Overall Weight Per Day)
Motion Speed	15	Parts Handling Number (Time Per One Part or One Tool)
Lateral Bending	15	Working Posture
Forward Bending	5	(Painful Posture)

perfectly appropriate evaluation items applicable to the 'parts selection judgment' that comes into play when selecting separate parts for every differing products and then attaching them to the products accordingly. Similarly, we chose the following evaluation items (in quotation marks) from the figures given by Saitoh: from a 37% decline in leg strength, the evaluation item 'Up-and-Down Height,; from a 25% decline in gripping power and a 20% decline in arm strength, the evaluation item 'Pulling the Trigger Tools'; from a 25% decline in back muscle strength, the evaluation item 'Lifting of Heavy Weights' and 'Total Heavy Weights a Day'; from a 15% decline in working speed, the evaluation item 'The Total Handling Number of Parts'; and from a 15% decline in spine bending lateral and a 5% decline in spine bending forward, the evaluation item 'Holding of working Posture'.

In the second experiment, we added the item of 'difficulty' to the item of 'weariness'. Because 'difficulty' varies with the degree to which a worker possesses experience and technical know-how, it is associated with the degree of difficulty of an operation, and is expressed in operational speed. 'Accurate installation' and 'adjusting operations' are both applicable operations. In the third experiment, we added the item of 'awkwardness ' to the item of 'weariness'. 'Awkwardness' is when operations come to be performed more lately with aging; dexterity and physical strength are influenced by such things as 'memory and judgment', 'installation of minute or thin plate' and 'elaboration tasks that require physical force'.

The following results became clear after the three experiments had been carried out. In the first experiment, there were some operations that could not be evaluated solely upon the basis of 'weariness', and the evaluation items were insufficient. In the second experiment, when the prolonged period of proficiency of elderly workers was taken into consideration, it was thought best to avoid operations reshuffling as much as possible. When it was taken into consideration that elderly workers, who had been continuing at the same job for a long period of time, were already proficient, experienced, technically capable, skilled in the inner workings of their tasks, and possessed of a sense of how their tasks should be carried out, the evaluation item of 'difficulty' was found to be unnecessary, as it bore only a slight direct relation with elderly workers. In the end, we identified operations of 'awkwardness', with a focus on the third experiment, and based upon the information in the previously mentioned 'Saitoh Calculation Chart' [4] on an age-related function decline of 37% loss of visual acuity, we selected as an evaluation item the ability of 'reading of the lettering' on the part picked up, which is associated with the sizes of alphanumeric characters, while from a 15% decline in fingertip motion we chose the evaluation item 'installation of minute or thin plate'. Our decision was finalised thereafter by the inclusion of 'speed on foot', and 'distance on foot' in a day.

The reaction of jobsite supervisors to these experiments was that there was an agreement that the third experiment came closest to what actual operating conditions are really like. For instance, there were 'zero impediment factors' as the job evaluation results of operations that elderly workers could perform in a continuous task.

For the criteria values for each item, the criteria values set forth in 'Safety and Health Standards' were reconsidered for the previous experiments, while for items for which no criteria values existed, we conducted an overall study of the measurement values and conventional values at the production jobsite, and then corrected and decided upon criteria values in consideration of the experiment results and the values desired by jobsite supervisors . For example, for 'lifting of heavy weights', in the first experiment there was an increase in claims of 'work left over due to pain in shoulders and/or waist' among middle-aged and elderly workers who participated in the experiment, vis-à-vis the general adult criteria value of 10 kg set forth in the 'Safety and Health Standards' so that in the end we set the criteria value at 7kg instead of 10 kg. The criteria values of other items, as well, were agreed to after having been adjusted towards the conventional jobsite values during the trial stage of the jobsite evaluations. This is because the 'Safety and Health Standards' are mainly from a labour hygiene disease prevention standpoint, and the kind of occupations is limited.

On the other hand, it is difficult to quantitatively set a criteria value for 'awkwardness', so for the elderly worker job evaluation standards [5, 6] for workers less than 60 years of age, we made our decision based upon the presence or absence of certain operations. In these countermeasures, as well, there is the need for methods in which the upper age limit for representative operations and operational proportioning for each kind of occupation is decided, with the applicable operations being reduced as much as possible for new products in response to the age limits of workers.

4 JOB EVALUATION CRITERIA FOR ELDERLY WORKERS

For the setting of evaluation criteria values for workers 60 to 65 years of age [7], separately from the conventional under-60 job evaluation standards, for 10 persons aged 63 to 73 (average age 66.8) with an average physical strength age of 62.3, we were able clearly to set criteria values for items for which it was possible to assign a numerical value by performing operational ability measurements using movements similar to those of the actual work operation. Those persons being tested were workers with forty or more years' (average of 43.3 years) experience engaged in mainly heavy operations at production sites, six of whom continue to work at light operations in operations which require physical strength. Using this method, we were able to set criteria values for up to 65 years of age for the following seven categories: lifting weight (lifting and lowering materials), walking distance, walking speed, up-and-down height, holding of working posture, number of selective fittings, installation of minute or thin plate. Furthermore, for the overall numbers in one day, after measuring operational ability we estimated numerical values based upon the criteria under-60 values for one-day continuous operations. The result of this was that

Table 3 Job Evaluation Items and Criteria Values or Workers up to the Age of 65

	Measurement mean value	Standard deviation	Criteria value
Age of Measured Person (yrs)	66.8	2.9	
Lifting Weight (kg)	5.1	1.37	5
Gripping Power (kg)	38	6.83	
Distance on Foot (km)	3.6	2.5	4
Speed on Foot (km/h)	4	0.43	4
Up-and-Down Height (cm)	20.4	2.5	20
Holding of Working Posture (s)	2.9	0.85	3
Selective Fitting (kind)	3.4	0.52	3
Font Size of Lettering (pt)	17.6	11.4	17
Installation of Minute/Thin Plate (mm)	4.6	0.74	5

each measurement value decreased 10–30%. For example, the measurement value for materials lifting weight was in the range of 3 to 7 kg per single item; we used the 5.1 kg median and average value. The criteria value of lifting weight for workers up to 65 years of age, as shown in Table3, was set at 5 kg or less per single item. This value is equivalent to a roughly 30% reduction of the 7 kg criteria value for workers 60 years of age or younger. n addition, gripping power came to an average 38.0 kg; as this value was an approximate 10% reduction from the value for workers up to the age of 60, we also set the tool-use value to a level roughly 10% lower than that for workers younger than 60 years of age.

In addition, using the same method it was possible to set a criteria value for 'awkwardness'. For instance, for the size of the lettering on the label of a parts box in the case of inputting at the time of installation parts, we prepared labels with several different font sizes and then took the 17-point font as our criteria value by referring to the average value at instantaneous inputting. Also, for the minute or thin parts, we prepared circular disks of several different thickness, and then set the criteria value at 5 mm by referring to the thickness at which it was possible for a worker to instantaneously pick up a part by hand while wearing white cotton work gloves.

5 CLASSIFICATION METHOD BETWEEN COUNTER-MEASURES AGAINST AGING IN THE JOB EXPANSION FOR ELDERLY WORKERS AND PERSONNEL-REDUCTION COUNTERMEASURES THROUGH AUTOMATION

5.1 Setting of Job Redesign Extent

Corporations and industries are now implementing job redesign aimed at putting elderly workers to active use in order to make up for shortages of workers, but the main focus now is the introduction of production-enhancing robots used as personnel-reduction countermeasures. A comparison of the results of these dual approaches is unfeasible as their terms cannot be expressed as masses, but from the standpoint of elderly workers, countermeasures against aging are necessary and indispensable, and there is certainly a

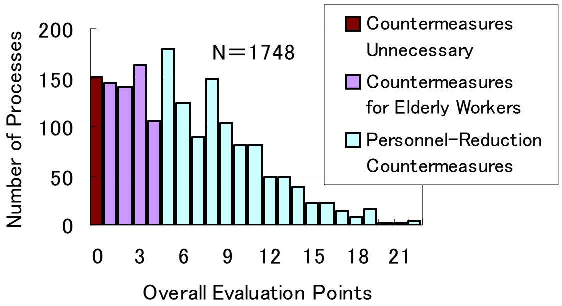


Figure 2 Distribution of Overall Evaluation Points by Score for Workers Aged 60 and Younger

sense in which elderly workers are becoming the sacrificial victims of personnel-reduction countermeasures. For example, because of the cost-reduction of facilities investment, designers in charge of unmanned robot fabrication sometimes replace the operations that are easy and inexpensive to be robotized with new robots even if they are suitable for elderly workers. The result of this is that elderly workers lose a place to work, and get reassigned to other jobs, which never motivate them disregarding their humanity, experience, and technical know-how. Therefore, we considered it to be appropriate to make a clear distinction between countermeasures against aging, which can be conducted at relatively low cost, and personnel-reduction countermeasures, in which the introduction cost of the automating robot is very large. It is necessary to decide in advance the extent for the expansion of jobs accommodating of elderly workers, and to take into consideration ways in which personnel-reduction countermeasures (unmanned robotisation), even if mistakenly implemented, will not act as hindrances to the expansion of jobs accommodating of elderly workers. In other words, we made clear distinction based on results so that a choice could be made, for each different kind of operation, whether to choose one kind of countermeasure or another.

In this study, we propose using the 'Improvement Level-of-Difficulty Points' [5,8] shown in Table 4 as a classification index. The 'Improvement Level-of-Difficulty Points' quantitatively express the level of difficulty of facility improvements according to an operation's 'degree of weariness' and 'degree of awkwardness'. By examining the distribution of the 'Overall Evaluation Points' shown in Fig. 2, which are the sum total of the 'Improvement Level-of-Difficulty Points' found for each different operation, it becomes possible to select easy-to-perform counter-measures aimed at elderly workers if the overall evaluation points are low, and to choose personnel-reduction countermeasures if the overall evaluation points are high, thus allowing for clear distinction between, and clear classification of, both choices.

For the 'degree of weariness' (degree of physical strength) Improvement Level-of-Difficulty Points, we used standard deviation and equidistribution to divide into three groups those operations whose evaluation-criteria values exceeded those of workers aged 60 or younger, and assigned to them a ranking of either aa, ab or ac, such that those operations with an evaluation criteria value of less than the minimum evaluation criteria value were at 0 points, while those operations at the maximum value were at 3 points. We totalled these for each operation of which an individual was in charge, taking this total as the weariness evaluation point A.

Next, for the 'presence or absence of awkwardness' (degree of arduousness) Improvement Level-of-Difficulty Points, we assigned a 0 point minimum Improvement Level-of-Difficulty

Table 4 Overall Evaluation Points Table for Elderly Workers up to the Age of 65

(a)<Weariness due to Aging>

Evaluation Item \ Rank	aa	ab	ac	ad	ae
Improvement-Difficult Point	0	0.5	1	2	3
Lifting of Heavy Weights	0～5	5.1～7	7.1～10	10.1～13	13.1～(kg/piece)
Total Heavy Weights a Day	0～5	5.1～7	7.1～9	9.1～11	11.1～(ton/day)
The Total Handing Number	0～10000	10001～15000	15001～17000	17001～19000	19001～(piece/day)
Pulling the Trigger	0～4500	4501～5000	5001～6500	6501～8000	8001～(times/day)
Using a Vibration Tool	0～1.8	1.9～2	2.1～2.5	2.6～3	3.1～(h/day)
Using a Special Tool	0～1300	1301～1500	1501～1800	1801～2100	2101～(times/day)
Distance on Foot	0～4	4.1～7	7.1～8.5	8.6～10	10.1～(km/day)
Speed on Foot	0～4	4.1	4.2～5	5.1～6	6.1～(km/h)
Up-and-Down Height	0～20	21～30	31～50	51～70	71～(cm/tact)
Holding of Posture	0～3	4	5～22	23～40	41～(s/tact)
Selective Fitting	0～3	4	5～9	10～14	15～(kinds/tact)

(b)<Awkwardness due to Aging>

Evaluation Item \ Rank	ba	bb	bc
Improvement-Difficult Point	0	0.5	1
Font Size of Lettering Labels or Parts Lists to Read	17～	15～16	13～14(pt)
Installation of Minute or Thin Plate	5～	4	3 (mm)

(note) improvement-difficult point of the other items: 5 point

Points if the operation was noting, and 5 points if the operation was exist; we took as the awkwardness evaluation point B the sum that was added to the operational units. The criteria upon which these 5 points were decided are as follows: Taking the total of the weariness evaluation point A and the awkwardness evaluation point B as the overall evaluation points, the number of operations for each different overall evaluation point is shown in Fig.2. Then, we discovered that it is best to adopt as a job redesign target an operational extent in which overall evaluation points are at 4 points or less in order to create 600 operations suitable for middle-aged and elderly workers for estimating from future employee aging. Thus, taking the Improvement Level-of-Difficulty Points as 5 or more for job redesigning that has been deemed difficult (presence of awkward), such job redesign would not be included in the job redesign target extents. Ultimately, an overall evaluation points score of 4 was the upper limit of the extents for countermeasures aimed at elderly workers; personnel-reduction countermeasures came to fall within extents at scores of 5 or more.

As shown in Table 4, regarding the Improvement Level-of-Difficulty Points for elderly workers aged 65, for the 'degree of weariness' for the table for 60-year-old workers [5, 6], the point for values at or below the criteria value for 60- year-old workers were revised from 0 points to 0.5 points, and a score of 0 points for values at or below the criteria value for 65-year-old workers was newly added. Furthermore, regarding the 'degree of awkwardness', we classified into 3 ranks those criteria values for which it was possible to convert to numerical values.

Equation (1) is the equation used to compute the overall evaluation points. We divided the awkwardness evaluation points into what is capable of improvement and what has a difficulty in improvement.

$$\text{Overall evaluation points} = \text{Evaluation point A} + \text{Evaluation point B'} + \text{Evaluation point C} \qquad (1)$$

where, evaluation point A is the total points of the Improvement Level-of-Difficulty Points of the 'weariness items'; evaluation point B' is the total points of the Improvement Level-of-Difficulty Points of the 'awkwardness items' that are capable of being converted to numerical values; evaluation point C is the total of the Improvement Level-of-Difficulty Points of 'presence or absence of awkwardness' that is not capable of being converted to numerical values.

Regarding the classification of personnel-reduction countermeasures and countermeasures against aging, we attempted a sample experiment to determine how the overall evaluation points distribution of 1748 operations, as shown in Fig.2, would change with the addition of 0.5 points of the criteria value of workers aged 60. The result of this sample experiment was that, in general, there were many operations of 4 overall evaluation points or less, in which case it is possible that the classification points for both countermeasures could change somewhat. In any event, it is necessary to choose countermeasures such that it will be possible to maintain, out of the 1748 operations, operations suitable for middle-aged and elderly workers.

5.2 Preparation and Evaluation of Job Redesign Advancement Plans

Job redesign plans are estimating operations carried out for the purpose of countering the impediment factors brought to light during job evaluations. One must consider the feasibility of countermeasures, consultations between and among supervisory departments regarding whether to enact countermeasures on pre-existing production lines or on newly built facilities, roughly-calculated costs and items to be implemented.

For the order of priority of operations to be performed during countermeasures, final decisions on operations of 4 overall evaluation points or less, which are part of possible countermeasures, are to be made after taking the following conditions into overall consideration.

(1) Operations for which overall evaluation points are low, and for which improvements are simple
(2) Operations for which it is simple to eliminate 'awkwardness' impediment factors
(3) Operations with a high degree of on-site urgency
(4) Operations which can be performed at or below the minimum facilities cost, requiring reduced labour, or even an individual worker

After implementing job redesign, a five-step evaluation is to be performed upon the investment results and the streamlining results for each item that has undergone improvement, with improvement follow-up performed according to the following:

(1) Results of lessening mental and physical burdens
(2) Efficiency improvement
(3) Results of expanding workplaces accommodating of elderly workers (Number of impediment factor reductions)

The incorporation of job redesign into new facilities is carried out through prior study involving turning the workplace evaluation standards into intra-company regulations. After job redesign incorporation is complete, a check-mark is made in the 'countermeasures finished' column of a pre-action study's accuracy report form.

6 CONCLUSION

Japanese society has already become 'an aging society with a fewer children', facing a pronounced shortage of labour force. At the Council on Economic and Fiscal Policy held in April of this year, the government announced a 'New Employment Strategy' incorporating important employment counter-measures for the coming three-year period. This strategy fixed the numerical value index at 56–57% of workers aged 60 to 64 by the year 2010, publicised an intention to create one million new jobs, and declared a desire to involve the government in realising a 'full-participation economy' in which everyone can work easily.

Job redesigns are countermeasures as a complement to the declining functions of elderly workers as they age. Job redesigns do not help to increase in productivity, and are categories of investment which do not make a profits. However, corporations are facing labour shortages due to 'an aging society with fewer children', and the time has come when investment must be made in providing employees with a sense of happiness and job satisfaction. When corporations take it upon themselves to realise production systems that pay attention to the human respect, employees are able to enjoy more stable lives, job satisfaction is created, and workers are able to feel a sense of joy and pride in their work, all of which contribute towards bringing forth jobs that are high in creativity. This study is a proposition for allowing corporations to implement these indispensable counter-measures with as little burden as possible.

The results of this study are summarized as follows:

(1) If the above-mentioned ways of thinking are implemented, it will be impossible to consider the necessity of establishing countermeasures for transforming overall operations to operations suitable for elderly workers. Those minimum-level counter-measures that are in accord with the number of elderly workers are sufficient. Therefore, unmanned and/or personnel-reduction robots are available for increasing productivity, and a production site in which both these entities coexist is preferable.
(2) In order for robots and elderly workers to coexist, it is necessary to clearly delineate the respective territories of each. Job redesign cannot be completed simply through partial and superficial countermeasures. Job redesign must be carried out in a planned fashion, and in terms of overall structure. To this end, it is necessary to grasp in a multifaceted the factors that impede the operations of elderly workers based on job evaluations. There is a wide array of different kinds of declines in function due to aging, and it is necessary to find countermeasures for each of them individually.
(3) In order for job evaluations to tie in with job redesign, it is necessary that job evaluations be carried out as objectively as possible. It is necessary to consider movements that are as close as possible to actual operations, and to perform operations ability measurements thereby.
(4) It is difficult to conduct improvement of 'awkwardness' with job redesign of pre-existing operations; if possible, methods that eliminate 'awkwardness' from new product settings are preferable.
(5) When considering the prolonging of employment from age 60 to age 65, it is necessary to study more on the classification of countermeasures towards unmanned production and countermeasures against aging in a variety of cases for workers aged 65 and younger. We plan to study conducting simulations in which each kind of case is tested.
(6) We also have a plan to start to investigate medium- and small-scale enterprises that receive many separate production orders.

7 REFERENCES

[1] National Institute of Population and Social Security Research, 2006, http://www.ipss.go.jp/ppnewest/j/newest03/ h11.html (in Japanese).

[2] Koichi Ono, 1986, Job redesign as seen from the standpoint of humanisation of labour, Asia University Management Papers, Vol.22, No.2, 3-34 (in Japanese).

[3] The Japan Machinery Federation, Osaka Science & Technology Center, 1991, Survey on Construction of New Production Environment for Machinery Industry in Kinki Area (In 1990), 5-65. (in Japanese).

[4] Hajime Saitoh, 1980, Health management for middle-aged and elderly workers as seen from a labour science perspective, Rouken Ijikai Shiryou, Dual No. 864-865, 9 (in Japanese).

[5] Kiyohiko Hori, 1991, Creating automobile production lines accommodating of middle-aged workers, 7th Symposium on Human Interface,139- (in Japanese).

[6] Japan Society of Mechanical Engineering, 2005, Examples of conditions criteria for operations suitable for middle-aged and elderly workers, JSME Mechanical Engineers' Handbook, Design β7 Manufacturing System, β7-83 (in Japanese).

[7] Kiyohiko Hori, Tsuneo Kawano, Keiichi Shirase, 2007, Reassessment of Job Evaluation Standard for Elderly Workers in Assembly Line, JSME Annual Symposium, Vol. 4, 315-316 (in Japanese).

[8] Kiyohiko Hori, Tsuneo Kawano, Keiichi Shirase, 2008, Reassessment of Promotion Method of Job Redesign for Elderly Workers in Assembly Line, JSPE Annual Symposium, 661-662 (in Japanese).

Human-Oriented Dynamic Task Reallocation and Rescheduling in Cellular Manufacturing Systems

Yoshitaka TANIMIZU[1], Yoshiyuki SAKASHITA[1], Koji IWAMURA[1], Nobuhiro SUGIMURA[1]

[1] Dept. of Mechanical Engineering, Graduate School of Engineering, Osaka Prefecture University, Osaka, Japan

Abstract

Cellular manufacturing systems are well known as flexible and reliable manufacturing systems for assembly processes. A suitable control method is required for the cellular manufacturing systems to keep the production capacity in the case where unforeseen changes, such as delays of assembly processes, addition of new orders, and unscheduled workers' absences, occur during the progress of assembly processes. This research proposes a dynamic task reallocation and rescheduling method in the cellular manufacturing systems, consisting of a set of assembly cells in which all the assembly processes are carried out by single workers. The method proposed here dynamically modifies an initial production schedule and minimizes the total tardiness by a genetic algorithm and a heuristic rule, called as EDD (Earliest Due Date), to cope with unforeseen changes in the cellular manufacturing systems. A prototype of dynamic task reallocation and rescheduling system was developed and applied to some computational experiments in order to evaluate the effectiveness of the proposed method.

Keywords:

Cellular Manufacturing; Task Allocation; Genetic Algorithm; Tardiness; Learning Curve

1 INTRODUCTION

Flexible manufacturing systems have been required to cope with dynamic changes of market requirements, product mix and production volumes. Cellular manufacturing systems are well known as more flexible and reliable manufacturing systems for assembly processes than the traditional mass-production systems. Individual products are assembled by single workers or a group of workers in the cellular manufacturing systems.

The cellular manufacturing systems need a suitable control method to keep the production capacity in the case where unforeseen changes, such as delays of assembly processes, addition of new orders, and unscheduled workers' absences, occur during the progress of assembly processes. The workers carry out their assembly processes cooperatively in the actual cellular manufacturing systems. For instance, untrained workers are often supported by neighbour workers or factory managers, when the assembly processes of the untrained workers are delayed. The cooperation among the workers may locally contribute to the flexibility and the reliability of the assembly processes. However, the effectiveness of the cooperation has not yet been evaluated systematically from the viewpoint of the whole manufacturing systems.

This research proposes a dynamic task reallocation and rescheduling method in the cellular manufacturing systems, consisting of a set of assembly cells in which all the assembly processes are carried out by single workers. The method proposed here dynamically modifies the initial production schedule to cope with unforeseen changes in the cellular manufacturing systems, in order to minimize the total tardiness. A learning curve effect is considered in the modification process to generate a suitable production schedule. A prototype of dynamic task reallocation and rescheduling system was developed on a personal computer and applied to some computational experiments in order to evaluate the effectiveness of the proposed method.

2 CELLULAR MANUFACTURING SYSTEMS

2.1 Characteristics of cellular manufacturing systems

The cellular manufacturing system is a kind of lean manufacturing systems. Single workers or a group of workers assembles individual products in the cellular manufacturing systems. The cellular manufacturing systems are recognized to have the following characteristics.

- To provide more flexibility for variations of product mix and production volumes.
- To be motivational for the workers due to be in charge of assembling the whole products.
- To improve productivity due to high motivation of workers.
- To reduce work-in-process inventories and workspace.

The most popular cellular manufacturing systems locate working tables around a worker for assembling products, as shown in Fig. 1. The working area constructed by the working tables is called 'cell'. Each worker is provided for assembling tools and jigs in the cell. The workers assemble various products, based on a predetermined production schedule.

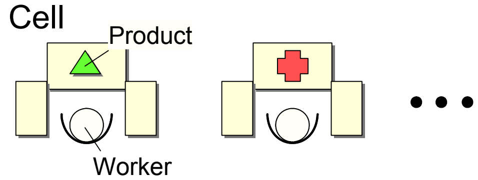


Figure 1: Target cellular manufacturing systems.

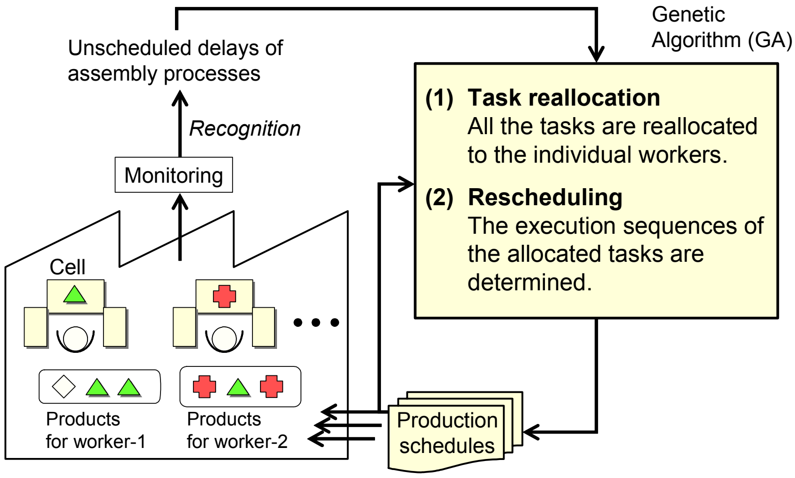


Figure 2: Whole processes of task reallocation and rescheduling.

2.2 Flexibility in cellular manufacturing systems

The performances of cellular manufacturing systems depend on human capabilities to enhance productivity and flexibility. However, it is difficult to control the production capacity of the cellular manufacturing systems, since human capabilities are not stable and controllable. For instance, unforeseen changes, such as delays of assembly processes, addition of new orders, and unscheduled workers' absences, occur during the progress of assembly processes in the actual cellular manufacturing systems. On the other hand, the processing time of workers is gradually improved by repeating the same tasks.

Generally, untrained workers are often supported by neighbour workers or factory managers, when the assembly processes of the untrained workers are delayed. The cooperation among the workers may contribute to the flexibility and the reliability of the assembly processes. However, the effectiveness of the cooperation has not yet been evaluated systematically from the viewpoint of the whole manufacturing systems. Therefore, the cellular manufacturing systems need a suitable control method to keep the production capacity, without interrupting the progress of assembly processes of workers, in the case where the capabilities of cells unexpectedly change during the progress of assembly processes in the cellular manufacturing systems.

3 HUMAN-ORIENTED DYNAMIC TASK REALLOCATION AND RESCHEDULING METHOD

3.1 Genetic algorithm based task reallocation and rescheduling method

This research proposes a dynamic task reallocation and rescheduling method in the cellular manufacturing systems, consisting of a set of assembly cells in which all the assembly processes are carried out by single workers, as shown in Fig. 2. All the tasks are firstly assigned to the individual workers and the execution sequences of the tasks are determined in the initial scheduling processes, before starting the assembly processes.

The method proposed here dynamically changes the initial task allocation for workers and modifies the execution sequences of the tasks, in the cases where unforeseen changes such as delays of assembly processes occur in the cellular manufacturing systems. The objective function considered here is minimization of the total tardiness, as shown in the following equation.

$$Minimize \quad \sum_{i=1}^{n} \max(0, ft_i - dd_i) \tag{1}$$

Where,

ft_i : finishing time of an assembly process of the task J_i.

dd_i : due date of the task J_i.

n : total number of tasks.

The proposed task reallocation and rescheduling procedures consist of the following two steps.

1. All the tasks are reallocated to the individual workers real-timely by using the GA (genetic algorithm) [1], and
2. The execution sequences of the allocated tasks in the individual workers are determined by using a heuristic rule, called as EDD (Earliest Due Date) [2].

Reallocation of tasks

The task allocation for workers is represented as a chromosome in the GA based task reallocation method. The GA is a probabilistic search technique based on the evolution mechanism. Figure 3 shows an algorithm of the task reallocation using the GA. The algorithm starts with a population of parent individuals from which offspring are generated. Each individual has a chromosome, and it is evaluated based on a fitness value. The genes in the chromosome represent names of workers in this research. The names of workers are allocated to an array in the order of the number of tasks, as shown in Fig. 3. This array represents the chromosome of the first individual. Other individuals in the initial population are created by randomly changing the positions of genes in the first individual, until the number of individuals is equal to the predetermined population size.

Genetic operators, such as selection, crossover and mutation, are applied to the first individuals, in order to create

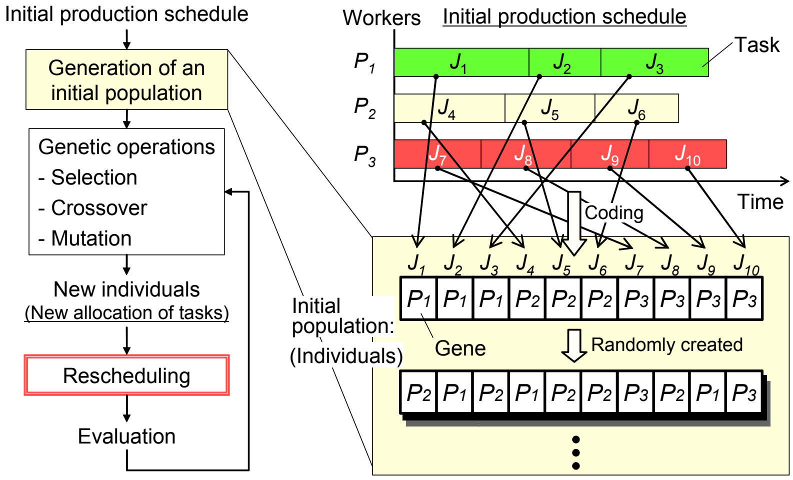


Figure 3: Algorithm of task reallocation using the GA.

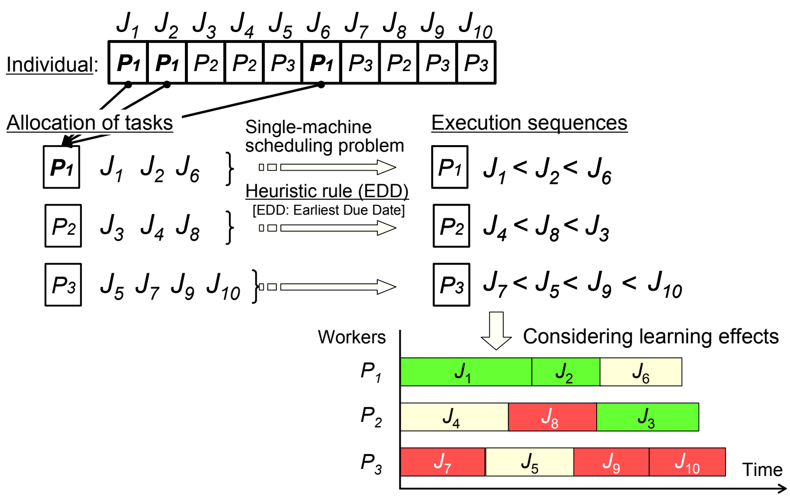


Figure 4: Rescheduling process.

new individuals representing the improved task allocations. Selection operator selects an individual which has the best fitness value in the population, and the selected individual survives to the next population. Crossover operator called multi-points crossover randomly selects many points of chromosomes and exchanges genes of the selected points of two parent individuals, in order to create new offspring individuals. Mutation operator called multi-points mutation considered here randomly selects and changes a few genes in single individuals.

Determination of execution sequences of allocated tasks

Decoding processes of the individuals provide new task allocations for workers, as shown in Fig. 4. A heuristic rule, called as EDD (Earliest Due Date), is applied to the allocated tasks, in order to generate suitable execution sequences of the tasks, since the sequencing problem of the tasks is recognized as a single-machine scheduling problem. Then, feasible production schedules are newly generated from the individuals.

A learning curve effect is considered in this research in generation of a production schedule. The learning curve effect shows that the workers are trained by repeating same assembly processes. We apply a typical learning curve which shows that the processing time by the workers is improved in an exponential manner by repeating the same tasks, as shown in Fig. 5. A learning rate β is defined as the following equation.

$$\beta = \frac{1}{2^{\alpha}} \tag{2}$$

$$\log H_N = \log H_1 - \alpha \log N \tag{3}$$

Where,

H_N : cumulative average time of tasks assembled.

H_1 : processing time required to assemble the first task.

N : cumulative number of tasks assembled.

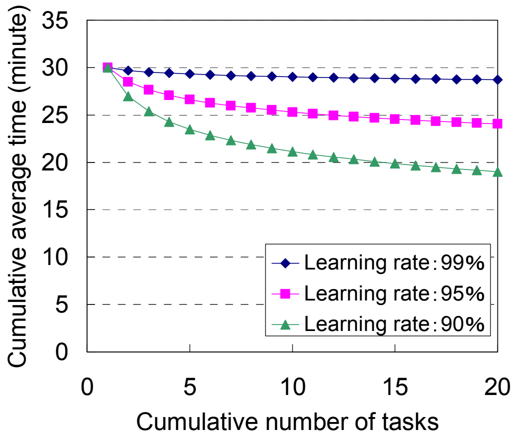


Figure 5: Learning curve.

The proposed method tries to allocate same tasks to the workers using the GA, aimed at reducing the processing times of the tasks based on the learning curve effects. The method also considers the load-balance among all the workers in order to minimize the total tardiness.

Evaluation of new production schedules

A task reallocation and rescheduling process generates production schedules based on the individuals in a population. The most suitable production schedule is selected from the viewpoint of the total tardiness, in order to replace a current production schedule with the new production schedule. The assembly processes are carried out according to the selected production schedule, if the selected production schedule improves the total tardiness of the current production schedule. The task reallocation and rescheduling process is repeated continuously, until the total tardiness is reduced to zero, or until all the tasks have been already started assembling. Therefore, the proposed method can improve the disturbed production schedule without stopping the progress of assembly processes.

3.2 Dynamic task reallocation and rescheduling process

The task reallocation and rescheduling process is activated during the progress of assembly processes at the present time T_0, only when the delays of assembly processes occur and the predetermined initial production schedule does not satisfy the given constraints on delivery times. It is assumed that a cycle of task reallocation and rescheduling process takes computation time *dt* to generate a new feasible production schedule. Therefore, the schedule of the tasks that are started assembling after (T_0 + *dt*) can be modified in the task reallocation and rescheduling process [3] [4].

The following steps represent a task reallocation and rescheduling process proposed in this research.

Step 1 Setting up of the present time T_i

The present time T_i (i=0, 1, ...) is set up.

Step 2 Creation of an initial population

Two cases are considered in the creation of the initial population. They are,

1. First activation of the task reallocation and rescheduling process at time T_0
2. Second or later activations of the task reallocation and rescheduling process at time T_1 or later.

Step 2-1 First activation

Individuals in the first population are created based on the task reallocation method as described in Section 3.1.

Step 2-2 Second or later activations

Some individuals created in the previous task reallocation and rescheduling process can be inherited to the next initial population in the iteration of the task reallocation and rescheduling process, as shown in Fig. 6. Two cases are considered for the inheritance process of the population as shown in the followings.

Case-A No tasks start between T_i and (T_i + dt)

If no tasks start between T_i and (T_i + *dt*), all the individuals of the last population of the previous task reallocation and rescheduling process are inherited to a new task reallocation and rescheduling process between T_i and (T_i + *dt*).

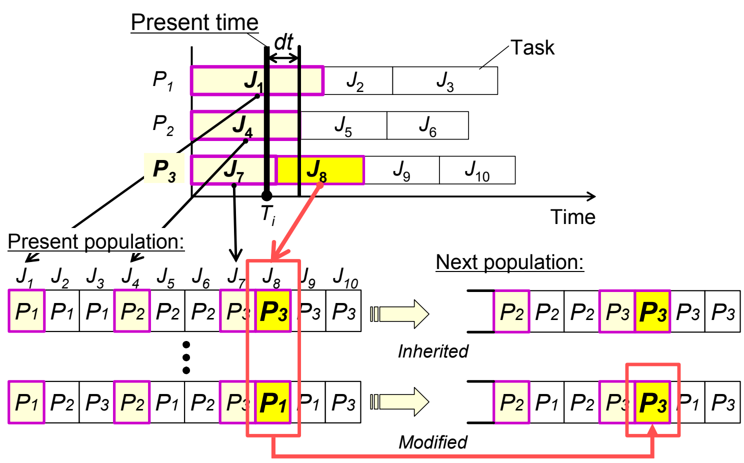


Figure 6: Iteration of task reallocation and rescheduling process.

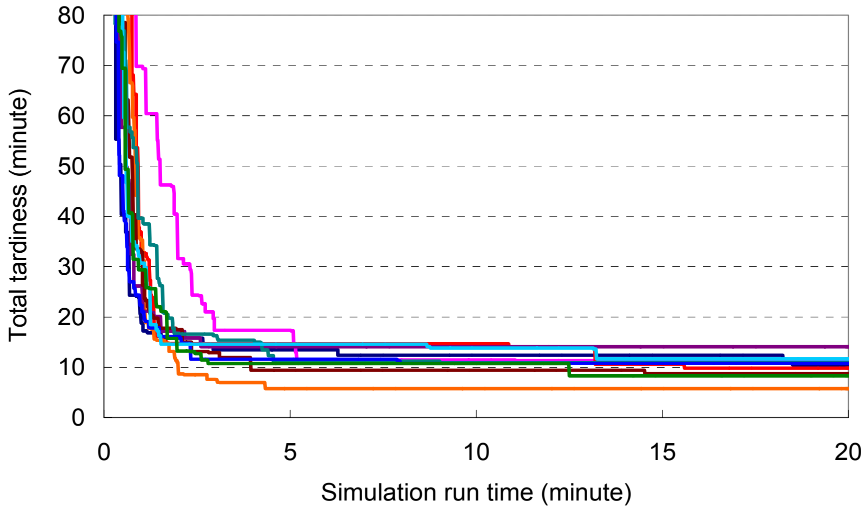


Figure 7: Variations of the total tardiness in experiments.

Case-B Some tasks start between T_i and (T_i + dt)

If some tasks start between T_i and (T_i + dt), the production schedules of these tasks should be fixed. Therefore, a new task reallocation and rescheduling process can inherit only the individuals, which are consistent with the schedules of the fixed tasks, from the last population created in the previous task reallocation and rescheduling process.

Step 3 Application of genetic operators to the population

Production schedules are generated from the individuals by applying EDD and the learning curve effect, in order to calculate the total tardiness which has to be minimized as the fitness value of each individual. Based on the fitness value, genetic operators, such as selection, crossover and mutation, are applied to the individuals of the population created in Step 2, in order to create new individuals of the next population.

Step 4 Evaluation of modified production schedules

The new individuals created in Step 3 generate production schedules. If one of the newly generated schedules is better than the current schedule from the viewpoint of the total tardiness, the current schedule is substituted by the new schedule. All the task reallocation and rescheduling processes are terminated, if the total tardiness of the new production schedule is zero, or all the manufacturing operations have been already started. Otherwise, the steps from Step 2 to Step 4 are repeated.

4 COMPUTATIONAL EXPERIMENTS

4.1 Experimental conditions

A prototype system of dynamic task reallocation and rescheduling was implemented by using an object-oriented language, Smalltalk. It was developed on a personal computer operating under the Windows system.

Computational experiments were carried out to evaluate the effectiveness for such unforeseen disruptions as a significant delay of an assembly process. A cellular manufacturing system had a suitable production schedule at the initial condition. Experimental conditions are summarized as follows.

- Number of workers: 12.
  - Trained workers: 3.
  - Untrained workers: 9.
- Number of tasks: 145.
  - Types of tasks: 10.
  - Standard processing time: 5 ~ 60 [min.]
  - Due date: 120 or 240 [min.]
- Total tardiness at the initial condition: 0.

Any worker can assemble all the tasks. The workers take a standard processing time in the case where the workers assemble each type of tasks for the first time. However, the processing time of tasks decreases based on the learning rate of workers in the case where the workers have assembled the same type of tasks before.

A significant delay of an assembly process occurred in a cell after the assembly processes had been started in the cellular manufacturing system. The total tardiness of the production schedule was increased to 258 [min] due to the delay. Then, the prototype system activated the task reallocation and rescheduling process in order to improve the disturbed production schedule.

4.2 Experimental results

Computational experiments have been carried out ten times on the same experimental conditions, since the GA based method is a probabilistic search technique and has random computational operations in creating a population. Population size, crossover rate and mutation rate were 100, 0.4 and 0.01, respectively.

Figure 7 shows a part of experimental results. The horizontal axis and the vertical axis are the simulation run time and the total tardiness of production schedules, respectively. Lines on the figure show the improvements of the total tardiness of the current production schedules. The proposed method improved the disturbed production schedules and generated suitable production schedules from the viewpoint of the minimization of the total tardiness, considering the learning curve effects of the workers. The results of the experiments are summarized in Table 1.

4.3 Comparisons of experimental results

The effectiveness of the proposed method was verified through the comparison between the experimental results of the GA based method proposed in this research and the one of a rule-based method. The rule-based method reallocates a task with the largest tardiness to a worker that has the shortest total processing time. This kind of method might be used in the actual cellular manufacturing systems, in the case where unscheduled disruptions occur in the cellular manufacturing systems.

The experimental result of the rule-based method is also shown in Table 1. All the experimental results are summarized in Fig. 8. As shown in this figure, the total tardiness of the production schedule improved by the proposed method is shorter than the one of the production schedule improved by the rule-based method. Therefore, the proposed method is superior to the rule-based method from the viewpoint of the minimization of the total tardiness.

5 CONCLUSIONS

This research proposed a dynamic task reallocation and rescheduling method in the cellular manufacturing systems. The proposed method real-timely changes the task allocation for the workers by using the GA and modifies the execution sequences of the tasks by using EDD, in the cases where delays of assembly processes occur in the cellular manufacturing systems. Some computational experiments verified the effectiveness of the proposed method from the viewpoint of the minimization of the total tardiness.

6 REFERENCES

[1] Hollanad JH, 1975, Adaptation in Natural and Artificial Systems. University of Michigan Press.

[2] Jackson, J.R., 1955. Scheduling a Production Line to Minimize Maximum Tardiness, Research Report 43, University of California, Los Angeles.

[3] Tanimizu, Y. Sugimura, N., 2002, A study on reactive scheduling based on genetic algorithm, Proceedings of the 35th CIRP-ISMS, 219–224.

[4] Tanimizu, Y., Sakaguchi, T., Iwamura, K., Sugimura, N., 2006, Evolutional Reactive Scheduling for Agile Manufacturing Systems, International Journal of Production Research, 44, 18–19, 3727–3742.

Table 1: Experimental results.

		Total tardiness (min.)
Proposed method	Average	9.8
	Maximum	10.8
	Minimum	5.3
Rule based method		24.2

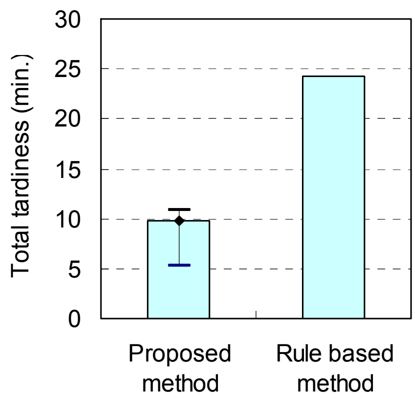


Figure 8: Summary of experimental results.

Shape Data Registration based on Structured Light Pattern Direction

Tatsuya Ogino[1], Yoshihiro Yasumuro[2] and Masahiko Fuyuki[2]
[1] Graduate School of Engineering, Kansai University, Osaka, Japan
[2] Faculty of Environmental and Urban Engineering, Kansai University, Osaka, Japan

Abstract
Structured light is used for measuring 3D shape efficiently by active stereo methods. The measurement resolution of this method depends on not only the interval of structured light but also the pattern direction. In this paper, we propose a method for acquiring homogeneous data by structured light measurements. The method uses multiple structured lights with different directions and then registers the data based on measurement density over the object surface. Our experimental results showed that the proposed method is capable of reproducing 3D shape of measurement objects more precisely.

Keywords:
Active Stereo Method; Structured light; Data Registration

1 INTRODUCTION

Technologies for measuring three-dimensional (3D) shapes of physical objects have gathered more and more attention in various fields in recent years. Reverse engineering, for instance, has been introduced in the manufacturing business for a swift production process. While pre-existing digital shape models are available, the product process goes smoothly, once, however, new models are required, costly designing labour is needed even by up-to-date computer-aided design (CAD) or computer-aided modelling (CAM) software packages. Designing object shape with a physical clay model is often more flexible than modelling with CAD/CAM software, however, digital design data is suitable for further production processes. 3D shape capturing is capable of connecting physical and digital designing phases. Moreover, prior to the mass-production, recent rapid prototyping machines and physical test exemplars are available [1, 2, 3]. Consequently, a smooth transition from physical to digital designing has a great potential to shorten the total time for production development. 3D shape scanning is also actively used for the field of science and education. For instance, a number of cultural properties, including historical relics, are targeted for a 3D scan before they deteriorate [4]. Large-scale objects, such as a whole building structure or a section of a town are also scanned and archived in a digital media form, recently called “e-monuments”. E-monuments have a great potential for preserving existing properties of those cultural heritages and reusing them for exhibition via computer graphics with additional information for easy-to-understand display. Those digital archives enable not only quantitative but qualitative analysis from an archaeological point of view.

One of the major techniques used for 3D measurement is optical triangulation with an active stereo system. A combination of a light emitter and an imaging sensor (a camera) is used in active stereo methods. For emitting light, spotlight or slit light is generally used. The measurement with spotlight employs a focused laser spotlight and elaborately controls the light emission to scan densely and uniformly. Consequently, spot- or point- based scanning takes time to sweep the surface to complete the measurement. Structured light is used for time-effective 3D scanning. The simplest structured light is line shaped or a slit beam, which is adopted in many commercial laser range scanners for a tabletop range scale. The slit beam forms a 3D plane that intersects the surface of the object. This allows capturing the range of points on the intersection line, and thus, rapid measurement is possible compared with the spot-light. Moreover, by coding with the structured light, spatial division becomes more effective. Consecutive illumination with several patterns of structured light is equivalent to hundreds of slit-lines scanning resolution. Most of the 3D scanning systems for structured light are implemented as projector-camera systems and a variety of coding methods are proposed [5,6].

A common usage of 3D scanning systems is to take multiple range images and put them together in a coordinate system for reconstructing a target object shape. Multiple measurements with different positions and angles allow capturing the whole shape of the object as well as compensating for missing parts occluded from the other views. Thus, the quality of the captured data is supposed to be ensured by redundant multiple scans.

We focus rather on the quality of a single scanned data set. Intrinsically, visual distortion of a projected pattern on the object surface causes parallactic displacement that enables optical triangulation. However, the parallactic displacement also gives an uneven distribution of the measured points. This paper proposes a method for acquiring homogeneous density data from a fixed viewpoint with multiple structured light patterns.

2 PATTEN DIRECTIVITY IN STRUCTURED LIGHT

In general, 3D shape measurement with structured light scans along the structured line pattern. Figure 1 shows an example of the measurement data (right) with vertical line patterns (left).

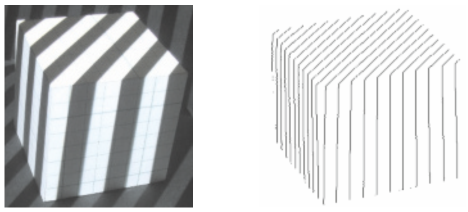


Figure 1: Density difference of range data over the object surface

The observed intervals of the captured lines are different because the faces vary face-by-face. Ridgelines, especially, are sampled with a different number of points. In Figure 1, the projected lines are almost parallel to the vertical ridges on which only one scan line traverses. The ridges have the characteristic of forming the whole shape of the cube, however, they are easily miss-sampled in the captured point set. The characteristic parts tend to contain higher curvature, which requires more scanned points to reconstruct the original shape.

3 PROPOSED APPROACH

We focus on the fact that the same surface shape can be sampled with a different density by varying the scan line directions. We employ multiple structured lights with different pattern directions and monitor the density of captured points for each scan. Characteristic regions can be detected as the ones whose data density varies by changing scan line directions. Then measurement results with higher density are selected for characteristic regions and registered for the final data set.

We categorize the edges in objects into three types: jump, roof, and smooth edges (see Figure 2). A jump edge is observed as a large gap in a range distribution from the measurement viewpoint. The jump edges can be detected as discontinuities along the scan lines. They appear in the contour of objects' silhouettes in general. A roof edge and a smooth edge have a continuous series of points on the scan lines. The normal vector on the surface of the object continuously changes as well. A roof edge and a smooth edge can be detected by searching for high variance in the normal vector distribution. Creating a mesh from the point sets without overlap allows finding adjacency relationships among the points. A normal vector in each facet on the mesh can be calculated and the variance of the direction of the normal vector within a small region shows the shape complexity of the region [7]. Comparing the densities of range data between data sets acquired by different structured lights, higher density data can be selected for each region. The measurement data can be supplemented by registration based on measurement density.

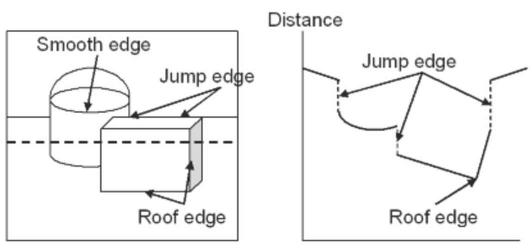


Figure 2: Edge types

4 SHAPE DATA REGISTRATION SCHEME

Since there are no overlaps of the range data from a single view of a camera, range data can be mapped on the image plane of the camera. Hereafter, the image plane of the camera is called the 'screen coordinates'. Since the projector-camera system is fixed while changing pattern directions to project, mapping 3D to 2D screen coordinates allows a simpler process for comparing, selecting, and registering the measured points. The registration process is as follows:

1. Measure the 3D shape with multiple structured lights of different pattern directions. Examine the continuities of the scan lines to detect jump edge regions.
2. Form a Delaunay mesh from a range data set and acquire a connecting structure of the data points.
3. Calculate a normal vector at each vertex on the Delaunay mesh. Make a normal map by mapping the normal vectors onto the screen coordinates.
4. Examining variance of the normal directions over the normal map.
5. Detect high variance regions of normal vectors to find roof and smooth edges.

Calculate the density of each range data set for each region containing jump, roof, and smooth edges, and register the higher density data for the region.

5 IMPLEMENTATION

5.1 Structured light with gray code

In this paper, we implemented a projector-camera system for 3D scanning by structured light with a gray code system [8]. A set of gray codes can be expressed by consecutive lighting patterns. Each pattern is composed of black and white stripes. N patterns generate N bits of gray code. To establish the correspondence between the pixels of a projector and those of a camera, every pixel on the camera image is estimated as either shaded by a black-stripe area or lighted by a white-stripe area from the projector. As shown in Figure 3, for example, lighting with a projector with 3 patterns: A, B and C generates 3 bits of gray code. The lighted area is segmented to be read as 1 and the shaded area as 0. The segmented area is coded independently by a "space code". Referring to the space code, a ray from a projector pixel and camera pixel are coupled and form an epipoler triangle, and thus, the 3D coordinates can be calculated as the intersection of the 2 rays from a calibrated projector and camera system.

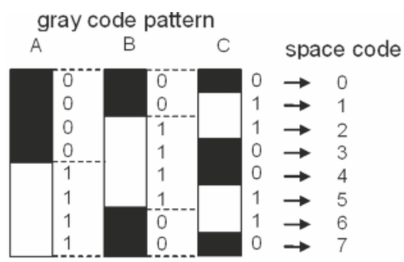


Figure 3: Gray code pattern (3 bit)

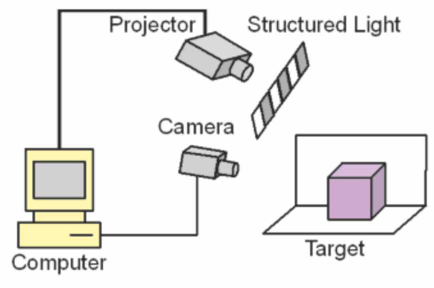
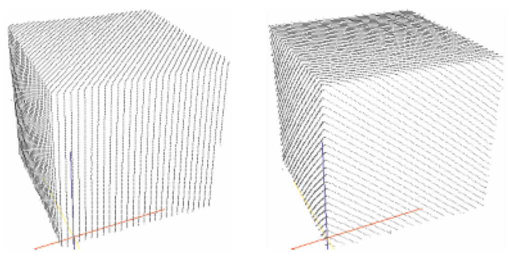


Figure 4: Measurement System Setup

5.2 Measurement system

We implemented a measurement system to execute the process mentioned above. This system uses a single projector and camera system as shown in Figure 4 and is capable of triangulation by structured light with gray code and registration of the acquired data. A set of structured patterns are sequentially projected onto a target object from the projector and the camera captures the scene for each projection. The system uses two kinds of directional patterns: vertical and horizontal. 3D measurement results for the two patterns are shown in Figure 5. These data are a whole series of points that have 3D coordinates. The following is the process to acquire a registered result from these range data. First, map the measurement data shown in Figure 6 onto the screen coordinates (Figure 6(a)). Next, form a Delaunay mesh and calculate the normal vector of each vertex. Third, make a normal map on the screen coordinates by separating each component of a normal vector by colour (Figure 6(b)). Because the calculated normal vectors might contain some noise, moving the average filtering smoothes the normal map. The variance distribution of the normal vector is calculated in each small region of 7x7 pixels over the normal map. Then high variance regions of the normal vector (Black region of Figure 6(c)) are detected. The threshold for detection was empirically determined. The data density is computed by counting the measurement points within the small region of 7x7 pixels. Finally, by comparing the density measured by vertical pattern with the one measured by horizontal pattern, a total data set is filled with higher density points.

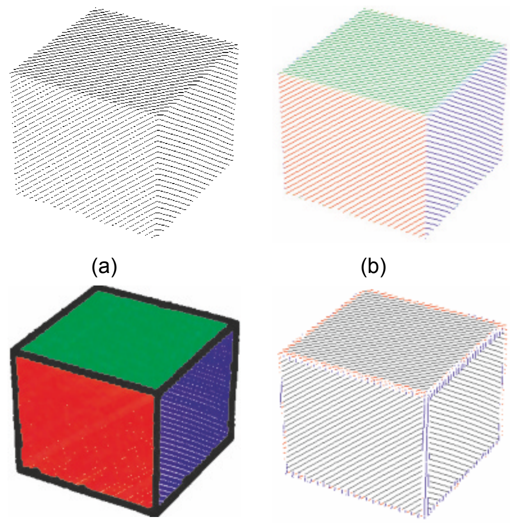


(a) (b)

Figure 5: Scanning results with vertical pattern (a) and horizontal pattern (b)

5.3 Pattern boundary interpolation

Range values are computed from camera images that capture projected binary patterns onto the object surface. This image is sampled as a digital image, on which boundaries of the binary



(a) (b)

(c) (d)

Figure 6: Measurement System

stripes are used as scan lines. Though the projected patterns are continuously distorted due to the object surface shape, stripes are used as scan lines. Though the projected patterns are continuously distorted due to the object surface shape, detected pattern lines are jagged because of aliasing with the limited spatial resolution of pixels. The jagged boundary edges directly deteriorate the resulting shape data. Therefore, a sub-pixel process is needed for detecting boundary edge lines. In this paper, positive and negative patterns are used for each stripe width [7]. A negative pattern is the dual of the inverted pattern of the positive pattern. While pixel intensity has an up-ward slope in the positive projection, there must be a down-ward slope in the dual projection image. In the same vicinity of boundary lines, the pattern border can be computed as a set of intersecting points of the up-ward and down-ward curves, by comparing between the dual images. Continuous boundary lines are acquired by interpolating the positions of the intersection points in the image plane. Figure 7 shows a comparison between shape reconstruction results with and without sub-pixel interpolation for boundary line detection.

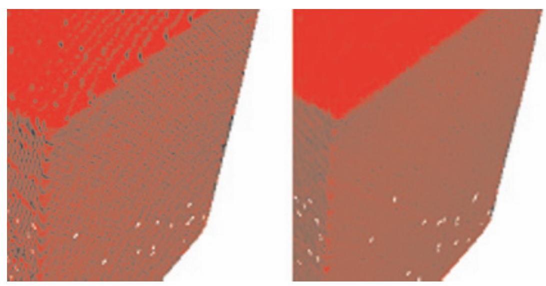


Figure 7: non-interpolated result (left) and interpolated result (right)

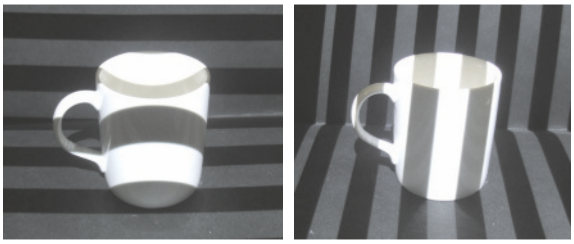
6 EXPERIMENTS

6.1 Experimental conditions

We conducted an experiment, implementing the process described in the previous section. We measured a cup shown in Figure 8(a) and a plaster statue shown in 8(b). The positions of the projector, the camera and the target object are shown in Figure 10. Figure 9(a) shows a scene where a horizontal pattern was projected onto the cup. Figure 9(b) also shows a scene with a vertical pattern. We used eight sheets of patterns for each vertical and horizontal direction. This pattern is capable of segmenting the space into 256, which is equivalent to 256 scanning lines. The projector we used is a DLP type and the maximum brightness is 2500 lumens. The capture size from the camera is 640x480 pixels.

(a) (b)

Figure 8:Target Object, (a) cup and (c) plaster statue



(a) (b)

Figure 9: (a) Horizontal Projection and (b) Vertical projection

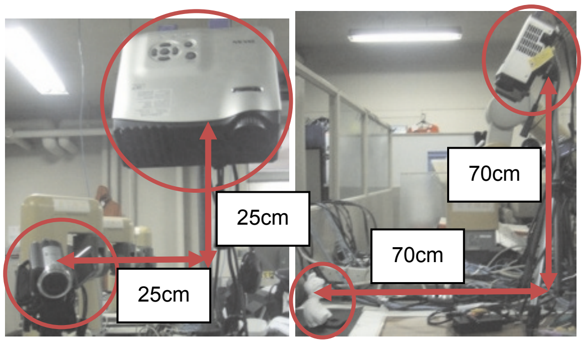


Figure 10: Position of projector, camera and target

6.2 Results and Discussion

Figure 11(a) and 18(a) show results measured by the vertical pattern. Figure 12(a) and 19(a) show results by the horizontal pattern. In a complex characteristic region, such as the handle edges of the cup and the face and the body parts of the plaster statue, the measurement density varies along the surface direction. Figure 15 and Figure 20 shows the selected regions for registration by the proposed method. In characteristic regions, result data measured by either vertical or horizontal pattern are selected based on measurement density. In non-characteristic regions, both results have similar quality. The horizontal pattern data is used in the Figure. Figure 16(a) and 21(a) show registered results of these data. The lower density areas are complemented through registration. Figure 11(b), 12(b), 17, 18(b), 19(b) and 22 show reconstructed 3D surfaces. In 3D surface rendering representation, the registration result complements the low-density regions. This shows that the registration on the screen coordinates effectively functioned for 3D shape reconstruction. Figures 13 and 14 show close-up of the handle part. The large curvature regions are represented as continuously shape surface with comparatively dense measurement points.

7 SUMMARY

In this paper, our aim was to make the most use of projector-camera system resolutions that are degraded by the relative geometrical conditions, such as pattern directions and edge shapes. To clear up this problem, we proposed a method for acquiring homogeneous density data by multiple structured light measurements. Focusing on measurement density, an automatic scheme for managing the data resolution is achieved. The experimental results showed the advantageous capability of reproducing 3D shapes of measurement objects more precisely. Our future work addresses how to design more flexible patterns instead of only patterns in two directions.

REFERENCES

[1] Andrew Willis, Jasper Speicher, David B. Cooper, 2007, Rapid prototyping 3D objects from scanned measurement data, Image and Vision Computing, 25/7 :1174-1184.

[2] Krause, F.-L., Kimura, F., Kjellberg, T., Lu, S.C.-Y., 1993, Product Modelling, Annals of the CIRP, 42/2:695-706.

[3] T. Varady, R. R. Martin and J. Cox, 1997, Reverse engineering of geometric models. An introduction, Computer-Aided Design, 29/4:255-268.

[4] A. Banno, T. Masuda, T. Oishi, and K. Ikeuchi, 2008, Flying Laser Range Sensor for Large-Scale Site-Modeling and Its Applications in Bayon Digital Archival Project, International Journal of Computer Vision, 78/2-3:207-222.

[5] Samet, H., 1990, Applications of Spatial Data Structure, Addison-Wesley, Reading, MA.

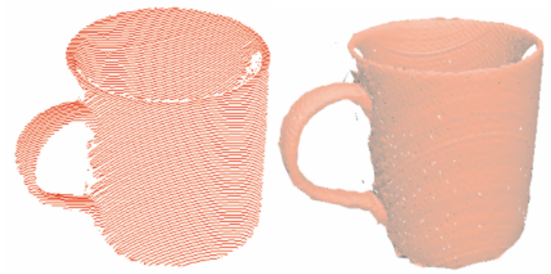
[6] E. Mouaddib, J. Batlle, and J. Salvi, 1997, Recent Progress in Structured light in order to solve the correspondence problem in Stereo Vision, IEEE Int. Conf. on Robotics and Automation, 130-136.

[7] Michikazu Matsumoto, Masataka Imura, Yoshihiro Yasumuro, Yoshitsugu Manabe, Kunihiro Chihara: Support System for Measurement of Relics based on Analysis of Point Clouds, Proceedings of the Tenth International Conference on Virtual Systems and MultiMedia, p.195 Gifu, Japan, 2004.

[8] S. Inokuchi, K. Sato, F. Matsuda, 1984, Range imaging system for 3-D object recognition, in: Proc. of the International Conference on Pattern Recognition, 806-808.

(a) (b)

Figure 11: Vertical pattern measurement
(a) scanned points and (b)3D reconstructed result

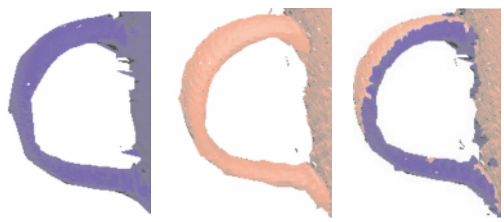


(a) (b)

Figure 12: Horizontal pattern measurement
(a) scanned points and (b)3D reconstructed result

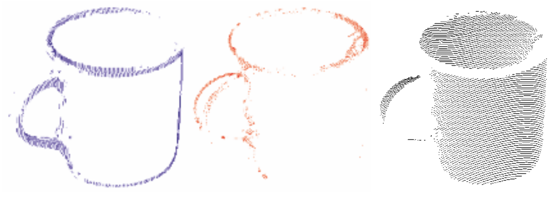
(a) (b) (c)

Figure 13: Measurement results (close-up of handle part):
(a) vertical pattern, (b) horizontal pattern, and
(c) registration result



(a) (b) (c)

Figure 14: 3D rendered result (close-up of handle part):
(a) vertical pattern, (b) horizontal pattern, and
(c) registration result



(a) (b) (c)

Figure 15: selected pattern for registration
(a) vertical pattern (b) horizontal pattern
(c) non-characteristic regions

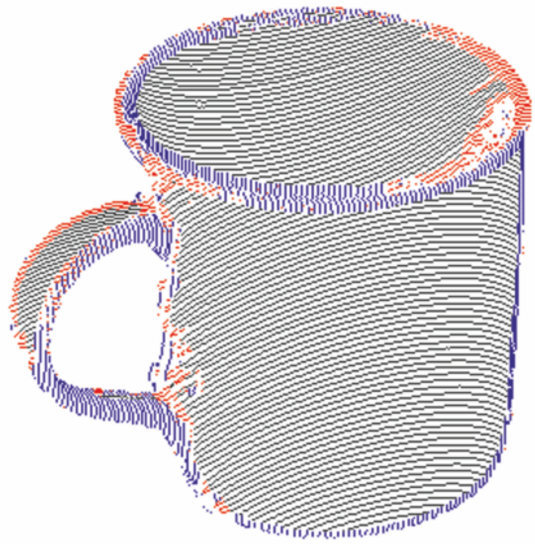


Figure 16: Registration result(camera plane)

Figure 17: Registration result(rendered 3D surface)

(a) (b)

Figure 18: Vertical pattern measurement
(a) scanned points and (b)3D reconstructed result

(a) (b)

Figure 19: Horizontal pattern measurement
(a) scanned points and (b)3D reconstructed result

(a) (b) (c)

Figure 20: selected pattern for registration
(b) vertical pattern (b) horizontal pattern
(c) non-characteristic regions

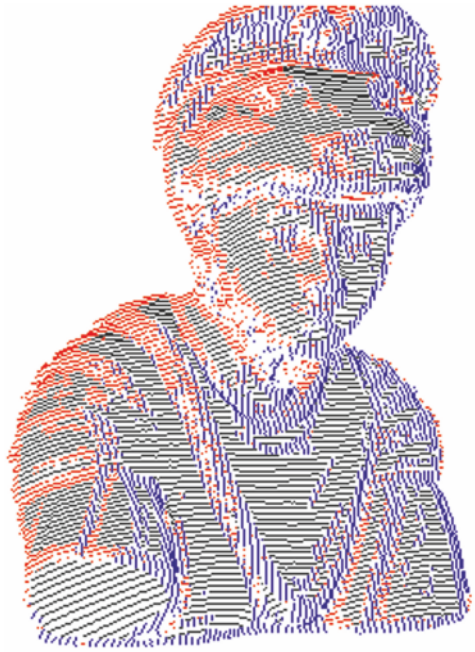


Figure 21: Registration result(camera plane)

Figure 22: Registration result (rendered 3D surface)

Object Tracking System Using Pan-Tilt Cameras and Arm Robot

Hiroyuki Ukida[1], Yasuyuki Yamanaka[1]

[1] Dept. of Mechanical Engineering, Faculty of Engineering, The University of Tokushima, Tokushima, Japan

Abstract
In this paper, we propose an object tracking system using an arm robot and two pan-tilt cameras. Here, we use the active search method to detect an object in images, and we estimate 3D coordinates of the object using the binocular stereo method. By using 3D object locations, the robot and cameras are rotated toward to the object. To realize the high speed performance, we use the parallel processing by thread functions for the object detection, the 3D coordinate estimation, the camera control, and the robot control. We perform object tracking experiments and confirm the efficiency of our proposed method.

Keywords:
Object Tracking; Pan-Tilt Camera; Arm Robot; Active Search; Stereo Method; Parallel Processing

1 INTRODUCTION

For autonomous mobile robots, it is important and fundamental techniques to acquire information and to trace a moving object surrounding of robots. Sensors to investigate the surrounding of robots are developed, for example, ultrasonic waves, lasers, and TV cameras. In particular, TV camera is effective to acquire information in a large area at one time.

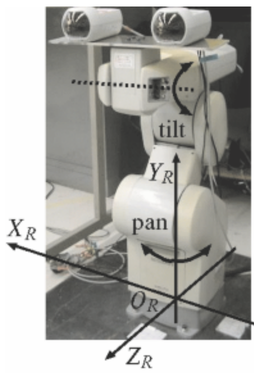
But, to acquire information about surrounding of robots, it is necessary to move cameras. Recently, a "pan-tilt camera" is also used in the teleconference systems, etc. This camera can rotate its lens and imaging device to horizontal (pan) and vertical (tilt) direction quickly [1],[2]. But, the ranges of pan and tilt rotations are still limited. Hence, we need another locomotive mechanism of cameras to inspect about surrounding of robots [3].

On the other hand, an "omni-directional camera", which can take an image of its surroundings all at once without moving by using a hyperbolic mirror, has been developed [4]. However, the resolution of captured image is coarse because the large area is taken by one CCD sensor. And, since the CCD sensor take an image on the hyperbolic mirror, its image is distorted. Hence, it is necessary to perform distortion correction process. Moreover, the omni-directional camera can take image at horizontal surroundings, but the vertical range is limited.

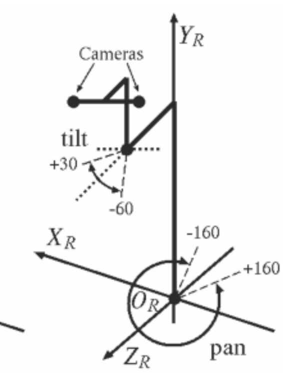
In this research, we propose a new tracking system using an arm robot equipped with two pan-tilt cameras. The pan-tilt cameras can move quickly but its ranges are limited. The speed of the arm robot is slow, but its degree of freedom is high. By using this system, we discuss the high speed and wide range object tracking method.

In addition, we use "the active search method" [5], [6], [7] to find an object in camera images. This method uses color information of objects and it can search faster than the ordinary template matching method. In this paper, we improve its scanning path, and show more fast performance.

This paper is organized as follows: In section 2, we show the configuration of proposed system. Section 3 denotes our tracking method. In section 4, we show experimental results of object tracking. And we conclude our paper in section 5.



(a) Overview.



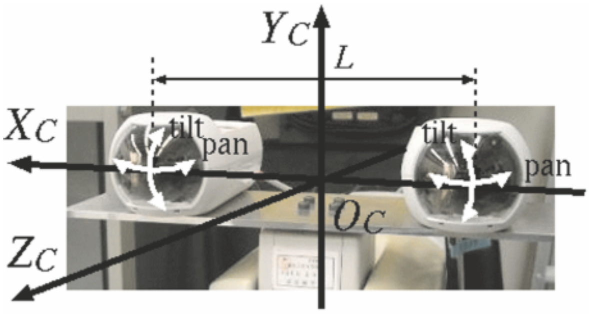
(b) Robot coordinate system.

Figure 1: Arm robot and pan-tilt cameras.

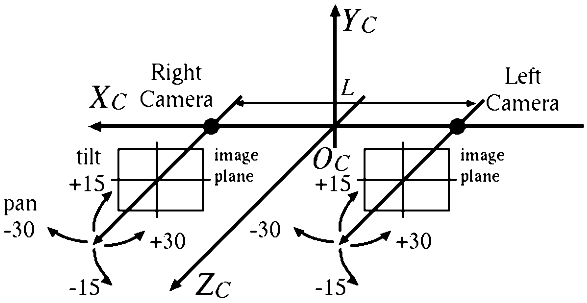
2 SYSTEM CONFIGURATION

Figure1(a) shows the arm robot and pan-tilt cameras we use. Two pan-tilt cameras are equipped on the upper part of the arm robot. This robot has six joints (six degrees of freedom). But we use only the first and fifth joints. The range of horizontal (pan) rotation by the first joint is from -160 to +160[deg.], and the vertical (tilt) rotation by the fifth joint is from -60 to +30[deg.] (Figure1(b)). The rotational speed of each joint (the angular velocity) is about 25[deg./sec.]. $O_R - X_R - Y_R - Z_R$ in Figure1(b) is called the robot coordinate system attached to the first joint.

Figure2(a) shows the pan-tilt cameras we use. The rotation angles of the imaging device in these cameras are from -30 to +30[deg.] (the horizontal (pan) direction) and from -15 to +15[deg.](the vertical (tilt) direction) (Figure2(b)). The angular velocity of each rotation is about 45[deg./sec.]. The distance L between these cameras is 200[mm]. $O_C - X_C - Y_C - Z_C$



(a) Overview of cameras.



(b) Camera coordinate system.

Figure 2: Pan-tilt cameras.

in Figure2(b) is called the camera coordinate system attached to the center of them.

3 OBJECT TRACKING METHOD

We divide tracking processes into two modules and execute them in parallel (Figure3). And we also prepare the shared memory to transfer some data between these modules. The following shows the overview of each module:

1. Object Detection and Camera Control Module

   Taking images by the right and left cameras, and detecting the object in these images by the active search method. Then, estimating the 3D coordinates of this object by the stereo method [8]. And, moving cameras in which the center of the camera image plane locates at the object position in the input image.

2. Robot Control Module

   Moving the arm robot to face to the object.

The parallel processing of these modules is realized by using the thread mechanism [9] provided by the computer operating system function. By using this mechanism, when the waiting time for movement of the robot is long, it becomes possible to perform another processing. In the following sections, we describe the details of the object detection modules.

3.1 Object Detection and Camera Control Module

This module consists of three processes as follows:

1. For each (right and left) camera, detecting the object location in the captured image.
2. From the object locations in the right and left camera images, estimating the 3D coordinates of the object by the stereo method.
3. Move cameras in which the center of camera locates at the object position.

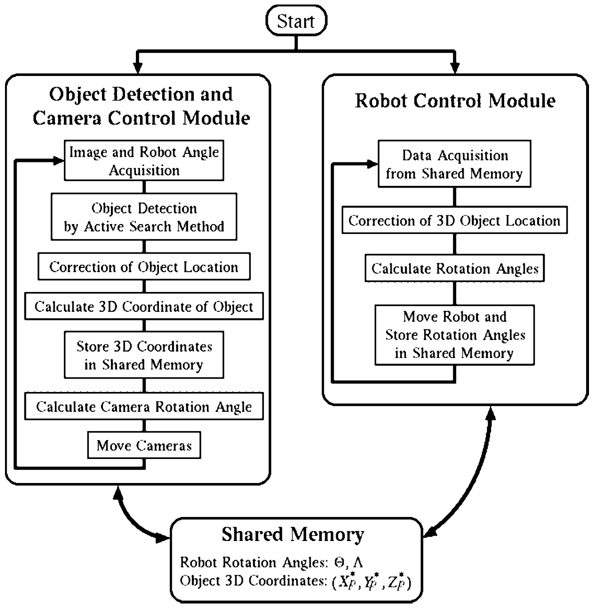


Figure 3: Parallel processing modules.

To detect an object pattern in an input image, the template matching method is famous and is often used. In this method, the sum of absolute differences or the normalized cross-correlations of the intensities between the input image and the object pattern image are used as the similarity measure [10]. But, since this method must perform the matching process changing parameters of the location, pose and size in round robin, it takes much computational time.

To avoid such problems, we use "the active search method" to detect the object in the input camera images. This method compares the object image (we call it "the reference image") with the input image evaluating the similarity calculated from the histogram of the color data in these images. So, this similarity is invariable about the object pose. And, from the histogram calculation, when the similarity of a region can be calculated:

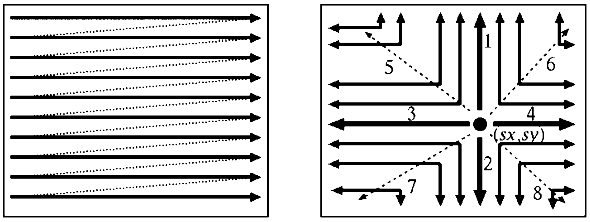
1. In the neighbor regions of this region, the maximum value considered as the similarity can be also estimated without calculating the similarity directly, and,
2. We can estimate a range around the region in which the matching process does not need to perform.

From these characteristics, the active search method does not perform the matching process all over the input image like the template matching. And, because of the histogram calculation, this method guarantees the accuracy as same as the template matching. In other words, this method performs:

- The dense matching processes in the high similarity places, and,
- The rough matching processes in the low similarity places,

and reduces the number of matching processes. Therefore, this method can detect the object more efficiently and fewer computational time than the template matching.

Moreover, the performance of the active search method depends on the object scanning method in the images. If a high similarity can obtain at the begin of scanning, the



(a) Normal path(raster scan). (b) Proposed path(cross scan).

Figure 4: Scanning path

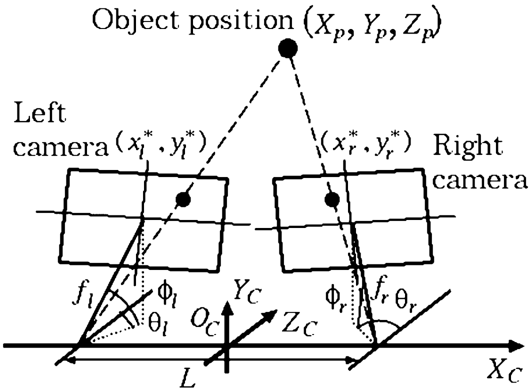


Figure 5: 3D coordinates in camera coordinate system.

number of matching is reduced. Hence, instead of the normal scanning path (raster scan) in Figure4 (a), we employ "the cross road" scanning path (cross scan) as shown in Figure4 (b) where (sx, sy) denotes the object position detected in previous search.

The following shows the processes of the object detection module using the active search method.

Step 1-1. Image Acquisition

Acquire the input images by right and left cameras simultaneously. In this step, the rotational angles of the robot (Θ, Λ) are also acquired from the shared memory.

Step 1-2. Object Detection by Active Search Method

For each input image, detect the object location in the image by using the active search method. Let the center of gravity locations of the object areas in these images as (x_r, y_r), (x_l, y_l). If the object location can not be detected in one or both images, return to Step 1-1.

Step 1-3. Correction of Object Location

The object locations in input images (x_r, y_r) and (x_l, y_l) are not always corresponding coordinates because of the difference of the appearance in left and right images. But, when (x_r, y_r) and (x_l, y_l) are not corresponded, we can not calculate the 3D coordinates accurately in next Step 1-4. Hence, we correct the object location by the template matching using sub images around (x_r, y_r) and (x_l, y_l). Let corrected object locations as (x_r^*, y_r^*) and (x_l^*, y_l^*).

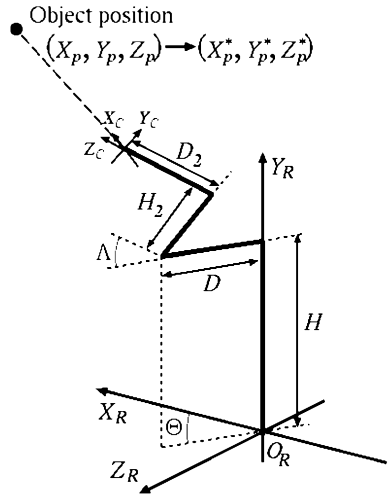


Figure 6: 3D coordinates in robot coordinate system.

Step 1-4. Calculate 3D Coordinates

First, the 3D coordinates (X_p, Y_p, Z_p) of the object in the camera coordinate system are calculated by using the object location in images (x_r^*, y_r^*) and (x_l^*, y_l^*) (Figure5):

$$X_p = \frac{L}{2\Delta}\left\{\left(y_r'^2 x_l'^2 - x_r'^2 y_l'^2\right) + \left(z_r'^2 x_l'^2 - x_r'^2 z_l'^2\right)\right\}, \quad (1)$$

$$Y_p = \frac{L}{2\Delta}\left\{2y_r' y_l'\left(y_r' x_l' - x_r' y_l'\right) + \left(y_r' z_l' + y_l' z_r'\right)\left(z_r' x_l' - x_r' z_l'\right)\right\}, \quad (2)$$

$$Z_p = \frac{L}{2\Delta}\left\{2z_r' z_l'\left(z_r' x_l' - x_r' z_l'\right) + \left(z_r' y_l' + y_r' z_l'\right)\left(y_r' x_l' - x_r' y_l'\right)\right\}, \quad (3)$$

where,

$$\Delta = \left(y_r' x_l' - x_r' y_l'\right)^2 + \left(z_r' x_l' - x_r' z_l'\right)^2 + \left(z_r' y_l' - y_r' z_l'\right)^2, \quad (4)$$

$$\begin{cases} x_r' = x_r^* \cos\theta_r + f_r \sin\theta_r, \\ y_r' = -x_r^* \sin\theta_r \sin\phi_r + y_r^* \cos\phi_r + f_r \cos\theta_r \sin\phi_r, \\ z_r' = -x_r^* \sin\theta_r \cos\phi_r - y_r^* \sin\phi_r + f_r \cos\theta_r \cos\phi_r, \end{cases} \quad (5)$$

$$\begin{cases} x_l' = x_l^* \cos\theta_l + f_l \sin\theta_l, \\ y_l' = -x_l^* \sin\theta_l \sin\phi_l + y_l^* \cos\phi_l + f_l \cos\theta_l \sin\phi_l, \\ z_l' = -x_l^* \sin\theta_l \cos\phi_l - y_l^* \sin\phi_l + f_l \cos\theta_l \cos\phi_l, \end{cases} \quad (6)$$

θ_r, ϕ_r and f_r are the pan angle, the tilt angle and the focal length of the right camera, θ_l, ϕ_l and f_l are those of the left camera.

Next, the 3D coordinates (X_p^*, Y_p^*, Z_p^*) of the object in the robot coordinate system are calculated by the following equations:

$$X_p^* = \left(\left(Z_p + D_2\right)\cos\Lambda - \left(Y_P + H_2\right)\sin\Lambda + D\right)\cos\Theta + X_p \sin\Theta, \quad (7)$$

$$Y_p^* = \left(Z_p + D_2\right)\sin\Lambda + \left(Y_P + H_2\right)\cos\Lambda + H, \quad (8)$$

$$Z_p^* = \left(\left(Z_p + D_2\right)\cos\Lambda - \left(Y_P + H_2\right)\sin\Lambda + D\right)\sin\Theta - X_p \cos\Theta, \quad (9)$$

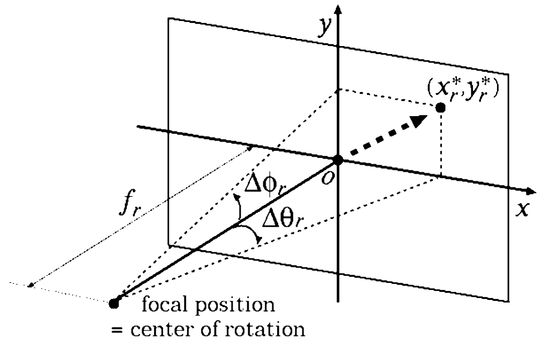
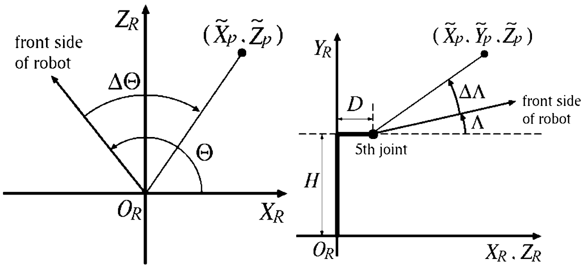


Figure 7: Camera rotation angle.



(a) Pan rotation angle. (b) Tilt rotation angle.

Figure 8: Robot rotation angle.

where Θ is the robot pan angle, Λ is the robot tilt angle, *D*, *H*, D_2 and H_2 are the lengths of arms of the robot (Figure6). The estimated 3D coordinates of the object (X_p^*, Y_p^*, Z_p^*) are stored in the shared memory.

Step 1-5. Rotate Camera Direction

In this paper, the policy of the camera movement is to move the center of the camera image to the location of the object obtained in Step 1-3. By this policy, our system will be able to track the object which moves anywhere.

As the characteristic of this camera, the focal position is corresponding to the center of the rotation. From this characteristics and the location of the object in the input images, the relative rotation angle of the pan and tilt direction ($\Delta\theta_r$ and $\Delta\phi_r$: the right camera) are calculated by the following equations (Figure7):

$$\Delta\theta_r = \arctan\left(x_r^* / f_r\right) \tag{10}$$

$$\Delta\phi_r = \arctan\left(y_r^* / f_r\right) \tag{11}$$

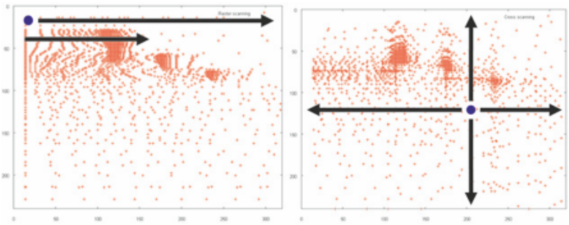
In case of the left camera, $\Delta\theta_l$ and $\Delta\phi_l$ are also calculated by the same way. And after the camera rotation, iterate from Step 1-1.

3.2 Robot Control Module

In this paper, the policy of the robot movement is to move the front of the robot facing to the object 3D coordinates obtained by the object detection and camera control module. By using this policy, our system will be also able to track the object which moves anywhere.

(a) Reference image. (b) Input image. (c) Detected position.

Figure 9: Example of object detection.



(a) Raster scan. (start from (0,0)) (b) Cross scan. (start from (200,120))

Figure 10:Comparison of scanning path.

Table 1: Comparison of matching process.

	Raster scan	Cross scan
Number of matching	2,346	1,642
Computational time [sec.]	0.203	0.156

Step 2-1. Get Data from Shared Memory

The 3D coordinates of the object (X_p^*, Y_p^*, Z_p^*) are acquired from the shared memory.

Step 2-2. Correction of 3D Object Position

Since the 3D object position (X_p^*, Y_p^*, Z_p^*) includes some errors, we correct its coordinates using some of previously estimated 3D object positions. Let $(X_{pi}^*, Y_{pi}^*, Z_{pi}^*)$ $(i = 1 \cdots n)$ the previously estimated 3D coordinates, t_i the time when 3D coordinates are calculated, we estimate the corrected 3D coordinate $(\tilde{X}_p, \tilde{Y}_p, \tilde{Z}_p)$ at this time (t_p) by follows:

$$\begin{cases} \tilde{X}_p = a_x \cdot t_p + b_x, \\ \tilde{Y}_p = a_y \cdot t_p + b_y, \\ \tilde{Z}_p = a_z \cdot t_p + b_z, \end{cases} \tag{12}$$

where (a_x, b_x), (a_y, b_y) and (a_z, b_z) is calculated from the following equations by the linear least square method:

$$\begin{bmatrix} t_1 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{bmatrix} \begin{bmatrix} a_x \\ b_x \end{bmatrix} = \begin{bmatrix} X_{p1}^* \\ \vdots \\ X_{pn}^* \end{bmatrix}, \begin{bmatrix} t_1 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{bmatrix} \begin{bmatrix} a_y \\ b_y \end{bmatrix} = \begin{bmatrix} Y_{p1}^* \\ \vdots \\ Y_{pn}^* \end{bmatrix}, \begin{bmatrix} t_1 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{bmatrix} \begin{bmatrix} a_z \\ b_z \end{bmatrix} = \begin{bmatrix} Z_{p1}^* \\ \vdots \\ Z_{pn}^* \end{bmatrix} \tag{13}$$

Step 2-3. Calculate Robot Rotation Angles

From the corrected 3D coordinates of the object, the relative rotation angle of the pan and tilt direction of the robot ($\Delta\Theta, \Delta\Lambda$) are calculated by the following equations:

$$\Delta\Theta = \arctan\left(\tilde{Z}_p / \tilde{X}_p\right) - \Theta \tag{14}$$

$$\Delta\Lambda = \arctan\left(\frac{\tilde{Y}_p - H}{\sqrt{\tilde{Z}_p^2 + \tilde{X}_p^2} - D}\right) - \Lambda \tag{15}$$

where Θ and Λ are current pan and tilt angles of the arm robot, and H and D are the height and length at fifth joint (Figure8).

Step 2-4. Move Robot and Repeat

The robot is rotated by the calculated pan and tilt angles. While rotating, the rotation angles are stored in the shared memory one by one. And iterate from Step 2-1.

4 EXPERIMENTAL RESULTS

4.1 Scanning Path in Active Search Method

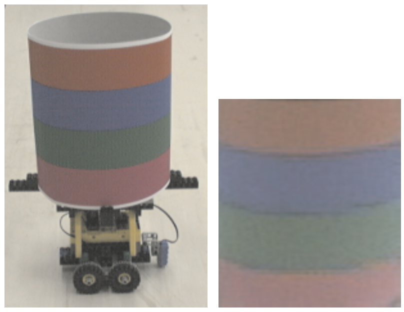
In this experiment, first, we show the comparison of scanning paths. Figure 9 (a) and (b) show examples of input and reference images. We apply the active search method to these images by the ways of raster scan and cross scan. Figure 9 (c) shows the result of detection. Both scanning paths can be found same position. Figure 10 shows positions of matching processes and Table 1 shows the number of matching and computational time for each scanning path. As shown these results, the cross scan can reduce the number of matching and computational time more than the raster scan. But, the cross scan is not always faster than the raster scan. It depends on the start position. If the start position is near the actual object position, the cross scan can search the object position quickly. In the object tracking process, because the object tends to locate at the center of image, the cross scan would be work effectively.

4.2 Object Tracking Results

Next, we show results of the object tracking. In this experiment, we use a line tracing robot as the object for the tracking and its marker as the reference image of the active search method (Figure11). This robot moves along the black line course drawn in a floor (Figure12). Its moving speed is about 70 [mm/s]. Figure13 shows images of right and left cameras and detected object locations in the tracking process. In these figures, the color in the area of detected object is reversed, and center points of rectangles in these areas denote corresponding points for the stereo method.

Figure14 shows the 3D trajectories of the tracking object (the line tracing robot). The blue line shows the object 3D positions estimated by the object detection module, and the red line shows the corrected 3D positions estimated by the robot control module. From this result, almost object 3D coordinates can be obtained along the black line course and this system can track the object.

In Figure14, some 3D positions estimated by the object detection module are different from the actual coordinates. This is because that in the Step 1-3 of the object detection and camera control module, the corresponding coordinates (x_r, y_r) and (x_l, y_l) are not estimated accurately, because the object size is much smaller than the reference image (Figure 15 (a)) or caused motion blur (b). However, in the Step 2-2 of the robot control module, the 3D position is corrected, so the 3D positions are estimated more accurately.



(a) Line tracing robot. (b) Reference image.

Figure 11: Tracking object.

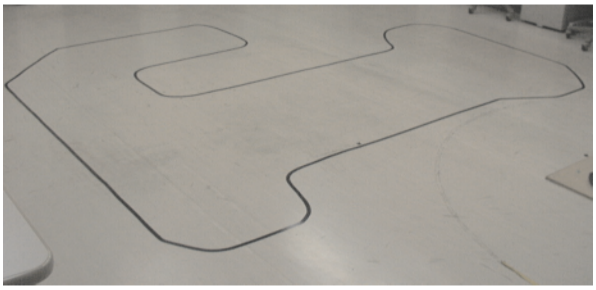
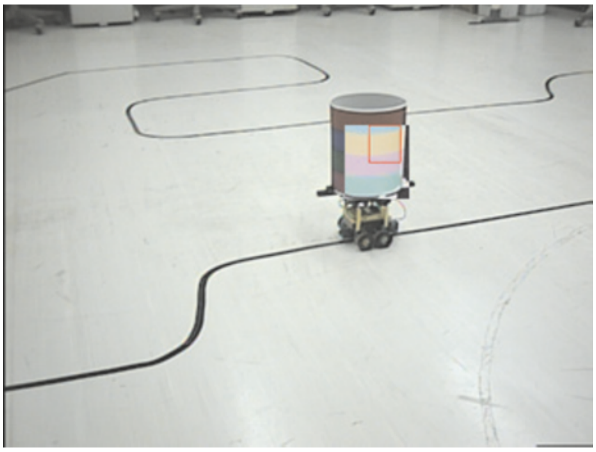
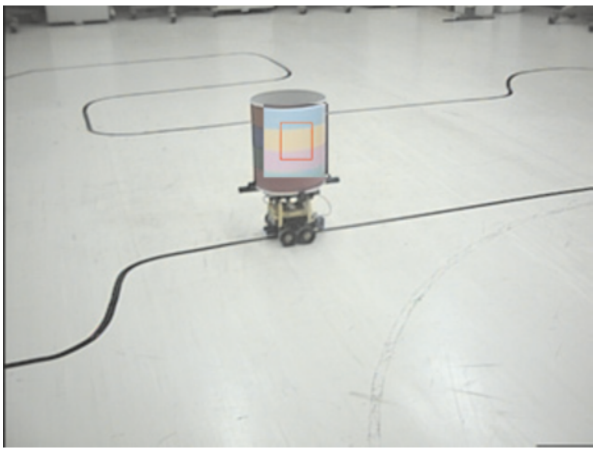


Figure 12: Object moving course.



(a) Left camera.



(b) Right camera.

Figure 13: Camera images and detected locations.

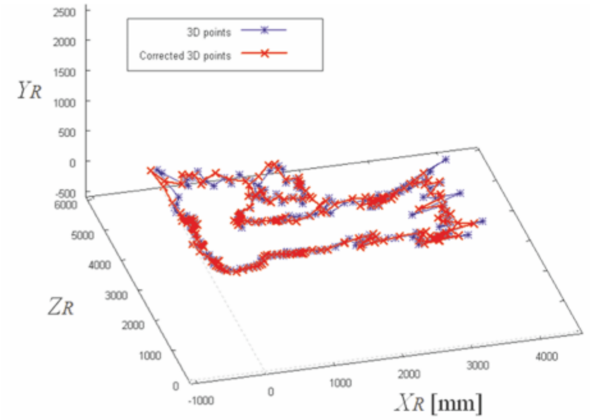


Figure 14: 3D object trajectory.

(a) Small size appearance.

(b)Motion blur.

Figure 15: Case of unmatched corresponding points.

5 CONCLUSIONS

In this paper, we propose the high speed and large range object tracking system using two pan-tilt cameras and the arm robot. In this system, we use the active search method to detect the object in the input images. And we configure two modules about the tracking process and execute them in parallel for high speed processing. We also show the effectiveness of the proposed system in the experiments using a line tracing robot.

As the future works, in order to calculate the 3D object positions accurately, we must improve the method to estimate the corresponding points in the left and right camera images. And we will discuss the more fast tracking method for the object which moves more quickly using the motion prediction and for the object which has various appearances like a human head using the adaptive reference images.

6 REFERENCES

[1] T.Wada, N.Ukita, T.Matsuyama, 1998, Fixed Viewpoint Pan-Tilt-Zoom Camera and Its Applications, Trans. IEICE D-II, J81-D-II/6:1182-1193 (in Japanese).

[2] S.Hiura, K.Murase, T.Matsuyama, 2000, Real Time Object Tracking Using Dynamic Memory, IPSJ Journal 41/11:3082-3091 (in Japanese).

[3] P.I.Corke, 1996, Visual Control of Robots: high-performance visual servoing, Research Studies Press ltd., Somerset, England.

[4] Y. Yagi, 1999, Omnidirectional Sensing and Its Applications, IEICE Trans. on Inf. & Syst., E82-D/3:568-579.

[5] V.V.Vinod, H.Murase, 1997, Focused color intersection with efficient searching for object extraction, Pattern Recognition, 30/10:1787-1797.

[6] H.Murase, V.V.Vinod, 1998, Fast Visual Search Using Focused Color Matching - Active Search -, Trans. IEICE D-II, J81-D-II/9:2035-2042 (in Japanese).

[7] T.Kawanishi, H.Murase, 2001, Quick Object Search Using Color Histograms and Pan-Tilt-Zoom Camera - Dynamic Active Search -, Trans. IEICE D-II, J84-D-II/8:1722-1730 (in Japanese).

[8] O.Faugeras, 1993, Three-Dimensional Computer Vision - A Geometric Viewpoint -, MIT Press, Cambridge.

[9] B.Nichols, D.Buttlar, J.P.Farrell, 1998, Pthreads Programming, Trans. M.Sakaki, O'Reilly Japan, Tokyo (in Japanese).

[10] R.Jain, R.Kasturi, B.G.Schunck, 1995, Machine Vision, McGraw-Hill, Inc., New York.

Influence of Reaching Actions on Driving Performance

Takafumi Asao[1], Kentaro Kotani[1], Ken Horii[1]

[1] Dept. of Mechanical Engineering, Faculty of Engineering Science, Kansai University, Osaka, Japan

Abstract

Drivers have various additional actions while driving, which do not belong to operations of a vehicle, for example, eating, drinking, pressing some buttons, and exchanging compact discs in a car audio system. These actions cause traffic accidents, because they used to distract driver's vision and decision making. Moreover posture of the driver during the actions influences driving operation seriously. Therefore, we focused on reaching actions while driving. In this paper, we investigated driving performance with reaching tasks using a simplified driving simulator. Furthermore, we examined influences of the additional actions on the driving performance with parameters of a driver-vehicle model.

Keywords:

Driving Posture; Steering Maneuverability; Driver Model

1 INTRODUCTION

A wide range of driver support system has been proposed and a number of operations for these devices are increased [1]-[3]. Beside primary driving tasks, we have various actions, for example, we push some buttons and look for positions of them. These actions distract human driving.

A driving requires to repeat cognition, decision making, and operation. If workloads of the cognitions, the decision makings, and the operations for a driver become greater, they distract the driver from driving. If the workloads are higher than managemental limit of the driver, traffic accidents may occur because of human errors.

The workloads that may influence on the driver are classified into three as follows [4]; 1) visual workloads, 2) mental workloads, and 3) execution workloads. The visual workloads are caused by visual searching. The mental workloads are psychological workloads caused by thinking. The execution workloads are caused by execution of any operation in a car. When the drivers want to push some buttons, they look at the buttons and consider how to operate the equipment, which contains the visual and mental workloads. Then they reach their hand to the button to push. It is important to clarify influence of the execution workloads on the driving behaviors, because the execution workloads may results in adding the visual and mental workloads.

Although trains, airplanes, and ships are controlled by expert operators, the cars are operated by ordinary people. The ordinary drivers may have unexpectedly dangerous action while driving, for example, picking a compact disc (CD) from a dashboard to exchange another CD in a car audio system. In this case, first, the drivers confirm a location of the CD in the dashboard with the visual workloads. Second, they have to select one CD with the mental workloads, which they want to listen. Finally, they reach their hand to the dashboard changing their driving posture with the execution workloads. Therefore it is necessary to educate the drivers the risk of doing such action, and to develop support systems of decreasing the execution workloads as well.

Several researches have dealt with comfortable driving posture and maneuverability influenced by the sitting postures [5][6]. However driving performance according to the posture was not investigated. Therefore, in preceding study, influences of the reaching actions with the visual and execution workloads on driving performances were investigated [7][8]. As the results, it was more dangerous for conditions with only the execution workloads than those with the visual and execution workloads. In this paper, we focused only on execution workloads to clarify the influence of the posture of drivers on the driving performances. The performances were examined by using a driving simulator (DS) under emergency driving. Moreover, we analyzed the influence of the reaching actions by using a driver model. To describe the driving behavior into the driver model, quantitatively evaluation of the performance as values of parameters of the model can be conducted.

2 DRIVING SIMULATOR

In this study, a driving simulator (DS) was employed for experiments. The DS was consisted from a computer, a display, and a steering wheel as shown in Figure 1. The steering wheel was GT Force RX (logicool). As the steering

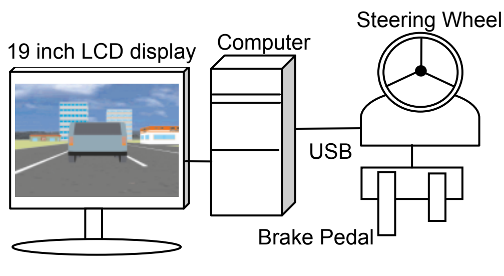


Figure 1: Overview of driving simulator

wheel was for race game, maximum angle of the steering wheel was small. Also, acceleration and brake pedals were set up. A 19 inch LCD display was placed at 900 mm from the drivers. Graphics of a driving scene due to driver's operations was programmed by using C/C++ language libraries of World Tool Kit (Sense8).

2.1 Vehicle model

A vehicle model calculated in the DS was equivalent to two wheels model as shown in Figure 2. The vehicle model had simple two degrees of freedom, which were a lateral movement and a yaw rotation, as follows [9];

$$M\ddot{y} = 2F_f + 2F_r \tag{1}$$

$$I\ddot{\psi} = L_f F_f - L_r F_r \tag{2}$$

where the variable y was a lateral position and ψ was a yaw angle of the vehicle, which were in a ground coordinate system. The constant M and I denoted vehicle mass and yaw moment of inertia of sprung mass, respectively. The constant L_f and L_r denote longitudinal distances from sprung mass center of gravity to front and rear axles. The constant F_f and F_r were side forces of front and rear axles combining side forces of left and right tires as follows;

$$F_f = -2C_f\left(\frac{\dot{y} - V\psi + L_f\dot{\psi}}{V} - \delta\right) \tag{3}$$

$$F_r = -2C_r\left(\frac{\dot{y} - V\psi - L_r\dot{\psi}}{V}\right) \tag{4}$$

where the variable δ denoted steering angle as driver input. The constant C_f and C_r were cornering powers of the front and rear tires, which meant stiffness between tires and ground. The V was vehicle velocity as constant value. In experiments, specifications of the vehicle model were determined as shown in Table 1, which were adjusted as same as a small sedan car. In Table 1, steering gear ratio meant a ratio of steering angle divided by steering wheel angle. The ratio was smaller than usual value, because the maximum steering wheel angle was too small.

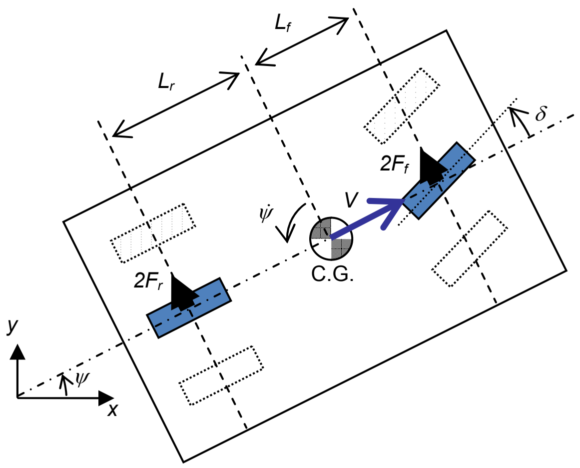


Figure 2: Two wheel vehicle model

Table 1: Specification of vehicle model

Parameters	Values
K_f [N/rad]	80000
K_r [N/rad]	112000
L_f [m]	1.0
L_r [m]	1.3
M [kg]	1000
I [Nm]	1500
Steering gear ratio [-]	3.5

3 DRIVER MODEL

A driver's steering operation model is assumed to reduce the deviation of the course from a front view point, in the following expression shown in Figure 3. A maneuver model consists of a gain K. The transfer function of a whole of the model is shown by an equation as follows;

$$\frac{\delta}{\varepsilon} = K \tag{5}$$

where

$$\varepsilon = y_r - y - VT\psi \tag{6}$$

where the variables y and ψ denote car lateral position and yaw angle, respectively. The variable y_r is lateral position of the target course. The constant value V means velocity of the vehicle. The T denotes preview time. The drivers view at the preview point from the cockpit for controlling the vehicle. A preview point is defined as the point located distance, expressed as $V*T$, ahead from the center of gravity. If the driver's maneuver model is simple as equation (5), it is easier to estimate the parameters K and T from experimental data and to consider the results of the estimation than a complex model.

As the whole, the driver-vehicle closed loop model is as shown in Figure 4. If the drivers perform reaching tasks, the variables K and T change compared with that with no reaching task.

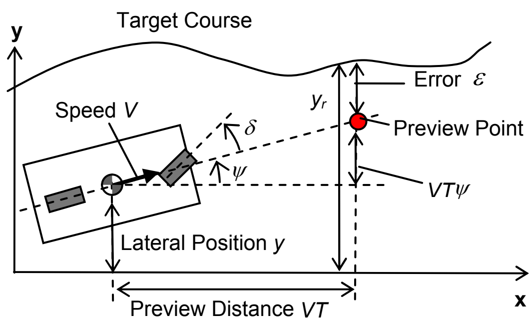


Figure 3: Block diagram of driver-vehicle model

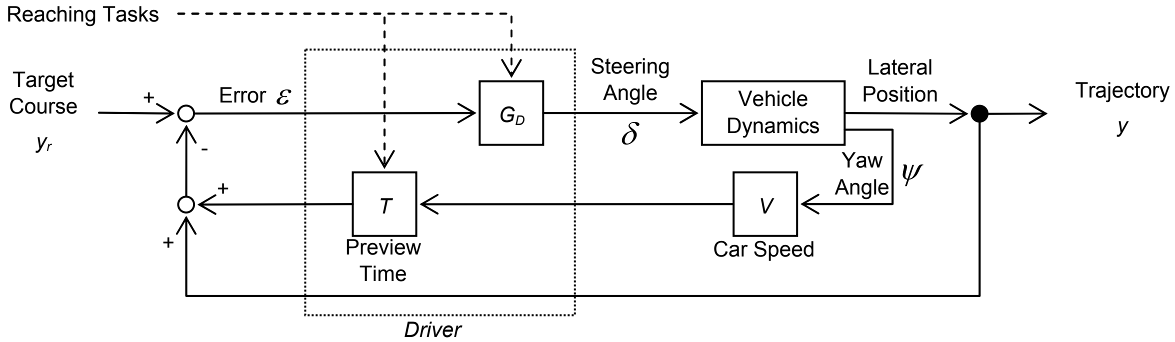


Figure 4: Block diagram of driver-vehicle model

4 EXPERIMENTS OF EMERGENCY AVOIDANCE

4.1 Reaching tasks

There were six targets of the reaching tasks, as shown in Figure 5. The targets of P_{S1}, P_{S2} and P_{S3}, denoted as Side Targets, were on a left side of a driver's body. The positions of the Front Targets, shown as P_{F1}, P_{F2} and P_{F3} in the figure, were 300 mm from the front of the Side Targets. The height of all targets was set to 820 mm from the floor. In the experiments, drivers grasped the top of the tripod stand as shown in Figure 6. Then they operated steering wheel with maintaining their postures as reaching tasks. As control task conditions, two of no reaching tasks were tested, which were single-handed and two-handed driving.

4.2 Driving conditions

The driving conditions were a car-following and emergency avoidance with the steering. Figure 7 shows driving conditions. If the experimental running was started, subject's car and a forward car run at 40 km/h with 15 m of the head way distance. The driver can operate only the steering wheel. If subject's car run through a line A, 100 m from the start position, the forward car suddenly decelerated at a random distance between 0 m and 200 m from the line A. After that, the forward car stopped at 10 m from the line B where the forward car started to decelerate. Then the drivers must avoid the forward car by the steering operation. Finally, the drivers run straight for 150 m from the line B. The sequence above was set to one trial.

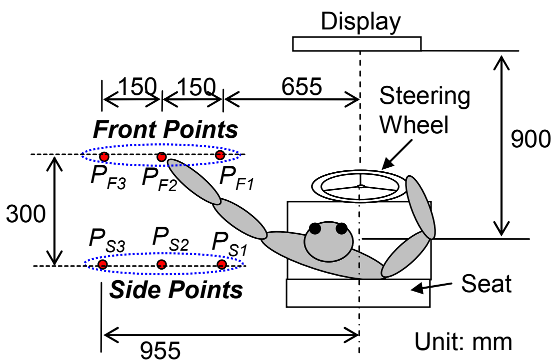


Figure 5: Target positions of reaching tasks

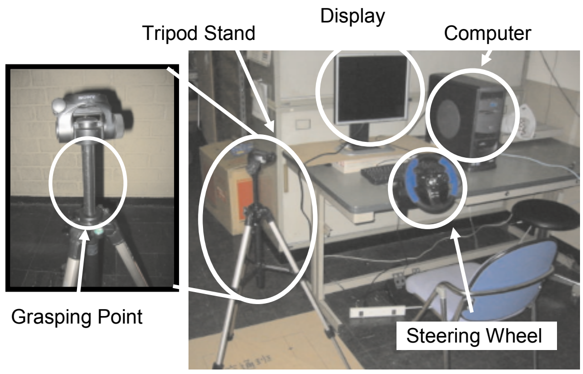


Figure 6: Experimental setup

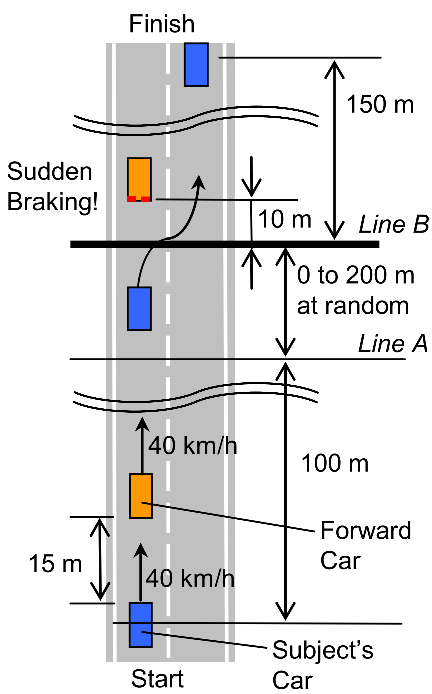


Figure 7: Driving conditions

4.3 Experimental Procedure

1) Drivers reached their left hand to the target, and then maintained their posture.
2) The drivers started to follow a car along a long straight road. The head way distance was 15 m, and speeds of both cars were set to 40 km/h. The driver followed the forward car only operating steering wheel without requiring any foot pedal operations.
3) The forward vehicle suddenly decelerated at the rate of 0.6*9.8 m/s^2 after a randomly given duration. Then the driver deflected the forward car to the right by using the steering. At this time, the driver changed the posture to normal driving posture.
4) The drivers run 150 m along the straight road.
5) Return to 1). The reaching target was set to other one.

One set of the experiments included six targets' conditions and two control conditions. There were 5 sets of experiments for each subject. Six university students participated in the experiments.

4.4 Results

Figure 8 shows the average steering angles for each subject by the different target. The steering angles in case of the control conditions are shown in the figure as well. According to the figure, amplitude of the steering angles became large, when the targets position became far from the drivers. Moreover, when the drivers changed lanes with reaching tasks, amount of overshoot of steering angles was found to be larger than that without the tasks. This implied that magnitude of the steering operations became large due to the generation of the reaching tasks.

Figure 9 and 10 show Lissajous of yaw rate and steering wheel for single-handed driving with no reaching tasks and reaching target of Front Points P_{F3}, respectively. Each line in these figures shows average of five trials for each subject. These Lissajours figures looking like an ellipse indicated a turning ability of a vehicle. If the ability was not satisfactory, the shapes of Lissajous became thick and long. Major axes of Lissajous in Figure 10 were longer than those in Figure 9. In other Lissajous figures for other tasks, major axes of Lissajous became long according to distance of targets position. This implied that steering angle became large because of reaching tasks.

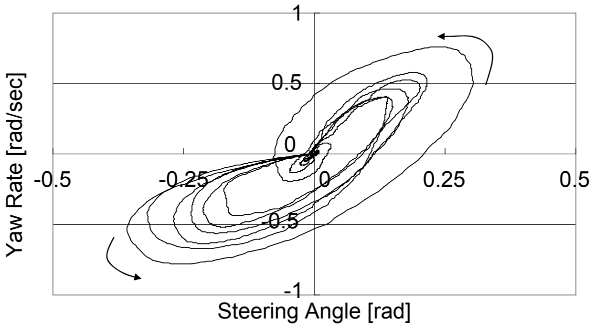


Figure 9: Lissajous of yaw rate and steering angle for single-handed driving with no reaching tasks.

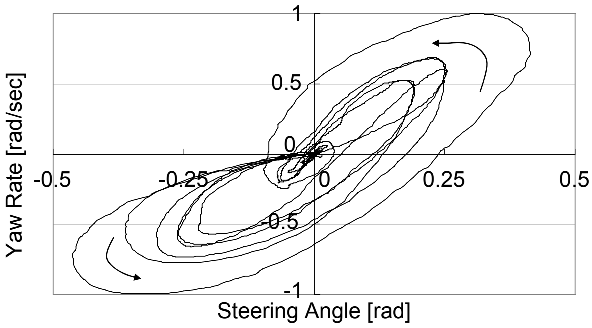


Figure 10: Lissajous of yaw rate and steering angle for reaching target of Front Points P_{F3}.

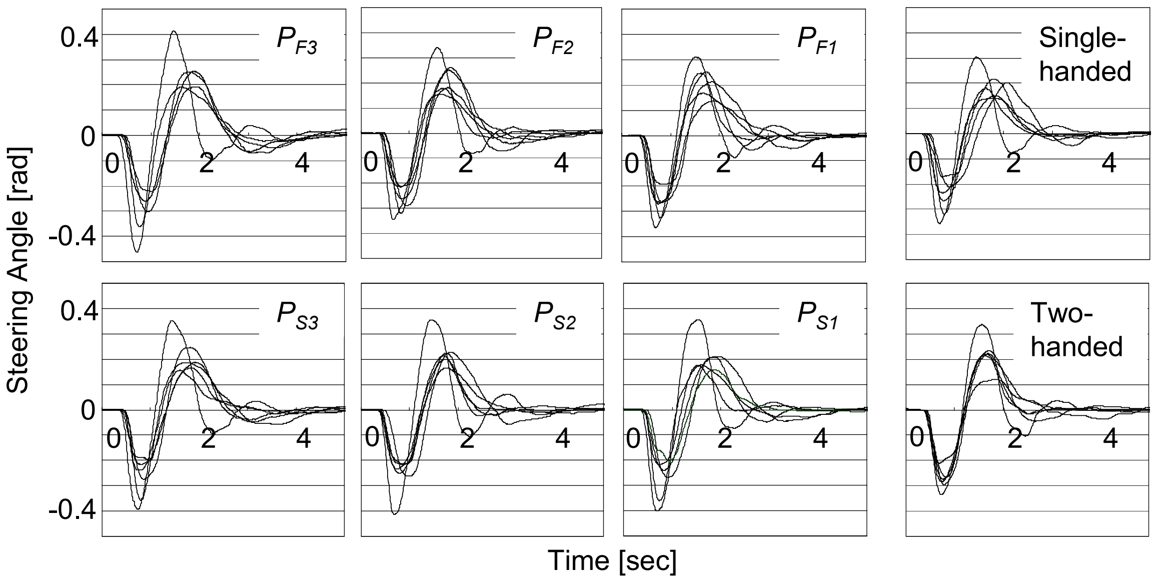


Figure 8: Steering Angles in Emergency Avoidance

5 PARAMETERS OF DRIVER MODEL

5.1 Estimation of Driver Model Parameters

As a help of understanding the driving behavior, the parameters *K* and *T* of the driver model were estimated as following optimization problem;

$$\text{minimize} \quad J = \int_0^{t_e} \{\delta - K[y_r - (y + VT\psi)]\}^2 dt \tag{7}$$

where t_e denotes end time of experiments. The parameters were estimated by the least square method.

5.2 Results of the Estimation

Figure 11 shows the average gain *K* according to the reaching targets for each subject and average of all subjects. The gain *K* became large according to the distance from the driver to the targets, in order of P_{S1}, P_{S2}, P_{S3} or P_{F1}, P_{F2}, P_{F3}. They were larger than that for no targets conditions. Moreover, the values *K* of Front Targets were larger than those of Side Targets. If an arm posture was stretching, stiffness at an end effector, that is a hand in this case, changed compared with that when the posture was not stretched [10][11]. If the stiffness was small, drivers were able to move their hand rapidly. Consequently, the *K* became large according to the distance to the target.

The gain *K* for subject D was larger than that of the other. Thus, physiques of the subjects were investigated to clarify the phenomena as a post-hoc analysis. Table 2 shows the upper limb length and body mass index (BMI) for each subject. The upper limb length and BMI for the subject D were the largest among all subjects. Therefore the gain *K* for the subject D was largest, since the stiffness for the subject D was the largest due to his arm length.

Figure 12 shows the relationship between parameters *K* and *T*. The gains *K* were large and the preview times *T* were small in the order of the targets P_{S1}, P_{S2}, P_{S3}, or P_{F1}, P_{F2}, P_{F3}. This tendency of the gain was caused by extending right arm according to the targets' position. Additionally, preview times became small, because lateral velocity of the vehicle became high due to large gain. In other words, drivers have to control their preview times due to their gains according to the posture. Furthermore, the gains for Front Targets were larger than those for Side Targets. The posture of the right arm for Front Targets was more extended than that for Side Targets. All in all, it was concluded that reaching a driver's hand forward tend to be more dangerous than that laterally.

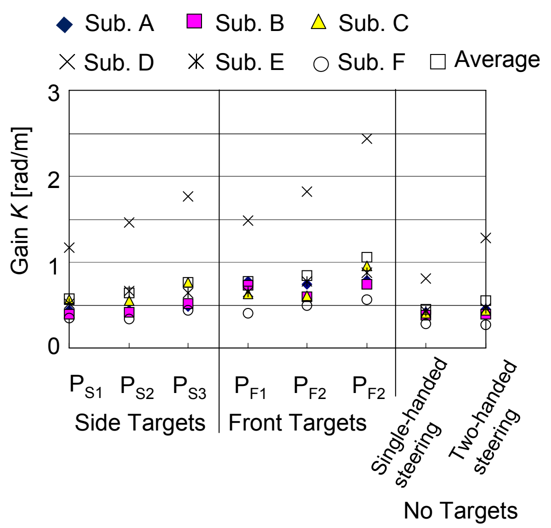


Figure 11: Parameters *K* accorfing to the reaching targets

6 CONCLUSIONS

In this paper, it was investigated that the steering performance with the reaching tasks using the driving simulator in emergency avoidance. Moreover the parameters of the driver model were estimated as the help of detail analysis. As the results, it was clarified as follows;

1) The steering angle became large according to the changes of the driving posture from the ordinary driving posture.
2) The gain of the driver model became large according to the degree of the stretching arm. Moreover, the gain was larger with the Front Targets than that with the Side Targets.
3) The preview time became small according to the distance to the reaching targets.

Driving with reaching actions is not safe. Especially, to reach our hand to the forward becomes steering angles large. It is necessary to educate drivers the danger reaching actions while driving.

Table 2: Physique of subjects

	Upper limb length [cm]	BMI
Sub. A	54	22
Sub. B	53	18
Sub. C	56	24
Sub. D	59	28
Sub. E	58	18
Sub. F	58	22

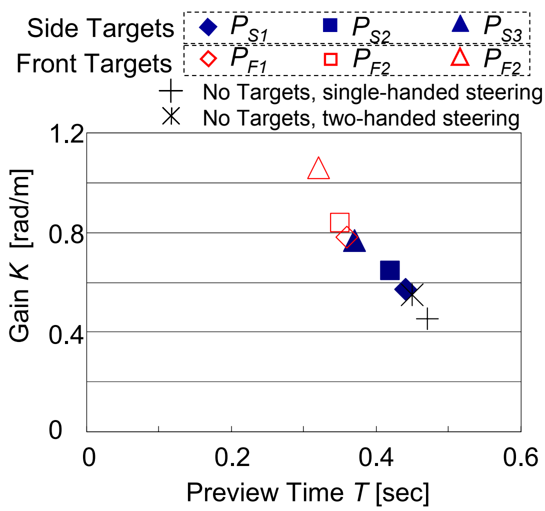


Figure 12: Relationship between the parameters *K* and *T* of the driver model with reaching tasks

As the future works, it is necessary to investigate the driving performance with reaching tasks for various target positions. Moreover, driver model has to be reconstructed with higher order than that in this paper for detailed considerations of control strategies under the reaching tasks.

REFERENCES

[1] Horrey, W. J., Alexander, A. L., Wickens, C. D., 2003, Doise Workload Modulate the Effects of In-Vehicle Display Location on Concurrent Driving and Side Task Performance?, Procs. of DSC.

[2] Reed, M. P., Green, P. A., 1999, Comparision of driving performance on road and in a low-cost simulator using a concurrent telephone dialling task, Ergonomics, 42/8:1015-1037.

[3] Makiguchi, M., Tokunaga, H., Kanamori, H., 2003, A human factors study of switches installed on automotive steering wheel, JSAE Review, 24:341-346.

[4] Gelau, C., Jahm, G., Krems, J. F., Uno, H., Kircher A., Östlund J. and Nilsson L., 2003, State-of-the-art of the SNRA/JARI/BAST joint research on driver workload measurement within the framework of IHRA-ITS, Procs. of ESV, Paper No.242-O.

[5] Park, S. J., Kim, C., Kim, C. J., Lee, J. W., 2000, Comfortable driving postures for Koreans, Industrial Ergonomics, 26:489-497.

[6] Hada, M., Amano, Y., Fukui, K., Doi, S., 2001, An Evaluation of Vehicle Steering Arrangement with Dexterity Measures of Virtual Human, SAE Paper, 2001-01-2109.

[7] Asao, T., Wada, T., and Doi, S., 2005, The effect of physical workloads on driving performance, Procs. of ESV, Paper Number 05-0435.

[8] Asao, T., Wada, T., Doi, S., and Tsukamoto, K., 2006, Influence of Physical Workloads on Driving Performance, Review of Automotive Engineering, 27/2:361-364.

[9] Rajamani, R., 2006, Vehicle Dynamics and Control, Springer, United States.

[10] Mussa-Ivaldi, F. A., Hogan, N., and Bizzi, E., 1985, Neural, Mechanical, and Geometric Factors Subserving Arm Posture in Humans, J. Neurosci., 5/10 :2732-2743.

[11] Tsuji, T., Morasso, P. G., Goto, K., Ito, K., 2006, Human hand impedance characteristics during maintained posture, Biol. Cybern, 72:475-485.

Measurement of a car driver's pulse interval

while driving with one hand

Hideki Tomimori[1], Yoshio Ishida[1], Ken Sasaki[1], Yasuhiko Nakano[2], Satoshi Sano[2]

[1] Department of Human and Engineered Environmental Studies, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa , Japan

[2] Fujitsu Laboratories Ltd.

Abstract

R-R interval measurement system for car drivers that allows measurement while the driver is driving with one hand has been developed. This system is intended for analysis of R-R interval fluctuation which is related to drowsiness and stress. R-R interval is the interval between R-waves in electrocardiogram (ECG). Our goal is to detect changes in R-R interval fluctuation and alert the driver to prevent drowsy driving. ECG was measured with a contact electrode on the steering wheel and a capacitive coupling electrode on the driver's seat. ECG measurement with capacitive coupling is very sensitive to body motion. Baseline fluctuation can be as large as hundred times of the R-wave. In order to utilize differential amplification to reduce the fluctuation, an additional seat electrode was placed on top of the seat electrode with a paper insulator in between the two electrodes. The outputs of the upper seat electrode and the steering wheel electrode showed similar baseline fluctuation with respect to the lower seat electrode, while R-waves appeared only in the signal from the steering wheel electrode. Baseline fluctuation was reduced by taking the difference between the two signals. The proposed measurement method was evaluated in a real car. The system was capable of measuring R-waves of a passenger who held an electrode in the hand to simulate one hand driving while driving on a highway.

Keywords:

Drowsy driving; Electrocardiogram; R-R interval; Capacitive coupling; Signal processing;

1 INTRODUCTION

Drowsy driving is one of the main causes of traffic accidents, and automatic drowsiness detection systems may help prevent these accidents. The key technology in these systems is detection of drowsiness. Methods for drowsiness detection can be classified into three categories; a)Analysis of eye movements and facial expressions by image analysis, b)Analysis of steering wheel, gas pedal, and brake pedal, c)Analysis of driver's physiological signals, such as heart rate and electroencephalography.

We have focused on analysis of physiological signals, specifically on R-R interval fluctuation which is related to driver's drowsiness level [1]. This method requires reliable non-invasive R-R interval measurement of a car driver.

R-R interval can be measured by taking electrocardiogram (ECG) or measuring pulse wave. ECG is usually measured by attaching electrodes on the body. Pulse wave is usually measured optically. Since both methods require attaching either electrodes or optical sensors on the body, they cannot be directly applied to car drivers.

Using two electrodes on the steering wheel is a simple way to measure ECG of a car driver [2]. However, we cannot expect the driver to hold the wheel with both hands all the time, especially when the driver is drowsy.

There are non-invasive ECG measurement methods that use capacitive coupling [3][4]. One major drawback of this method is that it is very sensitive to body motion.

This paper proposes an ECG measurement method that uses one contact electrode on the steering wheel and one capacitive coupling electrode on the driver's seat that enables R-R interval measurement while driving with one hand.

2 MEASUREMENT METHOD

2.1 Measurement of driver's R-R interval

R-R interval is measured as the interval between R-waves in the electrocardiogram (ECG). R-wave has an impulse-like waveform with amplitude of approximately 1 mV. It is the most conspicuous wave form in the electrocardiogram that can be readily recognized as shown in Fig. 1.

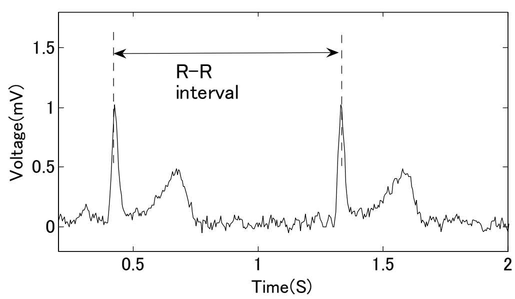


Fig.1 Typical ECG and Interval between R-waves

Although ECG measurement in medical diagnostics use 5 to 12 electrodes on the body surface to capture spatio-temporal characteristics of the electrical activities of the heart, we only need two electrodes for detecting R-wave. We can obtain R-waves by attaching two electrodes across the steering wheel of a car and measure when the driver is holding the wheel with both hands. The problem with this method is that we cannot expect drivers to be using both hands all the time. When a driver becomes tired or sleepy, it is likey that the driver is driving with one hand. When only one electrode on the steering wheel is in contact with a hand, we need to measure electrical potential difference between the lower part of the body and the hand on the steering wheel. For this purpose, we will use a capacitively coupled electrode on the driver's seat.

2.2 Measurement ECG using capacitive electrode

An electrode pressed on driver's clothes from outside will form a capacitive coupling between the human body and the electrode. The capacitance of this capacitive coupling is in the order of several hundreds of pico Farad, depending on the area of the electrode, thickness of the clothes, and relative dielectric constant of the fabric.

One of the drawbacks of ECG measurement using capacitive coupling is that the signal is easily degraded by body motion as shown in Fig.3. This signal was taken when the driver repeatedly moved his foot up and down on the floor, which is similar to motions such as pressing a gas pedal or a brake pedal. This baseline fluctuation is mainly due to the static electricity generated when the driver makes or breaks contact with the pedals. The baseline fluctuation in the ECG signal can be as large as several hundred milli-volt, while typical amplitude of R-wave is approximately 1 mV.

Quantity of this static electricity was estimated from the voltage and the current at the input resistance of the amplifier. The electric charge was in the order of several tens of nano Coulomb, which was about the same with that of the static electricity generated when a person stood up from a chair.

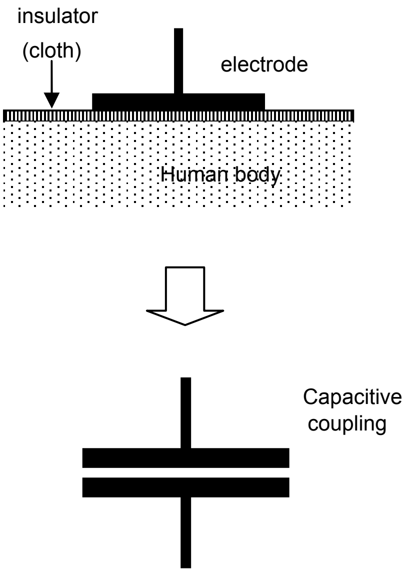


Fig.2 Capacitive coupling for ECG measurement

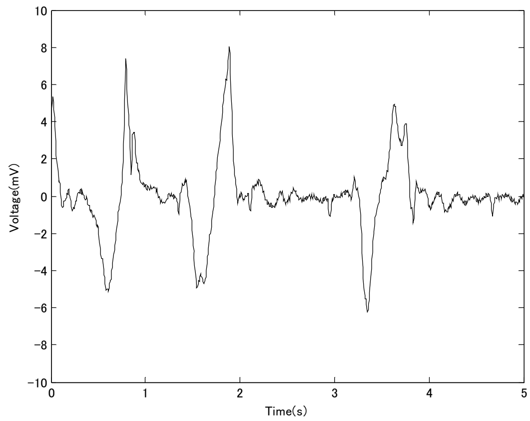


Fig.3 Degraded ECG signal by body motion

3 EXPERIMENT

3.1 Measurement in laboratory setup

R-R interval measurement method using contact electrode held in one hand and a capacitive coupling electrode on the lower part of the body was first evaluated in an indoor laboratory setup as shown in Fig. 4. Steering wheel, driver's seat and pedals were adapted from driving game controller and set up on a plywood board (1800mm by 900mm, thickness 30mm). This driver's cockpit was lifted 500 mm above the laboratory floor by using a block of styrene foam (1800mm by 900 mm, height 500mm) in order to reduce electromagnetic interference from commercial power line. We also covered the setup with stainless steel screen in order to simulate static charge balance between the driver and the chassis of a real car.

We embedded two contact electrodes along the right and the left outer circumference of the steering wheel. The electrodes were made of stainless steel, and the width was both 10 mm and their length was both 150 mm along the circumference. A piece of carbon fiber cloth was placed on the driver's seat for the capacitive coupling electrode. ECG signal between the right arm and the lower part of the human body is called "Lead II" in ECG analysis. Area of the carbon fiber cloth was 800cm^2, and electric resistance was several ohms across the piece. This resistance did not affect the measurement system because the input impedance of the pre-amplifier was 10 MΩ. Pre-amplifier output was put through a band pass filter (pass band: 0.2Hz to 30Hz) and further through an amplifier. The total gain of the amplifier was 30dB. Signals were recorded by a battery driven data recorder (Hioki 8807) to avoid coupling through commercial power line as well as to reduce electromagnetic noise.

We also measured ECG by using ordinary disposable contact electrodes to capture R-waves for reference. We used separate battery driven amplifiers, and used a data recorder (Hioki 8807) whose inputs were all electrically isolated to one another, in order to ensure that measurement for reference did not affect the measurement system to be tested.

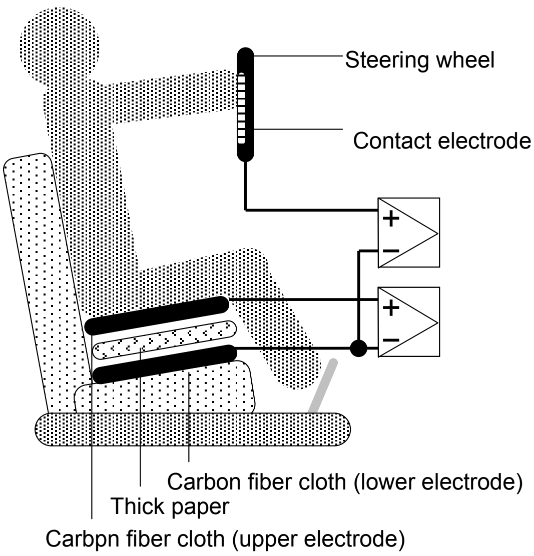


Fig.4 Measurement system with layered seat electrode

3.2 Reduction of Baseline Fluctuation

Differential amplification is frequently used to reduce common mode noise in measurement system. This method assumes that the noise in the two signals is common with respect to a certain reference level so that the noise can be cancelled by taking the difference between the two signals. In our case, there was no obvious reference level that we could utilize for differential amplification method.

We placed a secondary carbon fiber cloth on the carbon fiber cloth and inserted a sheet of thick paper as an insulator between the two. The lower electrode, which is farther from the human body, was used as the common electrode as shown in Fig.4. The two outputs, one from the contact electrode on the steering wheel and the other from the upper electrode on the driver's seat, were measured. The signal from the upper seat electrode did not contain ECG signal and had a baseline fluctuation pattern similar to that obtained from the steering wheel electrode but with smaller fluctuation amplitude. The fluctuation amplitude of the contact electrode was approximately 1.5 times larger than that from the upper seat electrode. Fig. 5 shows the result after adjusting the amplitude of the two signals. The two signals have the same fluctuation pattern while only the signal from the contact electrode on the steering wheel contained R-waves.

Fig. 6 shows the difference between the two signals and the reference ECG measured by ordinary contact electrodes. The signal was further put through a band pass filter to remove fluctuation and high frequency noise. The result is shown in Fig. 7 accompanied by reference ECG. The two signals are almost indistinguishable. The small delay with respect to the reference ECG comes from the computation of finite impulse response (FIR) digital filter. (Note: It is not a computational delay but a delay inherent in FIR filter. FIR filters usually have delay equal to a half of the window length.)

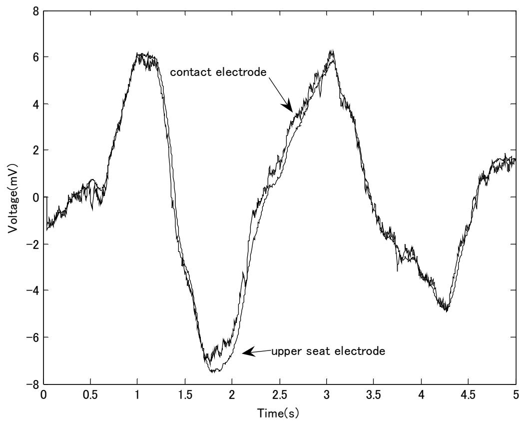


Fig. 5 Comparison of signals from the steering wheel electrode and the upper seat electrode

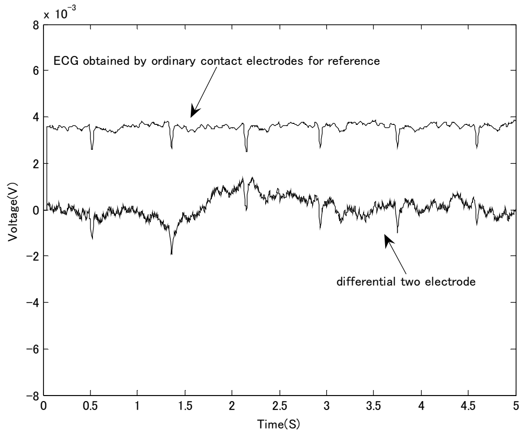


Fig. 6 Differential output compared with reference ECG

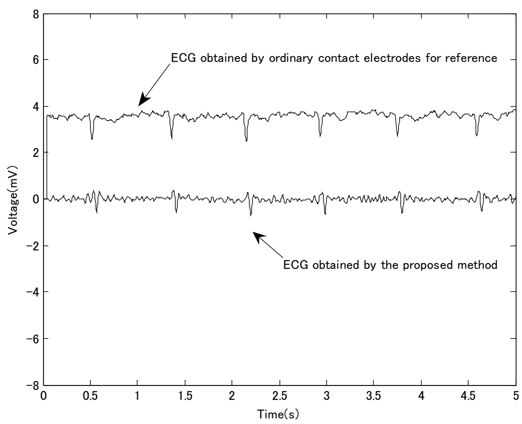


Fig. 7 Final output compared with reference ECG

3.3 Discussion

The reason why the lower seat electrode serves as the common level for the contact electrode on the steering wheel and the upper seat electrode needs further investigation.

When only two electrodes are used, i.e. the steering wheel electrode and the seat electrode, the static charge accumulated in the body do not necessarily cause fluctuation of potential difference between the two electrodes, because the human body and the measurement circuit forms a closed circuit, and the location of charge generation is between the foot and the floor, which is not part of the circuit. This suggests that there are capacitive couplings between the measurement circuitry including the electrodes and cables, and the surrounding environment.

It is our speculation that the lower seat electrode, i.e. the common electrode, reinforces the capacitive coupling with the surroundings and hence serves as the common level electrode for the upper seat electrode and the steering wheel electrode.

4 EVALUATION BY DRIVING A REAL CAR

4.1 Driving in closed area

We have evaluated the proposed measurement system by driving a real car. First, we drove in a closed university campus where no other car was driving. The maximum speed was no more than 20 km/h.

Chrome plated stainless steel film (10mm by 80mm) was attached on the right side of the steering wheel for the contact electrode.

Fig. 8 shows an example of the measured signals. The upper graph is the signal of the steering wheel electrode, the middle graph is the signal of the upper seat electrode, and the bottom graph is the difference between the two signals.

The fluctuations were caused by moving the foot and leg for pedal operation. Noises caused by vibration of engine or car motion were negligible.

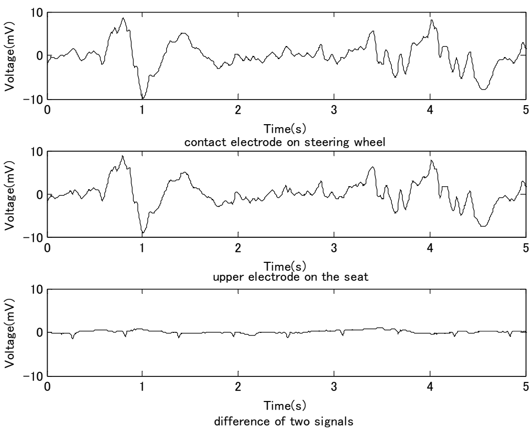


Fig. 8 R-wave measurement while driving in a closed area

4.2 Driving on a highway

Measurement on a highway was conducted differently. Since we could not attach electrodes on the steering wheel for safety reasons, we decided to assign a passenger as the subject. A subject on the passenger seat held a stainless steel bar as the contact electrode on the steering wheel. The rest of the setup was the same.

Fig. 9 shows the result of the measurement. Although the noise became larger because of the increased car vibration, we could successfully cancel the noise by taking the difference between the two electrodes.

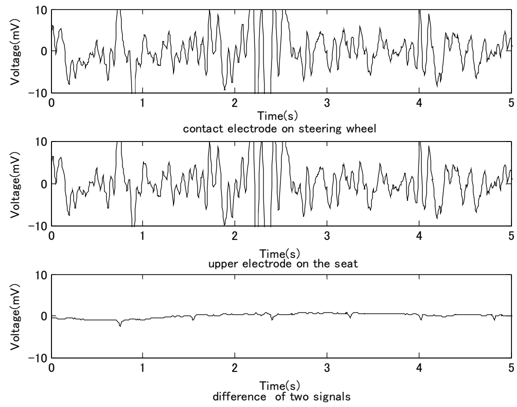


Fig. 9 Measurement of R-waves while driving on a highway

5 SUMMARY

R-R interval measurement system for a car driver that is capable of measuring while driving with one hand has been presented. ECG was measured with a contact electrode on the steering wheel and a capacitive coupling electrode on the driver's seat. Baseline fluctuation caused by body motion, which can be as large as hundred times of the R-wave, inhibited extraction of R-waves to obtain the R-R interval. In order to apply differential amplification to reduce the fluctuation, an additional seat electrode was placed on top of the seat electrode with a paper insulator in between the two electrodes. The outputs of the upper seat electrode and the steering wheel electrode showed similar baseline fluctuation with respect to the lower seat electrode, while R-waves appeared only in the signal from the steering wheel electrode. Baseline fluctuation was reduced by taking the difference between the two signals. The proposed measurement method was evaluated in a real car. The system was capable of measuring R-waves of a passenger who held an electrode to simulate one hand driving while driving on a highway. Further investigation on electrical model of the measurement system is necessary to clarify the mechanism of baseline fluctuation reduction.

6 REFERENCES

[1] Toshio Fukuda , Yuki Wakuda , Yasuhisa Hasegawa , Fumihito Arai ,Mitsuo Kawaguchi , Akiko Noda , "Sleep Quality Estimation based on Chaos Analysis for Heart Rater Variability" , IEEJ Trans.EIS ,Vol.125 No.1 , 2005

[2] Akinori Ueno , Yoichi Furusawa , Hiroshi Hoshino , Yoji Ishiyama, "Detection of Electrocardiogram by Electrodes with Fabrics Using Capacitive Coupling ", IEEJ Trans. EIS , Vol.124, No.9, 2004

[3] Masa Ishijima , "Monitoring of Electrocardiograms in Bed Without Utilizing Body Surface Electrodes" , IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, Vol.40, No.6,1993

[4] Masatoshi Yanagidaira, Mitsuto Yasushi, Koji Sugiura, Masaaki Minamino. Masato Mochizuki, "Developmentof a Driver's Condition Monitor", PIONEER R&D Vol.11 No.2, 2001

Validation of performances on attendant propelled wheelchairs with assisting control based on autonomous propelling model

Tatsuto Suzuki[1], Hironobu Uchiyama[2], Junichi Kurata[2], Yoshihiro Murakami[2], Masako Baba[2]

[1] Dept. of Mechanical Engineering, Maizuru National College of Technology, Kyoto, Japan

[2] Dept. of Mechanical Engineering, Faculty of Engineering Science, Kansai University, Osaka, Japan

Abstract

For the demands of assisting and energy saving on attendant propelling wheelchairs, we proposed new controller to generate assisting force when attendant's force exceeds attendants' autonomous propelling performance in daily life. On the decision for assisting, the controller has the model of the autonomous propelling activity. For the design of the controller, an attendant-wheelchair model is proposed in this paper and the difference of the performance of the proportional controller and our proposed model based controller were analyzed by simulation. The results shows that the model based controller is superior in terms of cooperation with attendant and energy saving.

Keywords:

Attendant Propelled Wheelchair; Assisting Control; Controller Design; Attendant-Wheelchair Model

1 INTRODUCTION

On high aging society, aged people with difficulty of walking, increase now. To take the disabled peoples for a walk by attendant's propelled wheelchairs contribute increase of disabled people's QOL greatly. The attendant has to handle a wheelchair with occupants in weight nearly 100kg and to protect occupants from any accidents and injuries. So hard physical burden for driving wheelchair, however, shoulders attendants because driving wheelchairs with disabled is tough work, like as pushing manual carts. Resnick[1] studied maximal and sub maximal cart pushing with different height of handles, Al-Eisawi[2] investigate the factors for minimum pushing force of cart such as weight, wheel width, diameter and orientation. Jansen[3] investigated the pushing force of cart with vertical push bar and large diameter caster. For reduction of the load propelling carts, the most effective solution is introduced by Cremers[4]. He proposed the detachable motorized driving unit for assisting attendant's propulsion. Now, there are many produced motorized assisting wheelchairs in a market. These wheelchairs have controllers to generate auxiliary driving force in proportion to attendants force. With increasing recently the combined desires for securing safety on propulsion and for saving electric energy under reducing the load of propulsion effectively, simple proportional controller hold problems to cover these demands.

To satisfy the demands, we propose new controller to generate assisting force when attendant's force exceeds attendants' autonomous propelling performance in daily life. On the decision for assisting, the controller has the model of the autonomous propelling activity. The wheelchair has hybrid driving forces consisting of only attendant's force in light load and additional motorized force in heavy load. In this paper, we clear the difference between the proportional controller and our proposed model based controller by simulation with attendants–wheelchair model. To estimate the performances of both systems, pushing force, wheelchair velocity and assisting force are simulated in two cases; One is the case when wheelchair's road resistance increases like the wheelchair entering upward slope, the other is when the attendant accelerates the wheelchair.

2 MODEL OF ATTENDANT-WHEELCHAIR MODEL

2.1 Overview of the system

Figure 1 shows overview of attendant–wheelchair system with assisting. In this paper, movement of an attendant and a wheelchair is assumed along with straight line, so the model has one dimension. The key physical values of the system are pushing force f(t), assisting force a(t) and wheelchair velocity w(t). Figure 2 also shows the block diagram of the attendant-wheelchair system with assisting system. The block diagram has mixed expression of time domain and s-domain. We use t-domain expression of static characteristic in the system, and s-domain for dynamics from stable points. Attendants generate propelling force f(t) regarding to wheelchair velocity w(t). We explain detailed model of attendants $G_H(s)$ later. The wheelchair in the system has its

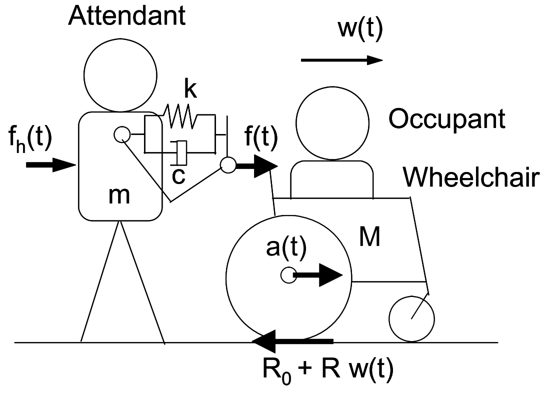


Figure 1: Attendant-wheelchair model

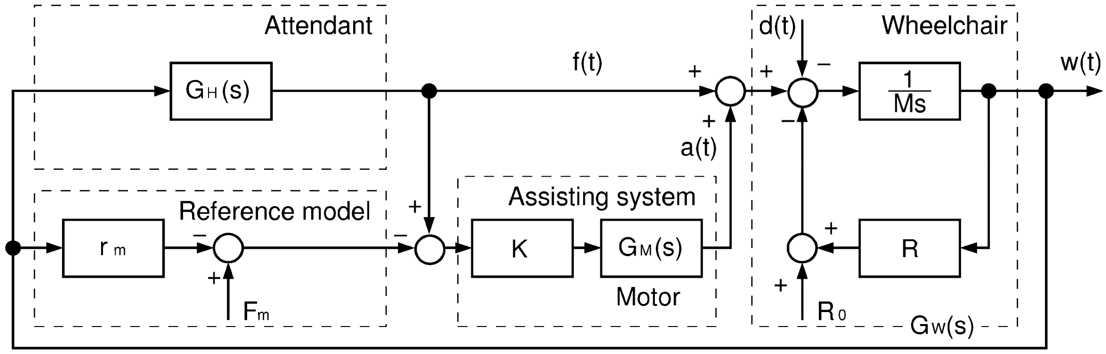


Figure 2: Attendant – wheelchair with assisting system

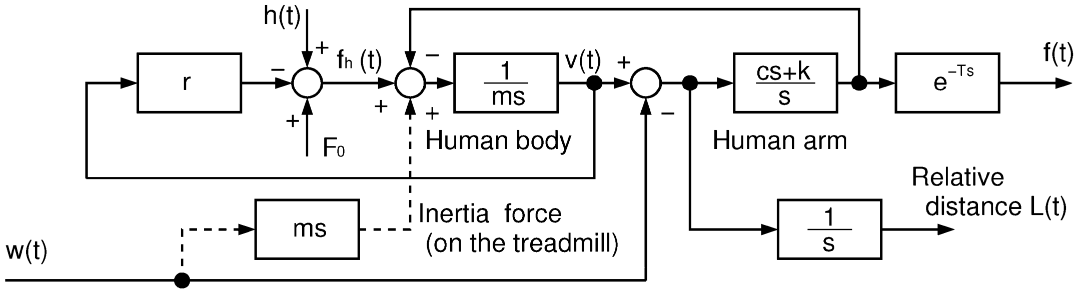


Figure 3: Detailed attendant system $G_H(s)$

mass M and road resistance force $R_F(t)$.

$$R_F(t) = R_0 + Rw(t) \quad (1)$$

Here, R_0 is static rolling resistance and R is coefficient against to w(t). And transfer function $G_W(s)$ is,

$$G_W(s) = \frac{1}{Ms + R} \quad (2)$$

The controller in the model consists of reference model described by F_m and r_m, proportional assisting raito K and motor dynamics $G_M(s)$. The parameters F_m and r_m of the reference model are determined according to attendants autonomous propelling performance show by Suzuki[5], which has maximal force F_0 at standing and decreasing ratio r against walking velocity. The controller can describe two type: the proportional controller (F_m =0, r_m =0) and the model based controller (0< F_m (=F_0 r_m /r) < F_0, 0< r_m <r). The electric motor to generate assisting force has typical dynamics of following.

$$G_M(s) = \frac{1}{0.2s + 1} \quad (3)$$

2.2 Detailed attendant model

Figure3 shows detailed attendants system $G_H(s)$. Attendants has its mass m, and generates forwarding force F_H= F_0-r v(t) by legs. Hence, v(t) is attendant's walking velocity. Attendant's arm which connects to wheelchair grips, works as parallel structure of spring k and dumper c, we assume. Lag element is connected to final place in the system because human has time lag T on responses against change of condition. Relative distance L(t) can be defined by calculating the integral of relative velocity v(t)-w(t). We estimate the parameters in the system on the treadmill as the first step. At the examinations on the treadmill, there is almost same kinematic condition against overground walking shown by Alton[6], but we must handle inertia force so that the condition under the model corresponds to the real one. From the model, we can describe f(t) at steady state:

$$f(t) = F_0 - rw(t) \quad (4)$$

On the treadmill, transfer function $G_H(s)$ is,

$$G_H(s) = \frac{F(s)}{W(s)} = \frac{-r(cs+k)}{ms^2 + (r+c)s + k} e^{-Ts} \quad (5)$$

For validating real condition with simulation, we use $G_H(s)$ excluding the inertia force.

$$G_H(s) = \frac{F(s)}{W(s)} = \frac{-(ms+r)(cs+k)}{ms^2 + (r+c)s + k} e^{-Ts} \quad (6)$$

2.3 Behaviour of operating point under assisting

From the system shown by figure 1, assisting force a(t) at steady state is,

$$a(t) = K(F_0 - F_m) + K(r - r_m)w(t) \quad (7)$$

Under the steady state, Eq.(1), (4) and (7) satisfy following relationship.

$$\begin{aligned} &f(t) + a(t) \\ &= \{(1+K)F_0 - F_m\} - \{(1+K)r - Kr_m\}w(t) \quad (8) \\ &= R_F(t) \end{aligned}$$

So we finally obtain operating point OP(fp, wp), which the attendants propel continuously under steady condition.

$$OP(fp,wp) \tag{9}$$

$$=\left(\frac{F_0R + rR_0 + K(rF_m - r_mF_0)}{R + r + K(r - r_m)}, \frac{F_0 - R_0 + K(F_0 - F_m)}{R + r + K(r - r_m)}\right)$$

2.4 Sensitivity to changes in h(t) and d(t)

In the system, increase of road resistance force attendants to increase propelling force against it, and to decrease walking velocity. In addition, attendants want to increase wheelchair velocity sometimes to increase its propelling force. The assisting system detects the change of the attendants force and the wheelchair velocity for generating assisting force for propulsion. So we investigate sensitivity of f(t), w(t) and a(t) in additional attendants force h(t) and road resistance d(t). We assume that h(t) and d(t) do not take place simultaneously. From calculating output F(s), W(s) and A(s) against H(s) and D(s) respectively, then use the final value theorem, we obtain following functions of sensitivity.

$$\begin{pmatrix} \Delta F_H \\ \Delta F_D \\ \Delta W_H \\ \Delta W_D \\ \Delta A_H \\ \Delta A_D \end{pmatrix} = \frac{1}{R + r + K(r - r_m)} \begin{pmatrix} R - Kr_m \\ r \\ K(R + r_m) \\ K(r - r_m) \\ 1 + K \\ -1 \end{pmatrix} \tag{10}$$

3 ESTIMATION OF MODEL PARAMETERS

To obtain the parameters in the model, we carried on experiments on a treadmill, which can simulate the wheelchair dynamics and the assisting system on the figure 2. Subject in this report was a 22years old man without any disorders. By using the same method in our previous study, which can solve autonomous propelling performance, we obtained the result of figure 4 and determined F_0=84N, r=74.2Ns/m on the subject. This result means that the subject propels a wheelchair at combination of propelling force and wheelchair velocity against road resistance. For example, the subject propels at about 40N and 0.5m/s when the road resistance R=80Ns/m. When the R became more large, the propelling force increased, and the wheelchair velocity slowed down. And these pairs of propelling force and wheelchair velocity were unconsciously determined by the subject, according subject's characteristics, such as physical, sensory, motor and pyschomotor performances. Also its depend on personality and experience. Exercising heart ratio, which shows subjective exercising level, was under 30%(light work) on all of experiments of figure 4.

Subject's weight was m=57kg. We used T=0.2s according to general studies of human response. Wheelchair's mass M was 100kg, which was total mass of rider, wheelchair's body and motorized assisting unit. We used the static road resistance R_0=40N, the increase ratio R=1.25Ns/m from experiments on upward asphalt slope (angle 1.2deg.) in advance.

For determining the stiffness k and the viscous damping c of the subject's arm, we carried on experiments on the treadmill under F_m=r_m=0, K=0. First, we waited until the subject became steady propelling on the treadmill under R_0=40N, then increased to R_0=65N. We measured the step responses of the propelling force f(t), the relative distance L(t) and the wheelchair velocity w(t). Figure 5 shows the experimental result of f(t), w(t) and L(t). When the road resistance increased, the propelling force f(t) increased and the wheelchair speed w(t) decreased with the distance L became narrow. Using $G_H(s)$ by Eq.(5), we estimate c=342Ns/m and k=1526N/m with parameters which we explained in previous section. On the result in the figure 5, periodic waves caused by two-leg walking were seen, but this estimation is good.

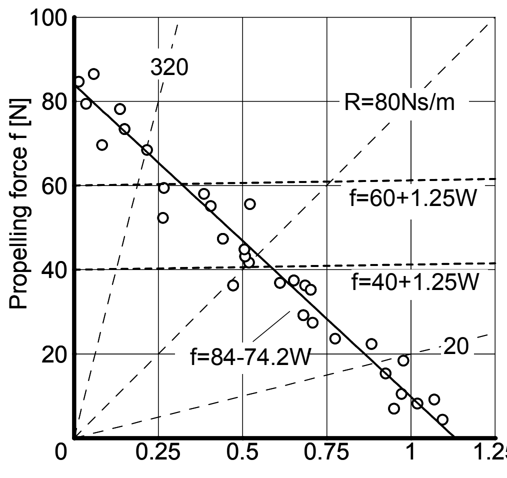


Figure 4: Autonomous propelling performance

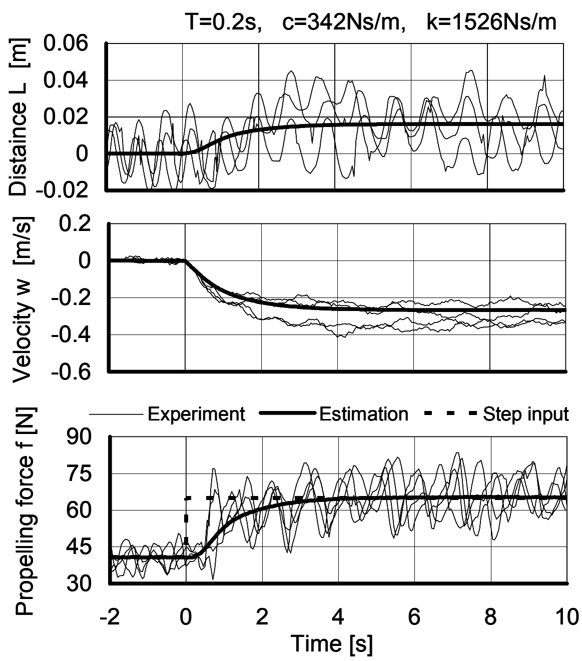


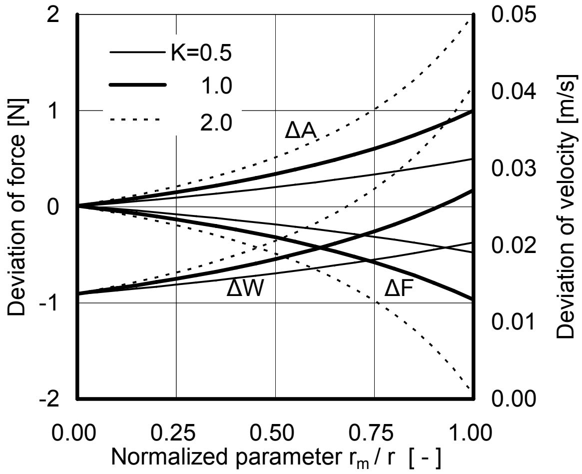
Figure 5: Step response of f, w and L

4 SYSTEM BEHAVIOR AGAINST MODEL PARAMETES

We investigated static and dynamic behaviour on the system regard to the assisting ratio K and the model parameter r_m. This is why that the parameters F_0, r, R_0 and R determined by the subject and the road conditions, are specific in Eq.(9) and (10). We use normalized r_m /r and F_m=$F_0 r_m$ /r in simulations. Figure 6 shows the change of the operating point against the normalized r_m and the K. At the model based controller r_m /r=1.0, the assist force was A=0, so the attendant had to exert force of equivalent road resistance. With decreasing r_m /r to 0 that is the proportional controller, the assisting force A increased with the propelling force F decreased and the wheelchair speed W increased. At r_m /r=0 under K=1, the propelling force F became half and the assisting force A shouldered remaining half of the F. This tendency was enhanced by increasing K. From these results, operating points, which subject power and wheelchair load are balanced, shifted to the direction with increasing the wheelchair velocity W and decreasing propelling force F, according to model parameter r_m decreased from the model based controller's to the proportional controller's. The increase of the assist ratio K accelerates this tendency.

Figure 7 shows the final value on the deviation of propelling force ΔF, wheelchair velocity ΔW and assisting force ΔA from the sensitivity F(s), W(s) and A(s) against additional pushing H(s) and additional road resistance D(s) respectively. Figure 7(a) shows attendant-wheelchair system behaviour against the road resistance D(s) increased. First, at the model based controller r_m /r=1.0, the deviation of the assisting force ΔA against the additional road resistance D(s) was 0, so no assist existed. With decreasing r_m /r to 0, the ΔA and the ΔW increased while the ΔF decreased, as the same tendency in the figure 6. Increasing the K enhanced this tendency. Figure 7(b) shows attendant-wheelchair behaviour when attendant additional propelling forces H(s) happen for acceleration. At the proportional controller r_m /r=0.0, the ΔF and the ΔA were 0 that no assist was obtained when the attendant try to rise the wheelchair velocity. With increasing r_m /r to 1.0 of model based controller, the ΔF decreased and the ΔA increased while the ΔW increased. That means the attendant can rises the wheelchair velocity easily with the assisting force. Increased K enhanced the tendency as the same in the mentioned results.

Figure 8 and Figure 9 shows trajectories against the additional propelling force H(s) and the additional road resistance D(s) of both controllers at K=1 and K=2 respectively. Initial condition is f(t=0)=w(t=0)=0. After the propelling force f(t) and the wheelchair velocity w(t) became steady state, we added additional propelling force h(t)=20N or additional road resistance d(t)=20N to the attendant-wheelchair system. In the figure 8(a) and figure 9(a), thick doted line is first road resistance and thin doted line shows the additional road resistance curve. In similar way at the figure 8(b) and figure 9(b) thick doted line is first propelling force and thin doted line shows additional propelling force. In the results, there were two of the steady operating points that



(b) Sensitivity against H(s)

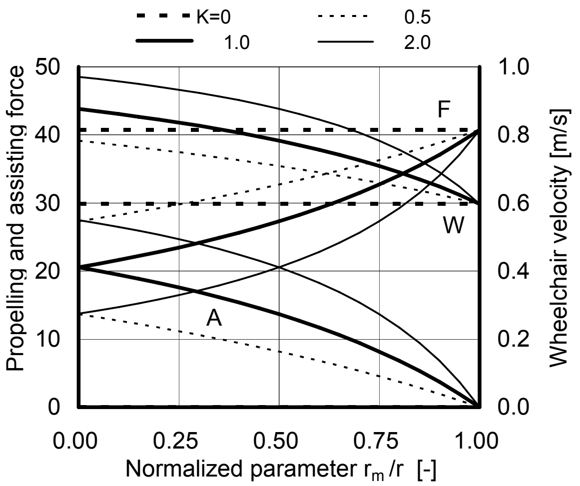


Figure 6: Change of operating point

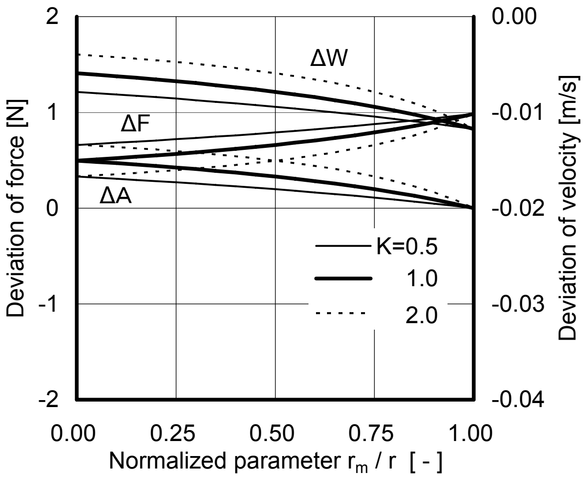
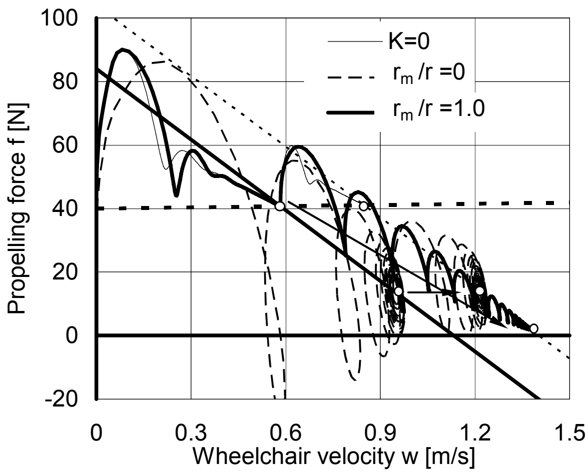


Figure 7:(a) Sensitivity against D(S)

are the one before impressed h(t) or d(t) and the other one after impressed. These two balanced points corresponded to the cross points between the propelling force curve and the road resistance curve. So after additional h(t) or d(t) impressed, the operating point moved from first one to another one. These phenomenons always occurred in each results according to the same rule on the figure 6 and 7. The trajectories went to the operating points with perturbation. Thin line of the trajectory is K=0, no assisting force. The trajectory (thick line) of the model based controller traced the similar way of the K=0 and took the same operating point of K=0. The trajectory (thin doted line) of the proportional controller had large perturbation; however, the controller took the superior operating point of low propelling force and fast wheelchair velocity. After additional d(t) impressed, the operating point shifted along with the line of propelling performance to the direction in low quality. The change of the operating point in the proportional controller was smaller than the one in the model based controller. By the way, after additional h(t) impressed, the operating point shifted on the curve of additional propelling force to the direction in high quality. The operating point of the model based controller became low propelling force and fast wheelchair velocity. The change of the operating force of the model base controller was larger than the one of the proportional controller. At the case of K=2, the tendency of change of the operating points are the same as K=1, but each trajectories at K=2 had larger perturbation than the ones at K=1. The deviation of trajectories in the model based controller r_m /r=1.0 at K=1 and K=2 was smaller than the one in the proportional controller r_m /r=0.0. This tendency became more clear at K=2 in the figure 9.

5 DISCUSSION

We compared the performance of two type of controller, one is the proportional controller and another is our proposed model based controller. In the static characteristic of the controllers, the proportional controller shows good performance against the additional road resistance. In the figure 4, the proportional controller can take high wheelchair velocity and low propelling force compared with the model



(b) Trajectories against H(s)

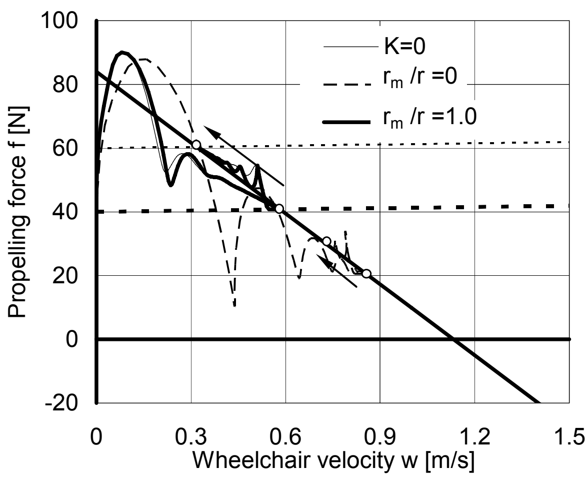
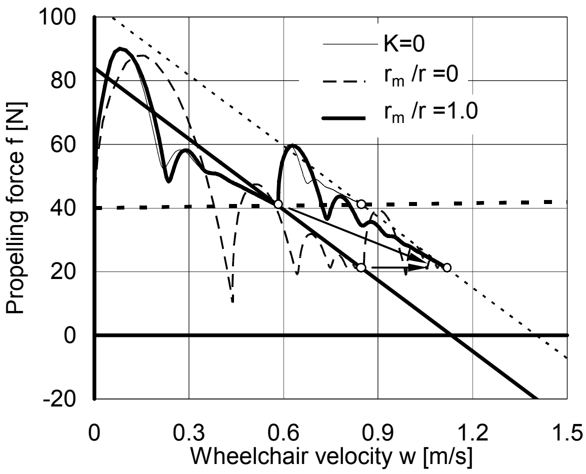


Figure 8:(a) Trajectories against D(s) :K=1



(b) Trajectories against H(s)

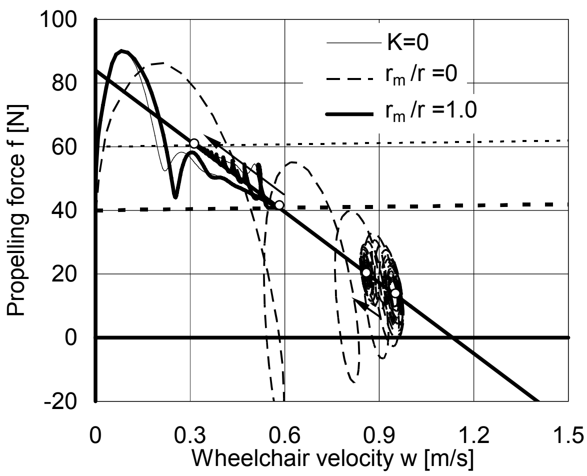


Figure 9:(a) Trajectories against D(s) :K=2

based controller, because of the assisting force. This shift of the operating points reduce the attendant's propelling load, however, large amount of the electric energy is used. The model based controller shows no shift of the operating point, because the controller does not work within the attendant's autonomous propelling performance in daily life. The autonomous propelling performance is very light load showing in the figure 4. So saving electric energy for assisting force, it is recommended if the subject has don't mind to carry out light work, or like some little exercise. The model base controller does not work against road resistance, but the region within the autonomous propelling, the load of attendants propelling is light.

Next, we handled the change of system behaviour against two type of situation; One is when the road resistance in the wheelchairs change higher than previously driving. It is the case that an attendant drives a wheelchair on flat surface, then the surface become upward slope. The other is when the attendant exerts additional propelling force for acceleration of the wheelchair. It is the case that we often encounter. From the figure 7(a), 8(a) and 9(a), the proportional controller showed stiffness against additional road resistance. The decrease of wheelchair velocity is lower than one of the model based controller. This is because that the proportional controller assists full range of the propelling force, so the assisting force shoulders some part of the propelling force. But the model based controller does not work at same situation because the attendant autonomously regulates the propelling force and wheelchair velocity within light work load. It is natural and the condition does not become propelling load, however, people like to keep own walking velocity. So some people may want to assist on this situation in the model based controller. In the case of acceleration of wheelchairs in the figure 7(b), 8(b) and 9(b), the model based controller gives more assisting force than the proportional controller. The proportional controller gives no assisting force when the attendant exerts additional force. This is because the proportional controller determines the assisting force only based on the propelling force. This means that the wheelchair with the proportional controller always has the same reduction from the view of the attendant. So the final balanced point between the attendant and the wheelchair shifts the direction of increasing wheelchair speed and the same propelling force in previous, though the assisting force is exerted at the while of acceleration. This is equivalent of the autonomous propelling performance in figure 4 rises up with additional propelling force, so final balanced point between the attendant and the wheelchair determine the cross point between propelling performance and road resistant curve. The wheelchair with the proportional controller always has same load curve, so the propelling force does not decrease. Against the additional propelling force, the proportional controller does not work, so we expect that attendants are not satisfied with the assisting performance at acceleration. But, the model based controller shows good performance in the same condition. It means that attendants feel cooperation with wheelchair moving at acceleration.

From the view of the dynamics at both controllers, the K determines smoothness of driving wheelchairs. The K=2 brings the large perturbation in attendant-wheelchair system, so attendants have to drive carefully. The perturbation of the model based controller, however, has smaller than the one of proportional controller. This is why the assisting force of the model based controller is limited under the border of the autonomous propelling performance of attendant. On the contrary, all range of the assisting force in the proportional controller is magnified. This perturbation provides mental burden to the attendants, so smaller perturbation is better, we expect easily. And the model based controller can employ higher assisting ratio.

Both controllers show the high quality in different situations. And we can combine with both characteristics by modification of the model parameter r_m. If the r_m is set to 0.5, the controller can have half benefits of both superior performances.

6 CONCLUSIONS

From our results, we finally found that our proposed model based controller for assisting wheelchair propelling, has three benefits compared with the proportional controller.

(1)Employment of high assisting ratio

(2)Cooperation with wheelchair moving according to attendant's operation.

(3)Savings of electric energy.

Also, the proportional controller has superior points below.

(4)Large assisting force against additional road resistance.

Finally, our final proposal from results is that the combination of the proportional controller and the model based controller keep quality and satisfy with the demand of the attendants.

7 ACKNOWLEDGMENTS

The studies were partially supported the Ministry of Education, Culture, Sports, Science and Technology (MEXT) to promote multi-disciplinary research projects.

8 REFERENCES

[1]Resnick, M. L., Chaffin, D. B., 1995, "An ergonomic evaluation of handle height and load in maximal and submaximal cart pushing," Applied Ergonomics, 26, pp. 173-178

[2]Al-Eisawi, K. W., Kerk, C. J., Congelton, J. J., Amendola, A. A., Jenkins, O. C., Gaines, W., 1999, "Factors affecting minimum push and pull forces of manual carts," Applied Ergonomics, 30, pp235-245

[3]Jansen, J.P. et al, 2002, "Evaluation of ergonomic adjustments of catering carts to reduce external pushing forces," Applied Ergonomics, 33, pp117-127

[4]Cremers, G. B., 1989, "Hybrid-powered wheelchair : a combination of arm force and electrical power for propelling a wheelchair," Journal of Engineering and Technology, 13, pp142-148

[5]Suzuki, T., Uchiyama, H., Kurata, J., Grattan, K.T.V., 2004, "I nvestigating the propelling wheelchair behavior against various loads", 2004 ASME International Mechanical Engineering Congress and RD&D Expo (IMECE2004) Proceedings CD-ROM

[6]Alton, F., Baldey, L., Caplan, S., Morrissey, M. C.,1998, "A kinematic comparison of overground and treadmill walking," Clinical Biomechanics, 13, pp.434-440

Hierarchy Genetic Algorithm to Solve Multi-Objective Scheduling Problems Involving Various Types of Assignments for Parallel Processing System

Masahiro Arakawa [1]

[1] Faculty of Environmental and Urban Engineering, Kansai University, Osaka, Japan

Abstract

This paper examines multi-objective scheduling problems involving various types of job assignments. A hierarchy type of genetic algorithm is proposed to search for Pareto solutions for multi-objective functions in an effective manner. The algorithm consists of job assignment control and control of condition for job assignments, and these controls are constructed hierarchically. The structure of controls in the proposed algorithm is adaptable to parallel processing systems to reduce computational time. In addition, various types of procedures are adoptable in order to control conditions for job assignments. In this study, a hybrid type of local search method is introduced as an effective procedure to search for solutions. The local search method is constructed of two local search procedures to optimize different objective functions, and these procedures are alternately executed to obtain optimal Pareto solutions. In this paper, the characteristics of the proposed hierarchy type of genetic algorithm are described. A parallel processing system for the algorithm is then developed and examined on a scheduling problem involving worker assignments and job assignments to evaluate its performance.

Keywords:

Genetic Algorithm; Multi-Objective Functions; Scheduling; Parallel Processing; Job Assignment; Assignment Condition

1 INTRODUCTION

In many of today's manufacturing plants, completion of products by a customer's due date is one of the most important obligations. Customers ordinarily require a short lead-time from orders to receipt of products, so deadline adherence becomes a difficult issue. Scheduling is one of the effective technologies to conform to tight due dates. Many types of scheduling procedures are proposed for due-date conformance [1-3]. When a due-date schedule is generated for a complex production environment, it is difficult not only to allocate job assignments to resources but also to control the conditions for job assignments: splitting of responsibility, worker assignment, control of resource capacity, and so forth. In addition, in today's production environment, various measures need to be introduced to evaluate the quality of scheduling because of increasingly constrained conditions.

It is ordinarily difficult to obtain high-quality solutions in practical computational time for problems involving many parameters. This is because multiple types of parameters necessitate the evaluation of multi-objective functions. Ordinarily, a genetic algorithm is adopted to effectively search for Pareto solutions for multi-objective functions. However, when a genetic algorithm is applied to a scheduling problem involving many parameters, chromosomes to control the parameters become large, and a lengthy calculation time is required to obtain solutions.

In this study, we propose a "Hierarchy Genetic Algorithm" (hereafter known as HGA) to solve a multi-objective scheduling problem involving various types of assignments. In the algorithm, the parameters for assignment controls are separated into parameters to control different types of assignments.

The separated parameters relate hierarchically to the other separated parameters for a more effective search. The hierarchy structure provides for control of the separated parameters under the conditions of the other separated parameters, through which significant results are obtained. Therefore, high quality solutions are attainable in a short computational time. Furthermore, the HGA incorporated is developed in a parallel processing system to reduce computational time.

This paper presents a categorization of assignments for a multi-objective scheduling problem involving various types of assignments. The characteristics of the HGA that enable it to operate chromosomes structured by the assignment categories are presented. The proposed algorithm is examined on a scheduling problem that includes job assignment and worker assignment in order to evaluate its performance. In addition, the performance of the HGA method incorporated in a parallel processing system is examined with regard to reducing the computational time.

2 CHARACTERISTICS OF OPERATIONS FOR VARIOUS TYPES OF ASSIGNMENTS IN THE SCHEDULING PROCESS

In order to manufacture products involving specific configurations by a short due date according to customer requirements, many factories apply complex assignment strategies as well as making decisions regarding the order of jobs assigned to resources: separation of jobs, job assignments to parallel processing machines, operation assignments of skilled workers, and so forth.

In this study, assignment procedures that arise in a scheduling problem are separated into two types of assignment controls: a control for the assignment of jobs to resources, and a control of the conditions to assign jobs to resources. These controls are constructed hierarchically and are executed individually to search for Pareto solutions to this multi-objective problem.

Ordinarily, job tardiness within a schedule is evaluated by using a unit of "job" completed according to the customer's order and/or type of product. Since control of the order of jobs to assign to machines is more effective than another control for reducing job tardiness, we make control of the condition for job and /or resource assignments belong to the control of order of job assigned to resources in the hierarchy structured control. Controlling the condition for job assignment is categorized into the following three types of controls:

(1) selection of resources for assignment control from different types of resources;

(2) assignment of operations of which jobs are composed;

(3) available time to assign operations to resources.

One (or a combination) of these categorized controls are then used to control the condition. A genetic algorithm is adopted for the operation of these controls for this study. The effectiveness of the controls (1) to (3) for seeking solutions is affected by the representation of chromosomes.

3 HIERARCHY STRUCTURED GENETIC ALGORITHM

We propose a hierarchy structured genetic algorithm, HGA, to resolve a multi-objective scheduling problem by using parameters for controls introduced in the previous section. In this paper, the variable "chromA" represents a chromosome to control the order of jobs assigned to resources, and "chromB" represents a chromosome to control the condition of job assignments. That is, these chromosomes are controlled individually, and chromB is controlled under the condition in which chromA is fixed.

Figure 1 shows a schematic diagram of structures to control chromosomes in an HGA and a simple genetic algorithm (GA). A simple GA is a conventional genetic algorithm to control a population of chromosomes, one of which generates a single schedule. In this figure, the chromosome is composed of chromA and chromB. As for the HGA, with regard to a single chromosome of chromA, multiple chromosomes of chromB are evaluated to search for Pareto solutions in the evolutional process.

For chromA to control the order of jobs assigned to resources, the determined number of job numbers are allocated in the chromosome. Data in the chromosome of chromA are controlled by the movement and replacement of job numbers within a chromosome array. On the other hand, chromB, used to control the condition of job assignments on different types of resources, is introduced to control the selection of resources and the amount of operations in split job lots. However, since data within the chromB chromosomal elements can be decided arbitrarily, a genetic operator composed of crossover and mutation based on swap and/or shift is difficult to identify to search for optimal solutions. The procedure of the HGA is shown below:

Process of HGA method

(1) Maximum generations and number of chromosomes in the population of chromA take value of MaxGene and value of popA, respectively.

(2) Generation is initialized as gene = 0.

(3) The initial population of chromA is generated, and is named set "z0."

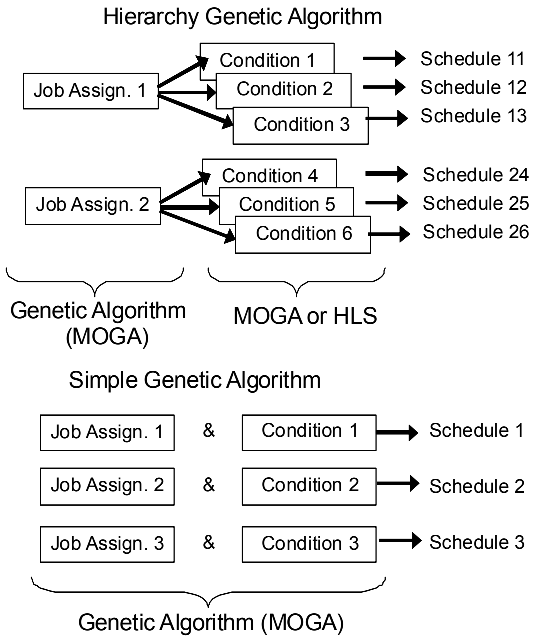


Fig.1 Schematic diagram of structures to control chromosomes in Simple GA and HGA

(4) A procedure "*Proc*" is executed for all chromosomes in the population.

(5) All chromosomes of chromB in chromA in set z0 are ranked.

(6) Generation is updated: gene = gene + 1.

(7) A new population of chromA is prepared as set "z1" and z1 is set empty: z1 = φ.

Chromosomes of popA are selected from set z0 and are added to set z1 according to the order of (7-1) and (7-3):

(7-1) Pareto preservation;

(7-2) selection using rank for sharing;

(7-3) parallel selection.

(8) All chromosomes of chromB in all chromosomes of chromA in set z1 are ranked.

(9) New chromosomes of chromA are generated according to the following process:

(9-1) the number of chromosomes is initialized, pop = 0, and all chromosomes in set z0 are removed: z0 = φ;

(9-2) chromosomes of chromA are selected as parent chromosomes from set z1 using the roulette selection process;

(9-3) crossover is executed to generate offspring chromosomes;

(9-4) mutation is executed using uniform random number;

(9-5) the offspring are added to set z0;

(9-6) the number of set z1, pop, is updated;

(9-7) when pop<popA, go back to (9-2). When pop>popA, chromosomes in set z1 are deleted to make pop = popA, then go to (10). Otherwise, go to (10).

(10) Procedure *Proc* is executed for all chromosomes of chromA in set z0.

(11) All chromosomes in set z1 are transferred into set z0.

(12) All chromosomes of chromB in all chromosomes of chromA in set z0 are ranked.

(13) When gene<MaxGene, go to (6). Otherwise, the procedure is complete.

Ordinarily, *Proc* includes the generation and evaluation of schedules. In this study, either the multi-objective genetic algorithm or a hybrid local search method is adopted as the procedure *Proc*. Properties of these procedures are explained in the next section. To select chromosomes of chromA for a successive generation, Pareto preservation and parallel selection are utilized. Here, “hybrid local search” denotes the hybrid type of procedure structured from the combination of local searches for different objective functions in the multi-objective problem. The Fonseca procedure[4] is used to calculate ranks of chromosomes. In the roulette selection in step (9-2), chromosomes of chromA are selected by using the rank of all chromosomes of chromB in all chromosomes of chromA in the population. The probability defined by Eq. (5) is used for the roulette selection process.

In the hybrid local search process, local searches for different objective functions are executed interchangeably under the condition of a single chromosome of chromA. Figure 2 shows a schematic diagram of the hybrid local search behavior on solution space. Each different local search starts from the temporal Pareto solutions obtained by the other local search in the hybrid local search procedure. The different local searches are calculated iteratively and alternately until Pareto solutions cannot be improved, or until a determined iterative number is achieved. The alternate process of the local searches denotes the combination of processes to search for single optimal solutions to different Cartesian coordinate directions composed of coordinates indicating the values of objective functions.

Since the HGA method includes different types of assignments constructed hierarchically and controlled individually, it is adoptable to a parallel computing system and would be effective at reducing calculation time. Here, the HGA method is re-proposed as a genetic algorithm to control different types of assignments hierarchically in the parallel processing system. Figure 3 shows a schematic diagram of a parallel computing system in which the HGA method is incorporated. A master PC is used to process a genetic operator for the population of chromA, and the chromosomes are distributed to slave PCs after the process is complete. In each slave PC, the population of chromB chromosomes is operated on to search for optimal Pareto solutions with regard to each chromosome that chromA distributed from the master PC. After population of chromosome chromB obtaining Pareto solutions is sought with regard to each chromosome in chromA, combinations of a single chromosome chromA and population of chromosome chromB are transferred to master PC from each slave PCs. After the master PC receives the combinations of each single chromA and population chromB from all slave PCs, the population of chromA is updated for the successive generation by using the rank of solutions for all chromB chromosomes of all chromA chromosomes. Then, the chromosomes of chromA selected are distributed to slave PCs after processing by the genetic operator in the master PC.

When the multi-objective genetic algorithm (MOGA) is adopted to search for optimal Pareto solutions obtained by controlling chromosomes of chromB under the single chromA condition, the chromB population transferred from slave PCs to the master PC corresponds to a population of chromosomes, the number of which is predetermined for the MOGA process. Therefore, the MOGA adopted for controlling the chromosomes of chromB in parallel processing system is effective to reduce the calculation time.

It is expected that the hybrid local search is able to seek Pareto solutions more effectively than the MOGA because of the direct search of an optimal solution for a single objective function from different initial points. However, when the hybrid local search is adopted to search for Pareto solutions by controlling the chromosomes of chromB, and the procedure includes a uniform distribution of chromosomes of chromA to slave PCs, the slave PCs would require different calculation times; parallel processing would not be effective in reducing calculation time. Therefore, developing a procedure to distribute chromA chromosomes for obtaining a uniform slave PC calculation time is required in a parallel processing system when a hybrid local search is introduced.

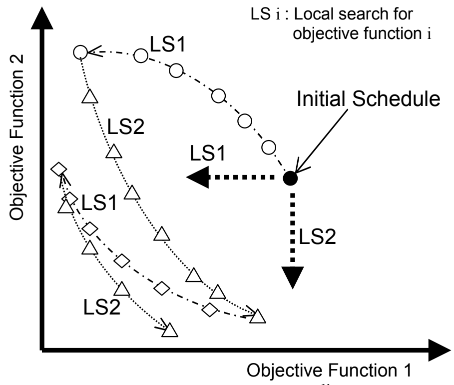


Fig.2 Schematic diagram of behavior of hybrid local search on solution space

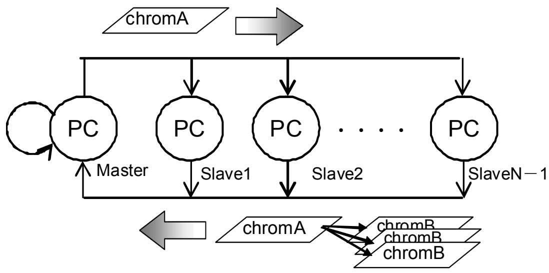
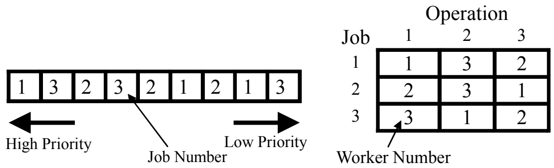


Fig.3 Schematic diagram of network structure of parallel processing system

4 REPRESENTATION OF CHROMOSOMES TO CONTROL JOB ASSIGNMENTS AND ASSIGNMENT CONDITIONS

For the chromosome of chromA to control the order of jobs assigned to resources, an operation-based representation [5] is adopted in which the priority of jobs assigned to resources is represented in a one-dimensional array. The chromB chromosome for the control of job assignments is defined as a three-dimensional array composed of multiple two-dimensional arrays. The two-dimensional array is defined as an array to control the conditions related to different resources, such as: the selection of machines, volume of lots separated from jobs, capacities of machines, and so forth. Figure 4 shows examples of chromA chromosome and chromB chromosome for the scheduling problem involving worker assignments. In this figure, chromB represents a two-dimensional array to control workers assigned to each operation.



(a) ChromA (b) ChromB

Fig. 4 Example of chromosomes (3 jobs, 3 machines, 3 workers)

5 MULTI-OBJECTIVE SCHEDULING CONSIDERING DIFFERENCE OF WORKER'S SKILLS

5.1 Characteristics of scheduling problem

When the skills which workers possess are assumed to affect the operation time, higher-skilled workers would cause a reduction in operation time. In a scheduling problem involving assignments of workers with various skills, past scheduling results indicate there is a trade-off relationship between job tardiness and processing time differences of operations assigned to skilled workers. In this scheduling problem, assignment control is separated into the control of two types of operations: the order of jobs assigned to resources, and the selection of workers assigned to the operations.

As for representing the chromosome of chromA, an operation-based one-dimensional array is introduced, and job numbers are allocated in the array. Here, the number of each job number is equal to the number of operations included in the job. For representing the chromosome of chromB, a two-dimensional array is introduced, with elements corresponding to the operations of jobs. Then, the job numbers and operation numbers are assigned to row and column of the array, respectively, and workers' numbers are allocated in the elements of the array.

5.2 Model

To evaluate the performance of the proposed methods, we used a numerical experiment on a simple job shop model. As the numerical model, a benchmark problem proposed by Fisher and Thompson, called "ft10," is utilized[3].

Further conditions are added to the original problem for evaluation purposes related to skilled workers and job due date. Due date of job (i, d_i), is defined as the following:

$$d_i = k_d \sum_j p_{ij} \tag{1}$$

Here, k_d denotes the due-date coefficient, which takes 1.5 in this experiment. In addition, three skill levels of workers are introduced to affect operation times assigned to workers, the number of workers is six, and each skill level of workers consists of two persons.

It is assumed that the skill levels depend on the workers alone, not on the job or resource. That is, the operational effective coefficient η_{ijk} for *j-th* operation of job i, O_{ij}, affects worker k. Here, the operational effective coefficient is equal to the ratio of operation time affected by skilled workers to the standard operation time, and corresponds to the worker's skill in this problem.

The operational effective coefficient η_{ijk} can be represented as $\eta_{ijk} = \eta_k$, and it takes one of a set of numbers corresponding to three levels: 0.8, 1.0, and 1.2. Therefore, the operation time of operation O_{ij} assigned to worker k, p_{ijk}, can be calculated from equation $p_{ijk} = \eta_k \cdot p_{ij}$. Here, p_{ij} represents the standard operation time of operation O_{ij}.

Two objective functions related to trade-off are introduced: average tardiness and the peak-to-peak value of operation time among all workers. In this paper, average job tardiness is denoted as L. The peak-to-peak value of operation time among the workers is referred to as "peak-to-peak value of operation time" and is hereafter denoted as δP. These objective functions are defines as follows:

$$L = \frac{1}{n}\sum_{i=1}^{n} \max(L_i,\ 0), \tag{2}$$

$$\delta P = P_{\max} - P_{\min} \quad , \tag{3}$$

$$L_i = C_i - d_i \ , \tag{4}$$

$$P_{\max} = \max\{P_k \mid k = 1, \cdots, K\} \quad , \tag{5}$$

$$P_{\min} = \min\{P_k \mid k = 1, \cdots, K\} \quad , \tag{6}$$

$$P_k = \sum_{i=1}^{n}\sum_{j=1}^{m_i} \bar{p}_{ij}\, \eta_{kij}\, \delta_{kij} \ . \tag{7}$$

Here, C_i and d_i represent the completion time of job i and due date of job i, respectively. Variables n and m_i are the number of jobs and the number of operations of job i., respectively. K denotes the number of workers. When worker k is assigned to j-th operation of job i, δ_{kij} takes 1. Otherwise, δ_{kij} takes 0.

5.3 Characteristics of procedures to control worker assignment as condition of job assignment

In the hybrid local search procedure, two types of local searches for different single objective functions are executed iteratively and alternately for deciding worker numbers in the chromosome of chromB. One of the two local searches includes a process to exchange workers assigned to operations of tardy jobs to workers assigned to operations of jobs completed earlier, after investigation of completion times of jobs in the schedules. This process represents a reduction in average job tardiness. The other local search includes a process to transfer operations assigned to a worker spending maximum operation time to a worker spending minimum operation time, in order to reduce the peak-to-peak value of operation times among workers. This process represents a peak-to-peak value reduction in operation time.

The MOGA involves the combination of Pareto preservation and parallel selection as a selection process for the successive generation in order to search for Pareto solutions by the chromB chromosomes operation. In addition, parents of chromosomes are selected from the population by roulette selection using Foresca's ranking procedure in the crossover process of the algorithm.

5.4 Evaluation of HGA method for multi-objective functions

In order to evaluate the Pareto solutions obtained by the HGA method, the proximity of solutions to the Pareto front, $CS(X_1, X_2)$, is introduced as a measure, and is defined as follows:

$$CS(X_1, X_2) = \frac{\left| \{ a_2 \in X_2 \mid \exists a_1 \in X_1 : a_1 \preceq a_2 \} \right|}{\left| X_2 \right|} . \tag{8}$$

Here, X_1 and X_2 represent the two sets of chromosomes used for comparison. The relationship $a_1 \preceq a_2$ indicates that solution a_2 is not inferior to solution a_1.

In order to evaluate the performance of the HGA method, we compared three types of scheduling procedures: the HGA-HLS method, the HGA-MOGA method, and the SGA method. The HGA-HLS method refers to the hybrid type of genetic algorithm in which a hybrid local search procedure is incorporated to control the chromB chromosome. The HGA-MOGA method refers to the hybrid type of genetic algorithm in which the MOGA is incorporated to control the chromB chromosome.

The SGA method refers to the simple genetic algorithm (conventional genetic algorithm). In the SGA method, a chromosome of chromA and a chromosome of chromB are combined into a single chromosome to generate a single schedule, and the combined chromosomes are operated on.

With regard to genetic operators, the crossover ratio and mutation ratio take values of 1.0 and 0.05 in all algorithms, respectively. The two-point crossover is used as the crossover process to operate the chromA chromosome. In the crossover process, when identical job numbers are involved in two exchanged partial chromosomes, the positions of the identical job numbers in the exchanged chromosomes are used, and the numbers are allocated at their positions. Then, unallocated job numbers in the original partial chromosomes are allocated at the space elements in the exchanged chromosomes according to the order of job priority in the original chromosomes. As for the mutation process to operate the chromA chromosome, job numbers in two arbitrary elements in the chromosomes are swapped iteratively. Here, the number of iterations is predetermined to be one-fourth of the length of the chromosome.

With reference to the crossover to operate the chromB chromosomes, a two-point crossover is utilized to operate workers' numbers in the elements of a one-dimensional array in a row with the same direction as the chromosomes. With regard to the mutation for chromB chromosomes, workers' numbers are changed to numbers decided by using a uniform random number in all the elements of chromB chromosome, while the probability to change the number is 0.25 at each element.

Table 1 shows the numerical conditions of genetic operators used for different methods. Table 2 shows the proximity of solutions to Pareto front and the computational time obtained from different methods. Figure 5 shows the distributions of Pareto solutions sought by different algorithms. Figure 5 and Table 2 show that both HGA methods obtained better Pareto solutions than the SGA method. In addition, the calculation time taken to obtain these solutions by both HGA methods is less than a half of that taken by the SGA method. These results indicate that the hybrid local search procedure is more effective than the multi-objective genetic algorithm based on parallel selection and Pareto preservation to operate workers' numbers in the chromosome of chromB with regard to the quality of solutions and reduction in computational time.

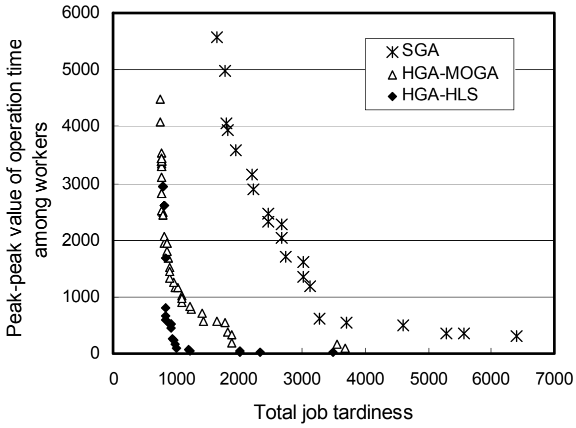


Fig. 5 Distributions of Pareto solutions obtained by different methods

6 EFFECTIVENESS OF PARALLEL PROCESS FOR THE HGA METHOD

In this section, the HGA method is incorporated into the parallel processing system and its effectiveness is evaluated for calculation time reduction. The network structure of the parallel computing system is shown in Fig. 3. The program used for parallel processing was generated by using MPICH. The chromosomes of chromA are distributed to all slave PCs from the master PC uniformly in the process.

Figure 6 shows the computational time, total parallel processing time for controlling the chromB chromosome, and the parallel processing ratio with regard to the HGA-HLS and HGA-MOGA methods. Here, the parallel processing ratio is defined as follows:

$$\eta = T_1 / pT_p . \tag{9}$$

The variable p denotes the number of nodes, T_p denotes the computational time at the parallel processing system in which the number of nodes takes p.

Job shop scheduling problem (ft10) shown in the previous section is used for these comparisons. The resultant computational time is the average value of elapsed times obtained from five tests on each node condition in the figure. The error bar represents the standard deviation of the computational times. PCs (with the same performance CPUs, Pentium 4, 2.66GHz) are used for the numerical experiment. RAM of between 512 MB and 1.2 GB are installed in the PCs, the capacities of which are sufficient to calculate the problem.

When the number of nods in the system increases, both the HGA-HLS method and the HGA-MOGA method generate a reduction in computational time. The HGA-MOGA method takes a larger parallel processing ratio at a large number of nodes than the HGA-HLS method. As for the HGA-HLS method, the system of five or ten nodes takes approximate half of the computational time required in the system of a single node.

Since the time elapsed for the parallel processing to operate the chromB chromosome is approximately equal to the total computational time in both methods, the HGA-HLS method is less effective than the HGA-MOGA method with regard to the system composed of a large size of nodes. The reason for this is that different slave PCs exhibit different elapsed times when searching for solutions after the chromosomes of chromA are distributed from the master PC uniformly in the HGA-HLS method. Therefore, a function is required to be installed in which the quantity of chromA chromosomes distributed from the master PC to each slave PC is dynamically controlled.

7 SUMMARY

In this study, a hierarchy type of genetic algorithm (HGA) was developed to resolve the multi-objective scheduling problem involving different types of assignment controls. A parameter control to search for Pareto solutions is separated into two types of controls: control of job assignment to resources, and control of condition of resources for job assignment. These controls are constructed hierarchically and are executed independently. Furthermore, to reduce computational time, an HGA was developed to process on a parallel processing system. A hybrid type of local search procedure was developed to control condition of job assignment.

These methods were examined on a job shop scheduling scenario problem involving the assignment of skilled workers to evaluate the effectiveness of each. Experimental results show that the HGA method incorporated into a parallel processing system is effective to search for Pareto solutions within a short computational time. These results indicate that the HGA method incorporated into a parallel processing system is useful for complex multi-objective scheduling problems. In particular, the HGA method, in which a hybrid type of local search procedure is incorporated, is effective for searching for high-quality Pareto solutions. However, the HGA method requires an additional function to distribute the chromosomes of chromA to slave PCs dynamically in order to reduce the large computational time used in the parallel processing system, a potential topic for future study.

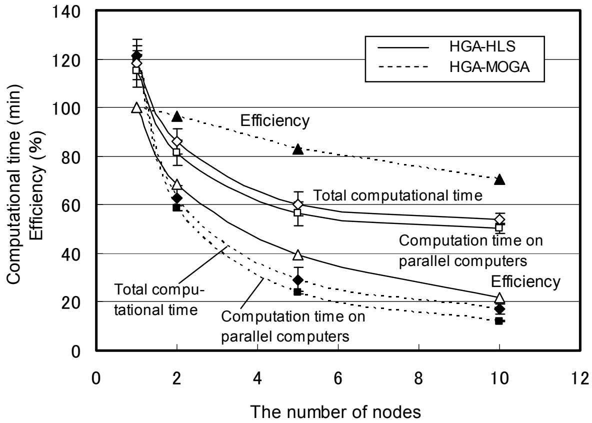


Fig. 6 Comparison of computational time on parallel process system between different HGA methods

8 REFERENCES

[1] Conway, R.W., Maxwell, W.L., Miller, L.W., 1967, Theory of Scheduling, Addison-Wesley.

[2] Brucker, P., 2003, Scheduling Algorithm, Springer-Verleg.

[3] Muth, J.F., Thompson, G.L. (eds.), 1963, *Industrial Scheduling*, Prentice Hall.

[4] Foresca, C.M., Fleming, P.J., 1993, Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generation, Proc. of 5th Int. Conf. on Genetic Algorithms, 416–423.

[5] Gen, M. and Cheng, R. (eds.), 2000, *Genetic Algorithms & Engineering Optimization*, Wiley-Interscience Publication.

Table 1: Conditions of genetic operators for different methods

	chromA		chromB	
	Population	Maximum Generation	Population	Maximum Generation
SGA	500	3000	–	–
HGA-MOGA	50	100	50	100
HGA-HLS	50	100	Max_pop* = 20	

* Max_pop denotes maximum population for local search to minimize peak-to-peak value of operation time.

Table 2: Proximity of solutions to Pareto front obtained by different methods and computational time

X_2 \ X_1	SGA	HGA-MOGA	HGA-HLS	Computational time
SGA	–	0.00	0.00	5:22’48’’
HGA-MOGA	1.00	–	0.37	2:01’34’’
HGA-HLS	1.00	0.83	–	1:54’30”

Productivity Analysis of Closed-loop Manufacturing System Which Performs Maintenance Activities

Yoshichika Tanaka[1], Hirohisa Narita[1], Ren Kanehira[2], Hideo Fujimoto[1]

[1] Nagoya Institute of Technology, Aichi, Japan
[2] Fujita Health University College, Aichi, Japan

Abstract
This study deals with the evaluation of the productivity of closed-loop manufacturing system which performs maintenance. The authors proposed the method for evaluating a productivity of theoretical evaluation of the productivity, and showed the effectiveness of parts diversions by analyzing the relationship between each parameter. Moreover, a simulation method which can show clearly the result of theoretical evaluation and the fluctuation between the relations of each parameter was proposed.

Keywords:
Closed-loop Manufacturing System; Maintenance; Long-term Use; Productivity Analysis; Simulation

1 INTRODUCTION

Long-term use and reuse of devices and equipments are required in order to construct a recycling-oriented society which can solve environmental issues such as resource and energy saving, and global warming prevention. As examples of the long-term use and the reuse, there are continuing operations of railway rolling stick or aircraft which are ensured by periodic inspection and maintenances in repair shops [1], [2], and there are also reuses of OA equipments by repair and upgrading according to the needs [3], [4].

In these cases, relevant parts of products are inspected and repaired after disassembly, and reassembled in a repair shop. Since the collected used products, replacement parts and power source are input, and the repaired products are output, this type of repair shop is considered as a closed-loop manufacturing system. Especially these manufacturing systems which reassemble the same products after inspection and repair of disassembled parts can be classified as maintenance or remanufacturing. This type of manufacturing system has generally lower environmental burdens than material recycling [5]. Thus, its contribution to earth environment preservation is also high, and its importance will continue to increase.

In conventional researches concerning the closed-loop manufacturing system, they discussed its economical efficiencies or environmental burdens due to manufacturing activities, and its cooperation with other manufacturing systems [6], [7]. However, there were not so many research works, which dealt with productivity of a single closed-loop manufacturing system performing maintenance or remanufacturing. Therefore, in this research, the authors studied on a productivity analysis of the closed-loop manufacturing systems which perform operating maintenance or remanufacturing, such as repair shops of rolling stock and aircraft, which are closed single manufacturing system.

2 FEATURES AND ISSUES OF THE SYSTEMS STUDIED

There are some issues of the closed-loop manufacturing systems which perform maintenance or remanufacturing. These Issues is as follows:

- The process time fluctuates because of the fluctuation of production load.
- The change of the process order is difficult.
- It is difficult to construct the manufacturing system adjusted to the products.

In these issues, a fluctuation of production load is the most

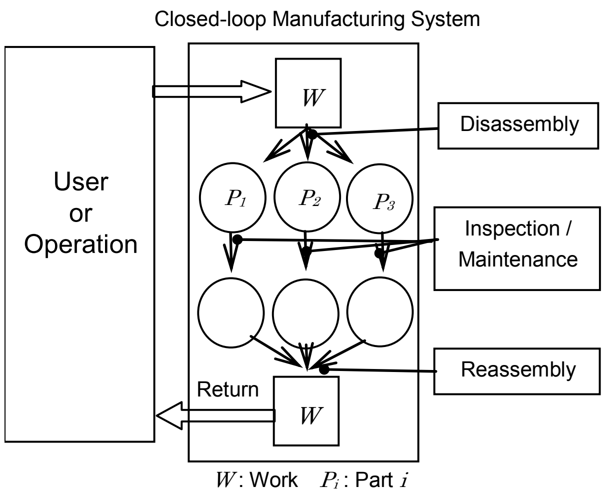


Figure 1: Model of a closed-loop manufacturing system

characteristic issue of these closed-loop manufacturing systems. In this research, we formulated the relation among indices relating to the production time and productivity, which are influenced by this fluctuation of production load. Moreover, we clarified relations between production time and productivity by a simulation based on this formulation.

The followings are given as features of the closed-loop manufacturing systems dealt with by this research. The inputted products, termed as a "work" henceforth, are deteriorated because of beforehand use. Furthermore, the deterioration level of the works may be fluctuating remarkably. The process of a closed-loop manufacturing system performing maintenance or remanufacturing can be divided roughly into disassembly, inspection / maintenance, and re-assembly as shown in Figure 1. Among these processes, by the fluctuation of deterioration states of these works, the process time fluctuates especially in the process of disassembly and inspection / maintenance, and productivity is influenced. Moreover, it is another feature of the closed-loop manufacturing system that the system is able to resume disassembled parts from other throwing-in products, not from the same product. For this reason, in this study we examined also the productivity of this system in case of reuse of the diverted parts. Figure 2 shows the image of parts diversion.

3 EVALUATION OF PRODUCTIVITY

In order to evaluate the productivity of the closed-loop manufacturing system, we examined indices which represent the productivity. Paying attention to indices representing working rate, delay rate, quantity of stocks and time productivity, we examined the fluctuation of the indices influenced by fluctuation of load.

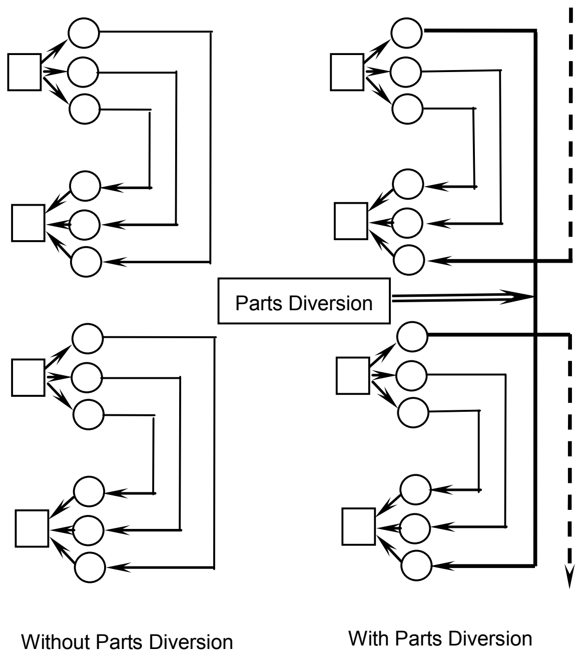


Figure 2: Image of parts diversion

3.1 Requisite assumptions

Based on the feature as of the systems described in the preceding chapter, assumptions which are required for examining the productivity of the closed-loop manufacturing system are listed as follows:

a. Actual times of disassembling and inspection / maintenance process (t_D, t_C) fluctuate.
b. Disassembling and inspection / maintenance process times T_D, T_C are set up in consideration of fluctuation of actual time t_D, t_C.
c. Actual time of a re-assembling process is not fluctuating.
d. Only one work exists in each process.
e. The parts which require long actual time of disassembling and inspection / maintenance process are planned to be diverted to other works.

3.2 Theoretical examination

Based on the features and preconditions arranged as mentioned above, in order to predict and evaluate the productivity according to the time fluctuation of the closed-loop manufacturing system, we examined the working situation, delay of process time and the fluctuation of quantity of stocks theoretically. In addition, fluctuations of these events due to performing parts diversion were investigated. Furthermore, we calculated working situation, delay of process time and quantity of stocks from the set up process time and the actual process time. Figure 3 shows the image of the process time.

A case without considering parts diversion

Working rate R_W is calculated by the ratio of set process time to actual process time. Therefore, the working rate R_W can be expressed by the equation (1).

$$R_W = \frac{\bar{t}_D + \bar{t}_C}{T_D + T_C} \qquad (1)$$

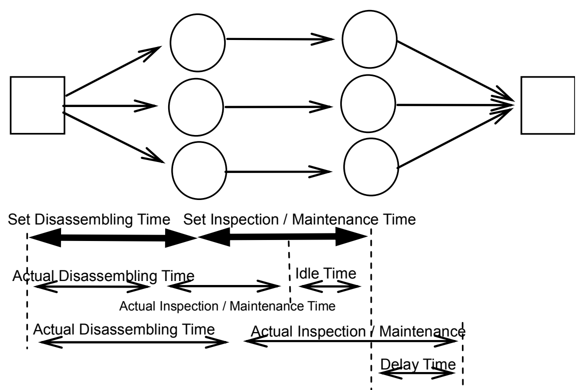


Figure 3: Image of process time

where

T_D : Set disassembling process time,

T_C : Set inspection / maintenance process time,

$\bar{t}_D$: Mean time of actual disassembling process,

$\bar{t}_C$: Mean time of actual inspection / maintenance process

Since the delay of each process occurs when the actual process time exceeds the set process time, we can consider the delay rate to be a probability that working rate will exceed one. Therefore, the delay rate R_D can be expressed by the equation (2).

$$R_D = \frac{N_{(R_W>1)}}{N_{ALL}} \qquad (2)$$

where

$N_{(R_W>1)}$: Number of patterns in which a working rate exceeds one

N_{ALL} : Total numbers of patterns

We can consider the quantity of stocks to be the total number of the parts, which exist in each process of the closed-loop manufacturing system. Therefore, the quantity of stocks St can be expressed by the equation (3).

$$St = n_D + n_C + n_A \qquad (3)$$

where

n_D : Number of parts in disassembling process,

n_C : Number of parts in inspection / maintenance process,

n_A : Number of parts in reassembling process

Furthermore, time productivity Pr is given by dividing the quantity of production by total set process time. Therefore, Pr can be expressed by the equation (4). This time productivity Pr is in inverse proportion to total set process time.

$$Pr = \frac{p}{T} \qquad (4)$$

where

p : Quantity of production

T : Total Set Process Time

A case with consideration of parts diversion

When parts diversion is considered, the working rate R_W needs to take into consideration the time required for diverting the parts to the destined works from the original works. For this reason, working rate can be calculated by the ratio of the actual process time of diverting the parts against the set process time which takes into consideration the work throwing-in interval of work and the number of skips. Therefore, it is expressed by the following equation (5).

$$R_W = \frac{\bar{t}_D + \bar{t}_C}{T_D + T_C + sT_I} \qquad (5)$$

where

s : Number of skips of parts diversion,

T_I : Throwing-in interval of work

Moreover, Figure 4 shows the explanation of throwing-in interval of work, and Figure 5 shows the explanation of number of skips when the parts diversion is executed.

Since the delay rate R_D is the probability that a working rate exceeds one, this value can be calculated in the same way as in the case without considering parts diversion. Thus, R_D is expressed by the equation (2).

Since the quantity of stocks St is a total number of the parts which exist in a closed-loop manufacturing system in case parts diversion is considered, parts are considered to exist in a manufacturing system even during being diverted to the destined work. For this reason, it is necessary that the number of skips of parts should be added to the number of parts which exists on each process. Therefore, St is expressed by the equation (6).

$$St = sn_D + sn_C + n_A = (1+2s)n \qquad (6)$$

Furthermore, time productivity Pr is the ratio of the quantity of production against set process time similarly to a case with considering parts diversion. Thus, Pr is expressed by equation (4).

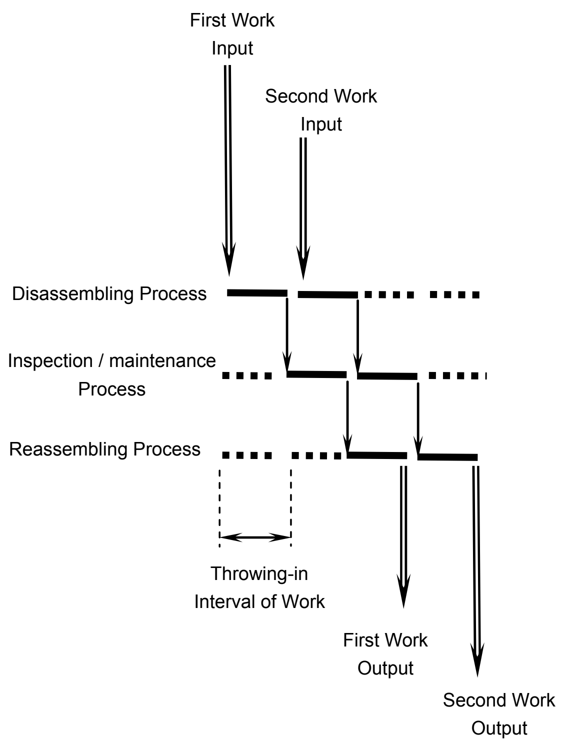
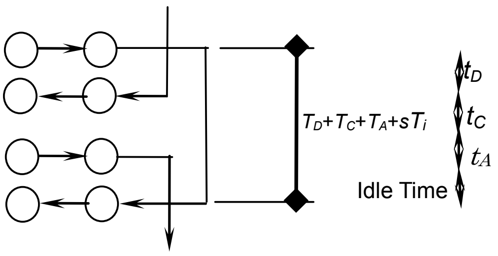
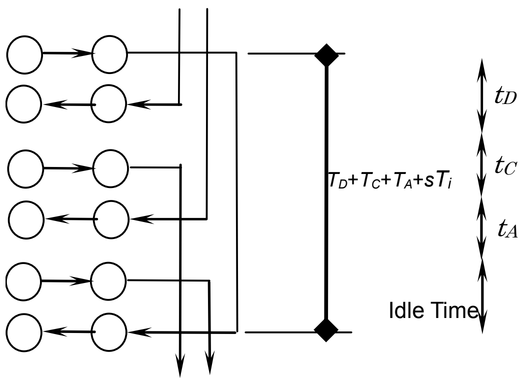


Figure 4: Explanation of throwing-in interval of work



Number of Skips s=1



Number of Skips s=2

Figure 5: Explanation of number of skips

3.3 Evaluation based on theory

A case with consideration of parts diversion

Based on the above-mentioned theory, we evaluated the relation between throwing-in interval of work and other parameters such as working rate, delay rate, time productivity, quantity of stocks in the case without considering parts diversion. These results are as follows:

a. The working rate of manufacturing system decreases when the throwing-in interval of works increases. This is because when the throwing-in interval increases, the set process time in the equation (1) increases compared with the actual process time, and the working rate decreases.

b. The delay rate decreases when the throwing-in interval of work increases. This is because when the throwing-in interval of work increases compared with set process time, number of pattern in which the working rate shown in the equation (2) exceeds one decreases.

c. The quantity of stocks does not fluctuate, when the throwing-in interval of work fluctuates. This is because even if the throwing-in interval of work fluctuates, the number of the works and parts which exists in a system does not fluctuate as shown in the equation (3).

d. The time productivity decreases when the throwing-in interval of work increases. This is because the quantity of production is in inverse proportion to the throwing-in interval of work, as shown in the equation (4).

A case with consideration of parts diversion

In the case where the parts diversion is considered, we evaluated the relations between throwing-in interval of work and other parameters such as working rate, delay rate, time productivity, quantity of stocks. These results are as follows:

a. If the throwing-in interval of work increases, the working rate will decrease. This is because the value of the denominator increases if the throwing-in interval of work increases, as the process time of diverted parts is added to the denominator of the equation (5).

b. Even If the throwing-in interval of work fluctuates, the delay rate does not fluctuate from zero. This is because the delay of inspection / maintenance process does not occur, because of performing parts diversion.

c. The quantity of stocks decreases, if the throwing-in interval of work increases. This is because the number of skips of diverted parts increases, and the quantity of works and existing parts in the manufacturing system decrease if the throwing-in interval of work increases, as shown in the equation (6).

d. The time productivity decreases when the throwing-in interval of work increases, similarly to a case without considering parts diversion.

As mentioned above, we could grasp and evaluate the relation between fluctuation of the actual process time and indices such as working rate, delay rate and quantity of stocks of closed-loop manufacturing systems, by theoretical examination. Moreover, we could confirm the effect of parts diversion by this evaluation based on theory. It is conceivable that optimization of the productivity of the closed-loop manufacturing system is attainable in consideration of the balance of these parameters and appropriate selection of the set process time and a throwing-in interval of work.

However, in an actual closed-loop manufacturing system, since the time productivity tends to decrease if a throwing-in interval of work increases, it is necessary to determine a set process time and throwing-in interval of work in consideration of demanded quantity of production. Moreover, when the set process time is prolonged, the number of equipments increases and sales opportunities decreases. Furthermore, if the number of parts diversion increases, the equipments for processing and the quantity of stocks increase. Therefore, it is necessary to consider these aspects in the practical examination.

4 SIMULATION

Regarding the productivity evaluation of the closed-loop manufacturing system as shown in the preceding chapter, we examined the simulation method based on the theory, in order that the evaluation results are made further concrete.

4.1 Simulation method

We examined the simulation method which will satisfy the following conditions:

a. Taking the set process time as constant, and the throwing-in interval rate of work which is calculated by dividing the throwing-in interval of work by set total process time as variable, we calculate the relations between a throwing-in interval of work and other parameters such as working rate, delay rate, quantity of stocks and time productivity.

b. Parts diversion is performed when the actual process time of disassembling and inspection / maintenance process exceeds the set process time.

Moreover, settings of the simulations were as follows:

a. The throwing-in interval rate of work is made to fluctuate from 0.25 to 1.

b. The actual process time of disassembling and inspection / maintenance process are made to fluctuate to 4 times in set process time.

With due consideration on the above-mentioned points, we performed simulations using the spreadsheet software of a personal computer.

4.2 Result of simulation

A case with consideration of parts diversion

Figure 6 shows the simulation result in the case without considering parts diversion. This figure shows the situation of other parameters according to fluctuation of throwing-in interval of work. As a result, it was apparent that as the throwing-in interval rate of work increases, the working rate decreases, the delay rate decreases remarkably, the time productivity decreases, and there is no fluctuation in the quantity of stocks.

The result of this simulation was consistent with the evaluation based on the theory as examined in Chapter 3. Moreover, since the situation of continuous fluctuation could be clarified, we could comprehend the tendency between the relations of each parameter.

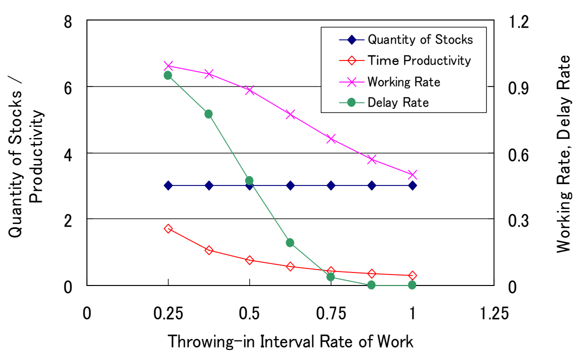


Figure 6: Result of simulation (without parts diversion)

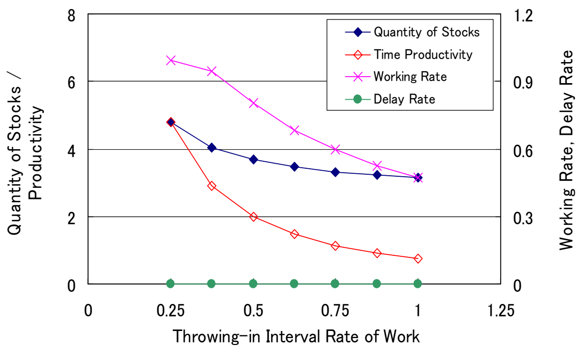


Figure 7: Result of simulation (with parts diversion)

A case with consideration of parts diversion

Figure 7 shows the simulation result in the case with consideration of parts diversion. As a result, it was apparent that as the throwing-in interval rate of work increases, the working rate decreases, and the delay rate is zero without any fluctuation, the time productivity decreases, and the quantity of stocks decreases slightly. Moreover, as compared with the case where part diversion was not performed, when the throwing-in interval of work fluctuated, it was concretely able to be confirmed that the generation of delay was suppressed, although quantity of stocks would fluctuate and the fluctuation of time productivity also increased.

Furthermore, this result of simulation also agrees with the evaluation result as examined in Chapter 3.

According to the above-mentioned result, the simulation of the closed-loop manufacturing system was able to indicate the same result as the evaluation result based on the theory, and to indicate the tendency of fluctuation of relation between parameters further clearly. Moreover, it is thought that we can optimize a manufacturing system by calculating an optimal ratio of set process time against throwing-in interval of work in consideration of the balance of delay time and working rate based on the simulation. However, when examining the optimization of the system using the simulation, it is necessary to consider the influence of extension of throwing-in interval of work and set process time, as in the case of theoretical evaluation.

5 CONCLUSIONS

In this research, the authors examined how to evaluate theoretically the productivity of the closed-loop manufacturing systems which perform maintenance and remanufacturing, and derived the simulation method based on this theory. As a result, we could draw the following conclusions:

a. In the closed-loop manufacturing system, it was possible to evaluate theoretically the relation between set process time, throwing-in interval of work and other parameters such as working rate, delay rate, quantity of stocks and time productivity. Furthermore, it was possible to evaluate the fluctuation of the relation between each parameter in the case with consideration of parts diversion which is the feature of closed-loop manufacturing system.

b. With the simulation based on the theory, it was possible to show clearly evaluation result obtained by theoretical examination, together with the tendency of fluctuation of relation between parameters.

c. By applying the theory and the simulation obtained by this study, it may be possible to optimize the closed-loop manufacturing systems effectively.

6 REFERENCES

[1] The Innovative New Technology Development Joint Research Committee of Maintenance Method, 1998, Consideration about the State of Future Railway Maintenance, The Institute of Electrical Engineers of Japan Electricity Railway Technical Committee Report.

[2] SAKAI, T., 1999, Aircraft Maintenance (Special Issue Life Cycle Maintenance), Journal of the Japan Society of Precision Engineering, 65/3:373-376.

[3] Nozawa, K., 2004, PC 3R and Business, Journal of the Society of Instrument and Control Engineers, 43/5:429-432.

[4] Sahashi, T., 2004, Life Cycle Management to Realize Reconditioned Machine, Journal of the Society of Instrument and Control Engineers, 43/5 :433-436.

[5] Kondo, S., 2004, Issues on a Closed-loop Manufacturing System, Journal of the Society of Instrument and Control Engineers, 43/5:401-406.

[6] Kondo, S. Soma, M. and Umeda, Y., 2002, Simulation of Inverse Manufacturing System, EcoDesign Japan Symposium, 154-157.

[7] Hatanaka, S., Maruyama, T., Narutaki, N. and Oba, F., 2001, Cost-Benefit Analysis of Recycling by Life-Cycle Simulation, Transactions of the Japan Society of Mechanical Engineers, 67/660C:2732-2738.

A study on optimization method with combinatorial auction

-Application to resource allocation problem of re-entrant flow shop-

Toshiya Kaihara[1], Nobutada Fujii[1], Hiroyuki Hasegawa[1], Shinji Kurose[2]

[1] Dept. of Computer Science and Systems Engineering, Graduate School of Engineering, Kobe University, Kobe, Japan

[2] Renesas Technology Corporation, Ibaraki, Japan

Abstract

In re-entrant flow shop, an optimal resource allocation becomes more difficult as the number of product types, equipment, and repetition of processes increases. Therefore, there is need to consider a new robust method to resolve this resource allocation problem. In this paper, we propose an optimization method that uses combinatorial auction to search solution space efficiently. The proposed method is a high-speed method based on a bidder's utility to solve the resource allocation problem in practical calculating time. Here, the utility means the levels on degree of content obtained by bids and is used to make the neighborhood efficiently in the combinatorial auction. We have applied this method to two objective problems and have obtained quasi-optimal solutions. Computation time of a model has reduced from 56 to 77 % and that of the other has reduced 91 %. This indicates that a combinatorial auction with utility is efficient method to search solution space.

Keywords:

Combinatorial Auction; Re-entrant flow shop; Resource allocation

1 INTRODUCTION

Effective production has become increasingly important in order to satisfy diversifying needs of consumers and reduce unnecessary costs. This is due to intensive competition among enterprises. Enterprise production systems include flow shops, job shops, re-entrant flow shops, and so on. In this paper, we take particular note of re-entrant flow shop because re-entrant flow shop is the most complex of these shops. Re-entrant flow shop is a very special type of flow shop. This flow shop processes a variety of jobs that are identical from a processing point of view. All jobs have the same routing in the facilities of the shop and require the same amount of processing time at each facility; individual jobs, though, may differ since they may have different tasks performed on them at a particular facility. Re-entrant means, all jobs go through the same facility several times as different facilities [1]. In re-entrant flow shop, an optimal resource allocation becomes more difficult as the number of product types, equipment, and repetition of processes increases. Therefore, there is need to consider a new robust method to resolve this resource allocation problem. Such a shop can be found in a semiconductor wafer fabrication and certain manufacturing production systems [2].

Combinatorial auction method gets the attention in various fields [3]. Combinatorial auction is one of the efficient resource allocation methods. In combinatorial auction, sellers market multiple goods for buyers. And bidders can tender for combinations of those goods. By using this combinatorial auction, we can consider complementarity and substitutability and allocate resources effectively. Complementarity means that values of goods increase their value by combining them as against the summation of their separate values. On the contrary, substitutability means that values of goods decreases as a result of such combinations. In this paper, due to the feature of the combinatorial auction which can handle many resources simultaneously by tendering a combination of those goods, we propose an optimization method that uses combinatorial auction. The proposed method is a high-speed method based on a bidder's utility to solve the resource allocation problem in practical calculating time. Here, utility means the levels or degree of content obtained by bids and is used to make the neighborhood efficiently in the combinatorial auction. We expect a resource allocation that considers complementarities of goods by using combinatorial auction.

This paper is organized as follows. In section 2, details of a objective system are described. In section 3, how a combinatorial auction applicate the resource allocation problems and proposed method are described. Experiment's detail is described in section 4. The results of several experiments are given in section 5. Conclusions are given in section 6.

2 OBJECTIVE SYSTEM

A Re-entrant flow-shop is shown in Figure 1. In this figure, each number indicates process, and several machines are in all processes. We target one process which all jobs must pass, and consider the one process re-entrant flow-shop problem. Jobs are considered resources and allocated to machines. This problem is a resource allocation problem. Two models are considered in the objective system. One is that the objective function is to minimize the number of holders under a machine bottleneck. The other is that the objective function is to maximize the number of the utility without a machine bottleneck. A set of holders is decided as the process of the job. Two models are shown in the next subsections.

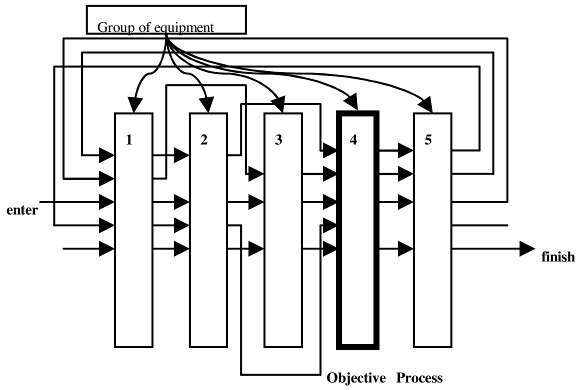


Figure 1: Re-entrant flow-shop.

2.1 Model1

A machine restriction means when a job passes the restricted process, the job must pass the same machine passed in the previous restricted process. In this model, processes and machines are divided shown below (Figure 2 is fully discussed in Model1)

Restricted process

When the job fails to pass the same machine passed in the previous restricted process, the quality of the job dwindles.

Normal process

Even if the job does not pass the same machine passed in the previous normal process, the quality of the job is not affected.

Restricted machine

Due to machine's high level process setting, this machine can handle above two types of process.

Normal machine

Due to the low level process setting, this machine can only handle normal processes.

2.2 Model2

The utility means the degree of satisfaction which is obtained by bids. This utility is how many jobs the restrict process fills. In this model, the restricted process means the job should process the restricted machine passed in the previous restricted process as an object. In this model, processes and machines are divided as model 1 (Figure 3 is fully discussed in Model2).

In the next section, we describe the combinatorial auction and how it applies to the objective system.

3 COMBINATORIAL AUCTION

In auctions bidders make bids for assets. An auctioneer collects these bids according to some rule and finally decides who wins the bid. There is an extensive literature on auction design, showing how different sets of rules affect the outcome of auctions and what strategies should be followed by the bidders [4].

We are interested in combinatorial auctions. In this type of auction bidders make bids for subsets of assets. The rationale behind this is that a bidder's valuation for a subset may not be equal to the sum of the valuations of individual assets. Reasons for that could be economies of scale or

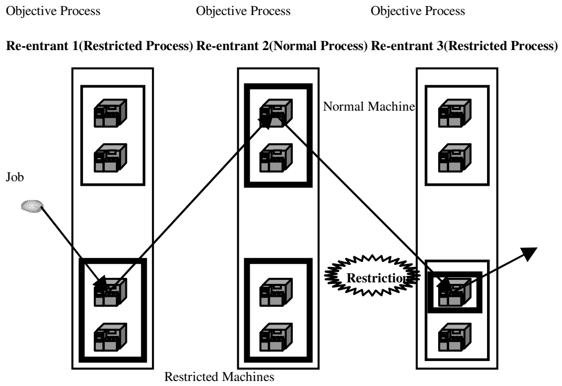


Figure 2: The state of model 1.

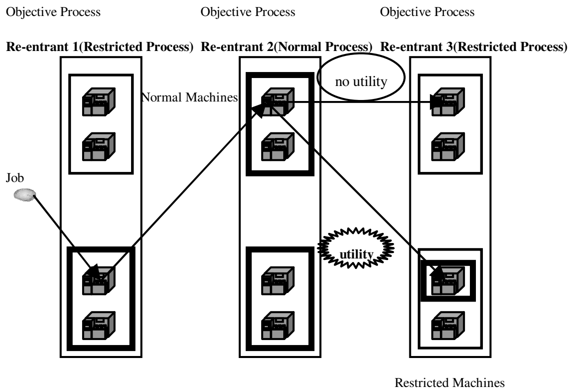


Figure 3: The state of model 2.

scope, i.e., combinations of bids may be more or less valuable than single bid, or winning a bid for a large set of assets would require expensive investments [5].

By using this combinatorial auction, we can consider complementarity. The bottleneck process put into the complementarity of jobs. We expect to allocate resources effectively.

This auction flow is shown below :

[STEP1] A coordinator shows the information of all jobs(the types of job, holder, process, number of processes) for all machines.

[STEP2] Bidders(machine i) make b bids, and bid for combinations of jobs.

[STEP3] The coordinator decides bids which minimize (maximize) the value of the objective.

[STEP4] If an end parameter is filled, all jobs are allocated to machines.

[STEP5] The result of the auction shows all machines and return to **[STEP2]**.

The flow of the combinatorial auciton is shown in Figure 4. In Figure 4, making the neighborhood is **[STEP2]**, and searching a quasi-optimal solution is **[STEP3].** We call **[STEP2]** a bid determination problem and **[STEP3]** a winner determination problem. And in **[STEP5]**, the result of the previous auction applies for the next auction. This mechanism is a process of the local search. In this paper, we

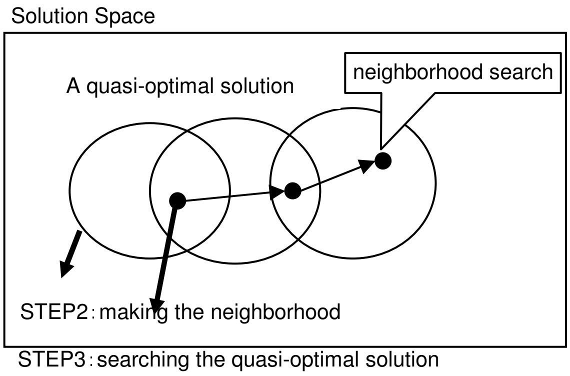


Figure 4: The flow of the combinatorial auction.

make use of utility as a novel approach in the bid determination problem of **[STEP2]**. In the bid determination problem, the restriction and the machine competency are appropriated as we mentioned in the previous section. In a winner determination problem, the best allocation, which is the best combinations of bids, is determined. Two problems are described in the next subsections.

3.1 Notations

The notations used in this paper are as follows :

i : number of machines $(i = 1, \cdots, n)$

d : number of bids $(d = 1, \cdots, b)$

u_d^i : utility in bid d of machine i

t_d^i : the number of holders in bid d of machine i

s_d^i : the number of jobs in bid d of machine i

δ : weight

U : threshold

p_d^i : objective value in bid d of machine i

C : the number of processable jobs

B_d^i : bid set in bid d of machine i. This encompasses the value of bids, job information in bids.

N : the number of normal machines

NB : bid set in normal process jobs

F : the number of bottleneck machines

FB : bid set in bottleneck process jobs efficiently

α : weight

β : weight

$Profit$: profit of allocating bottleneck processes efficiently

j : number of jobs $(j = 1, \cdots, W)$

S_d^i : job set in bid d of machine i

x_d^i : decision variable in bid d of machine i

3.2 Bid determination problem

In combinatorial auctions, when the optimal solution is obtained, bidders need to bid all combinations of goods. However, the solution space becomes larger as the number of bid increases, and the optimal solution can not be obtained in practical calculating time. In order to limit the solution space, bidder's utility is needed. This utility means the level or degree of content obtained by bids and is used to make the neighborhood efficiently. In a bid determination problem, the utility is decided and bidders need to bid under a utility constraint, the machine restriction and a max availability. Two models we described in section 2 are defined as this problem. Two models above are formulated as follows:

Model1

Bidder's utility

This utility is decided as the weighted number of jobs per holder, shown in equation (1). This equation is as follows:

$$u_d^i = \delta \frac{s_d^i}{t_d^i} \tag{1}$$

Utility constraint

The utility obtained by equation (1) so as to make the good neighborhood as the good threshold is used. This is the utility restriction, which is selected by

$$u_d^i \geq U \text{ .} \tag{2}$$

The value of bid

In this model, bidder expresses the value of bid, which is the number of holders. This equation is as follows

$$p_d^i = t_d^i \text{ .} \tag{3}$$

Max availability per a machine

All machines have maximum availability and this is the restriction. Bids are selected by

$$s_d^i \leq C \text{ .} \tag{4}$$

Machine restriction

The restricted process which we mentioned in the previous chapter is given as follows:

$$\forall B_d^{i \in F} \in FB \tag{5}$$

$$\forall B_d^{i \in N} \in NB \tag{6}$$

Equation (5) means restricted machines must bid for jobs contained in the restricted process. Equation (6) means normal machines must bid for jobs except those which are in the restricted process.

Model2

Bidder's utility

These utilities are different from the type of machines.

i) Normal machine

This is the same as equation (1).

ii) Restricted machine

This utility consists of 2 items. The first is the number of jobs per holder, and the second is the number of restricted process efficiently. This equation is as follows:

$$u_d^i = \alpha \frac{s_d^i}{t_d^i} + \beta Profit \tag{7}$$

where *profit* is the number of bid set all restricted process in the job.

Utility constraint

This is the same as equation (2).

The value of bid

In this model, bidder expresses the utility of bid, which is obtained from equation (1) and (7). This equation is as follows:

$$p_d^i = u_d^i. \tag{8}$$

Max availability per machine

This is the same as equation (4).

3.3 Winner determination problem

The bid allocation problem is called "Winner determination problem". This problem is very complex, i.e. , determining the items that each bidder wins is not difficult in the case of non-combinatorial auctions. It would take *O(nm)* time where *n* is the number of bidders and *m* is the number of items. But in the case of combinatorial auctions, this problem turns to be an instance of a weighted set packing problem and is known to be NP-hard [3].

The coordinator solves problem P1, which includes the following objective and constraints (9)-(10)

Objective function

$$\min(\max) \qquad \sum_{i=1}^{n} \sum_{d=1}^{b} p_d^i x_d^i \tag{P1}$$

The objective function minimizes (in model2, maximizes) the total revenue from accepted bids p_d^i , where $x_d^i = 1$ indicates bid i of bidder d has a winning bid.

Subject to:

$$\sum_{i | j \in S_d^i} \sum_{d | j \in S_d^i} x_d^i = 1 \tag{9}$$

Equation (9) guarantees that all jobs must allocate one of machines.

$$\sum_{d=1}^{B_i} x_d^i \le 1 \tag{10}$$

Equation (10) guarantees that machine's winning bid does not exceed 1.

4 EXPERIMENTS

Our experiments consist of simulating the execution of our model under carious conditions to analyze the mechanism's performance. The simulations use CPLEX in a winner bid determination. Simulation experiments are shown below. All the simulation results are in the average of 50 data patterns.

4.1 Experiments parameter

Experimental parameters are set below.

The number of bottleneck processes: 2

The number of normal processes: 3

The number of bottleneck machines F : 2

The number of normal machines N : 3

The type of jobs: 4

The number of holders: 10

Max availability per machines C : 5

The number of bids b : 200

The number of proximity search: 10

4.2 Experiments in two model

Experiment 1(Model 1)

In experiment 1, parameter U in equation (2) is changed, and its performance has obtained. U is increased by $U = U_k = \omega + U_{k-1}$. Variable i is the number of auctions minus 1, and ω is a step size. This value is set to 0.06 in this experiment. The fixed weight and the threshold parameter's patterns are shown below.

$\delta = 0.20$: Maximum value is 1 in equation (1).

$U = U_k = U_{k-1} + 0.06$, $U_0 = 0.00$: The threshold U is variable. The initial parameter is 0.00. The k is number of auctions minus 1, e.g., $U_1 = 0.06$ is the second auction.

Experiment 2(Model 2)

In experiment 2, U is fixed and α, β in equation (7) are changed. The fixed threshold and two weight parameter's patterns are shown below.

$U = 0.30$

$\alpha = 0.20$, $\beta = 0.00$: Maximum value is 1 in equation (7). In this parameter, the maximum value of the first term is 1 and the second term is 0. The first term, i.e., the number of jobs per a holder, is in the maximum weight.

$\alpha = 0.15$, $\beta = 0.25$: The maximum value of first term is max 0.75 and second tern is 0.25.

Table 1: Experimental results in model 1

U	$U_0 = 0.00$	$U_1 = 0.06$	$U_2 = 0.12$	$U_3 = 0.18$	$U_4 = 0.24$
Number of holders	12.44	12.44	12.16	12.14	12.12
Standard Deviation	0.90	0.99	1.01	1.03	1.02
CPLEX time	0.08	0.09	0.15	0.14	0.17
Standard Deviation	0.03	0.03	0.04	0.05	0.04
Efficiency in bottleneck process (%)	100.00	100.00	100.00	100.00	100.00
Standard Deviation	0.00	0.00	0.00	0.00	0.00
U	$U = 0.25$	$U = 0.26$	All Bids	DA	
Number of holders	12.14	No solutions	12.12	12.88	
Standard Deviation	0.42	No solutions	1.02	1.09	
CPLEX time	0.09	No solutions	0.39		
Standard Deviation	0.04	No solutions	0.32		
Efficiency in bottleneck process (%)	100.00	No solutions	100.00	100.00	
Standard Deviation	0.00	No solutions	0.00	0.00	

Table 2: Experimental results in model 2

α	β	0.20	0.00	0.15	0.25	0.10	0.50	0.05	0.80	0.00	1.00	Optimal Solution	DA
Number of holders		9.40		11.12		12.66		13.20		13.46		9.18	9.64
Standard Deviation		0.80		1.17		1.36		1.16		1.21		0.80	0.92
CPLEX time		0.36		0.66		0.49		0.50		0.61		3.85	
Standard Deviation		0.23		0.45		0.27		0.29		0.46		3.26	
Efficiency in bottleneck process (%)		7.00		61.00		86.50		93.50		94.00		6.50	3.00
Standard Deviation		11.33		16.87		12.58		11.07		10.78		12.17	8.20

$\alpha = 0.10$, $\beta = 0.50$: The maximum values α and β are 0.50 and 0.50, respectively.

$\alpha = 0.05$, $\beta = 0.75$: The maximum values α and β are 0.25 and 0.75, respectively.

$\alpha = 0.00$, $\beta = 1.00$: The maximum values α and β are 0.00 and 1.00, respectively. Parameter β is in the maximum weight.

4.3 Two compared methods

A dispatching auction and a combinatorial auction of all bids are used as referenced methods. In the dispatching auction, jobs are allocated in one dispatching rule. The order of holders is used in this experiment. A combinatorial auction of all bids means bidders must bid all combination of goods. The optimal solution is obtained in this method as searching all proximity contented the max availability per a machine.

5 RESULTS AND DISCUSSION

In this section, two experiment results are discussed below.

5.1 Experiment 1

The results from experiment 1 are shown in Table 1. Number of holders means value of the objective function. CPLEX time means calculation time per local search. The maximum value of efficiency in restricted process is 100%. In this model, this value is the constraint and never fails to obtain 100 %. “All Bids” means the optimal solution in the number of holders. DA means the compared method in the normal dispatching auction as we mentioned in the previous section. An analysis of the data in Table 1 indicates that the optimal solution has obtained as the threshold U is 0.24. Compared with DA, quality of solution increases as the utility threshold increases. This could be caused by the fact that the neighborhood has good solution space, which reduces the number of holders, as threshold rises. As threshold is 0.24, the optimal solution is obtained. However, when the threshold exceeds 0.24, bidders could not make bids in bid winner determination problem. This could be caused by the fact that good utility bids are reduced as the threshold becomes larger. Compared with all bids, using the utility restriction has produced quasi-optimal solutions in short calculating time. This auction required a reduction in calculating time from 56% to 77%.

Experiments in model1 using the neighborhood and the utility restriction have produced quasi-optimal solutions and obtained solutions in the shorter calculating time. As additional experiment, U are set to 0.25, 0.26. When U is 0.25, the solution is bad compared with as U is 0.24. And when U is 0.26, no solution exists in winner bid determination problem. The fact indicates that the solution space is too small and the solutions become worse or no solutions exist, when threshold exceeds appropriate value.

5.2 Experiment 2 result in model 2

The results from experiment 2 are shown in Table 2. In this model, the efficiency in restricted process is the utility and the restricted process means the job must be processed by the same restricted machine as in the previous restricted process. An optimal solution means the solution obtained in combinatorial auction of all bids, whose objective function is to minimize the number of holders. This solution is the minimized number of holders in this objective system. An analysis of the data in Table 2 indicates that the number of holders has reduced as weight α becomes larger and larger. As α is 0.2, the number of holders is 9.40. This is the better solution compared with DA and the quasi-optimal solution compared with the optimal solution. This is another aspect of the auction influencing calculating time. The reduction is 91% compared with the optimal solution. Another analysis of the data indicates that the efficiency in restricted process increases as weight of the second term becomes larger and larger. As β is 1.0, the efficiency in restricted process is 94 %. This could be attributed to the fact that neighborhood changes with the weight, and the neighborhood can be made efficiently. The flexible set of the utility can search solutions more efficiently.

6 CONCLUSION

This study has clarified that a combinatorial auction with a bidder's utility can be developed to accommodate the special needs of a traditionally negotiated environment. The mechanism is designed as a resource allocation problem using combinatorial auction. The resource allocation problem is divided into two problems, such as bid determination problem and winner determination problem. Furthermore the objective system is classified into two models. One is that the objective function minimizes the number of holders under machine restriction. The other is that the objective function maximizes the number of the utility without machine restriction. The utility is used to make neighborhood efficiently by sending combined bids. This is decided as bids. Experiment 1 results show a quasi-optimal solution obtained with this utility. Experiment 2 results show various solutions obtained as an objective function maximizes utility and this utility changes. In all patterns, a proposed combinatorial auction is calculated within shorter time than that of searching the entire solution space. Finally, the proposed method in this study has been proved to be effective and timely for an objective system. However, this method can not accommodate other systems and can not consider schedule time. These warrant future study on the method. Nevertheless, the proposed method can be a good method to search solution space efficiently.

7 REFERENCES

[1] S C. Graves, H C. Meal, D. Stefek, A H. Zeghimi, 1983, Scheduling of re-entrant flow shops, Journal of Operations Management, vol.3, issue 4, 197-207.

[2] JC-H. Pan, J-S Chen, 2003, Minimizing makespan in re-entrant permutation flow-shops, Journal of the Operational Research Society, vol.54, 642-653.

[3] Y. Narahari, P. Dayama, 2005, Combinational auctions for elector business, 2005, Sadhana, Vol.30, number 2-3, 179-211

[4] S V. Hoesel, R. Muller, 2001, Optimization in electronic markets: examples in combinatorial auctions, Netnomics, Vol.3, number 1, number 1, 23-33

[5] T. Kaihara, K Sashio, S. Fujii, K. Miura, 2007, A proposal optimization method based on combinatorial auction for resource allocation problem in cell manufacturing systems , Annual Conference of the Institute of Systems, Control, and Information Engineers, Kyoto, JAPAN, May. 2007.

Calibration of Kinematic Parameters of a Robot Using Neural Networks by a Laser Tracking System

Atsushi Kohama[1], Ryosuke Mori[1], Sho Komai[1], Masato Suzuki[1], Seiji Aoyagi[1],
Jun Fujioka[2], and Yoshitsugu Kamiya[3]

[1] Dept. of Mechanical Engineering, Graduate School of Engineering, Kobe University, Kobe, Japan

[2] Dept. of Mechanical Engineering, Faculty of Engineering Science, Kansai University, Osaka, Japan

[3] Dept. of National Science and Technology, Kanazawa University Graduate School, Kanazawa, Japan

Abstract

Almost the industrial robot tasks are performed by the teaching playback method. This method has problem that the laborious and time-consuming online manual teaching is inevitable. Thus, the offline teaching based on the high positioning accuracy is desired. However, a nominal kinematic model does not consider the geometric errors and non-geometric errors. Therefore, some method of calibrating precisely the parameters is required. The kinematic parameters are calibrated by minimizing errors between the measured positions and the predicted ones by nonlinear least square method. After that, the residual errors caused by non-geometric parameters are further reduced by using neural networks.

Keywords:

Robot calibration; Laser tracking system; Neural networks

1 INTRODUCTION

At present state, almost the industrial robot tasks are performed by the teaching playback method, in which the robot repeats the motion taught manually in advance using a teaching pendant, etc. This method is based on the comparatively high repeatability of the robot arm. The problem here is that the laborious and time-consuming online manual teaching is inevitable whenever the specification of the product is changed. It is desirable to teach the task quickly to the robot manipulator when the production line and the production goods are changed.

Considering these circumstances, the offline teaching based on the high positioning accuracy of the robot arm is desired to take the place of the online manual teaching [1]. In offline teaching, the joint angles to achieve the given Cartesian position of the arm's tip are calculated using a kinematic model of the robot arm. However, a nominal geometrically model according to the specification sheet includes the errors arising in manufacturing or assembly. Moreover, it also includes the non-geometric errors, such as gear transmission errors, gear backlashes, arm compliance, etc., which are difficult to be geometrically considered in the kinematic model.

Therefore, some method of calibrating precisely the geometric and non-geometric parameters in a kinematic model is required, in which the three dimensional (3-D) absolute position referring to a world coordinate system should be measured [2-6]. The parameters are obtained so as that the errors between the measured positions and the predicted positions based on the kinematic model are minimized by a computer calculation using a nonlinear least square method.

In the present paper, a laser tracking system was employed for measuring the 3-D position with high accuracy of approximately 5 μm [7-9]. As a target of calibration, a 7-DOF articulated robot (Mitsubishi Heavy Industries, Ltd., type PA10) was employed.

After the geometric parameters are calibrated, the residual errors caused by non-geometric parameters were further reduced by using neural networks, which is the major originality of this study.

2 MEASUREMENT APPARATUS

2.1 Robot arm and position measurement system

In this study, the seven degrees of freedom articulated robot (Mitsubishi Heavy Industries, product name: PA10-7C) is employed as a calibration object.

A laser tracking system (Leica Co. Ltd., product name: SMART310) is used as a position measuring instrument in this study. This system measures the distance by using a laser interferometer and measures the direction by using a tracking mirror, as shown in Figure 1.

A laser beam is emitted and reflected by the tracking mirror, which is installed in the reference point and is rotated around two axes, then, this beam is projected to a retro-reflector called Cat's-eye, which is fixed at the tip of the robot as a target (see Figures 2 and 3). The Cat's-eye consists of two hemispheres of glasses, which have same center and have different radiuses. A laser beam is reflected by the Cat's-eye and returns to the tracking mirror, following the same path as the incidence.

The horizontal and azimuth angle information of laser direction is obtained by optical encoders, which are attached to the two axes of the tracking mirror. The distance information of laser path is obtained by an interferometer. Then, using these information, the position of the center of Cat's-eye, i.e., the position of robot arm's tip, can be calculated with considerably high accuracy (the detail is explained in the following subsection).

2.2 Estimation of measuring performance

According to the specification sheet, the laser tracking system can measure three dimensional (3-D) coordinates with

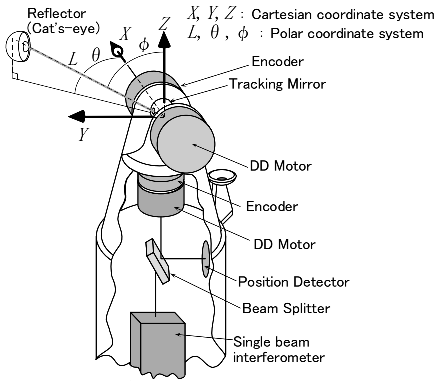


Figure 1: Principle of measurement of SMART310.

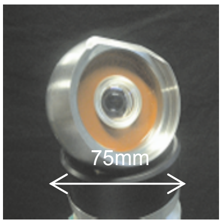


Figure 2: Cat's-eye.

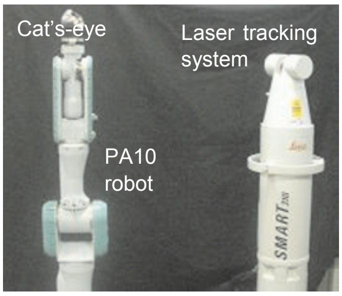


Figure 3: Experimental setup for measuring robot arm's position.

repeatability of ±5 ppm (μm/m) and accuracy of ±10 ppm (μm/m). In this section, these performances are experimentally checked.

First, Cat's-eye was fixed, and static position measurement was carried out to verify the repeatability of the laser tracking system. Figures 4, 5, and 6 show the results of transition of measured *x*, *y*, and *z* coordinate, respectively. Looking at these figures, it is proven that the repeatability is within ±4μm, which does not contradict the above-mentioned specification which the manufacturer claims.

Next, the known distance between two points were measured to verify the accuracy of the laser tracking system. Strictly speaking, the performance estimated here is not the accuracy, but is to be the equivalence. The scale bar, to both ends of which, the Cat's-eye can be fixed, as shown in Figure 7. The distance between two ends is precisely guaranteed to be 800.20 mm. The positions of Cat's-eye fixed at both ends were measured by the laser tracking system, and the distance between two ends was calculated by using the measured data. The results are shown in Figure 8. Concretely, the measurement was done for each end, and the difference between corresponding data in these ends is calculated off-line after the measurement. Looking at this figure, it is proven that the accuracy is within approximately ±10 μm, which does not contradict the specification.

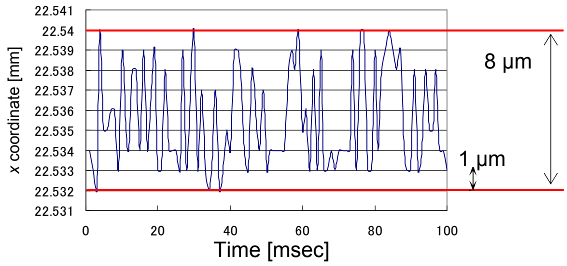


Figure 4: Result of transition of *x* coordinate.

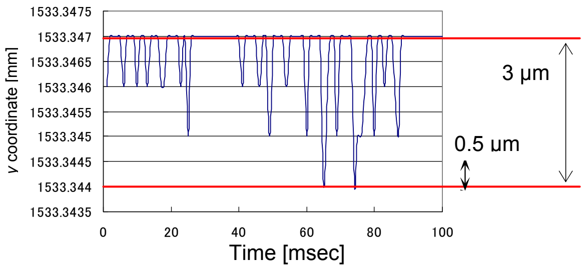


Figure 5: Result of transition of *y* coordinate.

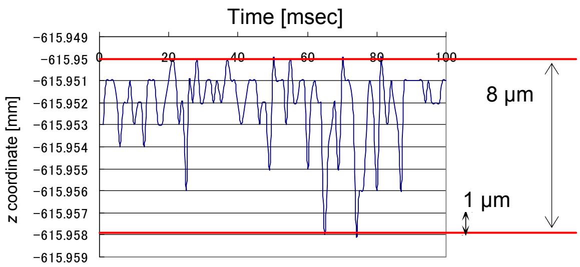


Figure 6: Result of transition of *z* coordinate.

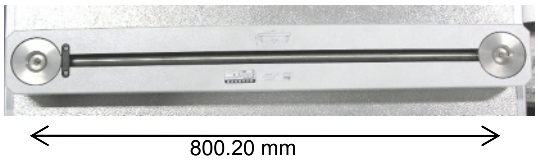


Figure 7: Scale bar.

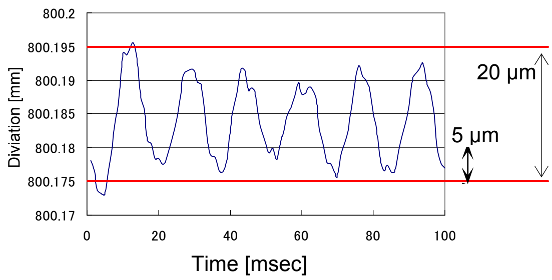


Figure 8: Result of measurement of scale bar.

3 METHOD OF CALIBRATING KINEMATIC PARAMETERS OF ROBOT ARM

3.1 Kinematic model using DH parameter

In this research, the kinematic model of the robot is constructed by using Denabit-Hartenberg (DH) parameters.

The outline of the DH notation is shown in Figure 9. In this modeling method, each axis is defined as Z axis and two common perpendiculars are drawn from Z_{i-1} to Z_i and from Z_i to Z_{i+1}, respectively. The distance and the angle between these two perpendiculars are defined as d_i and θ_i, respectively. The torsional angle between Z_i and Z_{i+1} around X_{i+1} is defined as α_i. The length of the perpendicular between Z_i and Z_{i+1} is defined as a_i. Using these four parameters, the rotational and translational relationship between adjacent two links is defined.

The relationship between two adjacent links can be expressed by a homogeneous coordinate transformation matrix constructed by using above-mentioned four parameters. Nominal values of DH parameters on the basis of a specification sheet are shown in Table 1.

The kinematic model of the relationship between the measurement coordinate system (i.e., SMART310 coordinate system) and the 1st axis coordinate system of the robot is expressed by a homogeneous transformation matrix using 6 parameters (not 4 parameters of DH notation), which are 3 parameters $\theta_r, \theta_p, \theta_y$ for expressing the rotation, and 3 parameters x_0, y_0, z_0 for expressing the translation.

These 6 parameters are estimated as follows: a circular path of the robot arm's tip around the 1st axis is measured by a laser tracking system. Then, the position and orientation of the 1st axis are calculated (estimated).

On the other hand, at the upright configuration, the position of the robot arm's tip is measured by a laser tracking system. The origin of the robot base coordinate system is defined so

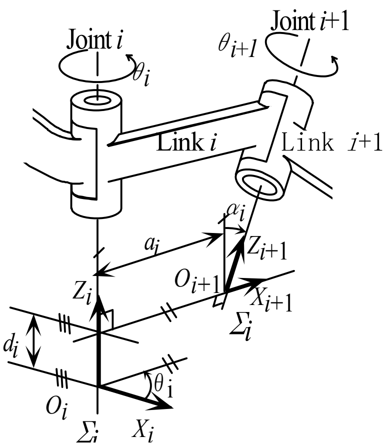


Figure 9: DH notation.

Table 1: DH parameter of PA10.

Joint	θ [deg]	d [mm]	a [mm]	α [deg]
1	0.0	315.0	0.0	-90.0
2	0.0	0.0	0.0	90.0
3	0.0	450.0	0.0	-90.0
4	0.0	0.0	0.0	90.0
5	0.0	500.0	0.0	-90.0
6	0.0	0.0	0.0	90.0
7	0.0	80.0	0.0	0.0

that it is on the estimated 1st axis, and it is apart from the abovementioned upright position by the distance defined by the nominal link lengths. Then, 6 parameters are determined using the direction of the 1st axis and the origin of the robot coordinate system.

The kinematic model from the robot base coordinate system to the 7th joint coordinate system is calculated by the product of homogeneous coordinate transformation matrices, which includes 4×7=28 DH parameters.

As for the relationship between the 7th joint coordinate system and the Cat's-eye coordinate system (i.e., endeffector coordinate system), it can be expressed by using translational 3 parameters x_8, y_8, z_8.

Thus, as the result, the kinematic model of the robot is expressed by using 6+28+3=37 parameters in total, which is as follows:

$$\boldsymbol{P} = (x_0, y_0, z_0, \theta_r, \theta_p, \theta_y, a_1, d_1, \alpha_1, \theta_1, \cdots, a_7, d_7, \alpha_7, \theta_7, x_8, y_8, z_8)^T \quad (1)$$

3.2 Nonlinear least square method for calibrating geometric parameters

The Cat's-eye is attached to the tip of PA10 robot, and it is positioned to various points by the robot, then the 3-D position of the robot arm's tip is measured by the laser tracking system.

The parameters are obtained so as that the errors between the measured positions and the predicted positions based on the kinematic model are minimized by a computer calculation using a nonlinear least square method.

The concrete procedure of calibration is described as follows: Let the joint angles be $\boldsymbol{\Theta} = (\theta_1, \theta_2, \cdots, \theta_7)$, designated Cartesian 3-D position of robot arm's tip be $\boldsymbol{X}_r = (X_r, Y_r, Z_r)$, measured that be $\boldsymbol{X} = (X, Y, Z)$, nominal kinematic parameters based on DH notation be $\boldsymbol{P}_n$ (see Eq. (1)), then, the nominal forward kinematic model based on the specification sheet is expressed as $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{\Theta}, \boldsymbol{P}_n)$.

By using the nominal kinematic model, the joint angle $\boldsymbol{\Theta}_r$ to realize $\boldsymbol{X}_r$ are calculated, i.e., the inverse kinematic is solved, which is expressed in a mathematical form as $\boldsymbol{\Theta}_r = \boldsymbol{f}^{-1}(\boldsymbol{X}_r, \boldsymbol{P}_n)$.

Note that in this calculation, the joint angle of 3rd axis is fixed to 0, to avoid the redundancy of the solution. This calculation is analytically performed, since the inverse kinematics can be solved analytically provided that the adjacent joint axes are accurately parallel or perpendicular to each other: this condition is of course satisfied in the nominal model.

By using the measured data $\boldsymbol{X}$, the kinematic model is calibrated by above-mentioned non-linear least square method. Here, let the calibrated parameter be $\hat{\boldsymbol{P}}$, and the predicted position based on the calibrated model be $\hat{\boldsymbol{X}}$, then the forward kinematic model using them is expressed as $\hat{\boldsymbol{X}} = \boldsymbol{f}(\boldsymbol{\Theta}_r, \hat{\boldsymbol{P}})$. The $\hat{\boldsymbol{P}}$ is obtained so as that the sum of the errors between the measured positions $\boldsymbol{X}$ and the predicted positions $\hat{\boldsymbol{X}}$ is minimized by a nonlinear least square method. These calibration procedures are shown in block diagram form in Figure 10.

3.3 Neural networks for compensating non-geometric errors

After the geometric parameters are calibrated as already explained above, the residual errors caused by non-geometric parameters are further reduced by using neural networks (abbreviated to NN hereinafter) [10].

NN is an algorithm to give a good solution for a non-linear problem. There are non-geometric errors in a real robot mechanism, such as gear transmission errors, gear backlashes, arm compliance, etc. Since they are non-linear, it is difficult to geometrically consider them in the kinematic model. However, by using NN, it may be possible to consider them in the kinematic model.

Typical three layered forward type NN is applied. The input layer is composed of 3 units, which correspond to Cartesian coordinates X, Y, Z . The hidden layer is composed of 100 units. The output layer is composed of 7 units, which correspond to compensation values of 7 joint angles, which is expressed as $\Delta\hat{\boldsymbol{\Theta}}_P = (\Delta\hat{\theta}_1, \Delta\hat{\theta}_2, \cdots, \Delta\hat{\theta}_7)$ and is added to the $\boldsymbol{\Theta}$ parameter in DH model.

In the learning of NN, measured data of robot arm's tip $\boldsymbol{X} = (X, Y, Z)$ is adopted as the input data to NN. Then, the parameter $\Delta\hat{\boldsymbol{\Theta}}_P$ to satisfy $\boldsymbol{X} = \boldsymbol{f}(\hat{\boldsymbol{\Theta}}_r, \hat{\boldsymbol{P}} + \Delta\hat{\boldsymbol{\Theta}}_P)$ is calculated numerically by nonlinear least square method, where $\hat{\boldsymbol{\Theta}}_r = \boldsymbol{f}^{-1}(\boldsymbol{X}_r, \hat{\boldsymbol{P}})$ is the joint angle to realize $\boldsymbol{X}_r$ based on kinematic model using calibrated parameters $\hat{\boldsymbol{P}}$ (see the previous Section 3.2).

Then, the obtained many of pairs of ($\boldsymbol{X}$, $\Delta\hat{\boldsymbol{\Theta}}_P$) are used as the teaching data for NN learning, in which the connecting weights between units, i.e., neurons, are calculated. For this numerical calculation, RPROP algorithm [11], which modifies conventional back-propagation method, is employed. These NN learning procedures are schematically shown in block diagram form in Figure 11.

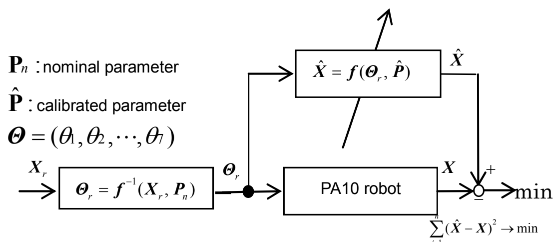


Figure 10: Calibration procedure by non-linear least squre method.

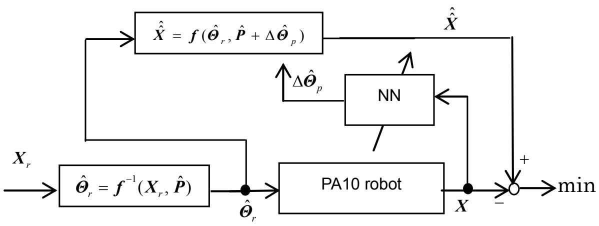


Figure 11: NN learning procedure.

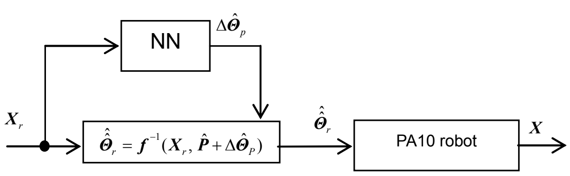


Figure 12: Implementation of NN.

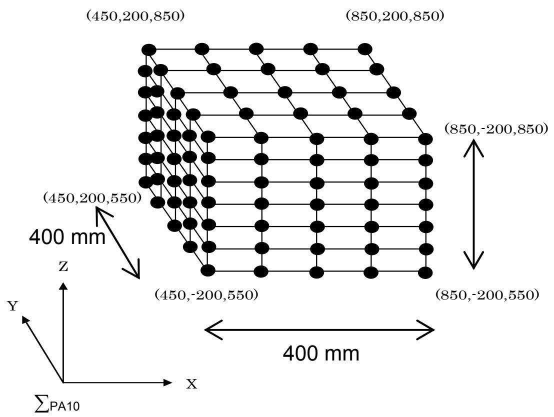
3.4 Implementation of neural networks for practical robot positioning

Figure 12 shows the implementation of NN for practical robot positioning. When $\boldsymbol{X}_r$ is given, the compensation parameter $\Delta\hat{\boldsymbol{\Theta}}_P$ is obtained by using NN, then, the accurate kinematic model includes $\Delta\hat{\boldsymbol{\Theta}}_P$ is constructed. Using this model, the joint angle $\hat{\hat{\boldsymbol{\Theta}}}_r = \boldsymbol{f}^{-1}(\boldsymbol{X}_r, \hat{\boldsymbol{P}} + \Delta\hat{\boldsymbol{\Theta}}_P)$ is calculated numerically, and it is positioned by a robot controller. Then, $\boldsymbol{X}_r$ is ideally realized.

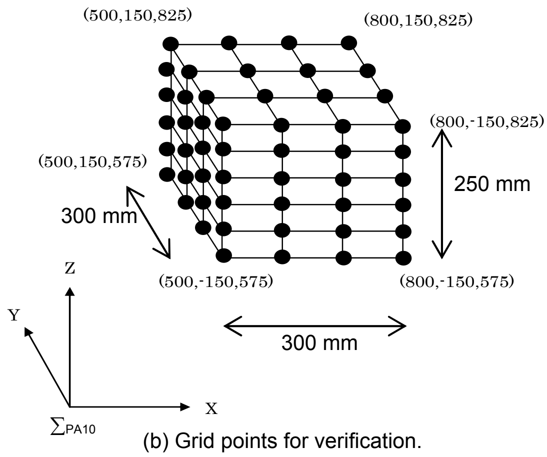
4 EXPERIMENTAL RESULTS OF CALIBRATION

4.1 Group of grid points for teaching and verification

Measurement area of 400×400×300 mm is set, as shown in Figure 13. This space is divided at intervals of 100mm for *x*, *y* coordinate, and that of 50 mm for *z* coordinate. As the result, 96 rectangular parallelepipeds and 5×5×7=175 grid points are



(a) Grid points for teaching.



(b) Grid points for verification.

Figure 13: Measurement area.

generated. The group of the grid points is used for teaching set for calibration.

On the other hand, the group of grid points, each of which is the center of each rectangular parallelepiped, is used for verification set of the calibrated kinematic model.

4.2 Procedure for verification of calibrated model

The joint angles to realize the verification set are calculated based on the calibrated model, and they are positioned by a robot controller. Here, the joint angle of 3rd axis is not fixed to 0, contrary to the case of teaching set.

Note that, this calculation of inverse kinematics is not solved analytically, so it should be numerically solved, since the adjacent joint axes in the calibrated model are no longer accurately parallel or perpendicular to each other.

Then, the Cartesian 3-D positions of the robot arm's tip, i.e., the verification set, are measured by a laser tracking system. By comparing the measured data with the designated data, the validity of the calibrated kinematic model is estimated.

4.3 Results of calibration in grid points for teaching

After the kinematic model is calibrated by using the measured data in the grid points for teaching, the robot arm's tip is again positioned to the same points based on the calibrated model. Here, the extent how the calibrated model can fit to the teaching points, i.e., can realize the positioning to the teaching points, is estimated.

Figure 14 shows the results. It is proven that the error is drastically reduced by calibrating geometric parameters using non-linear least square method. And, it is further reduced by calibrating non-geometric parameters using NN.

As the result, it is confirmed that the calibrated model is effective for improving accuracy at the gird points for teaching. The accuracy is improved from 3.53 mm to 0.76 mm by using non-linear least square method, and further improved to 0.24 mm by also using NN (see Table 2).

4.4 Results of calibration in grid points for verification

The robot arm's tip is positioned to the grid points for verification based on the calibrated model.

Figure 15 shows the results. The accuracy is improved from 3.33 mm to 0.81 mm by using non-linear least square method, and further improved to 0.40 mm by also using NN (see Table 3). Compared to the case for the teaching grid points (see the previous Section 5.3), the realized accuracy is somewhat degraded, especially for the case using NN.

The reason of the error in case using NN here supposedly arises from the limitation of generalization ability of NN, since the grid points for verification are considerably apart from those for teaching. To address this problem, increasing the number of teaching points is one countermeasure for obtaining the more accurate NN model, which is the projected work in the future. Another countermeasure is dividing the robot working space to comparatively small subspaces, and kinematic models corresponding to these subspaces are provided and they are calibrated by using the non-linear lest square method and NN. This countermeasure is also the future work.

5 CONCLUSIONS

A laser tracking system is employed for measuring the robot arm's tip with high accuracy. Using the measured data in grid points for teaching, the kinematic model is calibrated. First, geometric parameters based on DH notation are calibrated by using non-linear least square method. Second, non-geometric parameters arising from gear backlash, arm compliance, etc., is further calibrated by using neural networks. The accuracy is drastically improved from approximately 3.5 mm to 0.8 mm by the former method, and further improved to 0.2 mm by also using NN, in the case of calibrating PA10 robot manipulator in grid points for teaching. And the accuracy of positioning in grid points for verification is drastically improved from approximately 3.3 mm to 0.8 mm by the former method, and further improved to 0.4 mm by also using NN. This means that applying NN after calibration of kinematic model using non-linear least square method is efficient for various points. To conclude, it is confirmed that the proposed method of combining the usual kinematic parameter calibration with the NN has potential for calibrating the non-geometric parameters.

Table 2: Average of errors in grid points for teaching.

Before calibration	3.53 mm
Non-linear least squarw method	0.76 mm
Applying NN	0.24 mm

Table 3: Average of errors in grid points for verification.

Before calibration	3.33 mm
Non-linear least squarw method	0.81 mm
Applying NN	0.40 mm

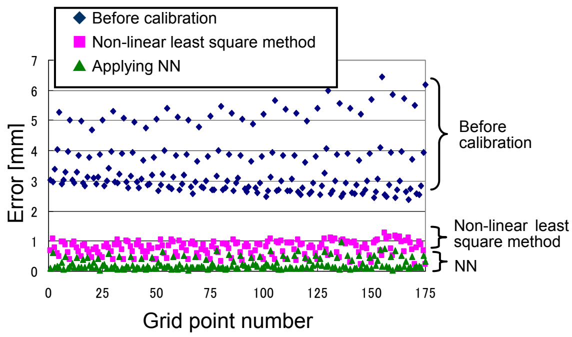


Figure 14: Result of Grid points for teaching.

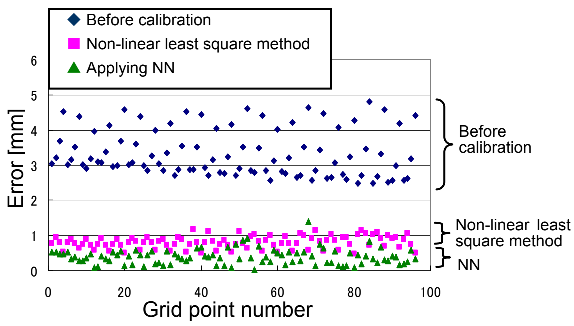


Figure 15: Result of Grid points for verification.

REFERENCES

[1] B.W.Mooring, Z.S.Roth and M.R.Driels: Fundamentals of Manipulator Calibration, Wiley & Sons, 3, (1991)

[2] B.W.Mooring and S.S.Padavala: The Effect of Kinematic Model Complexity on Manipulator Accuracy, Proc. IEEE Int. Conf. Robotics and Automation, (1989) 593-598

[3] D.E.Whitney, C.A.Lozinski and J.M.Rourke: Industrial Robot Forward Calibration Method and Results, J. Dyn. Sdyst. Meas. Contr., 108, 1 (1986) 1-8

[4] R.P.Judd and Al.B.Knasinski: A Technique to Calibrate Industrial Robots with Experimental Verification, IEEE Trans. Robotics and Automation, 6, 1 (1990) 20-30

[5] H.W.Stone: Kinematic Modeling, Identification, and Control of Robotic Manipulators, Kluwer Academic Publishers, Norwell, Mass., 1, (1987)

[6] S.Komai, S.Aoyagi: Calibration of kinematic model with motion capture system by neural networks, Journal of The Japan Society for Precision Engineering, 9(2007) 1151-1152

[7] Y.Koseki, T.Arai, K.Sugimoto, T.Takatsuji and M.Goto: Accuracy Evaluation of Parallel Mechanism Using Laser Tracking Coordinate Measuring System, J. Society of Instrument and Control Engineers, 34, 7 (1998) 726-733

[8] J.Fujioka, S.Aoyagi, K.Ishii, H.Seki and Y.Kamiya: Study on Robot Calibration Using a Laser Tracking System (2nd Report)-Discussion on How to Select Parameters, Number of Measurement and Pose of Measurement in Multiple Positioning Method-, Journal of The Japan Society for Precision Engineering, 67, 4(2001) 676-682

[9] J.Fujioka, S.Aoyagi, H.Seki and Y.Kamiya: Development of Orientation Measuring System of a Robot Using a Gyroscope (2nd Report): Proposal of Position and Orientation Calibration Method of a Robot Using Both Laser Tracking, Journal of the Japan Society for Precision Engineering, 67, 10(2001) 1657-1663

[10] K.Maekawa: Calibration for high accuracy of positioning by Neural Networks, J. Robotics Society of Japan, 13, 7(1995)35-36

[11] M.Riedmiller, H.Braun: A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm, Proceedings of the IEEE International Conference on Neural Networks, 1, 1(1993)586-591

Construction and Evaluation of a Robot Dance System

Kuniya Shinozaki[1], Akitsugu Iwatani[2], Ryohei Nakatsu[3]
[1] School of Science and Technology, Kwansei Gakuin University, Sanda, Japan
[2] Universal Studio Japan, Osaka, Japan
[3] Interactive & Digital Media Institute, National University of Singapore, Singapore

Abstract
Dance is one form of entertainment where physical movement is the key factor. The main reason why robots are experiencing a kind of "boom" is that they have a physical body. We propose a robot dance system that combines these two elements. First, various factors concerning entertainment and dance are studied. Then we propose the dance system by robot using motion unit and the synthetic rule referring the speech synthesis. Also we describe the details of the system by focusing on its software functions. Finally we show the evaluation results of robot dance performances.

Keywords:
Humanoid Robot; Dance; Entertainment; Dance Generation; Text to Speech

1 INTRODUCTION

The research and development of various kinds of robots is actively being carried out, especially in Japan [1][2][3][4][5]. Several reasons explain the current robot boom. One main reason is that robots have physical bodies, and so human-robot interaction extends beyond human-computer interaction.

Although in the future these robots are expected to support various aspects of our daily life, so far their capabilities are very limited. At present, installing such a task in robots remains very difficult. To break through such a situation, entertainment might be a good application area for robots.

Developing a dancing robot would be remarkable from various points of view. First, it might become a new form of entertainment, activates both the body and brain. Watching humans dance is already one established type of entertainment. Second, we might develop a new type of communication with computers, because dance can be considered one of the most sophisticated nonverbal communication methods.

Based on the above considerations we started to research dancing robots. In this paper we clarify the relationship among entertainment, humans, and robots and propose a robot dance system by robot using motion unit and the synthetic rule referring the speech synthesis. Also we will describe an evaluation experiment carried out to test this basic concept's feasibility.

2 DANCE ENTERTAINMENT AND ROBOT

2.1 Entertainment

The role of entertainment in our daily life is very important. It offers relaxation and thus contributes to our mental health. Many aspects concerning entertainment must be considered and discussed [6]. One of the most important may be the existence of two sides: entertainer and audience. Although these two sides change positions depending on the case, the existence of performers and spectators is an absolute prerequisite for entertainment. Many entertainments have both entertainer and spectator characteristics. In the case of dance, people sometimes go to theaters to watch good dance performances, and they sometimes go to dance clubs or discos to dance themselves.

Furthermore, when viewed from a different aspect entertainment can be classified into two types. One is a real-time type that includes performers or entertainers performing live in front of an audience. Good examples include plays and/or concerts. Another is the non-real-time type; reading books and watching movies are good examples.

Following this classification, dance basically belongs to the real-time type of entertainment. For robot dancing, however, as described later, its position is somewhat special.

2.2 Dance Robot

One main reason why we choose dance as an entertainment for robots is that dance is quite sophisticated [7]. Based on the considerations described above, what is the role of robots in dance entertainment? Dance robots allow us to become both entertainers and spectators. When watching a robot dance, we are spectators. On the other side, many people will probably want to install dance motions on their robots and show these actions to others. In this case they are entertainers. For the classification between real-time and non-real-time entertainment, dance robots also have significant characteristics. If we want to show people the robot dance, we have to install the dance actions beforehand, meaning that the robot dance is non-real-time entertainment. At the same time, by developing interactive capabilities, the robot would show impromptu dancing behaviors. For example, it could change the dance depending on audience requests. Or it could sense the audience mood and could adopt its dancing behaviors to reflect the sensor results. A dance robot could provide flexible entertainment that ranges between real-time and non-real-time entertainment.

3 DANCE ROBOT SYSTEM

3.1 Basic concept

Based on the above considerations we want to develop a system that can generate various dance motions. Since different dance genres exist, it is necessary to restrict dance genres to a specific one. Then the system would generate by selecting several basic dance motions and by concatenating them. This basic idea resembles text-to-speech synthesis (TTS) [8], where by restricting the language to be synthesized and by selecting a basic speech unit, any kind of text described by the language can be generated. The following is the basic concept adopted in TTS:

(1) Speech consists of a concatenation of basic speech units.

(2) Selection of the speech unit is crucial.

(3) Connection of speech units is also crucial.

As basic speech units, various basic units such as phonemes, phoneme pairs, CV (consonant-vowel concatenation), CVC, VCV and so on have been studied [8]. Based on research of the last several decades, phonemes including variations that depend on previous and following phonemes are widely used as speech units. Taking these situations into consideration, the basic concept of dance generation is as follows:

(1) We restrict the generated dance to a specific genre.

(2) All dance motions consist of a concatenation of several basic dance motions.

(3) Deciding what to select dance units as basic dance motions is very important.

(4) Connecting dance units is crucial.

(5) Also it is crucial how to express a dance unit as robot motion.

In the following sections, we answer the above questions.

3.2 Dance genre

For basic dance motions, there are several researches on classic ballet [9]. The classification of ballet motions is based on several leg positions and movements called steps. Although each leg position and step has its own name, basically no rules describe the details of whole body motions. We chose hip-hop as the dance genre because all of its dance steps and motions are classified into several categories, so it is easier to handle the whole body motions of hip-hop than ballet.

3.3 Dance unit

Next we must decide the basic unit for dance motions. As described above, since each hip-hop step/body motion has its own name, it can be selected as a dance unit. However, it is difficult for an amateur to extract them from continuous dance motions. Therefore we collaborated with a professional dancer to simplify the extraction of basic motions from continuous dance motions. In addition, when constructing robot motions based on human motions, we must deform complicated human motions into rather simple robot motions. In this deformation process, a professional dancer's advice is also of great help.

3.4 Concatenation of dance units

The next question is how to connect each motion unit. One method interpolates the last posture of the previous motion and the first posture of the next motion. The difficulty in the case of a dancing robot is how to connect these two motions and prevent the robot from falling down. We introduced a method in which a neutral posture represented by a standing still pose is used as a transition posture between two dance units. In this case developing an algorithm is unnecessary to generate a transitional motion that connects two different motions.

3.5 Realization of robot dance motion

The next issue is transforming human dance motions into the motions of robots. One common method adopts a motion capture system that is used to generate the motion of CG characters. For a robot, however, due to the limitations of the degree of freedom at each joint, directly transforming the motion captured by the system into robot motion does not work well. Research that transforms captured motions into robot motions is described in [10] that treats a Japanese traditional dance whose motions include legs moving slowly and smoothly front/back and left/right instead of dynamically. In this case it is relatively easy to maintain balance. However, hip-hop motions include dynamic body motions, and therefore it is difficult to maintain balance. Taking these situations into considerations, we chose a method where each motion unit extracted from continuous motion is transformed manually.

3.6 System architecture

Based on the above considerations, we constructed the first prototype of a robot dance system, as shown in Fig. 1, that consists of dance unit sequence generation, a dance unit database, and dance unit concatenation.

(1) Dance unit database

A large amount of dance units are stored here; each one corresponds to a basic short dance motion and is expressed as robot motion data.

(2) Dance unit sequence generation

An input data that expresses a dance motion is analyzed and converted into a sequence of dance units by this part. At the present stage, a sequence of dance units is directly used as input data and fed into the system.

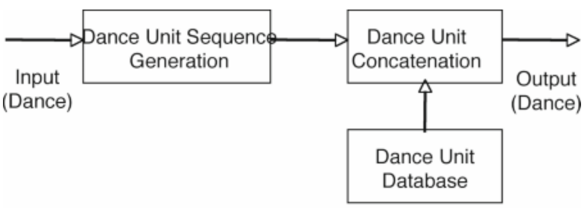


Fig. 1 Structure of dance robot system

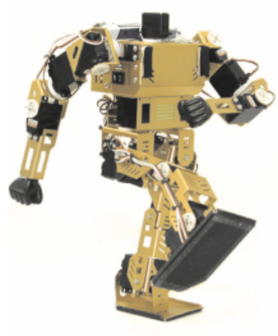


Fig. 2 Humanoid robot

(3) Dance unit concatenation

As is described in 3.4, a neutral posture is introduced as an intermediate posture between two dance units, and therefore, they can be easily connected.

4 SYSTEM DEVELOPMENT AND EVALUATION

4.1 Humanoid Robot

From the several humanoid robots already available on the market, we selected a humanoid robot developed by Nirvana Technology [11] and installed dance motions on it. Figure 2 shows its appearance, and Table 1 shows its basic specifications. Various robot motions can be designed and produced on PC using a “motion editor” realized by motion making and editing software.

Table 1: Specifications of humanoid robot

Size/Weight	34 cm / 1.7 kg
Degree of flexibility	22 (12 legs, 8 arms, 1 waist, 1 head)
CPU	SH2/7047F
Motor	KO PDS-2144, FUTABA S3003, FUTABA S3102, FUTABA S3103
Battery	DC6V

4.2 Development of dance unit database

As described above, we collaborated with a dancer to develop a dance unit database and conducted the following database generation:

(1) First, a typical hip-hop motion of several minutes long was recorded.

(2) Then we observed and discussed the dance sequence and selected about 60 motions as dance units that included almost all the representative hip-hop motions.

(3) We asked the dancer to separately perform each motion corresponding to each dance unit and recorded it. At the same time we asked him to start each dance motion from a “natural standing posture” and to finish in the same posture.

(4) By watching each dance motion being performed, we tried to create a robot dance motion that corresponds to human dance motion using motion editor.

4.3 Evaluation of robot dancing

Using the system described above we carried out simple evaluation experiments.

4.3.1. Comparison of the two types of robot dance units

We evaluated the two types of dance units; one was generated by the professional dancer (type 1) and the other by non-experts (type 2). First we classified all the dance motions into three categories according to the complications of the motions; primary, intermediate, and advanced. And we selected one representative motion for each category. These dance motions are "Lock"(primary), "Rolling Arm" (intermediate), and "Club"(advanced). Then we generated two types of robot dance motions for each of these motions.

Ten subjects were asked to compare these two types of robot dance motions by giving a score ranging from 1 to 5 to each dance motion (1 is the worst and 5 is the best). Figure 3 shows the comparison between the two types of dance motions; robot dance motions developed by the dancer himself (type 1) and those developed by non-experts of dancing (type 2) for three kinds of motions; (a) Lock, (b) Rolling arm, and (c) Crab. Also the live dance motions performed by the dancer is shown as references. Figure 4 shows the evaluation results for each of the three kinds of motions. The evaluation result and the consideration for each motion is described below.

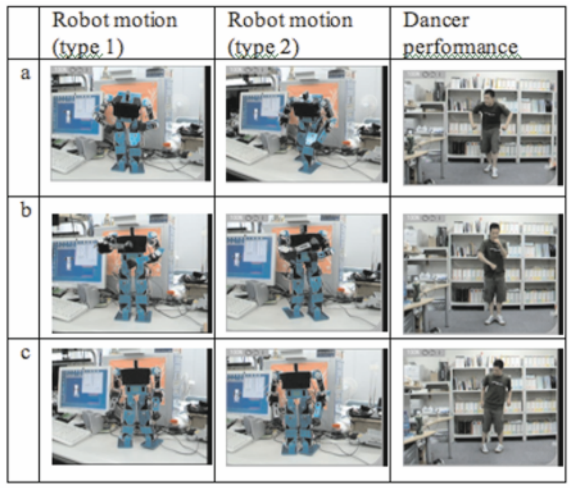
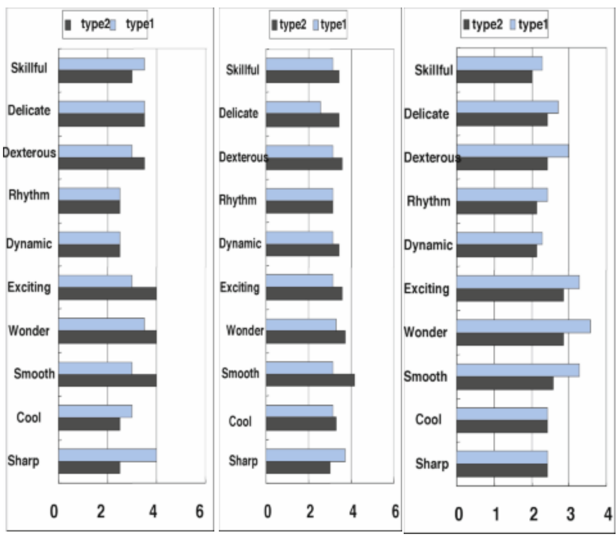


Figure 3: Comparison of three dance motions



(a) Lock (b) Rolling arm (c) Crab

Figure 4: Evaluation results for three kings of motions

(a)Lock

This is a repeating motion of moving the body and stopping it like being locked. In this move the sharpness of stopping motion is an important factor as a dance. For “sharpness,” type 1 motion (motion designed by a professional dancer) obtained the higher score than type 2 (motions designed by non-experts) as expected. On the other hand, for such evaluation items as “exiting,” “wonder”, and “smooth,” the type 2 motion got higher scores than the type 2 motion. It seems that the stop-and-go motion designed by the dancer was judged awkward by the subjects.

(b)Rolling arm

This is a motion of moving body while turning arms smoothly. For the “sharpness,” the type 1 motion obtained higher score than the type 2. But for other evaluation items, the type 2 motions generally got slightly higher scores. Especially for “smooth” type 2 received much higher scores against type 1. Originally this motion contains a step of raising legs, and the type 1 motion precisely simulates this process and in the case of “sharpness” it worked well and obtained the high score. On the other hand, the type 2 motion achieves this move by sliding legs without raising legs. As a result, it was judged that the type 2 motion looked smoother than the type 1, and this gave a influence to the result of smoothness evaluation and others.

(c)Crab

This motion is a move peculiar to the Hip-hop dance. It includes a move of sliding legs sideways without raising them and fixing their backside on floor and thus moving the body sideways. The motion designed by the professional dancer (type 1) receives higher scores than the motion designed by non-expert (type 2) for almost all evaluation items. Especially, important evaluation items for this move such as “exciting,” “wonder,” and “smooth,” the type 1 obtains fairly higher evaluation scores than the type 2.

These result shows that as the robot dance motions become more complex, they can get higher scores. The reason for this would be that the professional dancer understands so well the characteristics of each dance motion and his knowledge and now-how is reflected on the robot dance motion. Even though it does not appear so well in the case of simple motions, this characteristic reveals itself in the case of complicated motions. On the other hand, the motion designed by non-expert (type 2) obtained higher evaluation scores than the type 1 for simple motions. The explanation for this would be that the subjects got good impressions for the over-actions and the unstableness that the type 2 motions generally contain and express themselves. Contrarily, the type 1 motions designed by a professional dancer are sophisticated without containing such over-action nor unstableness. This characteristic sometimes leads to rather low evaluation scores as the subjects are non-expert of dances and thus could not understand the details of the dance motions where the knowledge and now-how of the professional are stored.

4.3.2 Evaluation of the continuous dance motion.

Then we carried out the experiment to evaluate the feasibility of the dance generation system. We compared two types of continuous dance motions. One is a continuous dance motion which is automatically generated by this system and has the length of about one minutes (type3).

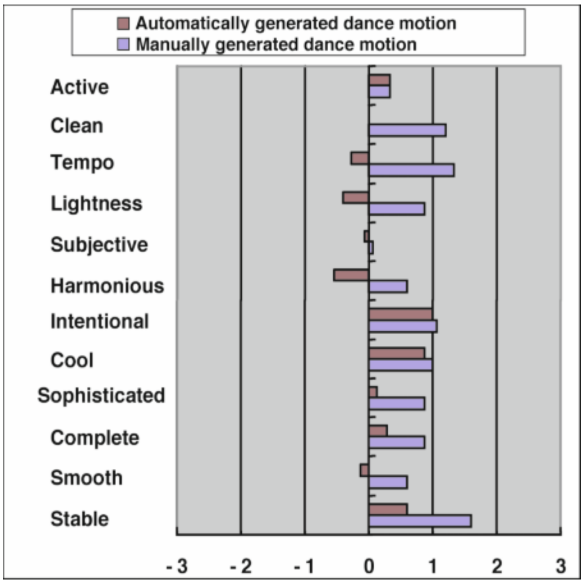


Figure 5: Comparison between automatically generated motions and manually generated motions

Another is the same dance motion where instead of automatic generation the professional dancer designed the whole continuous dance motion from scratch (type 4).

For evaluation twelve items generally used for the sensibility evaluation such as "stable," "soft", “smooth,” and so on were selected. Each evaluation item has a seven level score ranging from -3 to 3. For example, for the evaluation item "stable" the 0 means neutral, 3 means very stable, and -3 is very unstable. Figure 5 shows the evaluation result. The type 4 obtained fairly good results for most of the evaluation items. This means that the evaluation items were fairly well selected. Generally the dance motion generated by this dance generation system (type 3) obtained lower evaluation scores than the type 4 motion. Especially, for such evaluation items as "harmony," "lightness," and "tempo, " the type 3 motion obtained minus evaluation scores. This is because the subject felt unnaturalness due to the neutral posture effect used to connect the two dance units. This means that the system still needs further improvement to generate continuous dance motion, especially for the connection of two dance units. At the same time, however, the type 3 motion got plus scores for "stability", "cool", and "intentional." Especially for "cool" and "intentional" the evaluation results are almost as high as the results of the type 4 motion. This shows that the continuous dance motion generated by this system would be effective as far as it is used as a performance even at the present stage.

The difference between type 3 and type 4 motions are that in the case of type 3 motion it goes back to a neutral position at the point of the dance unit connection. It is necessary to improve this point by introducing better neutral posture or introducing multiple neutral postures.

5 CONCLUSION

In this paper we proposed a dance robot system as a new application area for humanoid robots. We clarified several distinctive entertainment characteristics and investigated the role of robots in entertainment.

Based on these basic considerations we proposed a dance robot system in which a humanoid robot performs various dance motions. We hypothesized that any dance motion consists of a concatenation of short dance motions called dance units. This basic idea was imported from TTS, where any text can be converted into speech by concatenating short basic speech called speech units. Based on this basic idea, we collaborated with a professional dancer. After recording and analyzing his hip-hop dancing, we extracted about sixty dance units and converted them into the motions of a humanoid robot. By concatenating these dance units we found that a huge amount of dance variations for the hip-hop genre could be achieved.

Then we carried out two types of evaluation experiments. First we compared dance motions designed by the professional dancer and the ones by non-experts of dancing. We found that as the dance motions become more complicated and sophisticated, the dance motions by the dancer got higher evaluation results. Then we compared a continuous dance motion automatically generated by this system and one fully manually designed. Although the automatically generated dance got lower evaluation results, for some evaluation items it got almost the same scores. This means that this system is promising from a point of automatic dance generation. Further studies must address the following issues. First we have to investigate how many dance units are enough to generate any type of hip-hop dance. Also we have to investigate the feasibility of a neutral posture that connects two dance units. As only one type of neutral posture was used so far, still there is some unnaturalness for the automatically generated continuous dance motion. We expect that by introducing several other neutral postures, continuous dance motions achieved by the robot would become more natural.

6 REFERENCES

[1] Golubovic, D., Li, B., Hu, H., 2003, A Hybrid Software Platform for Sony AIBO Robots, in RoboCup 2003: Robot Soccer World Cup VII, 478-486.

[2] Ishimura, T., Kato, T., Oda, T., Ohashi, T., 2003, An Open Robot Simulator Environment, in RoboCup 2003: Robot Soccer World Cup VII, 621-627.

[3] Kerepesi, A., Kubinyi, E., Jonsson, G. K., Magnusson, M. S., Kiklosi, A., 2006, Behavioural, Comparison of Human-Animal (Dog) and Human-Robot (AIBO) Interactions, Behavioural Processes, vol. 73, no.1, 92-99.

[4] Wama, T., Higuchi, M., Sakamoto, H., Nakatsu, R., 2004, Realization of Tai-chi Motion Using a Humanoid Robot, in Entertainment Computing, Springer LNCS, 14-19.

[5] http://www.expo2005.or.jp/en/index.html

[6] Callois, R., 1958, Les Jeux et les Hommes, Paris: Callimard.

[7] Laban, R., 1980, The Mastery of Movement, Macdonald and Evans; 4th ed., revised and enlarged edition.

[8] Kleijn, W. B., Paliwal, K. K. (ed.), 1995, Speech Coding and Synthesis, Elsevier.

[9] Lee, C., 2002, Ballet in Western Culture: A History of Its Origins and Evolution, Routledge, London.

[10] Nakaoka, S., Nakazawa, A., Yokoi, K, Hirukawa, H., Ikeuchi, K., 2003, Generating Whole Body Motions for a Biped Humanoid Robot from Captured Human Dances," in IEEE 2003 International Conference on Robotics and Automation.

[11] http://www.nirvana.ne.jp/

Evolutionary Artificial Neural Networks using Extended Minimal Simulation on Evolutionary Robotics

Masanori Goka[1] , Akira Tsumaya[1] , Toshiharu Taura[1]
[1] Dept. of Mechanical Engineering, Graduate School of Engineering, Kobe University, Kobe, Japan

Abstract
In this paper, we try to construct a simulation using evolutionary artificial neural networks models for building the robot controller and adopt an expansion that introduces noise to acquire precise and advanced processing for the robot. In ER approach addressing virtual spaces that contain noise, a simulation technique is called Minimal Simulation. In some experiments, it was shown that the controller was able to be acquired by simulation, yet it does not reach the level that is necessary for the development of a general simulation methodology. Therefore, we propose Extended Minimal Simulation method that introduces alternative coding techniques and show the effect of this technique.

Keywords:
Evolutionary Robotics; Evolutionary Computation; Minimal Simulation

1 INTRODUCTION

In a dynamic environment, it is proposed that a method in which a robot acquires appropriate emergent behaviour for environment by its recognizing surrounding environment through its body and by its interaction with the environment [1]. Contrary to classical artificial intelligence in which the designer makes the controller beforehand, there is embodied cognitive science as a field of research based on such a framework. In embodied cognitive science, there is an important concept of complete agents. In this case, 'agent' refer to an artifact, and complete agents refer to agents that continuously behave by themselves can keep like an animal or a human being. When this concept is applied to the autonomous robot, it is expected that the robot will become independent of human control, continuously interact with the environment and keep behaving appropriately. The robot should have the performance adaptation, autonomy, embodiment and self-sufficient in order to fulfil appropriate behaviour for the concept of complete agents.

Computational intelligence is mainly adopted in order to achieve these functions and construct systems using a bottom-up approach. Computational intelligence is a field of research where functions similar to those of the living things, adaptation, evolution, and learning are artificially realized using a computer. The goal of research In this field is to develop and use techniques similar to those used in artificial neural networks, evolutionary computing and reinforcement learning[4]. When the computational intelligence technique is applied in the development of a robot controller, the machine should learn and search in a vast solution space because a real environment is dynamic and difficult to predict. In this research study, evolutionary computation is adopted because it enables a solution search of the abovementioned type.

In embodied cognitive science, Evolutionary Robotics (ER) is a proposed methodology in which evolutionary computing is used to develop a robot controller. The term ER was advocated by I. Harvey and has been used widely in recent years[2][3]. Currently, artificial neural networks (ANNs) are the main models used in robot controller. Various methods based on ER have been proposed for developing robot controller. It is thought that ANNs are very effective as robot controllers because they are robust with respect noise and their function can be easily extended. Currently, evolutionary artificial neural networks (EANNs) in which evolutionary computation is used to obtain the feature of ANNs are used as robot controllers.

ANNs represent an information processing mechanism, and in those models, many elements of cranial nerves are connected with each other to form network. One of the characteristics of the human cerebral nerve network is that it is capable of nonsymbolistic and parallel distributed processing. Research that involves the application of such information processing, thinking process, and learning process to computer processing has been continuing for a long time. In EANN, it is important to identify which of the element of the ANNs the designer selects as an evolutional object and how the designer codes them to the genotype. As for the element that appears as the phenotype of the ANN, the synapse combination weights, the neuron model, the number of hidden nodes, and synapse formation are considered. Of these elements, the synapse combination weights are generally coded to the genotype, but the research still continues on how to model the other elements and on whether they are coded to match the genotype. Research studies in this area have achieved remarkable development in their respective fields over the last few dozens of years.

However, there are various problems that must be considered in ER. One such problem is the choice of environment in which to evolve the robot controller. In accordance with the concepts in embodiment cognitive science, in ER, It is preferable to evaluate all individuals in the search population using an actual robot in a real environment. However, the evolutional process is extremely time-consuming since evolutionary computing involves a search over a vast space. Therefore, approaches using computer simulation receive attention because of the large reduction in test time, and such approaches are used in many studies. However, in such techniques, how to model a real environment is still unknown. When computer simulation are used, it is relatively easy to model objects that do not change during an experiment (e.g., the field in which a robot is installed and the size of the robot), but it is difficult to model the sensor resolution and motor output accurately because of the influence of noise characteristics and hardware. Therefore, an existing problem is that the behaviour acquired by simulation cannot always be reproduced by an actual robot.

Moreover, when these models are used in simulation, the problem of how to develop an evolutional environment for a robot controller should be considered. In our current study, we have investigated this problem as a benchmark problem. In this case, we have not added any noise to the simulation in order to guarantee that the behaviour acquired by evolutionary computation can be reproduced by the robot. Therefore, it was difficult to investigate the feature of the EANN model because the problem is too simple for evaluating the sophisticated mapping ability of the model. In order to investigate the EANN model, it is necessary to construct a simulation environment that is similar to real environment including some noise. We adopt the minimal simulation technique (N. Jakobi) in this study, though various techniques have been proposed for the development of the computer simulation environment in ER. This technique, which is applied in the development of computer simulations, reflects a general theoretical and methodological framework that was developed from an experiment in which a robot controller developed using computer simulations was successfully used to make an actual robot acquire appropriate behaviour. Further, this technique can be used to develop an environment that can be implemented easily though a simple and inexpensive design. However, when a simulation is developed using this technique, an over-learning problem occurs in the final generation of evolution. Here, one problem is that the genotype that acquires the best fitness in the final generation of the evolutional process learns a peculiar noise characteristic that is specific to the simulation. In conventional minimal simulation, the robot controller is continuously adjusted after it is implemented in an actual robot in order to solve this problem. In this study, a new solution is necessary because only the computer simulation is investigated. Therefore, in this paper, we propose an approach involving extended minimal simulation in, where one genotype learns two or more noise patterns in order to avoid solve over learning. To verify the effectiveness of this extended simulation, we use the problem involving transport by one autonomous mobile robot as a benchmark test.

2 SIMULATION APPROACH

In the development of a robot controller, it is expected that the adopted approach that involve the least computational time and labour in developing the robot controller through computer simulation and implementing this controller in a real robot. However, even the best controller evolved using computer simulation may not perform satisfactorily when implemented in an actual robot because many uncertainties and some noise exist in a real environment. There are some differences between the real environment and the computer simulation that are difficult to model. Hence, it is necessary to invest time and labour in trial and error to model the noise characteristic in detail. The main challenge in the computer simulation approach in ER is to design the general theoretical and methodological framework that enables the development of a compact, cheap simulator with the ability to run at high speeds for the evolution of the behaviour of the robot in a real environment.

2.1 Minimal Simulation

N. Jakobi analyzed the differences between a real and virtual space by conducting some experiments in which a robot controller was evolved using computer simulations and succeeded in implementing an actual robot [5][6][7][8]. He developed the general theoretical and methodological framework in ER for the development of a controller and its implementation in an actual robot. This theoretical framework is called minimal simulation and a very important concept for the inexpensive construction of a compact, high-speed controller is included in this framework.

In minimal simulation, before the simulation is developed, it is necessary to establish the limits that separate the controller and the environment in order to distinguish the feature of the real world that should not be simulated from those that should be accurately and clearly simulated. In general, the division between the controller and the environment is established at the limit where the digital signal of the controller is converted into an analogue output and vice versa. As a result of this division, the controller turns out to be entirely digital. Both an actual controller and the controller developed using computer simulation can be considered systems with similar mechanisms that are executed by a computer. For a particular behaviour, the environment must contain a base set of environmental features so that the robot can interact appropriately with the environment. To decide whether or not the environment is capable of supporting the particular behaviour, we need to know whether it contains a base set of features that the robot can interact with.

The base set with respect to a behaviour will refer to the set of environmental features that must influence each other and interact with the controller's output in an appropriate manner defined by the behaviour. Whether a particular environment is capable of supporting a behaviour depends on whether that environment contains a suitable base set. A controller is said to reliably perform a particular behaviour if the interaction among the members of the base set and that between these members and the controller's output constitutes an instance of the behaviour for every fitness trial.

After understanding the abovementioned statements well, we must then undertake the task of including randomness in the aspects of the controller's input that are part of the implementation. In many cases, it is tempting to simply add a large amount of noise to all aspects that are not base set aspects. However, if this noise is reliable, i.e., developing controllers can operate on the assumption that the noise will always be present, the noise can be used to develop controllers with high fitness. Therefore, the form of the implementation feature of the controller input is varied at random from trial to trial. A method for determining the amount of random variation to apply in the base set feature in the simulation and the best manner in which to apply the variation is sufficient for developing reliable and appropriate individuals that are base-set robust. In order to favour the selection of controllers that can cope with slightly differing versions of each base set aspect of a simulation, a set of simulation noises need to varied between trials and not during trials. Noise levels should be varied among trials along with all the other base set aspects of a simulation. Noise levels should not be left constant at unrealistic levels throughout evolutionary process.

2.2 Extended Minimal Simulation

In this study, a simulator in which the framework of minimal simulation is used will be constructed in order to develop an evolutionary controller that is robust and widely applicable in environments with noise. The following features of the improved simulation are presented. In this case, it is necessary to consider both the features of the base set and the features of the implementation in order to study the problem of transport by the autonomous mobile robot. The cognitive ability to recognize other robots and the object that should be transported and the cognitive ability to determine absolute direction are features of the base set. Sensor inputs, motor outputs and an initial position and direction of the robot and of the object are implementation features; noise is added to these features.

2.3 Introduction of Noises and Guaranteeing Replication

In our existing simulation, when the obtained controller was set up in the same condition as in the simulation, the controller operation was guaranteed to be the same as that in the simulation since no noise was introduced in the environment. However, after the noise is introduced, the guarantee in which the task can be performed by creating an identical environment for operation might be lost because of the uncertainties in the environment. Therefore, to verify the robustness of the controller, its ability to reliably complete task is determined by operating the robot in the environment multiple times in the presence of different forms of the noise and calculating the average number of tasks completed.

Further, the completion of the task in a fixed generation is not guaranteed even if the experiment has very similar settings to those in the simulation because evolutional computing is essentially a heuristic search technique. Therefore, in the experiment without any noises, we tested multiple evolutionary processes with the same settings and determined the ability reliably complete the task by calculating the ratio of the task in which the robot obtained the controller that finally completed a task. In our improved technique, the existing for determining reliability is not important because each robot controller already has already achieved the task completion rate. Therefore, after having performed a trial several times under identical condition, the task completion ability is statistically determined by calculating the arithmetic average and the standard error of the mean of the task completion rate. In other words, the arithmetic average of the task completion rate $\bar{x}$ is shown in (1), where x_i is the task completion rate of the controller i, and n is the number of the repeated trials.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

In this case, the variance is given by (2).

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(\bar{x} - x_i) \tag{2}$$

The standard error of the mean is given by (3).

$$SEM = \frac{\sigma}{\sqrt{n}} \tag{3}$$

Minimal simulation is a simulation technique that enables the straightforward development of a robust evolutional robot controller; however, it is difficult to sufficiently reduce the difference between computer simulation and real experiments even when minimal simulation is used. In experiments where ANNs were evolved for a Khepera robot using minimal simulation [7], more than 20 generations had to be evolved in an actual robot after having obtained the robot controller by evolutionary computing. In other research studies, the large degradation in the performance of the robot caused even by slight change in the environment is because of the over learning, where a robot's behaviour is specialized for the simulation environment [9]. To address this issue, we considered factors leading to the difference between the robustness of the controller in the simulation and that in a real environment. Consequently, we hypothesized that this phenomenon is primarily due to the problem of the over learning and the behaviour of the robot controller is specific to the noise characteristics in final generation. To verify this hypothesis, we modify the conventional method of evaluating fitness and propose an evaluation technique in which one individual is not influenced by the specific characteristics of the specific noise. In the proposed technique, two or more evaluations with different noise characteristics are carried out for one individual, as shown in Figure 2; however, only one evaluation was carried out for one individual in the conventional fitness evaluation, as shown in Figure 1. In this evaluation technique, the value of fitness *F* of one individual is calculated as the summation over *N* evaluations, as shown in the following (4).

$$F = \frac{1}{n}\sum_{i=1}^{n} f_i \tag{4}$$

To increase fitness *F* of the individual, the evaluation should be high for the characteristics of any noises.

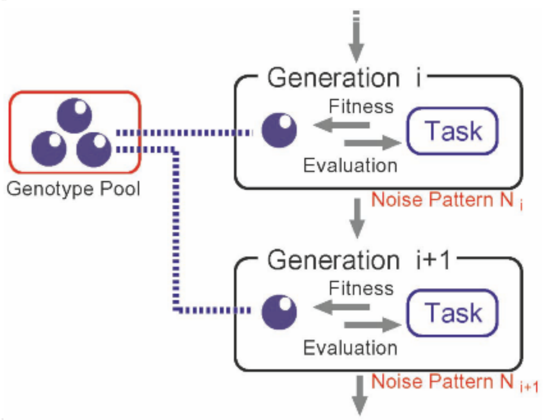


Figure 1: Established fitness coding method.

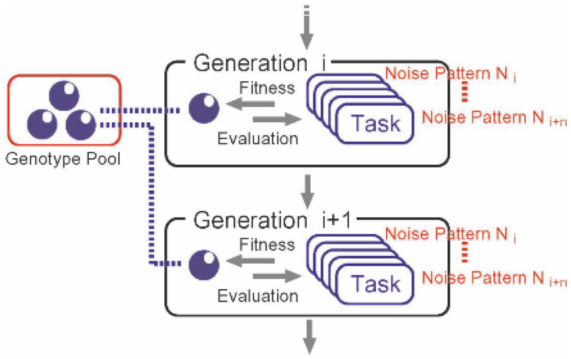


Figure 2: Proposed fitness coding method.

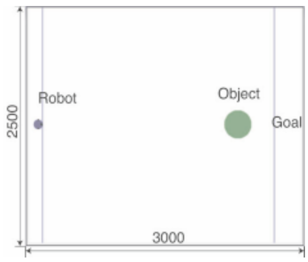


Figure 3: Transfer problem with a single autonomous robot.

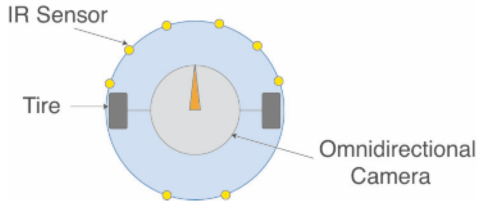


Figure 4: Simulated robot.

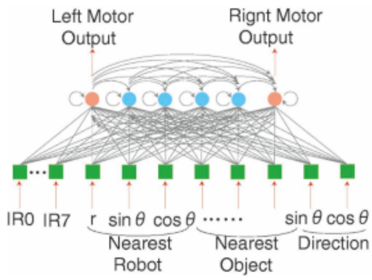


Figure 5: Topology of ANN.

3 EXPERIMENTAL SETTINGS

In this study, the problem of transportation by one robot is treated as a task. As shown in Figure 3, one robot and an object to be transported are placed in the field with the experimental environment; the object can be pushed by the single robot. If this robot pushes the object to the goal line at the end of the field, it is said to have completed the task. Figure 4 and 5 show the structure of the robot and the topology of ANN that is used as a robot controller, respectively. Further, the setting of the fitness function is shown in table 2; trials are performed until the 500th generation. A computer with Xeon 2.4-GHz dual CPU is used for all computer simulation. After evolution is over, we obtain the robot in the final generation; the task completion rate is evaluated by repeating the experiment 25 times in the environment with noise that is shown in table 1, and each task completion is assessed statistically.

In these experiments, we introduce the noise at the initial position of the object in the environment, sensor inputs, motor outputs and in the initial direction of the robot. The random noise has a maximum value equal to the one-fourth the diameter of the robot, the initial direction of the robot is assumed to be random. We added Gaussian noise (N(0, 0.1)) to the sensor inputs and the motor output. We do not know beforehand how strong influence of the magnitude of this noise; therefore, we conduct the experiment by the setting that doubled the maximum on the random noise and the variance of the Gaussian noise. We prepared the noise environment of the five noise patterns shown in table 1 in order to compare the influence of the noise at the initial position or in the initial direction with that of the signal noise.

In extended minimal simulation, the genotype is examined in two or more random noise environment so that it is not a specific to a specific noise pattern, and the summation over the evaluation is assumed to yield the fitness. However, the number of patterns of the noise in the environment that one genotype must learn will be the task-dependence. The trade-off problem occurred because a large number of tests with noise must be performed the robust controller, but the time required for computer simulation is directly proportional to the number of the tests. In this study, experiments are repeated 1, 5, 10, 15 and 20 times, and the correlation between the task completion rate and the execution time is verified.

4 RESULT

4.1 Effect of Minimal Simulation

Table 1: Noise mode.

MODE	1	2	3	4	5
Max initial noise level	0	x1	0	x1	x2
Max sensor noise level	0	0	x1	x1	x2

The experimental results obtained in the environment with the various noises are shown in Figure 6. As shown in Figure 6 (a), in an established technique, the controller that is obtained in the environment without any noises (Mode 1) as can complete the task at the rate of 100 % in the environment where the noise is not added during task repetition, but it can rarely complete the task in the environment with noise. In general, only when there are no uncertainties in a simulator, the robot controller can complete tasks in the real environment exactly as it did in simulations. As shown in Figure 6 (b), the task completion rate of the controller that is obtained in the environment with noises added at an initial position and in an initial direction is high when the controller repeats the task in an environment with a similar noise. Similarly, as shown in Figure 6 (c), the task completion rate of the controller that is evolved under condition where noise is added to the sensor input and the motor output is high when it the controller repeats the task in an environment without any noise or with a similar noise. From these results, it is inferred that the task completion rate of the controller that is evolved under conditions where only the input and output noise is high when it repeats the task in an environment with only input and output noise. These results indicate that the use of noise at the initial position and in the initial direction and the noise in input and output signals is exclusive to evolutionary computing, and both types of noise are necessary to obtain a widely applicable controller for the autonomous mobile robot.

Moreover, the controller which has been evolved in the environment that is added noises to an initial position and direction has the task achievement rate of 20 % in replay in the environment with only the noise in input and output signals, but the controller which has been evolved in the environment with only the noise in signals can hardly achieve the task in the environment in which there are noises in an initial position and direction. In this reason, it is clear that adding the noise to an initial position and direction is strongly related to acquiring the robot controller with high generality.

Figure 6 (d) shows the task achievement rate of the controller which was acquired by the evolutionary computation under adding the initial position noise, initial direction noise and input and output signal noise. This controller decreases the task achievement rate according to an increase in the kind of the noise, but it has about 70 \% task achievement rate even in the case of replay in the environment with each noise. In this task, it can be said that the technique of minimal simulation which simulates the noise environment is effective to obtain the robot controller with high generality.

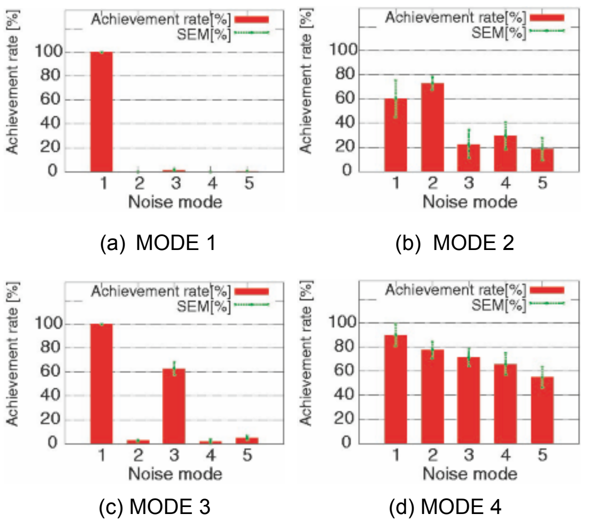
4.2 Effect of Extended Minimal Simulation

The results of extended minimal simulation, where one individual is evaluated five times, are shown in Figure 7. Figure 7 (b), (c) and (d) indicate that the individual evaluated multiple times with different types of noise in the environment is more widely applicable than the individual evaluated only once. The robot controller that is obtained in the environment with noise at the initial position, in the initial direction, and in signals has a task completion rate of 80% or more when tasks are repeated in any environment. However, Figure 7 (a) shows that repeated evaluation in the environment without the noise is ineffective in making controllers widely applicable. Figure 7 shows the relationship between the task completion rate and the number of the evaluations for one individual in the environment with noise at the initial position, in the initial direction, and signal noise. This graph shows that the task completion rate of the individual evaluated only once is less than 70 %, while that of the individual which is evaluated more than four times is more than 90 %; moreover, the task completion rate of the individual evaluated 20 times reaches 100 %. Further, the standard of the mean is greater than 8.8 % when the individual is evaluated only once, while it is less than 2.5 % when the individual is evaluated more than four times. It follows that it is possible to obtain a controller that is not influenced easily by the feature in a particular computer simulation by increasing the number of evaluations. These results show that a robot controller with high reliability is obtained by applying extended minimal simulation, if these trials are identical to the experimental settings.

Further, we consider the number of times one individual should be evaluated. The dashed line in figure 8 shows time required to complete the computer simulation (500th generation) for different number of evaluations. The simulation execution time increases linearly, as shown in figure 8. Therefore, for this task, it is appropriate to evaluate one individual five times. However, we should be note that the appropriate of evaluations number depends on the task. In a task with a different difficulty, it is necessary to investigate beforehand the required number of generations for the evolutionary computation and the relationship between the number of evaluations and the task completion rate. From the abovementioned results, it is found that extended minimal simulation, where one individual is repeatedly evaluated with different type of noise, is effective in overcoming the over-learning problem and obtaining a widely applicable robot controller.

4.3 Effect of Amount of Sample

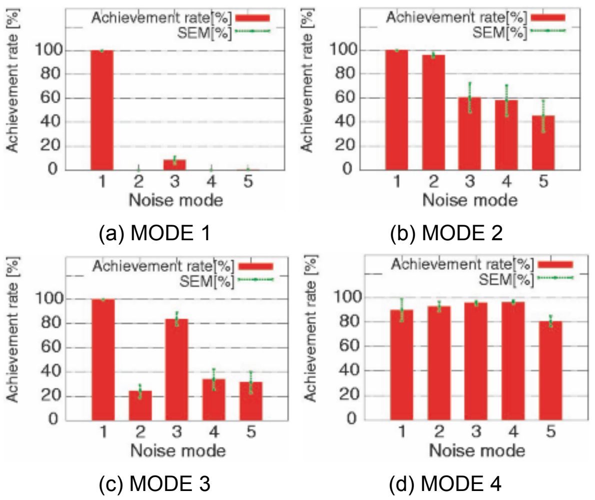
In our current experiments, we performed the computer simulation multiple times with the same experimental settings, obtained two or more robot controllers, and evaluated their performance by evaluating the average fitness level and average task completion rate.



(a) MODE 1 (b) MODE 2

(c) MODE 3 (d) MODE 4

Figure 6: Arithmetic average of task completion rate and SEM.



(a) MODE 1 (b) MODE 2

(c) MODE 3 (d) MODE 4

Figure 7: Arithmetic average of task completion rate and SEM(evaluated five times).

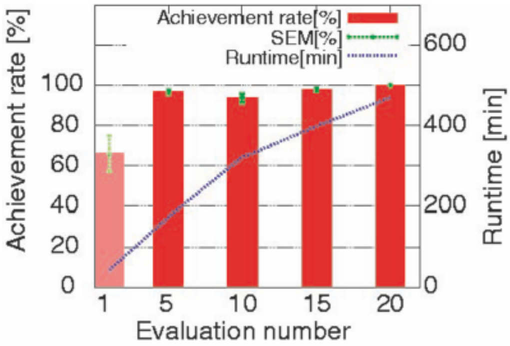


Figure 8: Relationship between coding evaluation number and arithmetic average of task completion rate and SEM.

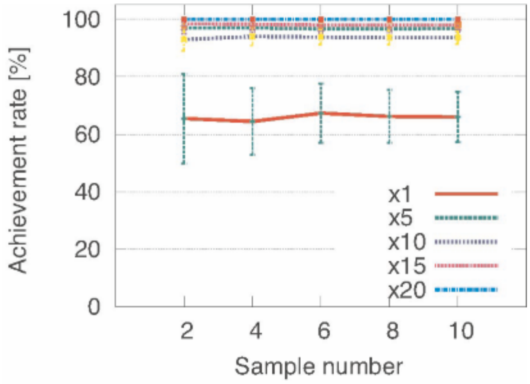


Figure 9: Relationship between coding evaluation number and arithmetic average of task completion rate and SEM.

In this study, to determine and verify the number of samples to be acquired by extended minimal simulation, we conducted experiments in accordance with the following procedures that are based on inferential statistics:

- Perform the computer simulation ten times, and collect ten sample of controllers.
- Select *n* controllers at random from among ten samples.
- Calculate the arithmetic average and the standard error of the mean from the task completion rate of the selected controller.
- Measure the standard error of the mean of the entire trial by repeating step 2 and step 3 hundred times.

Figure 9 shows the results of verification performed in accordance with procedure for *n* = 2, 4, 6, 8 and 10. In the experiment conducted for the verification, the individual is evolved in an environment with noise at the initial position, in the initial direction, and signal noise, and the individual repeats the tasks in an environment with a similar noise. Each dashed line shows the number of evaluations of the fitness of one individual in extended minimal simulation. As shown in the figure, in conventional minimal simulation (red solid line), as the number of samples increase, the task completion rate is almost constant, though the standard error of the mean decreases gradually. However, in extended minimal simulation, it is found that the standard error of the mean is very small, is almost constant, and does not depend on the number of the mean. In particular, we have succeeded in generating a widely applicable homogeneous controller with certainty, even if experiments are not repeated with identical settings, because the standard error of the mean is very small for two samples.

5 CONCLUSION

In this study, we introduced minimal simulation in order to add noise to simulation environments and to develop the environment that required the robot to behave in a careful and advanced manner. Further, we proposed extended minimal simulation in order to overcome the over-learning problem and to realize general versatility of the robot; the incorporation of these features was verified. When conventional and extended minimal simulation were compared experimentally, it was found that extended minimal simulation consistently generated a robot controller that had a task completion rate 30% higher than the corresponding task completion rate in the case of conventional minimal simulation. Further, it was shown that extended minimal simulation can be used to obtain a homogeneous and widely applicable robot controller. We would like to further develop extended minimal simulation by changing the target problem because it is thought that the characteristics depend of this simulation technique depend on the task being performed.

REFERENCES

[1] R. Pfeifer and C. Sheier, “Understanding Intelligence”, The MIT Press, 1999

[2] I. Harvey, P. Husband and D. Cliff, “Issue in Evolutionary Robotics“, Cognitive Science Paper CSRP219, 1999

[3] S. Nolfi and D. Floreano, “Evolutionary Robotics“, The MIT Press, 2000

[4] R. Sutton, “Reinforcement Learning“, The MIT Press, 1997

[5] N. Jakobi, “Half-baked, ad-hoc and Noisy: Minimal Simulation for Evolutionary Robotics“, Fourth European Conference on Artificial Life, pp.348-357, 1997

[6] N. Jakobi, “The Minimal Simulation Approach to Evolutionary Robotics“, Intelligent Robots to Artificial Life (ER ’98), 1998

[7] N. Jakobi, “Evolutionary Robotics and the Radical Envelope of Noise Hypothesis“, Journal of Adaptive Behaviour, 6(2): 325-368, 1998

[8] N. Jakobi, “Running across the Reality Gap : Octopod locomotion evolved in a Minimal Simulation“, Proceeding of the First European Workshop on Evolutionary Robotics : EvoRobot ’98, 1998

[9] O. Miglino, H. H. Lund and D. Nolfi, “Evolving Robots in Simulated and Real Environments“, Artificial Life 2, pp. 417-434, 1995

Table 2: Fitness table.

Transfering an object	Final position of an object	Touching an object	Elapsed time
+1000(x num)	+*x* coordinate	+100(x num)	*-step*

ROBO-BLOCK and Rational Formula of Robots

Hirofumi Niimi[1], Minoru Koike[1], Seiichi Takeuchi[2], and Noriyoshi Douhara[2]

[1]Dept. of Systems Design, College of Industrial Technology, Amagasaki-City, Hyogo-Pref, Japan

[2] Dept. of Mechanical Engineering, College of Industrial Technology, Amagasaki-City, Hyogo-Pref, Japan

Abstract

ROBO-BLOCK (robot block system) and the rational formula of robots were proposed. ROBO-BLOCK is composed of servomotors, the parts for servomotor rotor, the brackets for servomotor fixation, the board parts and the controllers. A robot can be assembled easily by ROBO-BLOCK. Meanwhile, it is convenient when the structure of the robot can be described easily as a character. The whole structure of the robot is expressed as rational formula of the robot to show molecule structure in chemistry. ROBO-BLOCK can be useful for not only the research but also the education. SANDY-1 was developed based on the structure formula of human and skeletal structure and the motion was confirmed.

Keywords:

Design Methodology; Robotics; Mechatronics; Humanoid Robot; ROBO-BLOCK

1 INTRODUCTION

Various robots were researched in the various laboratories. The production of one robot takes a high cost and labor. Therefore, it is difficult to develop many robots in a laboratory. Meanwhile, the main parts of the robot are motors, and how to control the motors decides the performance of the robot. One robot can be made by combining multiple servo motors. A robot can be made easily if multiple servo motor can be combined like a block. A block can be rearranged easily, and a cost and labor can be restrained. So, ROBO-BLOCK (robot block system) which assembled a robot like a block was developed. A trial production robot is made by using these, and a new robot can be developed . Meanwhile, it is convenient when the structure of the robot can be described easily as a character. The whole structure of the robot is expressed as rational formula of the robot to show molecule structure in chemistry. ROBO-BLOCK can be useful for not only the research but also the education. These were applied, and a humanoid robot SANDY-1 was made. In this paper, we propose ROBO-BLOCK, the rational formula of robots, application of ROBO-BLOCK and the structure formulas of human and humanoid.

2 ROBO-BLOCK

The small humanoid robot SANSYRO-1 [1] was developed in our laboratory in 2005. The leg of SANSYRO-1 was designed based on the human arm (Figure 1(a)(b)(c)) . An arm can be made by repeating rotary joint (1) and (2). The design of the part is simple because it is the repetition of the same part. The part of Figure 1 (d) was made after it was confirmed that it could be designed well with 3D-CAD. A leg was assembled like Figure 1 (e), and the part of the body and the arm were made, and SANSYRO-1 was completed like Figure 1(f). SANSYRO-1 is the robot of its weight 1.65kg, the height 427mm and the degree of freedom 24. It was difficult to make it walk because it was too big and heavy to the output torque of the servomotors. So, a small and light weight robot SANSYRO-2 [2] was decided to be developed.

In the development process of the first humanoid robot SANSYRO-1, it became clear that the most parts of the robot was assembled with 3 parts(Servo-motors, C form parts and box parts). So, when SANSYRO-2 was developed, the leg was designed with the combination of these 3 parts(Figure 2). When a leg was designed, we thought about the combination to bend a knee deeply. SANSYRO-2 is the robot

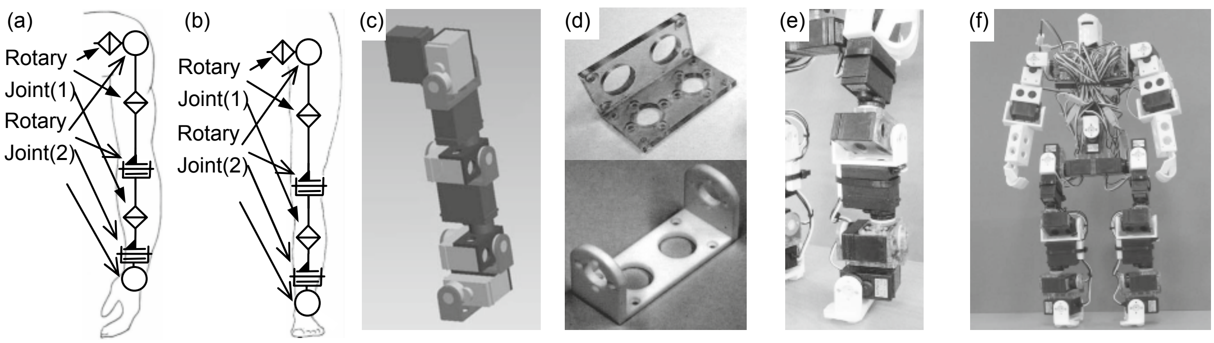


Figure 1: The development process of SANSYRO-1.

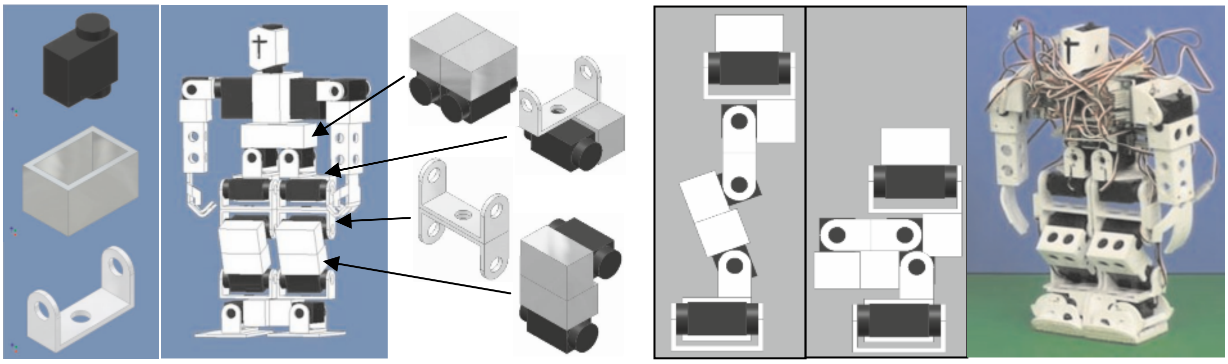


Figure 2: Design of SANSYRO-2.

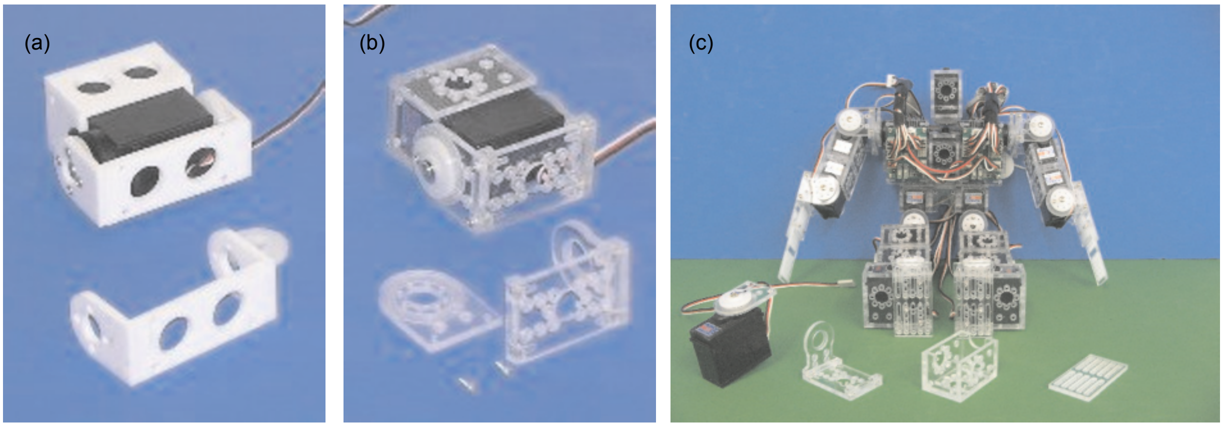


Figure 3: The humanoid by ROBO-BLOCK.

of its weight 1.25kg, the height 300mm and the degree of freedom 16. The experiment of SANSYRO-2 gettingt up was done and it could get up. The walking of SANSYRO-2 was tested, and it could walk. SANSYRO-2 could walk with 100mm/s.

SANSYRO-2 was designed by using the part of Figure 3(a). The part which could be actually assembled was devised newly (Figure 3(b)). Screws are put in the hole of a diameter 2mm, and parts can be fixed. The way to build up like blocks is called ROBO-BLOCK. ROBO-BLOCK is composed of servo motors, the parts for servomotor rotor, the brackets for servomotor fixation, the board parts and the controllers. Characteristics are to make the hole of diameter 2mm in a 10(5) mm pace. They can fix each other by using these holes with screws and nuts of M2. The various forms of robot can be made by using ROBO-BLOCK. The holes are available to fix sensors, exterior parts, and so on.

Humanoid robot can be made by using ROBO-BLOCK (Figure 3(c)). The trial production of the humanoid robot can be made easily with changing the arrangement of the motor. The controllers are the product of KONDO KAGAKU co., Ltd. An advantage is that a motion can be made with moving a robot like a doll. Therefore, even a beginner can make the motion of the robot easily.

The walking experiment of the humanoid robot by using ROBO-BLOCK was done (Figure 4). The center of gravity was moved to the left foot, and the right foot was advanced in front of one step. It confirmed that it could walk though the

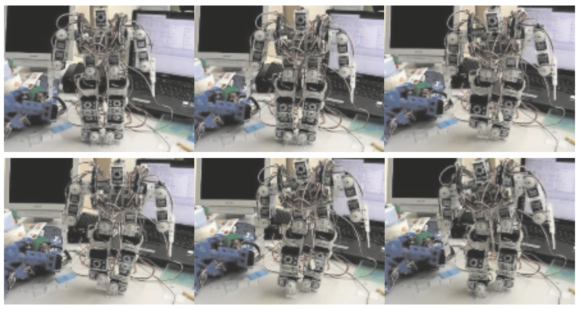


Figure 4: The walking of the humanoid by ROBO-BLOCK.

motion making was difficult because the feet were small. The motion making to stand up from a state to lie was tried, but did not go well. It became a problem that a mobile range of the rotor was small. Because the parts (Figure 3(b)) gave priority to compactness, the mobile range became small. When movement range was wide, it was easier to make the motion of the robot. The part that the mobile range is wide is going to be made in the next version.

3 RATIONAL FORMULAS OF ROBOTS

The method to show structure of a robot with columns is used well. It is convenient when the structure of the robot can be expressed like a chemical formula. So, the rational formula of robots was proposed. The axis direction of the motor which composes a robot is classified by Yaw axis, Roll axis

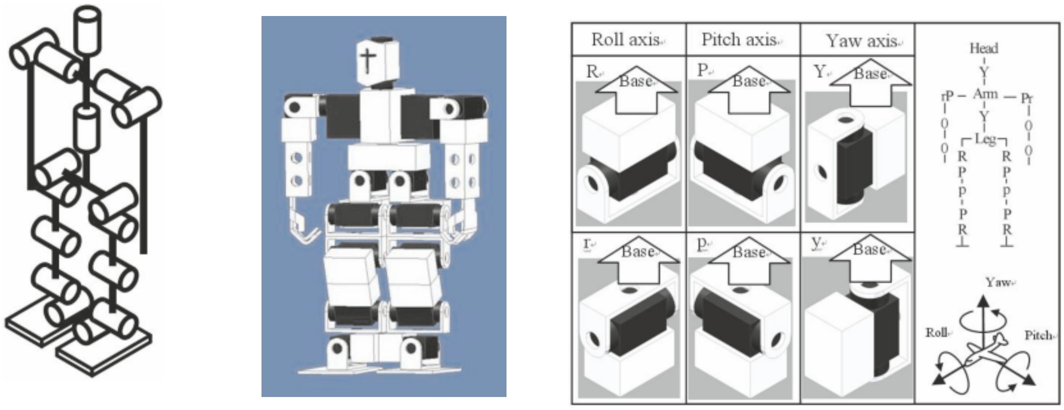


Figure 5: Humanoid -16DOF=Head-Y-(Arm=Pr-0-0)$_2$-Y-(Leg=RP-p-PR)$_2$.

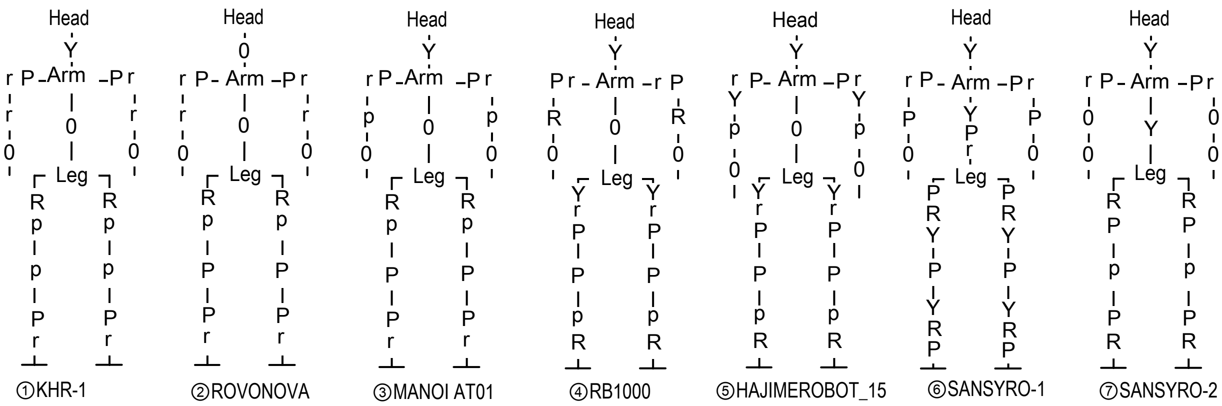


Figure 6: The structure formulas of robots.

Table 1: The rational formula of robots.

No.	Name	Rational Formula of Robots.
①	KHR-1	Humanoid-17DOF=Head-Y-(Arm=Pr-r-0)$_2$-0-(Leg=Rp-p-Pr)$_2$
②	ROBONOVA	Humanoid -16DOF=Head -0-(Arm=Pr-r-0)$_2$-0-(Leg=Rp-P-Pr)$_2$
③	MANOI AT01	Humanoid -17DOF=Head-Y-(Arm=Pr-p-0)$_2$-0-(Leg=Rp-P-Pr)$_2$
④	RB1000	Humanoid -19DOF=Head-Y-(Arm=rP-R-0)$_2$-0-(Leg=YrP-P-pR)$_2$
⑤	HAJIMEROBOT_15	Humanoid -21DOF =Head-Y-(Arm=PrY-p-0)$_2$-0-(Leg=YrP-P-pR)$_2$
⑥	SANSYRO-1	Humanoid -24DOF =Head-Y-(Arm=Pr-P-0)$_2$-YPr- (Leg=PRY-P-YPR)$_2$
⑦	SANSYRO-2	Humanoid -16DOF =Head-Y-(Arm=Pr-0-0)$_2$-Y-(Leg=RP-p-PR)$_2$

and Pitch axis. When there is a motor on the body side, the initial of axis is written with a capital letter, and written with a small letter in case of opposition. The leg of the humanoid robot can be written with “Leg=RP-p-PR”. The whole structure of the robot is expressed as rational formula of the robot to show molecule structure in chemistry. The humanoid robot is expressed as a rational formula of the robot with “Humanoid-16DOF=Head-Y-(Arm=Pr-0-0)$_2$-Y-(Leg=RP-p-PR)$_2$”.

The structure formulas and the rational formulas of the robots are collected about the robot which is in the market, and the robot of this laboratory (Figure 6, Table 1). When formulas of the robots are compared, the common points, difference of the robots become distinct. It's natural that the knees of robots are “p” or “P”. The hip joints, the ankle joints and the shoulder joints have “P(p)” and “R(r)”. The characteristics of the robots can be grasped. The shoulder joint is “Pr” with a robot except for PB1000. When a small and capital letter continues like “Pr”, it is shown that it is one unit with two servomotors. It is the characteristic point of RB1000 to make one unit with two servomotors (the hip joints, the ankle joints and the shoulder joints). The neck joint is “Y” with a robot except for ROBONOVA-I. In the case of this expression the legs of RB1000 and HAJIMEROBOT_15 are the same. In the case of HAJIMEROBOT_15, it is a characteristic that the two axes meet perpendicularly. When those characteristics are expressed, the leg is expressed with “Leg=Y(r⊥P)-P-(p⊥R)”.

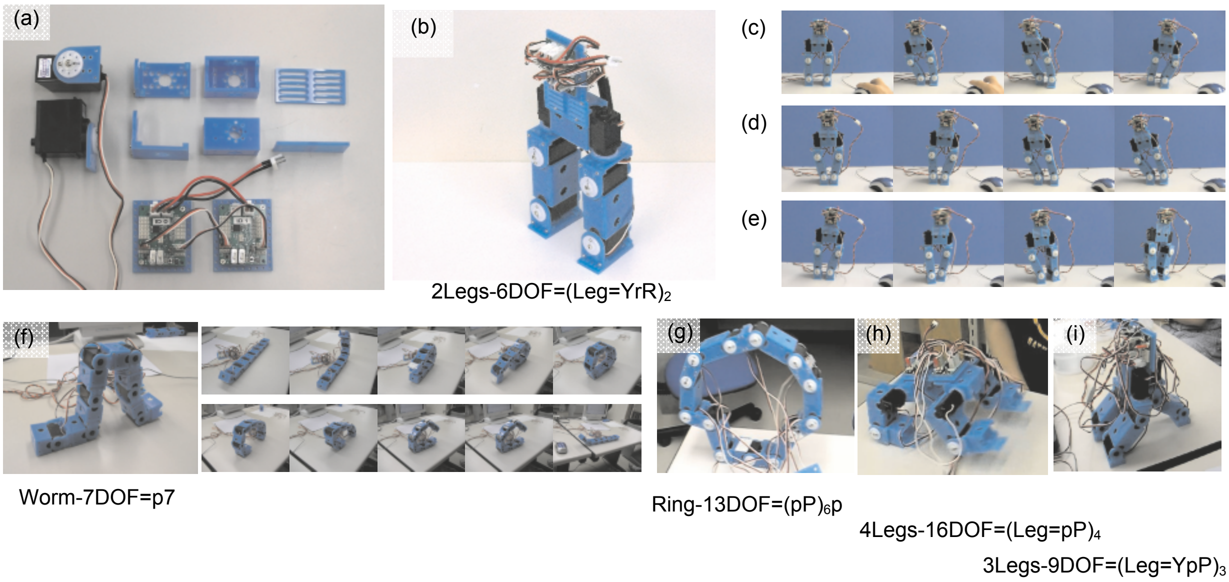


Figure 7: The robots which students assembled.

The characteristics of SANSYRO are to have a Yaw axis. It is characteristics that the robots can twist the body in comparison with other robots.

4 APPLICATION OF ROBO-BLOCK

ROBO-BLOCK can be useful for not only the research but also the education. Creative student experiment was done in the college of industrial technology. Robots were assembled in the group. First, students were assembled biped robots (Figure 7(b)) by using ROBO-BLOCK (Figure 7(a)). Then, the motions to move forward (Figure 7(c)), to move sideway Figure.7(d)) and to turn around (Figure 7(e)) were made. Next, original robots were made by the students (Figure 7(f)-(i)). Some groups assembled a measuring worm robot (Figure 7(f)). The robot was moved with inchworm motion and it possible that the robot turns the whole body and moves, too. Some groups assembled a ring robot (Figure 7(g)). The robot turns the whole body and moves. Some groups assembled robot with four legs(Figure 7(h)). Some groups assembled robot with three legs(Figure 7(i)). Students made original robots and moved them.

The rational formula of the robots was shown in Figure 7. The rational formula of the robots can be described the robot except for humanoid, too. The structure of the robot can be appreciated easily by comparing the rational formulas and the photographs.

5 THE STRUCTURE FORMULAS OF HUMAN AND HUMANOID ROBOT

5.1 The Analysis

A humanoid robot is made based on the human body. The modeling of human body has been studied intensively. We studied the modeling of a human body through the method of the structure formulas and the rational formulas of the robots. The symbols of joint, spine and muscle for the modeling of a human body are listed in Table 2. "B^{j}" is a ball and socket joint. "H^{j}" is a hinge joint. "C^{j}" is a condylar joint. "P^{j}" is a pivot

Table 2: The symbols for the human.

Symbol	Account
b^j	ball and socket joint
h^j	hinge joint
c^j	condylar joint
p^j	pivot joint
s^j	saddle joint
c^v	cervical vertebrae
t^v	thoracic vertebrae
l^v	lumbar vertebrae
M^u	Muscle unit

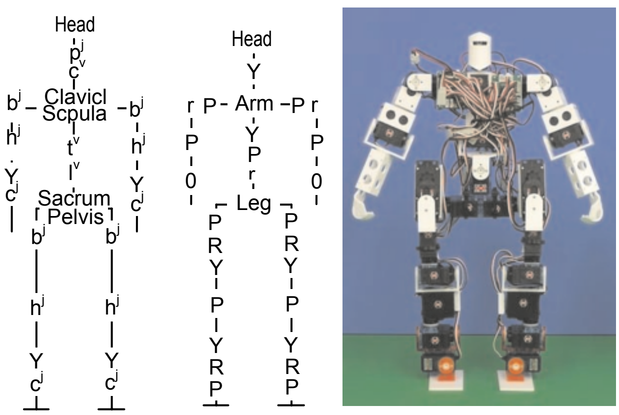


Figure 8: The structure formulas of a human and a humanoid robot.

joint. "S^j" is a saddle joint. "C^v" is cervical vertebrae. "T^v" is thoracic vertebrae. "L^v" is lumbar vertebrae. "M^u" is a muscle unit. Figure 8 shows the simplified structure formula of a human and the structure formula of SANSYRO-1.

The neck of a human has cervical vertebrae and pivot joint. The cervical vertebrae are flexible. So we can tilt the head forward, back and, to the right and the left. And because of the pivot joint, we can shake the head. Many small humanoid robots can shake the head. These are the modelings of the pivot joint that has wide range of movement.

The thoracic vertebrae and the lumbar vertebrae are flexible. So these have wide range of movement for bending and twisting. For the modeling of the thoracic vertebrae and the lumbar vertebrae, SANSYRO-1 had the ranges of movement around Yaw axis, Pitch axis and Roll axis. Considering the position where the range of movement is wide, the thoracic vertebrae was modelled with Yaw axis, boundary line between the thoracic vertebrae and the lumbar vertebrae was modelled with Pitch axis, and the lumbar vertebrae was modelled with Roll axis. SANSYRO-2 has the range of movement around the Yaw axis. And it is useful when SANSYRO-2 made the full body motion as getting up motion and rolling over motion.

The shoulder has a ball and socket joint. The humerus is jointed with the scapula. The scapula is jointed with the clavicle. For them, we can make the shoulders move, as raising, hunching up, turning. The shoulder of human has a ball and socket joint. But the shoulders of many small size humanoid robots are modelled to "Pr". And the shoulder of many humanoid robots is modelled to "PrY(Pry)", too. But that is the structure assembled three axes, not the ball and the socket joint. To bring motion of a robot close to human motion, a spherical motor or the structure that ball and socket joint was combined with muscle units is better than that.

The elbow has a hinge joint. When an arm is stretched and a palm is turned to the front, the above arm and a former arm are not straight lines, and it has an angle in the outside a little. It is good when the arm of the robot is designed in the same way. A former arm can rotate by a radius and a ulna being twisted. This motion can be modelled with a servomotor. The joint of the wrist has a condylar joint, and it can be moved in 2 direction. The range that a wrist is shaken in front and back is wide. The range that a wrist is shaken at the side is small. These motion can be modelled with two servomotor.

A hip joint has a ball and socket joint as well as the shoulder joint. The hip joint of many humanoid robots is modelled to the structure assembled three axes. To emulate a ball and socket joint, there are many robots which have structure of axis that run at right angles to one another, too. Mechanical interference is produced easily because the rotation axes are concentrated on one place. We must devise structures. The hip joints of many small size humanoid robots are modelled to the structures assembled two axes.

The combination of the Yaw axis, the Pitch axis and the Roll axis was discussed for the modeling of the hip joint. Paying attention to a straight walk and a walking on four limbs, the structure of hip joint was considered. The combination of "pry" is taken from both condition to open the legs. Figure 9 shows a hip looking from the oblique back . "Pr" is a hip and "y" is femoral region. The shape of the hips looks like human hips.

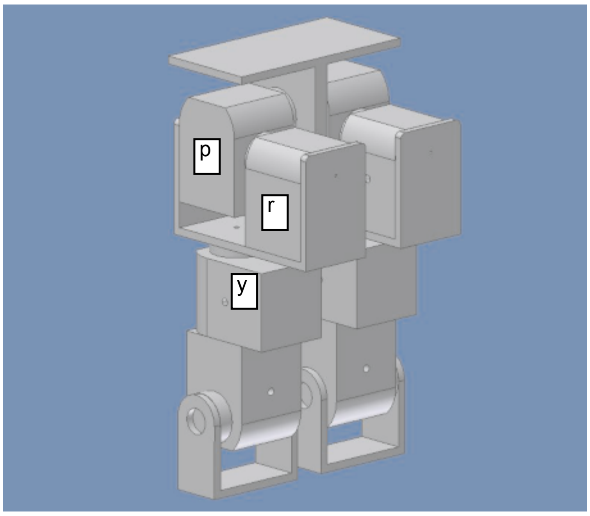


Figure 9: The hip of a humanoid.

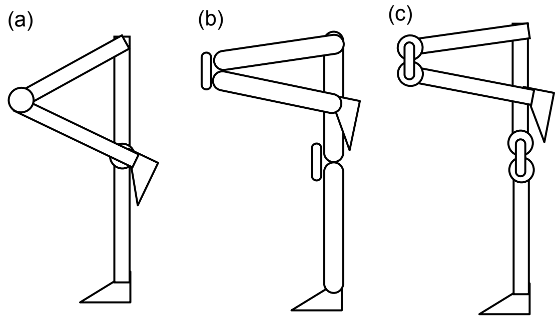


Figure 10: The knees of a humanoid and a human.

A knee joint has a hinge join as well as the elbow. The knee joint of many humanoid robots is modelled to "P" or "p" of one axis(Figure 10(a)). The contact face of the thighbone and the shinbone rolls, and a knee turns(Figure 10(b)). The knee joint can be modelled to two Pitch axes (Figure 10(c)). If the knee joint has been modelled to one axis, mechanical interference is produced easily. The knee joint is modelled to two axes and the knee can be bend deeply. And, large torque is necessary for the knee joint. Two servomotors are sometimes used to solve this problem.

Because movement range is small, there are many cases that the rotation of the lower thigh is omitted. The ankle of many humanoid robots is modelled to "Pr". The distance from the Pitch axis of the ankle to the reverse side of the foot gets comparatively long in case of small size humanoid. When it is compared to the human, a humanoid is the condition that it wears platform shoes, and therefore the humanoid can't move quickly any more. Because the distance between the Roll axeis of the left and the right ankles is long, it becomes the condition that humanoid is putting on snowshoes and walk. A cam was used in SANSYRO-2 to bring the Roll axes of the ankles close to each other. And it was easy to move the center of gravity to the right and the left.

5.2 The Humnoid Robot

New humanoid robot named "SANDY-1" was developed based on the above analysis and skeletal structure. The Pitch axis was assembled making angles of 30 degrees around the

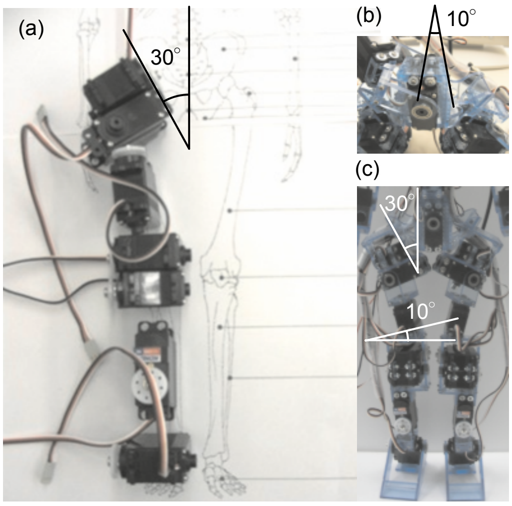


Figure 11: Design of the legs.

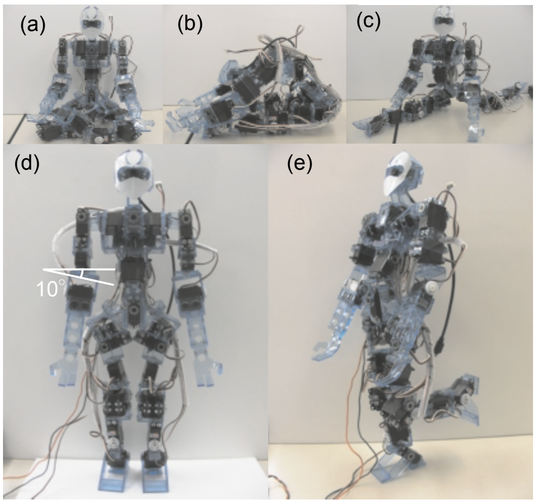


Figure 12: SANDY-1.

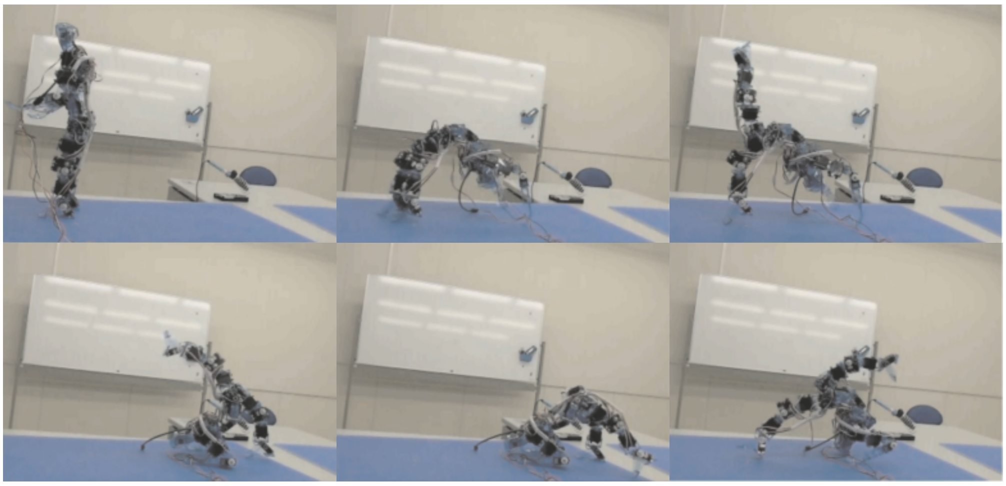


Figure 13: Backward rotation.

Yaw axis and angle of 10 degrees around the Roll axis(Figure11(a)). The lower thigh gave the angle of 10 degrees with the femoral region. SANDY-1 is the robot of its weight 2.4kg, the height 570mm and the degrees of freedom 23. SANDY-1 was made to take various poses to confirm movement range(Figure 12(a)-(c)). SANDY-1 can seat cross-legged because the leg of it is thin and the movement range is wide(Figure 12(a)). SANDY-1 can bend forward using hip joint and Pitch axis of the boundary line between the thoracic vertebrae and the lumbar vertebrae(Figure 12(b)). SANDY-1 can open the legs like a gymnast(Figure 12(c)). SANDY-1 was powered on, and it confirmed that SANDY-1 stood in both legs and the one leg(Figure 12(d)(e)). SANDY-1 was made to do back rotation like a gymnast(Figure 13). Pitch axis of the boundary line between the thoracic vertebrae and the lumbar vertebrae was useful for this motion.

6 SUMMARY

ROBO-BLOCK (robot block system) and the rational formula of robots were proposed. A humanoid robot can be assembled easily by ROBO-BLOCK. And another type of robot can be assembled easily by ROBO-BLOCK. Meanwhile, the whole structure of the humanoid robot is expressed as rational formula of the robot to show molecule structure in chemistry. Creative student experiment using ROBO-BLOCK was done. SANDY-1 was developed based on the structure formula of human and skeletal structure and the motion was confirmed.

7 REFERENCES

[1] Hirofumi Niimi, Noriyoshi Douhara, "Development of the Miniature Humanoid Robot" BULLETIN OF COLLEGE OF INDUSTRIAL TECHNOLOGY, vol.40,pp.83-86, 2006

[2] Hirofumi Niimi, Noriyoshi Douhara, "Development of the Miniature Humanoid Robot SUNSYRO-2" BULLETIN OF COLLEGE OF INDUSTRIAL TECHNOLOGY, vol.41,pp.115-118, 2007

Experimental verification of a mass measurement device under zero gravity with a prismatic variable stiffness mechanism

Ryota Ishibashi[1], Ryuta Ozawa[2], Sadao Kawamura[2]
[1] Graduate School of Science and Engineering, Ritsumeikan University, Shiga, Japan
[2] Dept. of Robotics, Ritsumeikan University, Shiga, Japan

Abstract
A new method to measure the mass of an object in microgravity is proposed in this paper. The proposed method uses a new system that consists of a prismatic variable stiffness mechanism and an actuator. The object to be measured is attached to the prismatic variable stiffness mechanism. The actuator vibrates the mechanism and the object. The stiffness is adjusted to satisfy the anti-resonance condition, and the mass is measured from the stiffness and the frequency. An advantage of the proposed measurement method is that the actuator vibration converges to zero as time tends to infinity. Moreover, the physical parameters, such as the inertia, and the damping parameters of the measurement system are not required to measure the mass value. Several experiments were executed to confirm the effectiveness of the proposed method.

Keywords:
Mass Measurement; Prismatic Variable Stiffness Mechanism; Microgravity Environment; Anti-resonance Mode

1 INTRODUCTION

Recently, experiments in biology and materials science have been conducted in microgravity environments, such as in orbiting space shuttles or space stations. For these investigations, it is important to measure the mass of an object in the experiment. However, conventional methods that require a natural gravity field cannot be applied in the microgravity environment. The mass of an object is usually measured by the signals such as position, velocity, and acceleration of the object in microgravity when a force is applied to the object. These measurement methods are classified into the following four categories.

(1) Methods based on the momentum conservation law

An impulse is given to the object and the momentum is measured. The mass is measured by the velocity of the motion, using the momentum conservation law. [1] [2] [3] [4] [5]

(2) Methods based on centrifugal force

The object is rotated with a constant period. The mass of the object can be measured by the centrifugal force and the radius of the rotation. [6] [7]

(3) Methods based on resonance

The object is attached to the tip of an elastic element. An actuator oscillates the object through the elastic element. The mass can be measured by the frequency of a resonance mode. [8] [9] [10] [11] [12] [13]

(4) Methods based on anti-resonance

The mechanism of this method is nearly identical to the method based on resonance. The object is attached to the tip of an elastic element, and an actuator oscillates the object through the elastic element. The main difference between this method and previous one is that the stiffness value is adjustable. The stiffness will be tuned to satisfy the anti-resonance mode. As the result, the mass is measured by with the stiffness value and the frequency. [14] [15] [16]

The previously proposed methods have two problems. The first is that the signals, such as acceleration, and the physical parameters contain noise, which decreases the accuracy of the mass measurement. The second is the vibration from the measurement systems of the proposed methods that use oscillation. The vibration should be eliminated to stabilize the space structure. In the methods based on anti-resonance, the motion of the actuator converges to zero as time tends to infinity because of anti-resonance. Thus, it is suitable for use in space shuttles or space stations. However, the first problem still remains in the method based on anti-resonance.

Ishibashi et al. proposed a new mass measurement system based on anti-resonance. [17] In this measurement system, the exact physical parameter values are not required to measure the mass. Moreover, a mass measurement device was developed that has a revolute joint and a variable stiffness mechanism. The stiffness is adjusted by changing the length of a blade spring. The mass of an object is measured with the desired period and the length of the blade spring in an anti-resonance mode. This method overcomes the above two problems. However, in this method, the mass center of the blade must be exactly estimated, which is a drawback of this measurement system.

This paper proposes a new mass measurement device for microgravity environments. In the proposed device, a prismatic variable stiffness mechanism can be utilized to move the object. The mass of the object is measured without estimating the mass center of the blade. In Section 2, a mass measurement method and an adjustment law to attain an anti-resonance mode are explained. Subsequently, the model of a prismatic variable stiffness mechanism is presented. In Section 3, the experimental setup is shown, and some experimental results are demonstrated.

2 MODELING AND CONTROL OF THE PROPOSED MASS MEASUREMENT METHOD

2.1 Modeling of the mass measurement system

Figure 1 shows a model of the mass measurement device proposed in this paper. The device consists of a prismatic variable stiffness mechanism and an actuator. A sample object is attached at the end point of the prismatic variable stiffness mechanism. The actuator is fixed at the base frame, and it can vibrate the prismatic variable stiffness mechanism and a sample object.

Let x_1 and x_2 be the displacement of the object and the actuator. Here, it is assumed that the dynamics of the stiffness control mechanism can be neglected. Then, the equations of motion of the system are given by

$$m_a\ddot{x}_2 + b_a\dot{x}_2 + \hat{k}(x_2 - x_1) = \tau, \tag{1}$$

$$m\ddot{x}_1 + \hat{k}(x_1 - x_2) = 0, \tag{2}$$

where $\hat{k}$ is the stiffness value of the variable stiffness mechanism, m is the mass of the sample object, m_a and b_a are the inertia moment and the viscosity of the actuator respectively, and τ is the control input of the actuator.

To oscillate the mechanism and the sample object, the following control input is given to the actuator:

$$\tau = -\hat{k}(x_2 - \alpha) - k_v\dot{x}_2. \tag{3}$$

Here, $\alpha = ce^{j\omega_d t}$ denotes the desired periodic trajectory of x_1, and k_v is the control gain. Then, the closed loop is obtained as follows:

$$m_a\ddot{x}_2 + b_a\dot{x}_2 + \hat{k}(x_2 - x_1) + \hat{k}(x_2 - \alpha) + k_v\dot{x}_2 = 0, \tag{4}$$

$$m\ddot{x}_1 + \hat{k}(x_1 - x_2) = 0. \tag{5}$$

It is next assumed that the system's natural frequency $\sqrt{\hat{k}/m}$ coincides with ω_d; thus, the displacements can be expressed as $x_1 = c_1e^{j\omega_d t}$ and $x_2 = c_2e^{j\omega_d t}$, and we can obtain the following relations:

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \frac{\hat{k}}{F(j\omega_d)}\begin{bmatrix} \hat{k} \\ -m\omega_d^2 + \hat{k} \end{bmatrix}c, \tag{6}$$

$$F(j\omega_d) = (-m\omega_d^2 + \hat{k})\left\{(2\hat{k} - m\omega_d^2) + j\omega_d(k_v + b_a)\right\} - \hat{k}^2.$$

If the system satisfies

$$m\omega_d^2 = \hat{k}, \tag{7}$$

then

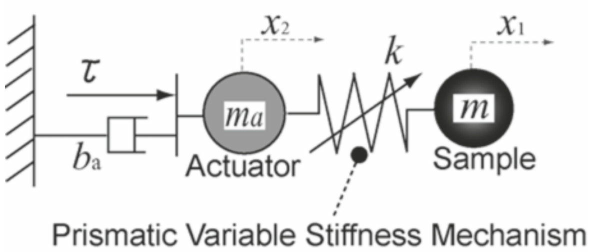


Figure 1: A mass measurement system with a prismatic variable stiffness mechanism.

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}c. \tag{8}$$

In this case, the motion of the actuator completely stops, and the motion of a sample converges to the anti-phase of α. In this paper, we call the condition (7) the anti-resonance condition.

If the closed loop satisfies the anti-resonance condition as described in equation (7), the mass of the sample m is measured from

$$m = \frac{\hat{k}}{\omega_d^2}. \tag{9}$$

The mass measurement only requires us to know the stiffness $\hat{k}$ because we can choose ω_d as a desired trajectory of α.

2.2 Control of variable stiffness

The following stiffness adjustment law is employed to attain the anti-resonance condition.

$$\begin{aligned}\hat{k}(t) = k_0 + \beta\int_0^t\Big[&-\{\dot{\alpha}(\tau) + \dot{x}_1(\tau)\}\alpha(\tau) \\ &+ \{\dot{\alpha}(\tau) + \dot{x}_1(\tau) - \dot{x}_2(\tau)\}\{\alpha(\tau) + x_1(\tau) - x_2(\tau)\} \\ &+ \dot{x}_2(\tau)x_2(\tau)\Big]d\tau.\end{aligned} \tag{10}$$

Here, k_0 is the initial stiffness, and β is a positive constant. It should be noted that the stiffness adjustment law (10) does not include any parameters of the system; thus, the parameter identification error does not affects the convergence of the system to the anti-resonance condition. It has been mathematically proven that the adjustment law (10) and the actuator input (3) constrain the system to a state that satisfies the anti-resonance condition [18].

2.3 A practical control method of variable stiffness

The second and the third terms in the integration part of Equation (10) become

$$\frac{\beta}{2}\left[\{\alpha(\tau)+x_1(\tau)-x_2(\tau)\}^2+x_2^2(\tau)\right]_0^t$$
$$=\frac{\beta}{2}\left[\{\alpha(t)+x_1(t)-x_2(t)\}^2+x_2^2(t)\right] \tag{11}$$
$$-\frac{\beta}{2}\left[\{\alpha(0)+x_1(0)-x_2(0)\}^2+x_2^2(0)\right]$$

The first term on the RHS of equation (11) converges to zero when time goes to infinity. The second term in equation (11) depends on the initial state. Therefore, these terms can be neglected when the adjustment law is applied to the real system, and we use

$$\hat{k}(t)=k_0+\beta\int_0^t\left[-\{\dot{\alpha}(\tau)+\dot{x}_1(\tau)\}\alpha(\tau)\right]d\tau \tag{12}$$

instead of equation (10). Equation (12) only requires the measurement of $\dot{x}_1$ and allows us to easily realize the system.

2.4 Prismatic variable stiffness mechanism

Figure 2 shows a model of the prismatic variable stiffness mechanism. This mechanism is based on a beam. Here, l and $\Delta x = x_1 - x_2$ are the effective length and deflection of the blade spring, respectively. The blade spring is assumed to be massless and its deflection Δx small enough so that the stiffness is linear in Δx . The sample mass m is fixed at the center of the blade spring. An actuator vibrates this mechanism periodically, and reaction force w caused by the mass is added to the blade spring.

Here, we assume that

$$w=\frac{4Ebh^3}{l^3}\Delta x, \tag{13}$$

then the stiffness $\hat{k}(l)$ can be modeled as

$$\hat{k}(l)=\frac{C}{l^3}, \tag{14}$$

$$C=4Ebh^3 , \tag{15}$$

where E , b , and h are the Young's modulus, the width, and the thickness, respectively. All of the parameters are constant, and C also becomes a constant.

Here, we assume that the line density of the beam ρ is linear in l . Then, m can be obtained from equations (9), (14), and l by the following equation.

$$m=\frac{C}{\omega_d^2 l^3}-\rho l. \tag{16}$$

We can choose ω_d as a desired frequency, and C is a constant. Therefore, the mass measurement only requires us to know the length l .

From equation (16), C can be calculated from

$$\overline{C}=(m+\rho l)\omega_d^2 l^3. \tag{17}$$

Thus, $\overline{C}$ can be calculated using the reference masses and the values of corresponding l .

3 EXPERIMENTAL RESULTS

3.1 Experimental device

Figure 3 shows the experimental system. The effective length of the blade spring is the length between the two supported points. The sample mass is fixed at the stage, which is placed at the center of the blade spring.

The length of the blade spring is controlled with a linear actuator in the variable stiffness mechanism.

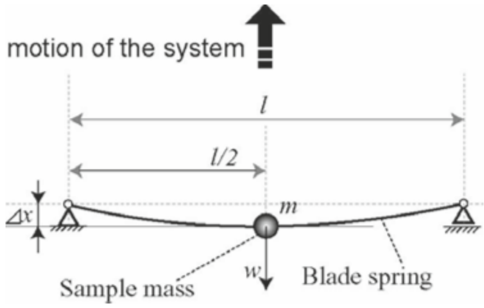


Figure 2: A model of a prismatic variable stiffness mechanism with a double support mechanism.

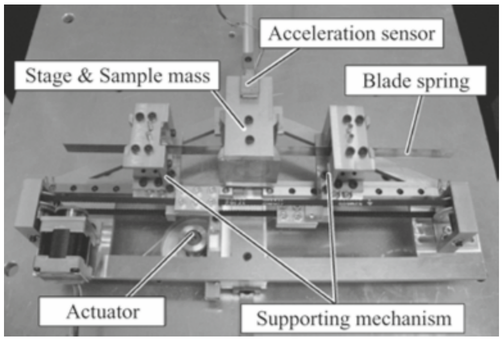


Figure 3: The developed measurement system with a double support mechanism.

Figure 4 shows an actuator for oscillation. An actuator unit is fixed at the base frame and drives the variable stiffness mechanism. The actuator unit is composed of a linear guide, a rack-and-pinion mechanism, and an AC servomotor with optical encoder (output: $200[\mathrm{w}]$).

Figure 5 shows the motion of the variable stiffness mechanism. The linear actuator controls the two supported points, where the effective length of the blade spring is adjusted from l_1 to l_2 and stiffness $\hat{k}(l)$ is changed from C/l_1^3 to C/l_2^3 . The linear actuator is composed of a linear guide, a pair of belt pulleys, and a pulse motor (angular resolution: $0.36\,[\mathrm{deg}]$). The resolution of the effective length is $1.13\times10^{-4}[\mathrm{m}]$.

Figure 6 shows the supporting mechanism of a blade spring. A pair of rollers is arranged at the supported point, placing the spring to attain point contact.

Figure 7 shows a center mechanism. The stage for the sample is fixed at the center of the spring as the point contact by the half-column (curvature radius R = $50[\mathrm{mm}]$). An acceleration sensor is attached to the stage in order to obtain the velocity of the sample (shown in Figure 3) by integration of acceleration signals. The weight of the center mechanism is $0.299[\mathrm{kg}]$.

The sample and the developed measurement mechanism lie in a horizontal plane and no gravitational force affects the system.

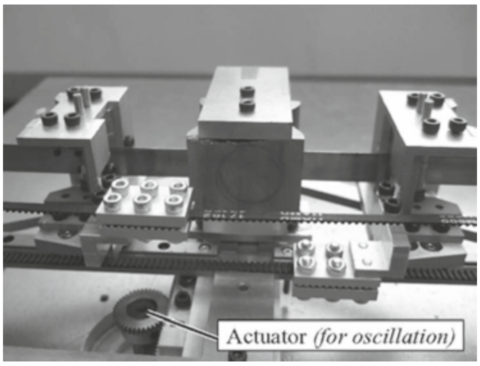


Figure 4: An actuator for oscillation.

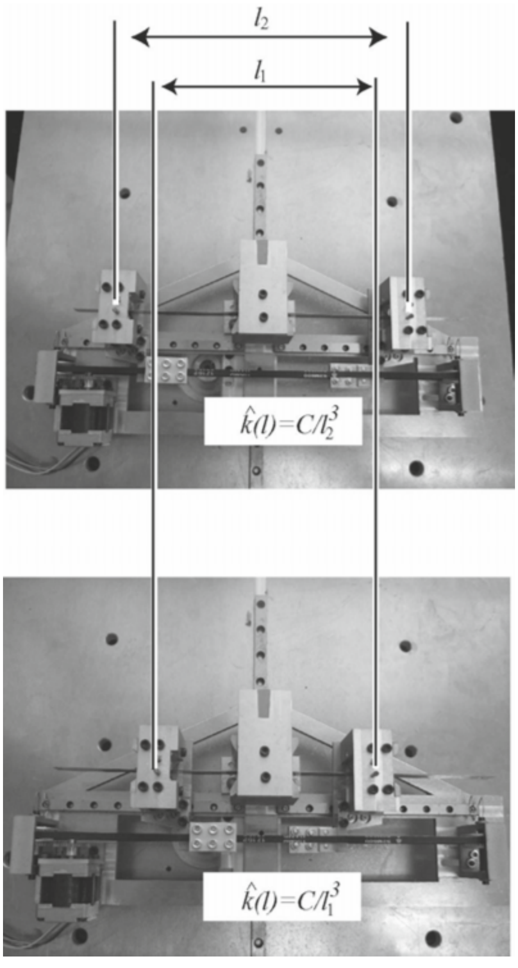
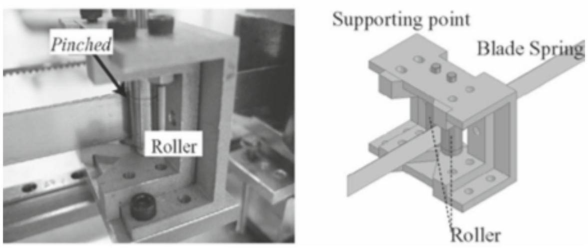


Figure 5: Phase of the stiffness adjustment.

3.2 Calculation of $\overline{C}$

Before the mass measurement that uses Equation (16), we calculate $\overline{C}$ with Equation (17).

The desired trajectory is given as $\alpha = 0.08\sin 12\pi t$. The initial length of the blade spring is $0.110[\mathrm{m}]$ and the stiffness is $501[\mathrm{N/m}]$. Nominal value of C is 0.666 .



(a) (b)

Figure 6: Supporting mechanism of a blade spring.

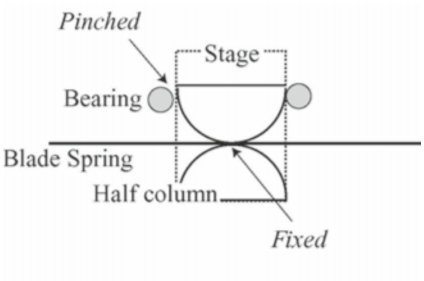


Figure 7: A center mechanism.

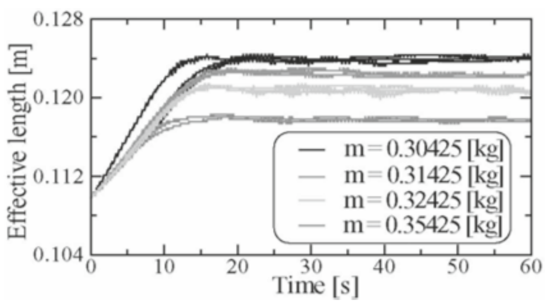


Figure 8: Effective length of the spring.

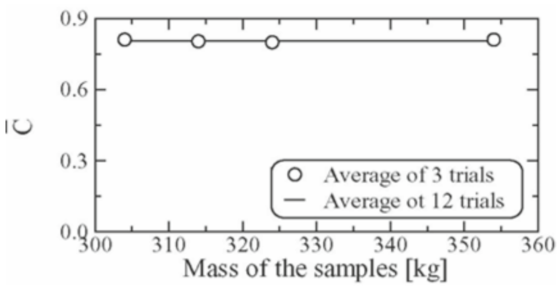


Figure 9: Calculated values of $\overline{C}$.

Figure 8 shows the effective length of the blade spring. Four calibrated samples were prepared, ranging from $0.304[\mathrm{kg}]$ to $0.354[\mathrm{kg}]$. For each sample, the experiment was repeated three times and the time response of the length for each sample was drawn with the same color. In every trial with the same sample, l converges to the same state.

Figure 9 shows $\overline{C}$ calculated with Equation (17). The circles are the calculated values $\overline{C}$ and the solid line is the average of it, which is 0.807 . The standard deviation of $\overline{C}$ is 5.9×10^{-3} . It can be observed that $\overline{C}$ is almost constant.

Table 1: Mass measurement results

Actual [kg]	Measured [kg]			
-	Trial1	Trial2	Trial3	AVG
0.304	0.305	0.304	0.304	0.304
0.309	0.309	0.309	0.309	0.309
0.314	0.314	0.315	0.314	0.314
0.319	0.317	0.318	0.319	0.318
0.325	0.326	0.326	0.324	0.325
0.329	0.330	0.330	0.330	0.330
0.334	0.336	0.334	0.334	0.335
0.339	0.340	0.339	0.341	0.340
0.344	0.347	0.346	0.345	0.346
0.349	0.349	0.351	0.351	0.350
0.354	0.354	0.354	0.354	0.354

Thus it can be concluded that the proposed control method works well and each l converges to the state satisfying the anti-resonance condition (7).

3.3 Mass measurement results

Ten calibrated samples were prepared, ranging from $0.305[\mathrm{kg}]$ to $0.355[\mathrm{kg}]$ in intervals of $0.005[\mathrm{kg}]$.

Resolution of this measurement is $0.9\times10^{-3}[\mathrm{kg}]$ to $1.3\times10^{-3}[\mathrm{kg}]$, obtained from the resolution of the blade length controller. $\overline{C}=0.807$, calculated with equation (17), was used for the measurement. For each sample, the experiment was repeated three times.

Table 1 shows results of the mass measurement, displayed in Figure 10 to verify the accuracy. The horizontal axis shows the mass of the samples, and the vertical axis shows the measured mass. The error bars represent the measured values of three trials. The solid line is the diagonal line, which represents the measured value coinciding with the actual one. The maximum error of the measured value is $1.9\times10^{-3}[\mathrm{kg}]$, and average of the standard deviation is 0.7×10^{-4} . It seems that the difference between the measured masses in three trials is within the arrangement error margin of the initial length of the blade spring.

Figure 11 displays the averages of the mass measurement results. Circles are averages of three trials. The maximum error of the averages is $1.6\times10^{-3}[\mathrm{kg}]$. The black line is the linear approximation line; the intercept is 0.006 , and the gradient is 1.02 . If we let m_{actual} be the actual mass, we obtain the relative error e as

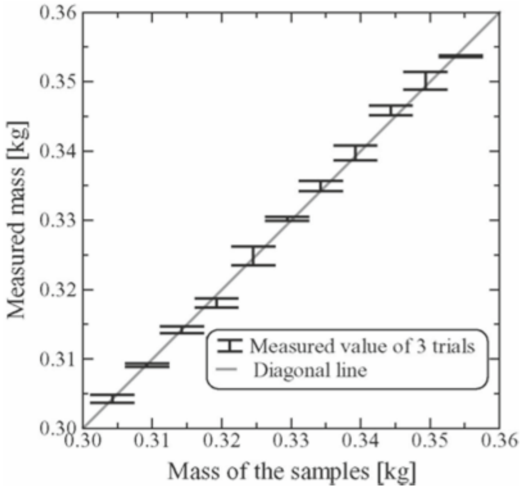


Figure 10: Mass measurement results (3 trials).

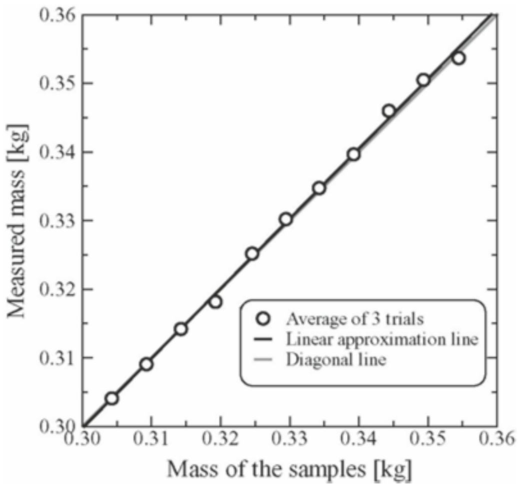


Figure 11: Mass measurement results (average).

$$e=100\times\frac{|m_{actual}-m|}{m_{actual}} \tag{18}$$

The maximum e of this measurement is $0.47[\%]$, and the average of e is $0.29[\%]$. Thus, we can say that mass is well-measured within the error margin despite the presence of backlash, the friction of the gear on the actuator unit, and the linear guide in the variable stiffness mechanism.

4 CONCLUSION

In this paper, we proposed a new mass measurement method for microgravity environments. The proposed method uses a prismatic variable stiffness mechanism, which adjusts the stiffness to satisfy an anti-resonance condition. Then, the mass of an object can be measured by the stiffness and the frequency.

In the proposed method, mass is measured from the mechanical stiffness of the spring. Therefore, it is very important to design a mechanism that can change spring stiffness and whose stiffness values can be exactly measured. In this paper, a prismatic variable stiffness mechanism using a blade spring was developed. Experiments were executed to confirm the effectiveness of the proposed method. In these experiments, measurement had been achieved within a $0.050[\mathrm{kg}]$ range, and the maximum error was $1.9\times10^{-3}[\mathrm{kg}]$.

5 REFERENCES

[1] Fujii, Y., Fujimoto, H., and Namioka, S., 1999, Mass measurement under weightless conditions, Review of Scientific Instruments, 70/1: 111-113.

[2] Fujii, Y., Fujimoto, H., Watanabe, R., and Miki, Y., 2001, Balance for measuring mass under microgravity conditions, AIAA Journal, 39/3, 455-457.

[3] Fujii, Y., 2002, Method for Measuring Mass of Non-rigid Objects under Microgravity Conditions, Transactions of the Society of Instrument and Control Engineers, 38/4, 337-344.

[4] Fujii, Y., and Shimada, K., 2006, Instrument for measuring the mass of an astronaut, Measurement science and technology, 17/10, pp.2705-2710.

[5] Mizuno, T., Minowa, J., 2000, Mass Measurement Based on the Law of Conservation of Momentum--Mass Measurement System Using Electromagnetic Impulsive Force--, Transactions of the Society of Instrument and Control Engineers, 36/12, 1059-1064.

[6] Rivetti, A., Martini, G., Alasia, F., Birello, G., Gatti, L., and Solitro, F., 1998, An Inertial, low-capacity balance designed to be accommodated on board the International Space Station, Proceedings of the 2nd European Symposium on the Utilization of the International Space Station, ESTEC, Noordwijk, The Netherlands, 93-98.

[7] Maeda, C., Masuo, R., Baba, K., 1993, Mass Measurement Using Centrifugal Force Under Weightless Condition, Proceedings of the 30th SICE Annual Conference, pp.899-900.

[8] Ono, T., Shimaoka, H., 1985, Dynamic Mass Measurement under Weightless Conditions, Transactions of the Society of Instrument and Control Engineers, 21/11, 1184-1190.

[9] Okugawa, M., Taguchi, T., and Hori, Y., 2003, Study on a mass measurement method under weightless conditions by using a smart cantilever beam, Journal of the Japan Society of Applied Electromagnetics, 11/2, 94-99.

[10] Thornton, W. and Ord,J., 1974, Specimen mass measurement, NASA Tech. Report, N74-11867.

[11] Oakey, W.E. and Lorenz, R., 1977, Survey and experimental testing of no gravimetric mass measurement devices, NASA Tech. Report, N77-26456.

[12] Beyer, N., Lomme, J., McCollough, H., Price, B., and Weber, H., 1995, Proposal for an astronaut mass measurement device for the space shuttle, NASA Tech. Report, N95-14854.

[13] Ono, T., Uozumi, H., Honda, O., and Nagata, K., 1997, Mass-measurement under weightless conditions by the frequency-controlled method, Measurement, 22, 87-95.

[14] Mizuno, T., and Araki, K., 1997, Active Dynamic Vibration Absorber with Automatic Frequency-Tracking Performance, Transactions of the Japan Society of Mechanical Engineers, Series C, 63/612, 2616-2621. (in Japanese)

[15] Mizuno, T., and Negishi, T., 1999, Development of a Mass Measurement System with an Active Dynamic Vibration Absorber, Transactions of the Japan Society of Mechanical Engineers, Series C, 65/638, 4019-4024. (in Japanese)

[16] Mizuno, T., Sato, S., Ishino, Y., 2002, Vibration-type Mass Measurement System with an Undamped Dynamic Vibration Absorber, Transactions of the Japan Society of Mechanical Engineers, Series C, 68/665, 37-43. (in Japanese)

[17] Ishibashi, R., Ozawa R., and Kawamura, S., 2007, Experimental Verification of Mass Measurement System with a Variable Stiffness Mechanism, Proceedings of the Movic, Tokyo, Japan, 480-483. (in Japanese)

[18] Ozawa, R., and Kobayashi, H., 2003, A New Impedance Control Concept for Elastic Joint Robots, Proceedings of the IEEE International Conference on Robotics and Automation, Taipei, Taiwan, 3126-3131.

Piezoresistive Rotation Angle Sensor in Micromirror for Feedback Control

Takuro Aonuma[1], Shinya Kumagai[1], Minoru Sasaki[1], Motoki Tabata[2], Kazuhiro Hane[2]

[1] Dept. of Advanced Science and Technology, Toyota Technological Institute, Nagoya, Japan

[2] Dept. of Nanomechanics Engineering, Tohoku University, Sendai, Japan

Abstract

An electrostatic micromirror is integrated with the piezoresistive sensor. The sensor detects the shear stress inside the torsion bar generated by the mirror rotation under the electrostatic driving. The sensor signal is confirmed to follow the static movement of the mirror. The short-term noise is improved from our previous level of about 0.1 degrees to 0.01 degrees. The main improvement is the modulation frequency (50 kHz) of the bias current setting higher than the mechanical resonant frequency of the micromirror (~10 kHz). With the electrical design for obtaining the lower impedance, the electrical band width of the sensor becomes larger than that of the mechanical one enabling the larger modulation frequency. The leak current is reduced with the passivation film. The characteristics of the sensor signal including accuracy, drift, sensitivity, become clearer.

Keywords:

Micromirror; Piezoresistive sensor; Feedback control; Modulation

1 INTRODUCTION

Micromirror is expected to decrease the size of equipments which scans the optical beam. The applications require the high accuracy in the mirror rotation angle. Supposing the optical setup of the optical crossconnect presented from Lucent technologies, the angular accuracy is estimated to be 0.03 degrees realizing the clipping loss less than 0.15 dB [1]. Not limited to this example, many applications need the angular accuracy. The integration of the sensor in the micromirror and the feedback control of the rotation angle based on a sensor signal are desired. Although the bulk galvanometric mirrors have such setups, realizing the similar setup in the micromirror is challenging.

Some methods for sensing the mirror rotation angle have been proposed or reported. One is assembling the lasers and the photodetectors at the backside of the micromirrors inside the device package [2, 3]. The tails of the reflected beam profile are monitored. Comparing the magnitudes of the sensor signals, the mirror rotation angle can be obtained. Kallweit et al. proposes the use of a part of the incident light beam to be steered by preparing the transmission type grating in the micromirror [4]. Onix Microsystems Inc. proposes detecting the capacitance between electrodes [5].

Another sensing principle used in MEMS devices is the piezoresistive method. Pressure sensors which consist of Si diaphragms and the piezoresistors are well-known. The strain introduced in the crystal Si changes the conduction and the valence band structures. The resultant carrier density or the mobility change is observed as the resistivity change. Bourouina et al. reported the magnetostrictive actuator integrated with the piezoresistive sensor [6].

There are two different gauges in the piezoresistive sensors [7]. One is the normal gauge. The resistance change is generated by the longitudinal piezoresistance effect under the normal stress. For the signal detection, Wheastone bridge is usually constructed with 4 resistances [8]. Another gauge is the shear gauge. Since this gauge is the single element, the resistance fluctuation or the temperature distribution does not directly generate the noise. The pressure sensor FP101 from Yokogawa Electric Corp. takes this advantage [9]. Considering that the mirror rotation generates the shear stress inside the torsion bar, the share gauge is appropriate. The single element is advantageous also for reducing the foot-print size, since preparing 2 or 4 elements inside the torsion bar is generally difficult due to the small size of the torsion bar. The shear gauge is adopted.

An electrostatic micromirror integrated with the piezo-resistive sensor was fabricated in our previous study. The sensor signal was confirmed to follow the static movement of the mirror. The noise in the measurement was large corresponding to about 0.1 degrees [10]. This noise can be attributed to the measurement method not to the device itself. In the previous study, the modulation frequency was 5 kHz. Since this frequency is lower than the resonant frequency of the micromirror, the modulation will cause the noise motion of the mirror. The higher modulation frequency for measuring the sensor signal can decrease the noise. In this study, the noise is improved taking the advantages of the new device design and the operation. The sensor signal is measured clearing the performance detail.

2 DESIGN

Figure 1 shows the schematic drawing of sensor. The sensor's appearance is similar to that of a Hall element. The sensor consists of an element and 2 pairs of terminals. One pair is for the bias current I_{bias} to flow through. The other is for detecting the signal of the transverse voltage. In contrast to the normal piezoresistive sensor which measures the resistance, the signal is the voltage between detection terminals. The signal V_{out} is expressed as follows.

$$V_{out} = \frac{f\ \rho_0 \pi\ \tau}{t} I_{bias} \tag{1}$$

f is a geometric factor that depends on the shape of the gauge. ρ_0 and t are the isotropic resistivity and thickness of

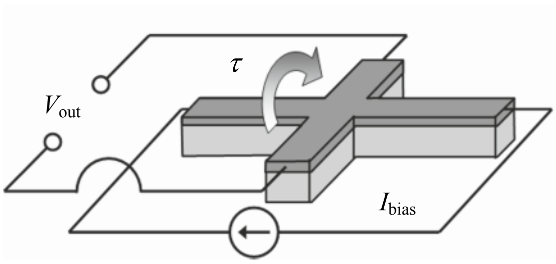


Figure 1: Schematic drawing of the basic sensor structure.

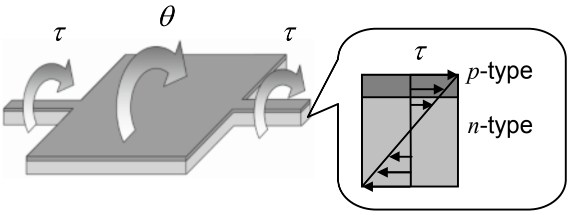


Figure 2: Vertical design of the micromirror.

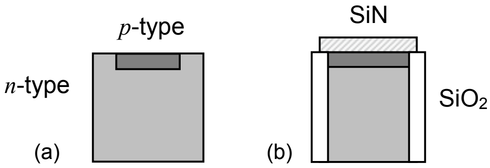


Figure 3: Crosssection of the sensor. (a) Design 1 without the passivation film. (b) Design 2 with the passivation films.

the sensing layer, respectively. τ is the shear stress, which is proportional to the mirror rotation angle. I_{bias} is the bias current. π is the shear piezoresistance coefficient. This value depends on the doping type, conductivity, and orientation of the crystal Si. The sensor is prepared from p-type Si due to the larger value of π than that of n-type Si. In the device Si layer, p-n junction is prepared.

Figure 2 shows the vertical design of the micromirror which has the close relation to the electronic structure. The device Si layer has p-n junction. The sensor is in p-type Si region at the top surface. When the mirror rotates, the shear stress is generated having the maximum magnitude on the surface as shown in the inset. The shear stress is proportional to the mirror rotation angle. The reverse bias is applied for confining the bias current in the p-type region. Since the leak current will decrease the sensitivity or increase the drift in the signal, the passivation film is prepared over the sensor structure as described in the following. Figure 3(a) shows the 1st design of the sensor showing the cross-section of the torsion bar. In the design 1, the p-type region is 3 μm inside from the sidewall. Figure 3(b) shows design 2 having the passivation films. Top and side surfaces of the sensor are covered with SiN and SiO_2 films. Although the design 1 retracts the position of p-n junction from the sidewall interface, the device surface is exposed to the ambient air. Since the design 2 is

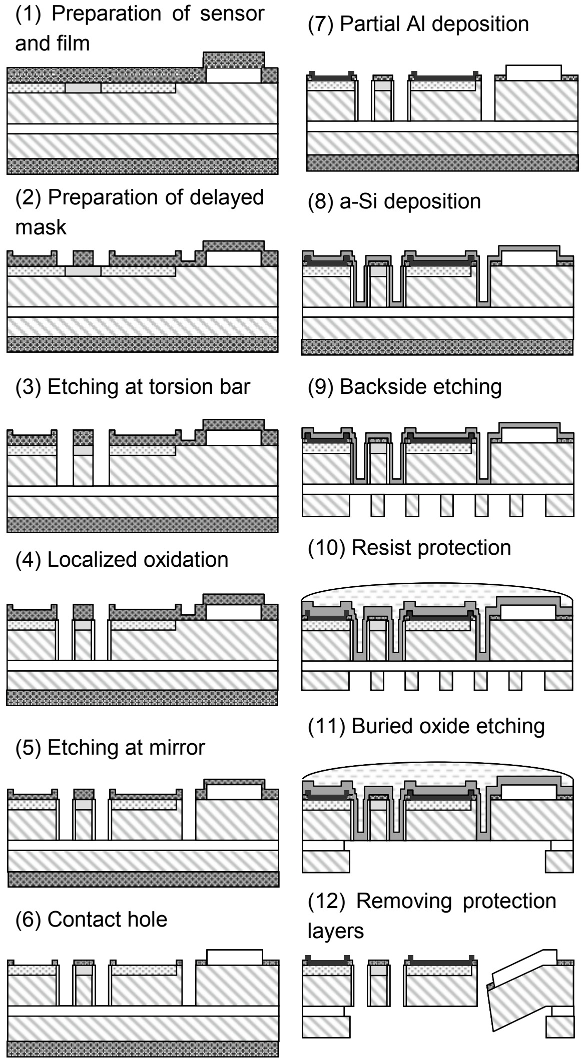


Figure 4: Fabrication sequence for the micromirror with the passivation film.

more difficult to realize the device fabrication, the following sections suppose the design 2 without mentioning the design.

3 FABRICATION

The device fabrication starts from the preparation of the sensor using the high temperature process followed by the deep reactive ion etchings for preparing the mechanical micromirror structure. Figure 4 shows the fabrication sequence. The mirror and sensor are prepared in the device layer of SOI wafer. The device Si (100) layer is 14.5 μm in thickness and n-type (>4000 Ω-cm). First, 1.4 μm-thick SiO_2 film is thermally oxidized. This SiO_2 film has the compress stress for bending the cantilever with the fixed comb. After the patterning of the oxide film, the boron ions are implanted (B^+ at 100 keV, 1.35×10^{15} atoms/cm^2 and BF^{2+} at 80 keV, 1.0×10^{15} atoms/cm^2) using the resist mask. The top 0.5-μm Si layer is considered to become p^+ region. This electrical structure is covered with 600 nm SiN film, which becomes the mask against the subsequent oxidation and plasma Si etching. The patterns of the SiN film defines the sensor shape and the mirror shape as the delayed mask (step 2). First, SiN film is opened as the shape of the torsion bar with sensor, and the vertical Si etching is carried out down to the buried oxide layer (step 3). Then the thermal oxide is grown on the sidewall of the torsion bar (step 4). This becomes the passivation film at the side wall. The top surface is protected by SiN film realizing the structure shown in Fig. 3(b). This selective oxide growth is same with that of LOCOS process. Then, SiN film is etched in a whole area. Due to the previous patterning in step 2, SiN is open as the delayed 2nd mask pattern. Using this mask pattern, the vertical Si etching is carried out preparing the mirror and the comb structures (step 5). SiN film is again etched opening the contact hole pattern(step 6). Al is deposited for the metal electrode using the stencil mask (step 7). The device front side is covered with a-Si using plasma CVD (step 8). The backside is patterned and etched for opening the space for the mirror rotation (step 9). After the additional protection of the front side using the photoresist against the vapor HF, the sacrificial buried oxide is removed (step 11). The a-Si cover is removed exposing the mirror structure (step 12).

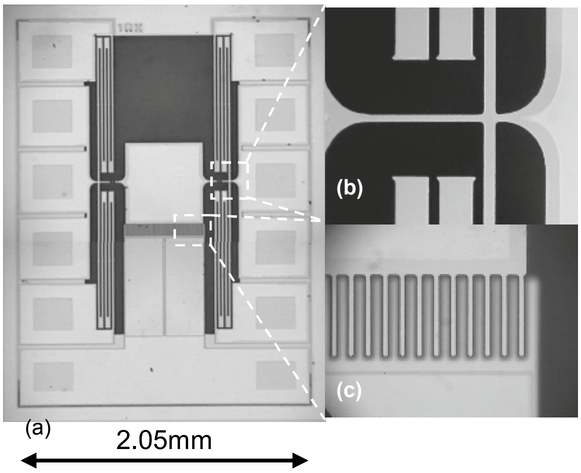


Figure 5: Micromirror device. (a) Whole view. (b) Sensor. (c) Vertical comb actuator.

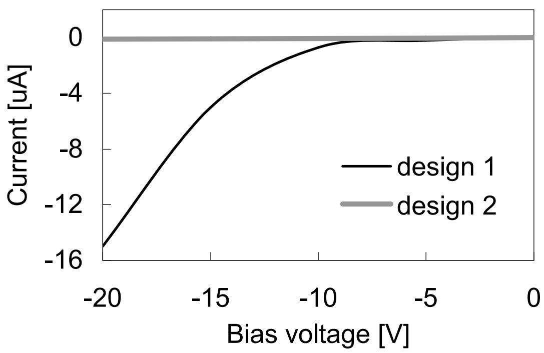


Figure 6: I-V curves from devices in designs with and without the passivation film.

Figure 5(a) shows an overview of a fabricated device. The total size including electrode pads is 2.05×2.7 mm^2. Figure 5(b) show the magnified image of the sensor. The position of the sensor is approached to the electrode pad for suppressing the temperature increase. The meandering Si spring has the resistance of 6.4 kΩ. These meandering structures have a rotational spring constant that is ~1/10 of that of the torsion bar. The center mirror size is 550×570 μm^2. Figure 5(c) shows the magnified image of the vertical comb. The width and the length of the comb finger is 5 and 80 μm, respectively. The gap between comb fingers is 4 μm.

Figure 6 shows I-V curves in the reverse bias region applying the voltage between *p*- and *n*-type regions of the sensor. Compared to the design 1 (Fig. 3(a)) showing the large leak current (15 μA at -20 V), the device of the design 2 (Fig. 3(b)) with the passivation film reduces the leak current (0.12 μA at -20 V).

4 EXPERIMENTAL SETUP

Figure 7 shows the electric circuits. Six electrical connections are obtained by probing the electrical wire using the micro-manipulators. The electrical connections are (1) one for setting GND level to the moving mirror, (2) one for setting driving voltage V_{drive} for actuating the mirror, (3) two for flowing the bias current I_{bias} inside the piezoresistive sensor, (4) two for obtaining the sensor voltage signals $V_{signal1}$, $V_{signal2}$ via meandering spring structures. The right side sensor is used in this drawing. The current source for the bias current of the sensor is selected realizing the higher band width. The modulation frequency is increased up to 100 kHz from 5 kHz used in the previous device. The resonant frequency of the micromirror is around 10 kHz (as going to be shown in Fig. 10). The sensor signals $V_{signal1}$, $V_{signal2}$ are amplified, and the difference of two signals is the input of the lock-in amplifier. The output amplitude is the signal of the rotation angle sensor.

The mirror rotation angle is measured using the optical lever method. The reflected laser beam is measured using the position sensitive detector (PSD). The equipments are set on the stage supported by the air suspension.

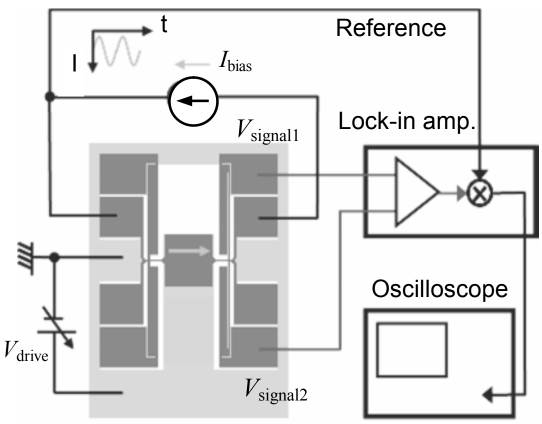


Figure 7: Electric circuits for driving the micromirror and for measuring the sensor signal.

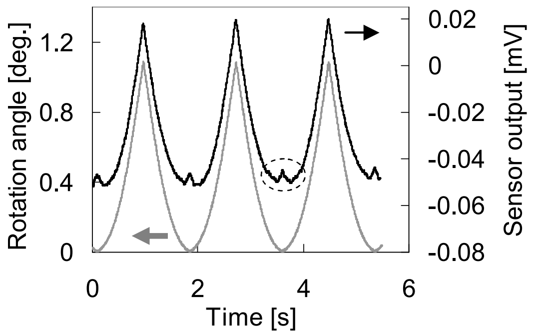


Figure 8: Time responses of 3 cycles of mirror movement.

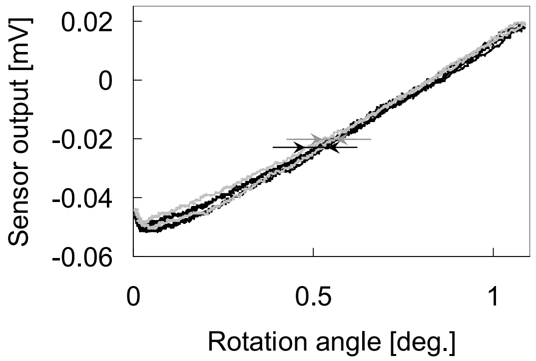


Figure 9: Sensor output as the function of mirror rotation angle for 3cycles of mirror movement.

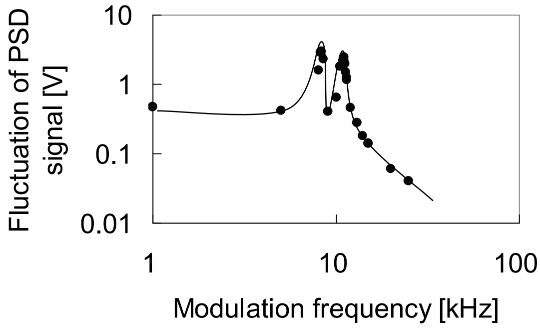


Figure 10: Fluctuation amplitude generated by the modulated bias current.

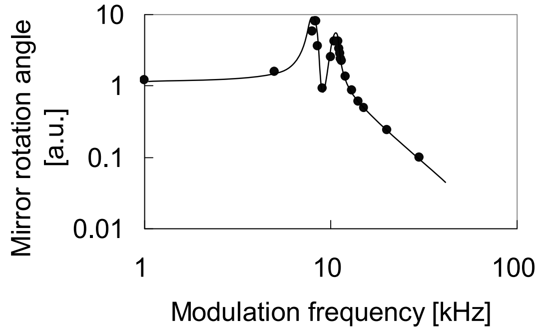


Figure 11: Amplitude of the mirror rotation angle as a function of the modulation frequency.

5 RESULTS

Figure 8 shows the typical time responses of 3 cycles of the mirror movement. The triangular voltage is applied for driving the electrostatic actuator. The black and gray curves are the sensor signal and the mirror rotation angle, respectively. The average bias current is 0.82 mA modulated at 50 kHz. The sensitivity and the time constant of the lock-in amplifier is 20 mV/V and 10ms, respectively. The as-measured curves has offset in time axis, due to the time delay of lock-in amplifier and the A/D conversion timing of the data logger used. This time offset is adjusted to have the peak at the same time between two curves. When the mirror rotates, the sensor output increases. Comparing two signals of the mirror rotation angle and of the sensor output, the signal shapes are approximately same. The shape difference appears when the mirror rotation angle is less than 0.1 degrees. The small peak appears. The reason of this small peak is not clear at now. The magnitude of this peak changes time to time implying the relation to the contact resistance between the sensor and the probe wire. There are also cases where such small peak disappear or becomes further outstanding. There is a possibility that the residual stress introduced by the passivation films (top tensile SiN and side compress SiO_2 films) at the initial state relate to this phenomenon.

Figure 9 plots the sensor signal as a function of the mirror rotation angle using the data shown in Fig. 8. There is a linear relation except the region of rotation angle smaller than 0.1 degrees corresponding to the small peak. The sensitivity against the rotation angle is 0.066mV/deg. in this torsion bar as the appearance (without converting the factor of amplifying ratio at the lock-in amplifier). Compared with the previous study [10], the short-term noise is improved form ~0.1 to ~0.01 degrees. The signal detail becomes clearer. Three cycles are included. The main noise source is the drift of the long-term noise at now. In this experiment, the accuracy of the sensor signal is about 0.05 degrees for 3 cycles. The drift during 1 cycle shown by gray curve is 0.02 degrees.

For clearing the sensor characteristics, the sensor signal and the mirror movement is examined changing the measurement conditions. Figure 10 shows the fluctuation amplitude of the mirror rotation when the modulation frequency of the bias current changes. The driving voltage is kept at the constant value of 50 V. The fluctuation amplitude of the PSD sensor signal is directly plotted without converting to the rotation angle. There are the peaks at around 10 kHz. There is an antiresonance in the center. At the frequency higher than these peaks, the fluctuation amplitude decreases. Over 50 kHz, the fluctuation is below the noise level. This fluctuation is considered to be generated by the change of the part of the potential in the moving mirror caused by the bias current. The mechanical resonance of the micromirror is examined. Figure 11 shows AC amplitude of the mirror rotation angle. The sinusoidal driving voltage is 10 V in peak-to-peak amplitude with 5.5 V offset. When the modulation frequency changes, the amplitude shows the curve shape similar to that of Fig. 10. These results confirm that the bias current can disturb the mirror movement when the modulation frequency is lower than the resonant

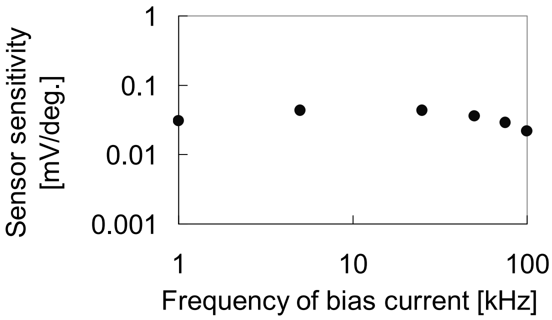


Figure 12: Sensor sensitivity as the function of the modulation frequency of the bias current.

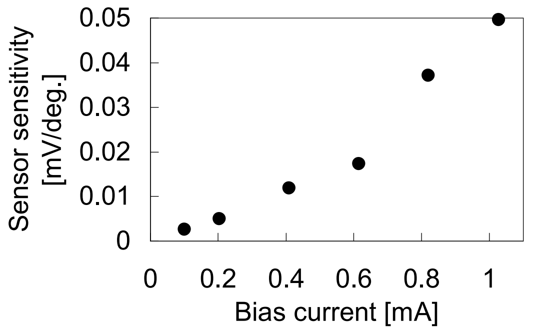


Figure 13: Sensor sensitivity as the function of the bias current.

frequency of the micromirror. The faster modulation frequency can remove this fluctuation.

Figure 12 shows the sensor sensitivity as the function of the modulation frequency of the bias current. The sensor signal starts to decrease at around 50 kHz, which corresponds to the cut-off frequency. Since the mechanical resonant frequency of the micromirror is at around 10 kHz, the rotation angle sensor realizes the broader band width. This is mainly obtained from the decrease of the resistance compared to the previous device. The resistance between the electrodes for flowing the bias current is decreased from 6.0 to 2.3 kΩ. This enable the modulation faster than the limit of the mechanical response.

Figure 13 shows the sensor sensitivity as the function of the bias current. The raw sensor sensitivity is shown counting the amplifying ratio of the lock-in amplifier. The sensitivity is measured from the change of the sensor output voltage divided by the mirror rotation angle obtaining the results similar to that of Fig. 9. The bias current changes from 0.1 to 1mA. The sensor signal rather increases linearly with the bias current as estimated in Eq. (1). Considering the facts that the current flow generates Joule heat and the piezoresistive coefficient π decreases with the temperature, Fig. 13 indicates that the temperature increase of the sensor is subtle at least up to 1 mA keeping the linear relation expected by Eq.(1).

6 SUMMARY

The piezo-resistive rotation angle sensor is examined. With the higher modulation frequency of the bias current over the mechanical resonance of the micromirror, the noise level of the sensor signal decreases to 1/10 compared to the previous study. Taking the advantage of these improvements, the detail of the sensor signal becomes clearer. The angular accuracy of ~0.01 degrees is demonstrated although the signal suffers from the drift which is the main noise at now.

7 ACKNOWLEDGMENTS

Authors thank to Profs. H. Sakaki and K. Furutani, Toyota Technological Institute, for the collaboration related to building the experimental setups. This research is supported by MEXT (Ministry of Education, Culture, Sports, Science and Technology) grant-in-aid for scientific research on priority areas (no. 19016003), MEXT HAITEKU (2006) and electro-mechanic technology advancing foundation. The facilities used for fabricating the device include the micro/nano-machining research and education center, at Tohoku University.

8 REFERENCES

[1] Alsyuk V. A., Arney S., Basavanhally N. R., et al., 2003, 238x238 Micromechanical Optical Cross Connenct, IEEE Photonics Technology Letters, Vol.15, No.4, 587-589.

[2] US patent 6714336, Japanese patent application 2002-49891.

[3] Akase K., Sawada R., Higurashi E., Kobayashi T., Maeda R., et al., 2007, Combined Device of Optical Microdisplacement Sensor and PZT-Actuated Micromirror, Proc. IEEE/LEOS optical MEMS and Nanophotonics, 199-200.

[4] Kallweit D., Zappe H., Fabrication of bulk-Si micromirrors with an integrated tilt sensing mechanism, 2006, J. Micromech. Microeng. Vol.16, 463-469.

[5] US patent 6629461, Japanese patent application 2001-571547.

[6] Bourouina T., Lebrasseur E., Reyne G., Debray A., Fujita H., Ludwig A., Quandt E., Muro H., Oki T., Asaoka A., 2002, Integration of Two Degree-of-Freedom Magnetostrictive Actuation and Piezoresistive Detection: Application to a Two-Dimensional Optical Scanner, J. Microelectromechanical Systems Vol.11, 355-361.

[7] Yalcinkaya A. D., Urey H., Brown D., Montague T., Sprague R., 2006, Two-Axis Electromagnetic Microscanner for High Resolution Displays, J. Microelectromechanical Systems, Vol. 15, No. 4, 786-794.

[8] Sandner T., Conrad H., Klose T., Schenk H., 2007, Integrated Piezo-resistive Positionsensor for Microscanning Mirrors, Proc. IEEE/LEOS optical MEMS and Nanophotonics, 195-196.

[9] Saigusa T., Aga T., Odohira T., Miyaji N., 1988, Yokogawa Giho, Single Gauge Type Silicon Pressure Sensor, Vol. 32, No. 3, 135-138 (in Japanese).

[10] Sasaki M., Tabata M., Haga T., Hane K., 2006, Piezoresistive Rotation Angle Sensor Integrated in Micromirror, Jpn. J. Appl. Phys. Vol. 45, No. 4B, 3789-3793.

Proposal of 2-axis Bulk-PZT Gyroscope

Kenta Maekawa[1], Yoshito Tanaka[1], Takayuki Fujita[1], Kazusuke Maenaka[1]
[1] Graduate School of Engineering, University of Hyogo, Himeji, Japan

Abstract
This paper reports a novel micro gyroscope with improved resistance to acceleration or shock, which can measure 2-axis angular velocity with a single device. The novel device has a completely solid structure made of piezoelectric material, bulk-PZT, which is quite simple and unlike any conventional MEMS gyroscopes. When a longitudinal vibration is excited in PZT rectangular prism, a differential output voltage, generated by the Coriolis force, is obtained at the top and bottom surface electrodes, which is proportional to the applied angular velocity. The paper describes the operation principle, device design, and preliminary experimental results.

Keywords:
Gyroscope; PZT; Piezoelectric material; Thickness shear vibration

1 INTRODUCTION

Miniaturized gyroscopes for measuring angular velocity have attracted attention for many applications [1-3]. Many types of MEMS gyroscopes have been reported and some are already commercialized. Almost all current MEMS devices are vibratory types that consist of vibration structures with a single beam, tuning fork vibration structures, rotational mass or gimbal ring structures [4-7]. However, their essential operating principal and basic structures are similar; i.e., they provide one or more masses and support springs. Because the spring mass system essentially acts as an acceleration sensor in such gyroscopes, acceleration or shock causes an unwanted movement of the electrodes and sensor body, which inhibits the transfer of the mass to the reference vibration, and causes permanent damage to the sensor device. Thus, no existing vibratory micro gyroscopes can operate under large acceleration or in shock environments.

We previously presented a PZT gyroscope for 1-axis angular velocity detection [8], which is composed of a rectangular prism-shaped bulk-PZT of size 3 × 4 × 5 mm^3. The PZT device has electrodes on its surface, as shown in Fig. 1. Because it has a solid body, PZT body, it is not influenced by high acceleration or shock. This very simple and sturdy structure increases not only resistance against the shock but productivity.

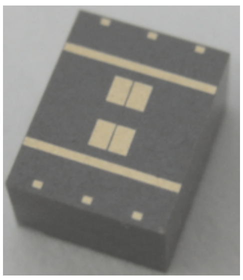


Figure 1: Photograph of the device fabricated in a previous study.

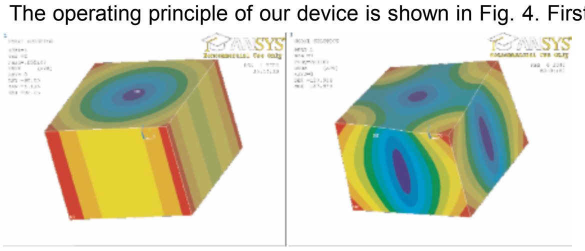
In this study, we improved the shape of our previous gyroscope and propose a novel 2-axis bulk-PZT gyroscope.

2 DEVICE CONCEPT

For vibratory type gyroscopes, a reference vibration is essential for generating a Coriolis force. Almost all MEMS gyroscopes have independent springs and lumped mass for generating the reference vibration. The transversal/bending vibration mode of the springs is used for operation. In these structures, applied acceleration or shock creates a large torque for the spring-mass system. Thus, unwanted reference and output vibration are obtained.

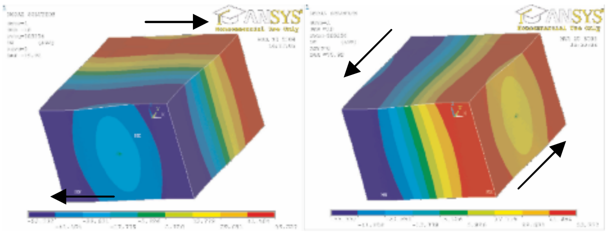
Our novel device has a rectangular prism shape, which has various longitudinal vibration modes that can be used as the reference vibration. Thus, the vibration is very resistant to shock and acceleration. In a PZT device of size 3 × 4 × 4 mm^3, the PZT block has many resonance modes. Figure 2 shows examples of the 6th and 7th longitudinal vibration modes. To achieve angular detection in two axes simultaneously, we find a suitable mode shape at the 10th resonance mode at 305 kHz, as shown in Fig. 3. The movement of the mass includes both the x and y directions, and a diagonal direction on a plane surface.

The operating principle of our device is shown in Fig. 4. First,



(a) 6th (4 kHz) (b) 7th (203 kHz)

Figure 2: Example of longitudinal resonance modes in an unrestrained condition.



(a) Displacement of the mass in the x direction (b) Displacement of the mass in the y direction

Figure 3: 10th resonance mode at 305 kHz for 3 × 4 × 4 mm^3 rectangular PZT prism.

our PZT device is polarized along the z axis. With excitation of the PZT prism at the 10th resonance mode, the reference vibration belonging to the x and y axes appear, as shown in Fig. 4(a). When an angular velocity is applied along the x axis, a Coriolis force is generated according to movement of the device, as shown in Fig. 4(b). Then, the Coriolis force shows thickness shear vibration, which differentially generates a piezoelectric voltage at the top and bottom surfaces of the device, as shown in Fig. 4(c). An output voltage proportional to the applied angular velocity is obtained. Because the device has a rotationally symmetrical structure, the angular velocity along the y axis simultaneously generates a perpendicular output voltage, as shown in Fig. 4(c).

3 DEVICE DESIGN AND SIMULATION

It is difficult to analyze the required properties of our PZT prism device. Thus, we used an FEM software (ANSYS) for device design. First, harmonic analysis was performed to determine the 10th resonance mode for the reference mode. Next, the response to the applied angular velocity was examined using ANSYS with a special Coriolis Macro program [9]. This analysis determined the isolation distance between the excitation and output electrodes.

3.1 Harmonic Analysis

Since it is necessary to use a higher-order resonance mode in our device, some other unwanted resonance modes may exist around the required 10th resonance mode. Harmonic analysis clarifies required resonance mode. Figure 5 shows the model and electrode configuration of our device. There are three kinds of electrodes at the surface; a driving electrode for longitudinal reference vibration; a detection electrode for PZT surface potential generated by the applied angular velocity; and a reference electrode for maintaining the reference vibration.

For estimating the device sensitivity, an excitation voltage for a reference 10 Vp-p and a mechanical Q factor of 200 are assumed. The amplitudes of the surface potential at the detection and reference electrodes are calculated as a function of excitation frequency, without applying the angular velocity. These amplitudes are shown in Fig. 6; the 10th resonance frequency is marked by an ellipsoid. The analysis showed that the 10th resonance mode has no peak for detection electrodes A – D and E – H, but it has a clear peak for reference electrode R. Fig. 7 shows an analysis of the phase difference between the excitation voltage and the output voltage of reference electrode R. At the 10th resonance frequency, we can find phase variation, which implies that reference electrode R can be used for finding and maintaining the required 10th resonance mode. Thus, maintaining the phase variation makes it possible to maintain a resonant condition. We can use the electrode at point R for self oscillation.

3.2 Response to angular velocity analysis

With excitation at the 10th resonance mode, an applied angular velocity of 1 rad/s was assumed. The surface potential of the PZT prism can be calculated by FEM, taking the Coriolis force into consideration. Figure 8 shows the results of the surface potential applied angular velocity in the x and y directions. The sensitivity was calculated to be about 34 μV/(rad/s). The detection electrodes should be located where a large output voltage is generated. The final design (size and formation of electrodes) is shown in Fig. 9.

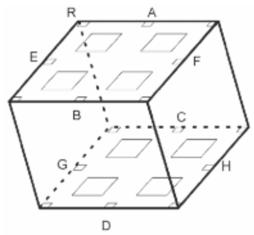
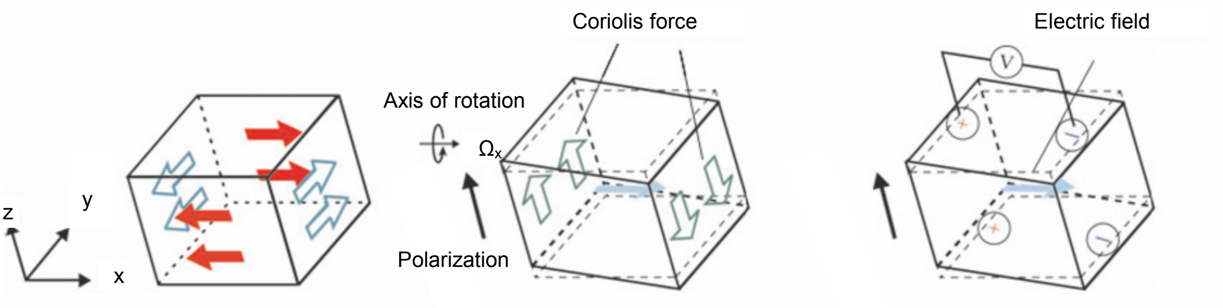


Figure 5: Model of the device.



(a) Reference vibration: movement of mass elements in the PZT body

(b) An applied angular velocity generates a Coriolis force

(c) Thickness shear vibration results in a surface potential due to the piezoelectric effect

Figure 4: Operating principle of our gyroscope.

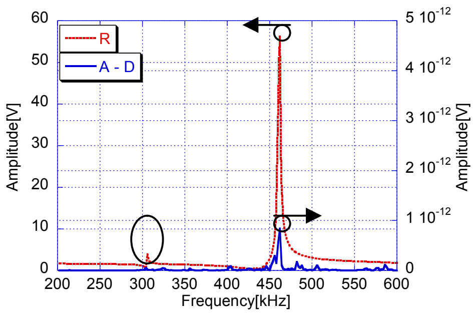


Figure 6: Calculated output amplitudes at A - D and R with no applied angular velocity.

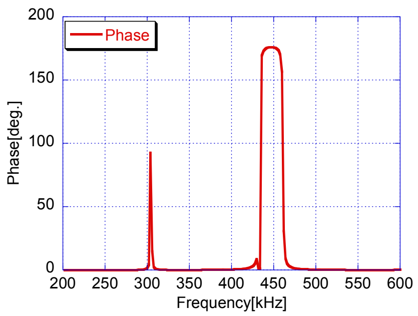
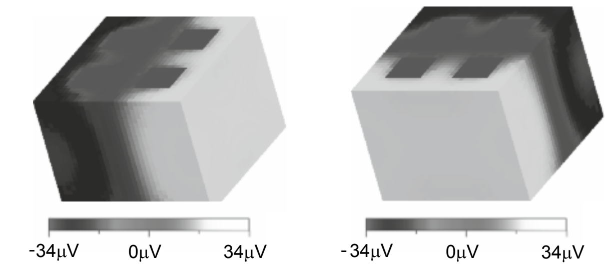


Figure 7: Calculated phase difference between outputs at R and excitation voltage.



(a) Applied angular velocity along the x axis

(b) Applied angular velocity along the y axis

Figure 8: Calculated output signal for an applied angular velocity.

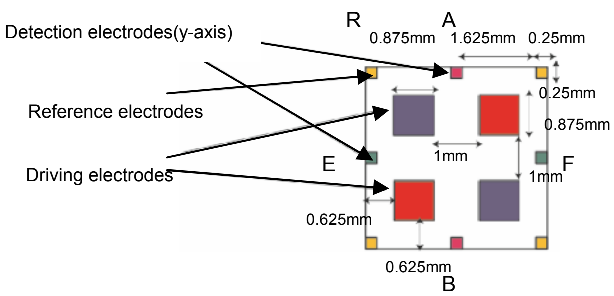


Figure 9: Arrangement of electrodes.

4 EXPERIMENTS

4.1 Preparation of prototype device and measurement circuitry

The fabrication steps are extremely simple, involving just the fabrication of electrodes and dicing. A photograph of the device is shown in Fig. 10. The device is clamped by four screws at the center of the side wall, where the amplitude of the vibration is very small. The electrodes are attached by the usual bonding methods. The measurement setup is shown in Fig. 11. This setup contains self oscillation circuitry using a Direct Digital Synthesizer (DDS), addition and subtraction circuits for removing common mode signals, a filter, an amplifier and a synchronous demodulator.

4.2 Resonance characteristics

We first measured the resonance characteristics for reference vibration from 200 to 600 kHz, as shown in Fig. 12. These correspond to the calculations shown in Fig. 6. The absolute values of mode frequencies are slightly different, by about 2 kHz, between calculations and measurements. However the analysis is sufficiently accurate. Figures 13 and 14 show voltage and phase comparisons of the calculation and measurement for reference electrode R. The differences between measurements and calculations are due to a lack of accuracy in estimating the Q factor for the analysis. The fabrication tolerance of the device is another influence.

5 DISCUSSION

In this state, the resonant frequency of the PZT device is slightly different from the designed value. Thus, the output signal tends to be masked with noise or the driving signal. To improve the resonance matching, we discuss the properties of the current device.

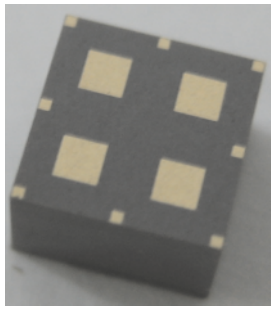


Figure 10: Photograph of the test device. Size is 3 × 4 × 4 mm^3.

In our device, because of the symmetrical arrangement of the driving electrodes, the common mode output signals on detection electrodes (shown in

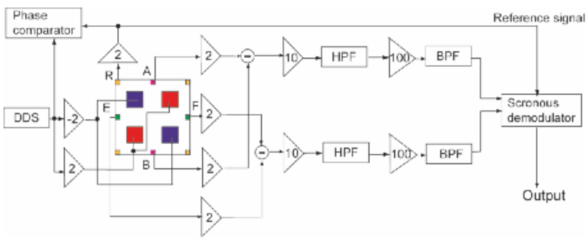


Figure 11: Measurement setup.

Fig. 5) A and B (C and D) should be equal. It also has a rotationally symmetric structure; the same should be true for detection electrodes E to H. In our experience, an asymmetric signal appeared at the detection electrodes because these electrodes had some misalignment, which can cause large offset signals in the output. Thus, the addition/subtraction circuits would not be effective, and we would not obtain clear output.

In Fig. 13, the output value at reference electrode differed from the calculated value. This difference was also caused by the device asymmetry. We are working to improve the process accuracy using MEMS technology, and will redesign the control circuitry to improve system sensing characteristics.

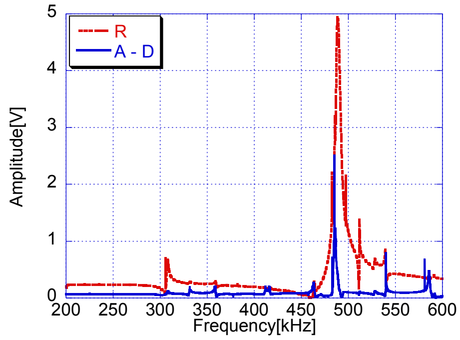


Figure 12: Measured resonance characteristics for electrodes A - D and R.

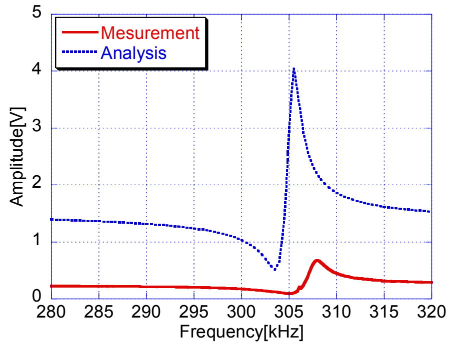


Figure 13: Comparison of measured and calculated output amplitudes for electrode R.

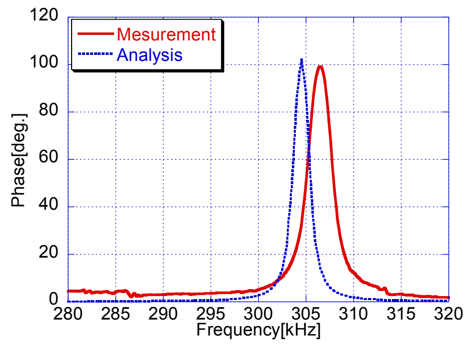


Figure 14: Comparison of measured and calculated phase differences between outputs for electrode R.

6 CONCLUSIONS

A 2-axis bulk PZT gyroscope was proposed. We explained the operating principle and device design, and examined the resonance characteristics for the prototype device. The rotationally symmetric PZT structure realizes a 2-axis gyroscope. The longitudinal vibrating gyroscope presented here will be a candidate for next generation gyroscopes that can operate in heavy acceleration and shock environments.

7 REFERENCES

[1] J. Sodeverkvist, "Micromachined gyroscopes", Sensors and Actuators, A 43, pp. 65-71, 1994.

[2] N. Yazdi, F. Ayazi, K. Njafi, "Micromachined Inertial Sensors", proceedings of the IEEE, 86, pp. 1640-1659, 1998.

[3] A. Lawrence, Modern Inertial Technology Navigation Guidance and Control, 2nd ed., Sepringer-Verlog, 1998.

[4] K. Maenaka, T. Fujita, Y. Konishi, M.Maeda, "Analysis of a high sensitive silicon gyroscope with cantilever beam as vibrating mass", Sensors and Actuators, A 54, pp. 568-573, 1996.

[5] S. Sassen, R. Voss, J. Schalk, E. Stenzel, T. Gleinssner, R. Gruenberger, F. Neubauer, W. Ficker, W. Kupke, K. Buer, "Silicon Angular Rate Sensor for Automotive Applications with Piezoelectric Drive and Piezoresistive Read-out", in Digest Tech. Papers Transducers '99 Conference, Sendai, pp. 906-909, 1999.

[6] T. Fujita, K, Hatano, K. Maenaka, T. Mizuno, T.Matsuoka, T, Kojima, T. Oshima, M, Maeda, "Vacuum Sealed Silicon Bulk Micromachined Gyroscope" , in Digest Tech. Papers Transducers '99 Conference, Sendai, 1999, pp. 914-917.

[7] F. Ayazi, K. Najafi, "A HARPSS Polysilicon Vibrating Ring Gyroscope", J. Micro Electro Mechanical Systems, 10 pp. 169-179, 2001

[8] K. Maenaka, H. Kohara, M. Nishimura, T. Fujita, Y. Takayama, "Novel Solid Micro-Gyroscope", Micro Electro Mechanical Systems, pp634-637, 2006

[9] M.Kita and K. Maenaka, "Harmonic Response Analysis of the Micromachined Gyroscopes with Coriolis Force", 2000 Ansys Conf., Aug. 28-30, The Pittsburgh, 2000

A Study on Atomization Characteristics of Surface Acoustic Wave Atomizer using Laser Doppler Anemometry

Jungmyoung Ju[1, 2], Yutaka Yamagata[2], Kozo Inoue[3], Toshiro Higuchi[1]

[1]The University of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo, Japan
[2]VCAD System Research Program, RIKEN (The Institute of Physical and Chemical Research),
2-1 Hirosawa Wako-shi, Saitama, Japan
[3]Fuence Co., Ltd., 1-11-2, 703 Hiroo Shibuya-ku, Tokyo, Japan

Abstract

In this report, to evaluate atomization characteristics of surface acoustic wave atomizer and electrostatic deposition method (SAW-ED), Laser Doppler Anemometry (LDA) is used. In the case of SAW-ED method, atomized sample is collected by electrostatic force, and measured by field-emission scanning electron microscope (FE-SEM). In LDA method, scattered beam is measured and drop speed is calculated. Applied RF power is in the range from 0.97 to 1.03 W. In the results of SAW-ED, particle size and input RF power is 0.36 μm (0.97 W), 0.38 μm (1 W), and 0.4 μm (1.03 W), respectively. In the case of LDA method, drop speed is 0.16 m/s (0.97 W), 0.19 m/s (1 W) and 0.26 m/s (1.03 W). As a result, atomization characteristics of SAW atomizer were sensitively changed by the variation of input power.

Keywords:
Surface Acoustic Wave Atomizer, Laser Doppler Anemometry, Ultrasonic Atomization, Micro droplets

1 INTRODUCTION

Generation techniques of nano-particles via liquid atomization methods have been expanding to the variety of fields, such as chemistry, physics, medical science and material science. Especially, dry-direct patterning using nano -particles and electrostatic deposition method showed possibility of high-resolution patterning [1]. Another merit is that physical and chemical damage to the sample is relatively lower than other deposition methods like PVD (thermal evaporation, e-beam evaporation, sputtering) or CVD, so that it is possible to use organic and bio-macromolecules (protein, DNA, cell).

The process of dry-direct patterning methods is normally composed of four steps; the first is sample charge using high-voltage; the second is atomization of micro-droplets; the third is drying process for the nano-particles generation; the final step is deposition of the charged particles by electrostatic force. In this process, atomization speed as well as drop size are one of the most dominant parameters, which limit patterning resolution, deposition time and so on.

Among variety of atomization methods, ultrasonic atomizer has merits on the compact equipment, narrow droplet distribution, high energy density and relatively high atomization speed. Based on the mechanism of ultrasonic atomization, drop size is reduced according to the operating frequency. In 1995, new ultrasonic type atomizer using surface acoustic wave device had been proposed [2]. The merit of proposed device is that it is possible to operate with higher frequency than ordinary ultrasonic type atomizer.

The authors have been studied about nano particle generation techniques as well as deposition methods using surface acoustic wave atomizer, which is possible to make tiny droplets with high frequencies ranged from 10 to 100 MHz [3-6]. In this report, standing wave type surface acoustic wave (SAW) atomizer is fabricated, and atomization performances are tested by means of an electrostatic deposition (SAW-ED) method and Laser Doppler Anemometry (LDA).

2 EXPERIMENTAL SETUP

2.1 Design of the Surface Acoustic Wave Atomizer

Mechanism of atomization is that surface acoustic wave is generated by applying an alternating current to a metal film electrode (Inter Digital Transducer; IDT) on the 128° Y-cut $LiNbO_3$. Then, generated SAW ultrasound is passed through a liquid located on the piezoelectric surface, capillary wave on the liquid surface is generated. If energy of capillary wave has enough level to overcome surface tension, liquid drop is atomized as shown in figure 1.

Figure 2 shows the designed electrodes of surface acoustic wave atomizer. For the standing wave vibration, two excitation electrodes are utilized [4]. And, each of excitation electrode, reflector electrode was utilized for the amplification of vibration energy by the reflection. Wave velocity is almost 4000 m/s, and resonance frequency is 10 MHz, thus wave-length of vibration is 0.4 mm. Therefore, width of electrode is fabricated with 100 micrometers. Electrode was fabricated by thermal evaporation using chrome and gold. Thickness of chromium and gold layer is about 100 nm respectively.

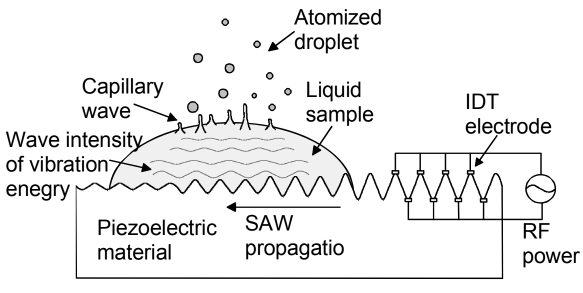


Figure 1: Concept of surface acoustic wave (SAW) atomizer

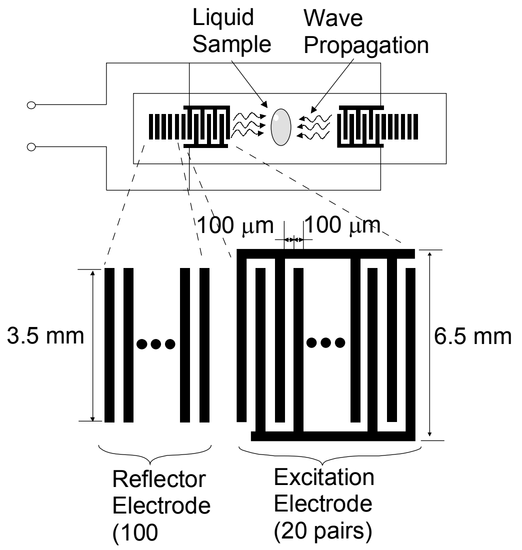


Figure 2: Electrode design of surface acoustic wave atomizer (128˚ Y-cut $LiNbO_3$ (Y-cut X-propagation), Thickness: 1 mm, Operation frequency: 10 MHz)

2.2 Driving circuits and operating conditions

In the case of surface acoustic wave device, impedance, resonance frequency and other mechanical and electronic characteristics are sensitively changed according to the boundary conditions. Especially, it is known that impedance variation lead to unstable atomization, when surface acoustic wave device used as an atomizer [7]. During atomization process, impedance variation is generated by the variation of temperature, power absorbance, weight of the liquid sample located on the vibration surface.

Figure 3 shows driving circuits of surface acoustic wave atomizer. Continuous sign signal is generated by PLL oscillator, low-pass filter, and programmable attenuator. For the stable power operation by the absorbing reflector power, -1 dB fixed attenuator is utilized. For the measurement of the forward and reflected power, SWR and power meter is connected after the attenuator. After that, impedance matching circuit is utilized for the purpose of 50 Ω matching.

For the finding operating condition of the RF power, we tested required minimum forward RF power for the atomization. As the results, we found that 0.97 W was minimum power, and other 1 and 1.03 W were also adopted for the following test.

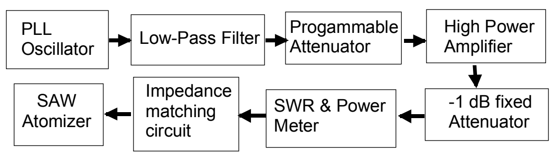


Figure 3: Driving circuits of surface acoustic wave atomizer for 10MHz operation

2.3 Surface Acoustic Wave Atomizer and Electrostatic Deposition (SAW-ED)

Figure 4 shows experimental apparatus of surface acoustic wave atomizer and electrostatic deposition (SAW-ED). Mechanism of SAW-ED is that liquid sample is charged by the high-voltage conductive wire, and atomized by the SAW atomizer. Atomized droplets are immediately dried and changed into nano-sized particles. These charged nano-particles are lifted up to the conductive substrate by the electrostatic force. Detailed experimental setup is as follows. Used material is bovine serum albumin (BSA) protein solution. Concentration of the sample was 5 mg/ml. Applied high voltage for the drop-charge was approximately 5 kV. Distance between SAW atomizer and conductive substrate was 15 cm. Aluminium coated plastic film was utilized between Teflon shield and conductive substrate. Deposited particles were measured by field-emission scanning electron microscope (FE-SEM), and particle size was calculated by the image processing software (Image-J, National institutes of Health, USA).

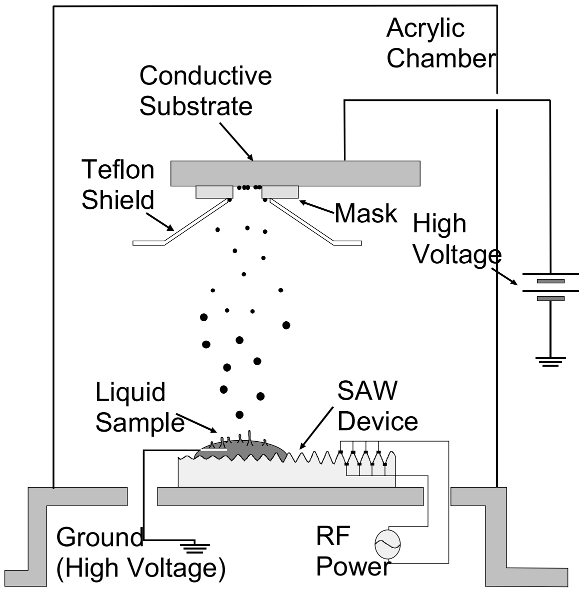


Figure 4: Experimental apparatus of surface acoustic wave atomizer and electrostatic deposition (SAW-ED)

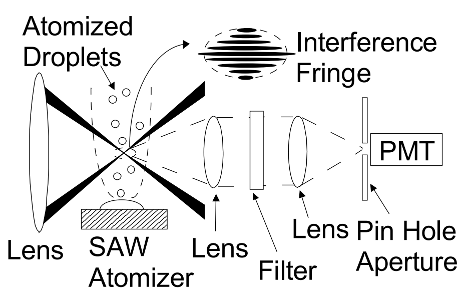


Figure 5: Experimental apparatus of Laser Doppler Anemometry (LDA)

2.4 Laser Doppler Anemometry (LDA) method

Figure 5 shows experimental apparatus of Laser Doppler Anemometry method. 532 nm laser was used for LDA system. Designed fringe size was 2.4 μm, interference spot size was 20 μm, and fringe number was approximately 8. Distance between interference fringe and SAW atomizer was 5 mm. When a drop passes through interference fringe, scattered light is generated and that signal has information about drop size and speed. Using two lens and filter, scattered laser beam was measured by photomultiplier tube (PMT).

3 EXPERIMENTAL RESULTS

3.1 SAW-ED method

Figure 6, 7 and 8 shows SEM image of deposited BSA particles by SAW-ED method. In the case of figure 6, relatively smaller particles are deposited compared with figure 7 and figure 8. Figure 9 shows size distribution of the deposited particles by SAW-ED method. In this graph, entire particle count is increased by the RF power rise. And, size range between minimum and maximum diameter is also increased by the power rise. Figure 10 shows mean particle diameter and standard deviation according to the forward RF power. In the results, particle size and input RF power is 0.36 μm (0.97 W), 0.38 μm (1 W), and 0.4 μm (1.03 W), respectively.

3.2 Laser Doppler Anemometry method

In the case of LDA method, drop speed is 0.16 m/s (0.97 W), 0.19 m/s (1W) and 0.26 m/s (1.03 W). Drop speed is increased according to the input RF power.

4 CONCLUSION AND DISCUSSION

In this report, particle size and drop speed was measured by the SAW-ED and LDA method. As the results, particle size and drop speed was increased by the RF power rise. This phenomenon may be explained by ultrasonic atomization theory [8-9]. Capillary waves on the liquid surface are generated by the absorbed vibration energy. Then, generated capillary waves are divided into droplet and liquid fountain if vibration of the capillary energy is enough to overcome surface tension of liquid sample. In this process, energy of the capillary may be strongly concerned to atomization process, increase of the capillary wave energy may lead to increase of drop size as well as drop speed.

In the previous research, we found that impedance of the SAW atomizer is changed during atomization process [7]. During atomization process, weight of the liquid sample as well as temperature was fluctuated. In that report, design of the impedance converter was focused to reduce variation of the impedance, and we obtained stable atomization. Compared with this report, we used same operation circuit and condition. In this report, we can conclude that operation power is also important parameter, and atomization characteristics may change by the input power. For the evolution of the atomization characteristics, further studies such as feed-back control of operation power or phase Doppler anemometry techniques are required.

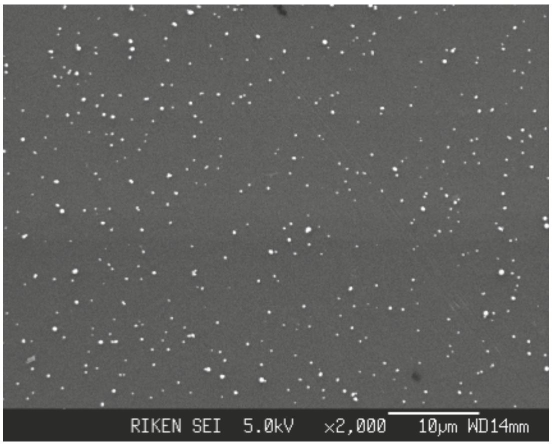


Figure 6: SEM image by surface acoustic wave atomizer and electrostatic deposition (SAW-ED) method, forward power 0.97W, magnification x2000

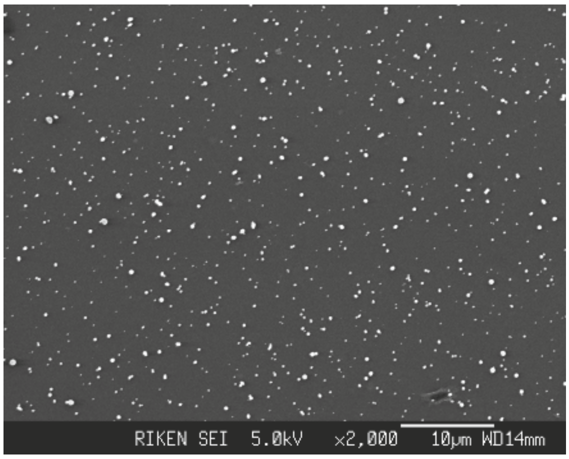


Figure 7: SEM image by surface acoustic wave atomizer and electrostatic deposition (SAW-ED) method, forward power 1 W, magnification x2000

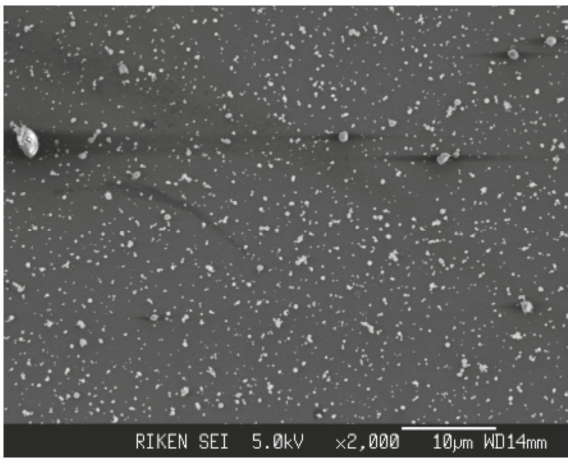


Figure 8: SEM image by surface acoustic wave atomizer and electrostatic deposition (SAW-ED) method, forward power 1.03 W, magnification x2000

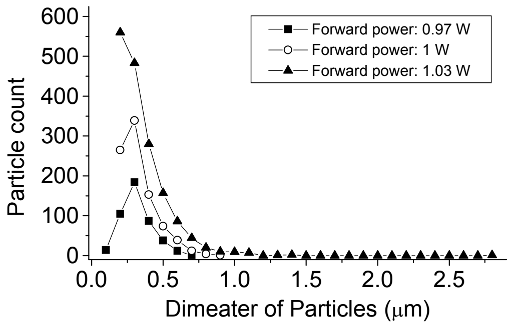


Figure 9: Size distribution of the deposited particles by SAW-ED method

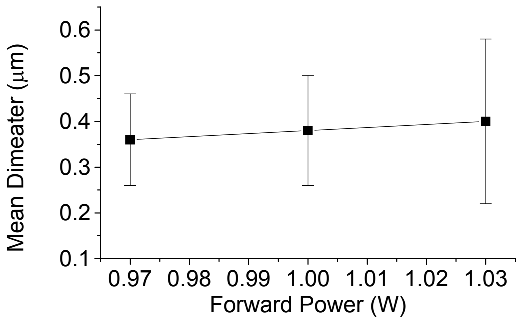


Figure 10: Mean particle diameter and standard deviation according to the forward power

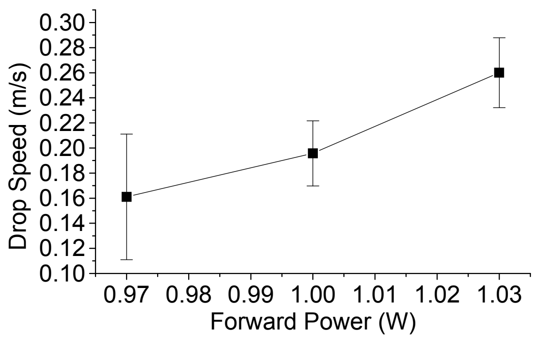


Figure 11: Drop speed measured by Laser Doppler Anemometry (LDA)

5 REFERENCES

[1] H. Kim, J. Kim, H. yang, J. Suh, T. Kim, B. Han, S. Kim, D-S. Kim, P. V. Pikhitsa, M. Choi, “Parallel patterning of nano particles via electrodynamic focusing of charged aerosols”, Nature Nanotechnology 1, pp. 117 – 121, 2006

[2] M. Kurosawa, T. Watanabe, A. Hutami, T. Higuchi, “Surface Acoustic Wave Atomizer”, Sensors and Actuators A 50, pp. 69-74, 1995

[3] J. W. Kim, Y. Yamagata, M. Takasaki, B. H. Lee, H. Ohmori, T. Higuchi, “A device for fabricating protein chips by using a surface acoustic wave atomizer and electrostatic deposition”, Sensors and Actuators B 107, pp 535-545, 2005

[4] J. Ju, Y. Yamagata, H. Ohmori, T. Higuchi, “Standing Wave Type Surface Acoustic Wave Atomizer”, Sensors and Actuators A, doi:10.1016/j.sna.2008.03.018

[5] J. Ju, Y. Yamagata, H. Ohmori, T. Higuchi, “High frequency surface acoustic wave atomizer”, Sensors and Actuators A, In press

[6] J. Ju, J.W. Kim, Y. Yamagata, H. Ohmori, K. Inoue, T. Higuchi, "Comparison Study on Electrostatic Deposition Methods using Electrospray, Surface Acoustic Wave Atomizer and Mesh-type Nebulizer", 10th International Conference on Liquid Atomization and Spray Systems (ICLASS), CD-ROM C4-10-211, 2006

[7] J. Ju, J.W. Kim, Y. Yamagata, H. Ohmori, K. Inoue, T. Higuchi, "A Study on Driving Condition of a Surface Acoustic Wave Atomizer", The 18th Symposium on Electromagnetics and Dynamics, pp. 325-328, 2006

[8] R. J. Lang, Ultrasonic Atomization of Liquids, J. Acoust. Soc. Am., 34 (1), pp. 6-8, 1962

[9] C. Chiba, Ultrasonic Spray, Sankaido, Japan, 1990, pp. 207-210, ISBN 4-381-10036-0 (in Japanese)

Scheduling Based Collision Avoidance for Multitasking Machine

Tatsuhiko Sakaguchi[1], Toshiaki Shimauchi[2] and Keiichi Shirase[2]
[1] Organization of Advanced Science and Technology, Kobe University, Japan
[2] Department of Mechanical Engineering, Graduate School of Engineering, Kobe University, Japan

Abstract
A scheduling based collision avoidance method for integrated or multitasking machine is proposed in this paper. This method consists of two procedures. One is the operation sequence sorting to find the most suitable machining schedule which has the least collisions during machining operation. The other is the dwell insertion to avoid the collisions which occur in the most suitable machining schedule. This method realizes the automatic generation of collision free NC program for integrated or multitasking machine. This method was implemented and applied to some computational experiments to verify the effectiveness of the proposed method. In the computational experiments, the proposed method is able to generate a collision free NC program automatically for integrated or multitasking machine.
Keywords:
Multitasking Machine; Integrated Machine; Collision Avoidance; Scheduling; CAM

1 INTRODUCTION

Integrated or multitasking machine has a lot of advantages such as reduction in the machining preparation time, accuracy enhancement in the loading/unloading operation, to complete complicated machining operations. However, because of complicated and multi-functional motions of the machine components of multitasking machine, the risky collisions among machine components and workpieces occur frequently. The collisions may cause the break down of the machines or the damage of the workpieces. Therefore, collision avoidance is absolutely important for perfect management of the multitasking machine. However, the generation of collision free NC program is very difficult. in order to generate a collision free NC program, the whole machining process is checked to find the collision and to modify the NC program in each case. This modification process is very time consuming. Therefore, a systematic collision avoidance method is required to modify the NC program automatically.

The objective of this research is to propose a scheduling based collision avoidance method for multitasking machine.

2 COLLISION PROBLEMS IN MULTITASKING MACHINIE

2.1 Multitasking machine

This research deals with the 2 spindles and 2 turrets type multitasking machine. Figure 1 shows the example of the target machine of this research. Both the turning and the milling process can be executed in this machine. The multitasking machine achieves the following effects.

- Reduction of the machining preparation time
- Reduction of the in-process inventory
- Accuracy enhancement by once chucking
- Need no carrier between processes
- Simultaneous machining operation

2.2 Collision problems in multitasking machine

In the multitasking machine, the turrets and the spindles move crisscross the inner space of the machine as shown in Figure 1. So the structure of this machine has the potential to collide frequently. Machine components, tools, chucks and workpieces are the object to make the collision. The collisions occurred in the multitasking machine are summarized as follows.

(1) Collision between machine components
(2) Collision between machine and tool
(3) Collision between machine and chuck
(4) Collision between machine and workpiece
(5) Collision between tool and workpiece

Meanwhile, the collisions are categorized as follows.

(a) Collisions which are derived from the machining shapes or the machining conditions.
(b) Collisions which are derived from the machines behavior.

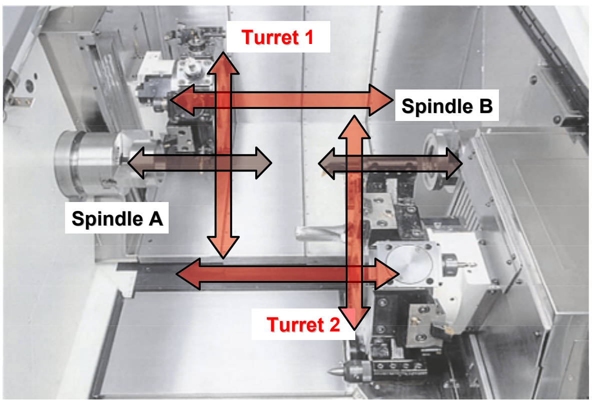


Figure 1: Example of multitasking machine

Figure 2 shows an example of collision for case (a). In this case, the tool conflicts with the workpiece because of the unsuitable tool posture. Such collisions often occur in 5 axis controlled machines[1]. In order to avoid such collisions, the machining conditions have to be modified before processing.

On the other hand, in case (b), the collisions occur because two structures in the machine run in the same time. In such collisions, by changing the process sequence, the collisions can be avoided.

In this research, only the collisions in case (b) were considered. The machining process sequence which has the least collisions during machining operation was searched by applying the scheduling method.

3 SCHEDULING BASED COLLISION AVOIDANCE

3.1 Description of machining process

The following terms are used to express the schedule of the machining process in this research.

(1) Machining process is the element to be executed in the multitasking machine, such as a rough processing, a finishing processing and a slot milling. Machining processes are represented by *MP i* (i = 1, 2, …, m).

(2) Spindle is the element that holds the workpieces. Spindles are represented by *SP j* (j = A, B).

(3) Operation represents the processing executed by a combination of a machining process and a spindle. Operations are represented by *OPij*. Operation *OPij* is the i-th process and is performed on the spindle SPj.

Process sequence represents the constraints on the sequence of machining process. Process sequences are described as shown in Table 1. All machining processes are written in the first row and the first column. The first row represents the post-process and the first column represents the pre-process. The cross points of the two processes represent the constraint of the sequence. If the sequential constraints exist between two processes, “1” is put in to the table. If it does not exist, “0” is put. For example, MP1 (Pre-process) has to be performed before MP2 (Post-process), as shown by “1” in Table 1.

3.2 Collision avoidance scheme

Two procedures are contained in the proposed method. One is the operation sequence sorting to find the most suitable machining schedule which has the least collisions during machining operation. The other is the dwell insertion to avoid the collisions which occur in the most suitable machining schedule. Figure 3 shows the flow chart of the proposed method.

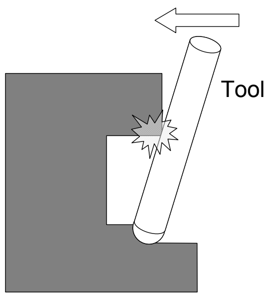


Figure 2: Collision example caused by tool posture

Table 1: Example of process sequence

Pre-process \ Post-process	MP1	MP2	MP3	MP4	MP5
MP1	0	1	1	0	0
MP2	0	0	0	1	0
MP3	0	0	0	1	0
MP4	0	0	0	0	1
MP5	0	0	0	0	1

In this research, all machining processes are predetermined before starting the collision avoidance process. Initial operation sequence is created by CAM software. CAM software does not consider the collision avoidance, so some collisions will be occurred in the initial sequence. Therefore, the operation sequence is sorted by applying the scheduling method. If no collision was occurred in the new operation sequence, this operation sequence is applied to the NC program. On the other hand, if some collisions were still occurred in the new operation sequence, the dwells are inserted for the process in order to have a wait for passing another process. If several operation sequences with no collision were found, the best operation sequence is selected based on the objective of machining, such as machining cost or processing time.

3.3 Sorting of operation sequence

Figure 4 shows an example Gantt chart of the machining process. Horizontal axis represents time and vertical axis represents spindles. The boxes and the numbers represent the machining operation and the process number, respectively. The sequence of the operations performed on each spindle can be changed. Therefore, the combinations of the operation sequences are searched by using the search tree as shown in Figure 5. In the search process, the constraints of the process sequence as shown in Table 1 have to be considered. In Figure 5, the nodes represent the operation and the arcs represent the flow of the operation. Node “0” represents the starting point of the search. All nodes whose operation can satisfy the constraints of the process

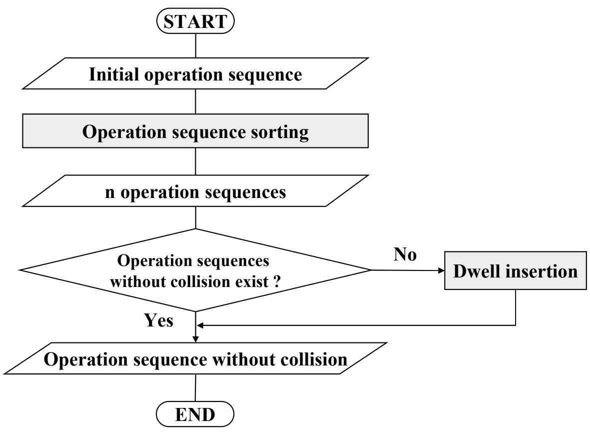


Figure 3: Flow chart of collision avoidance

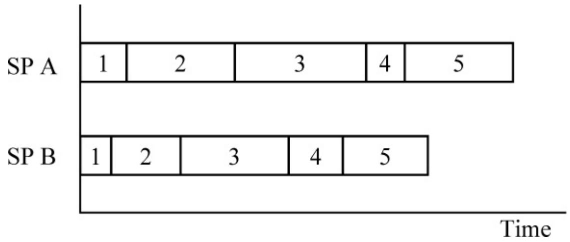


Figure 4: Example of Gantt chart

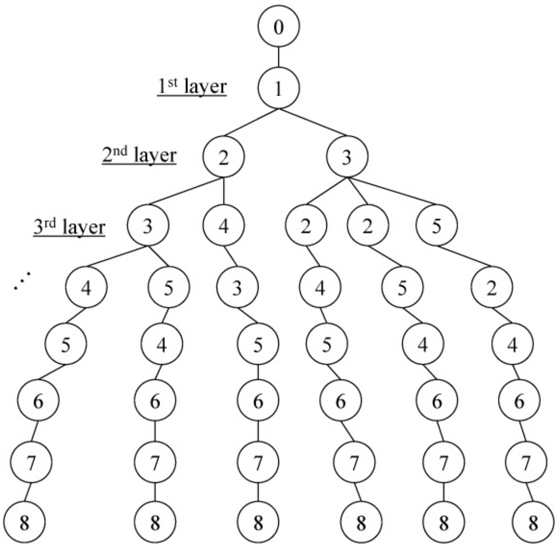


Figure 5: Search tree for operation sequence sorting

sequence are spreading the branch. For example, only node "1" can be spreading in the first layer in Figure 5, because only MP1 can satisfy the constraints of the process sequence based on Table 1. In the same way, the combinations of the operation sequences are searched. Finally, the connections of the vertical nodes are set as the candidate of the operation sequence. If the operation sequence which has no collision could not be found by the search, the operation sequence which has the least collision is selected. Then, the dwell insertion process will be carried out in order to avoid the remaining collisions.

3.4 Dwell insertion

The remaining collisions after the operation sequence sorting process are the collisions that can not be avoided by changing the operation sequence. Therefore, these collisions are dealt with the dwell insertion process.

Dwells mean the waiting time to avoid the confliction between structures in the machine. Dwells are inserted into the suitable position in the machining processes in order to avoid collisions. Figure 6 shows the work areas of each turret in the machine. The collisions occur in the over-lapping range of the work areas. Figure 7 (a) shows the Gantt chart of machining processes which is created by the operation sequence sorting. The shaded areas correspond to the time that the turret moves into the potential collision area. In order to avoid collisions, starting time of either operation should be delayed. In this research, the starting time of the operation with collision is shifted backward by the time which is calculated by the following equation.

$$LD = RFT_{a,i} - RST_{b,j} \tag{1}$$

where,

LD : Length of dwell

$RFT_{a,i}$: Finishing time of *i* th collision range at Turret *a*

$RST_{b,j}$: Starting time of *j* th collision range at Turret *b* (Turret *b* is the other side of Turret *a*)

Two patterns of the insertion of the dwell can be considered. One is to insert dwell into the operation sequence performed by Turret 1, and the other is to insert into the operation sequence performed by Turret 2. Therefore, in this research, the pattern which has the shorter machining time is selected. In the same way, dwells are inserted to avoid all collisions which remain after the operation sequence sorting.

4 COMPUTATIONAL EXPERIMENTS

4.1 Experimental conditions

The effectiveness of the proposed method was verified through the computational experiment.

In the experiments, the initial operation sequence and NC program were created by using ADMAC-Parts, which is the CAD/CAM software released by Okuma Corporation. The scheduling system which was used in the operation sorting process was implemented using Smalltalk, an object-oriented programming language. Figure 8 shows the 3-D models of

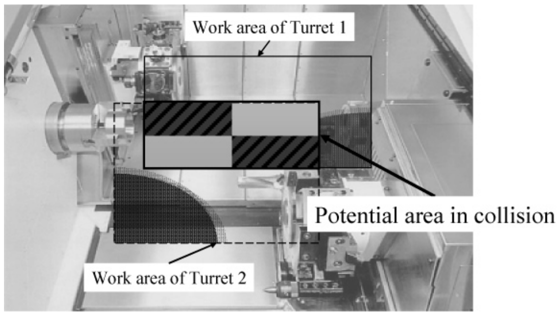


Figure 6: Potential area in collision

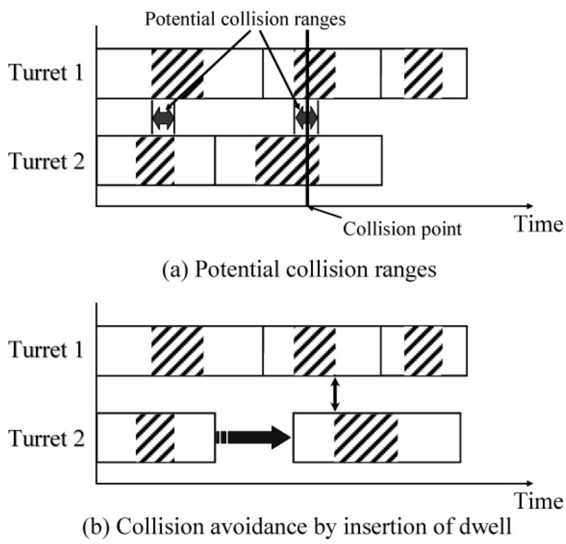


Figure 7: Collision avoidance by dwell insertion

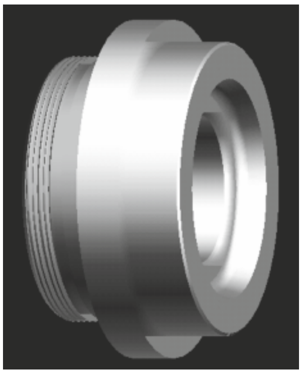


Figure 8: 3-D shape of subject product

Table 2: Machining processes in each spindle

Spindle A	Spindle B
Rough cutting (Outer shape, Edge face)	Rough cutting (Outer shape, Edge face)
Rough cutting (Outer shape)	Rough cutting (Outer shape)
Rough cutting (Inner shape)	Rough cutting (Inner shape)
Finishing (Outer shape)	Finishing (Outer shape)
Finishing (Inner shape)	Finishing (Inner shape)
Grooving	
Grooving (Finishing)	
Threading	

subject product. Table 2 shows the initial machining processes which were created by ADMAC-parts for each spindle. The collisions were detected by the machining simulator bundled with ADMAC-Parts. Exhaustive search was used as the scheduling method. The following constraints were considered in the scheduling method.

- Rough cutting for edge face of outer shape is processed first.
- Finishing processes are carried out after rough cutting process.
- Grooving processes are carried out after all finishing process.
- Threading and tapping processes are carried out at the end of all processes.

4.2 Experiment results and discussions

Table 3 shows the result of the computational experiments. Thirteen collisions were detected in the initial operation sequence. Therefore, the operation sequence sorting process was applied to the initial sequence. By applying the operation sequence sorting, the collisions were reduced by up to three. Table 4 shows the operation sequence after sorting. These collisions cannot be avoided by operation sequence sorting.

Therefore, the dwell insertion process was carried out. By this process, all collisions were avoided successfully.

From view point of the productivity, the total machining time after sorting the operation sequence did not increased much. However, the total machining time after inserting dwells increased about one minutes. However, in the real situation, the collision avoidance process was carried out manually by the skilled workers. It is a seat-of-the-pants operation and an optimal solution cannot be guaranteed. The proposed method is the systematic method to show the solution within the reasonable calculation time. Thus the increasing of total machining time should be considered tolerable.

5 CONCLUSIONS

Integrated or multitasking machine has a lot of advantages such as reduction in the machining preparation time, accuracy enhancement in the loading/unloading operation, to complete complicated machining operations. On the other hand, it faces the problem that collisions between the structures of the machine often occur, and the collision avoidance process is very difficult and time-wasting processes. In this research, a scheduling based collision avoidance method has been proposed. In this method, by sorting the operation sequence and by inserting dwells, suitable machining process is created efficiently. The effectiveness of the proposed method was verified through some computational experiments.

6 REFERENCES

[1] Nakamoto, K., Shirase, K., Morishita A., Arai, E., 2005, A New Architecture of Tool Path Generation for Five-axis Control Machining, Advances in Abrasive Technology VIII, Key Engineering Materials Vols. 291-292, Trans Tech Publications, 501-506.

Table 3: Results of computational experiment

	Number of collisions	Total machinig time
Initial sequence	13	3:38
After sorting	3	3:39
After insertion of dwells	0	4:35

Table 4: Operation sequence after sorting

Spindle A	Spindle B
Rough cutting (Outer shape, Edge face)	Rough cutting (Outer shape, Edge face)
Rough cutting (Outer shape)	Rough cutting (Outer shape)
Finishing (Outer shape)	Finishing (Outer shape)
Rough cutting (Inner shape)	Rough cutting (Inner shape)
Finishing (Inner shape)	Finishing (Inner shape)
Grooving	
Grooving (Finishing)	
Threading	

Development of a Support System of Operation Planning for Parallel Kinematic Machine Tool

Keiichi Nakamoto[1], Kouta Otake[2], Toshimichi Moriwaki[3], Keiichi Shirase[2]

[1] Dept. of Mechanical Engineering, Osaka University, Osaka, Japan

[2] Dept. of Mechanical Engineering, Kobe University, Kobe, Japan

[3] Dept. of Industrial and Systems Engineering, Setsunan University, Osaka, Japan

Abstract

Compared to conventional serial kinematic machine tools, parallel kinematic machine tools have advantages for high speed and multi-degrees of freedom positioning. However, they cannot simply achieve high speed cutting because of their relatively complex structures. In order to achieve high speed cutting on parallel kinematic machine tools, it becomes important to optimize the tool trajectories and the workpiece position. In this study, the strut motion controlling a spindle unit is simulated while cutting. And external forces, which are gravity, friction and inertia forces, acting each strut are estimated by solving the kinematic model. The tool posture, the feed direction of ball-end milling and the workpiece setup position are evaluated based on the estimated external forces during a cutting operation. Also, an evaluation method of vibrational stiffness is proposed by employing the unit vector of each strut direction obtained from the simulator developed above. The proposal method is validated by the hammering tests in various conditions. It is confirmed that the support system developed in this study can realize to reduce the variation of external forces and the influence of vibration during a cutting operation on a parallel kinematic machine tool.

Keywords:

Parallel Kinematic Machine Tool, Operation Planning, Tool Trajectory, Vibration Characteristics

1 INTRODUCTION

In aerospace, automotive and die/mold industries, high speed ball-end milling is expected to realize efficient and productive machining of sculptured surfaces. Compared to conventional serial kinematic machine tools, parallel kinematic machine tools have advantages for high speed and multi-degrees of freedom positioning. Therefore, parallel kinematic machine tools attract the attention of a lot of manufacturing engineers and researchers.

Parallel kinematic machine tools have begun to be accepted in industries through over one decade since the first appearance. There are many kinds of machine tools adopting parallel kinematic structures at the present day. The classification of realized structures and their applications in industries are reviewed with great circumstance [1]. Machining performances of parallel kinematic machine tools are investigated particularly and the results show the potential advantages for high speed cutting [2, 3].

However, operating technologies of parallel kinematic machine tools are not adequately established at this time. The reason comes from the fact that commercially available CAM systems, which not only generate NC programs but also include functionalities of the cutting simulation, focus on only avoiding collisions among the machined part, the machine tool and the cutting tool. The dynamics of cutting process including cutting forces, machine vibration and the smoothness of the cutting motion are not considered based on the specific structure of each machine tool.

Machine control method is one of essential differences between parallel kinematic machine tools and serial kinematic machine tools. Serial kinematic machine tools are generally controlled directly by NC programs in Cartesian coordinates (*X*, *Y*, *Z*, etc.) without any coordinate conversion. On the other hand, to operate parallel kinematic machine tools, tool locations and postures instructed by NC programs are converted to a specific coordinate based on the structure, such as the lengths of struts. Hence, the smoothness of the cutting motion is not guaranteed even though tool locations and postures are decided slickly in Cartesian coordinates. Furthermore, dynamic characteristics of parallel kinematic machine tools generally depend on the location and the posture of the spindle unit. In order to achieve high speed cutting on parallel kinematic machine tools, it becomes important to optimize the tool trajectories and the workpiece position.

In this study, the strut motion controlling a spindle unit is simulated while cutting. And external forces, which are gravity, friction and inertia forces, acting each strut are estimated by solving the kinematic model. The summation of external forces imposed on each strut is equal to each servo motor's driving load in non-cutting operation. The tool posture, the feed direction of ball-end milling and the workpiece setup position are evaluated based on the estimated servo motor load during a cutting operation.

Also, dynamic characteristics of a parallel kinematic machine tool differ considerably according to the tool location and posture even though the spindle unit exists within working space. Therefore, in this study, vibration characteristics like vibrational stiffness and eigen frequency are evaluated in various combinations of the tool location and posture in order to support operation planning for a parallel kinematic machine tool. An evaluation method of vibrational stiffness is proposed by employing the unit vector of each strut direction obtained from the simulator developed above. The proposal method is validated by the hammering tests in various conditions.

Figure 1: Hexapod type parallel kinematic machine tool.

Table 1: Major specification of “COSMO CENTER PM-600”.

Work space	mm	ϕ600 x 400
Tilt angle	deg.	±25
Rapid traverse speed	m/min	100
Cutting speed	m/min	40
Acceleration	m/s^2	14.7
Spindle speed	min^{-1}	30,000

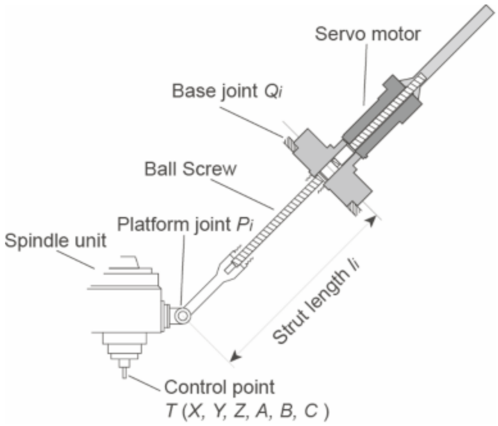


Figure 2: Schematic view of a strut and two joints.

Moreover, in some conditions, it is confirmed that effective tool postures can decrease the influence of vibration even at the same tool location.

As a result, it is confirmed that the support system developed in this study can realize to reduce the variation of external forces and the influence of vibration during a cutting operation on a parallel kinematic machine tool.

2 PARALLEL KINEMATIC MACHINE TOOL

A commercial Hexapod type parallel kinematic machine tool “COSMO CENTER PM-600” developed by Okuma Corp., shown in Figure 1, is used in our study. The machine tool consists of the Stewart platform. Experimental tests are conducted by using the actual machine tool and its major specifications are shown in Table 1.

The machine tool has six telescoping struts and each strut is driven by a built-in servo motor via ball screw simultaneously. One end of each strut is connected to a base plate by a base joint which has 2-DOF, and the other end is connected to a platform plate by a platform joint which has 3-DOF. Tool

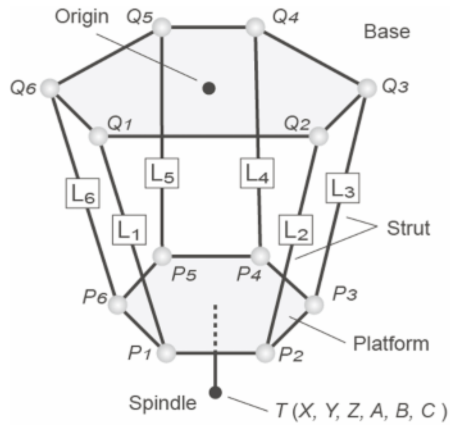


Figure 3: Hexapod type parallel kinematic machine tool.

locations and postures of a spindle unit are controlled by varying the lengths of six struts. Figure 2 shows a schematic view of a strut, with a servo motor and a spindle unit.

First of this study, the strut motion controlling a spindle unit is simulated while cutting [4], in order to optimize the tool trajectories and the workpiece position on the parallel kinematic machine tool. In a parallel kinematic machine tool, the tool locations and postures instructed by NC programs are converted to a specific coordinate in real time. The lengths of six struts correspond to the values in the specific coordinate to control the Hexapod type parallel kinematic machine tool. The conversion is called the inverse kinematic problem, which can be calculated geometrically as follows :

1. The relative distance Ri (i=1,2,···6) between the control point of the cutting tool (or the spindle unit) and each platform joint is decided.
2. The position in machine coordinate system of each platform joint Pi is obtained by above Ri and the tool locations and posture T (X, Y, Z, A, B, C), as shown in Figure 3.
3. The lengths of six struts L (l_1, l_2, ··· l_6) are calculated as the distance between Pi and the position in machine coordinate system of each base joint Qi.

The parameters Ri and Qi are contained within kinematic parameters of the parallel kinematic machine tool. The same calculation step of the conversion is also performed in real equipments.

On the other hand, the problem calculating the tool locations and posture from the strut lengths is called as the forward kinematic problem. This problem can not be solved analytically and is calculated numerically using Jacobian matrix as the following equation :

$$T_{k+1} = T_k - J(T_k)^{-1}(L_k - L) \tag{1}$$

where T_k is the k-th tool location and posture, L_k is the k-th lengths of six struts, and J is the Jacobian matrix. Equation. (1) is calculated iteratively until the difference between L_k and the target struts’ lengths L becomes sufficient small. To solve the forward kinematic problem is not required to control the machine tool, but it is extremely important for calibrating various kinematic parameters. Hence, Nakagawa and Ibaraki solve this problem in their studies [5 - 7]. In this study, the forward kinematic problem is also solved to intend to generate tool trajectories with a smooth strut motion.

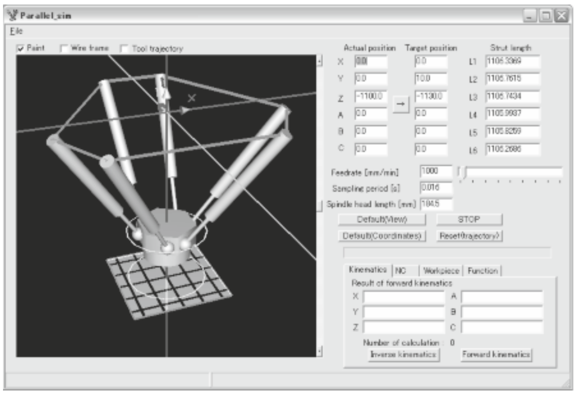


Figure 4: Developed strut motion simulator.

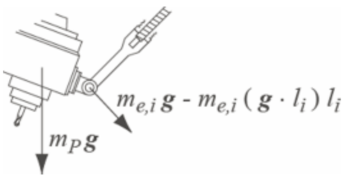


Figure 5: Force acting on the spindle unit and joint.

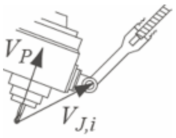


Figure 6: Vector toward centre of gravity.

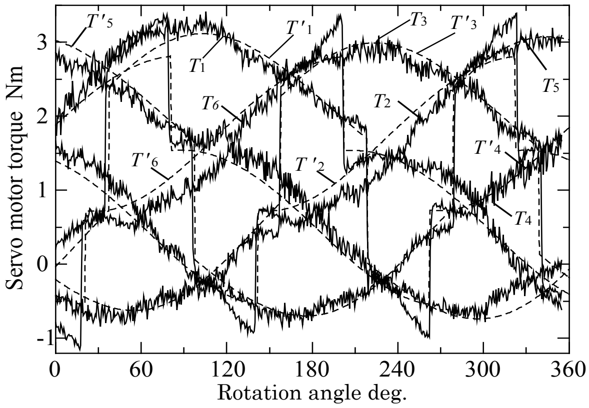


Figure 7: Comparison between measured (T_i) and estimated servo motor torque (T_i') in a circular operation. Centre location: (x,y,z)=(0,0,-1000), Radius = 150(mm)

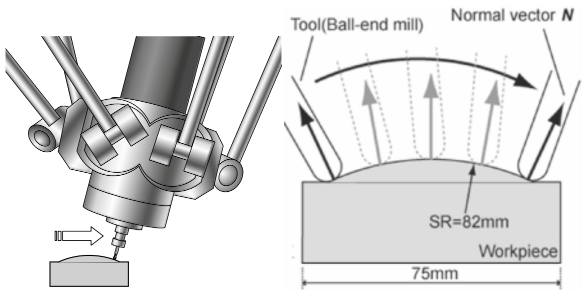


Figure 8: Overview of assumed cutting operation.

3 VARIATION OF EXTERNAL FORCES

As mentioned in Section 2, Nakagawa et al. have proposed the calibration methodology of various kinematic parameters of a Hexapod type parallel kinematic machine tool [5]. Also, Ibaraki et al. have improved the motion accuracy by considering gravity-induced errors based on the elastic deformation of each strut [6] and proposed a disturbance estimation methodology to monitor cutting forces [7]. Their researches will contribute to improve the static motion accuracy of a parallel kinematic machine tool. In this study, in order to evaluate the tool posture, the feed direction of ball-end milling and the workpiece setup position based on the estimated servo motor load during a cutting operation, their methods are applied in a developed strut motion simulator.

In the developed simulator, the strut motion controlling a spindle unit is simulated while cutting. And external forces, which are gravity, friction and inertia forces, acting each strut are estimated by solving the kinematic model [6, 7]. Figure 4 shows the GUI of the developed strut motion simulator in this study. The specific functions are as follows:

a. Calculation of each strut's length in various tool locations and postures of the spindle unit.
b. Simulation of cutting motions given by NC programs.
c. Calculation of telescoping velocity and acceleration of each strut.
d. Estimation of external forces (gravity, friction and inertia) acting on each strut.

Kinematic models used in this study are shown in Figures 5 and 6. m_p and $m_{e,i}$ represent mass of the spindle unit and equivalent mass of the strut and the ball screw etc., respectively. Also, V_p and $V_{J,i}$ represent position vectors from the tool tip to the centre of gravity of the spindle unit and from the tool tip to the centre of the platform joint. Gravity acting on each strut is given by solving equilibrium equations of force and moment. Friction and inertia are estimated in the similar manner [6, 7].

The summation of external forces imposed on each strut is equal to each servo motor's driving load in non-cutting operation. In order to verify this estimation method, a circular motion test is conducted. The comparison of measured (T_i) and estimated (T_i') servo motor torque profiles are shown in Figure 7. In this figure, T_i is the servo motor torque of the *i*-th strut. The mean of error between the measured and the estimated servo motor torque profiles is 0.10 Nm.

Tool trajectory and workpiece position on the machine table are evaluated by using the developed strut motion simulator. Figure 8 shows the overview of the assumed cutting operation, which is finish ball end milling (ϕ10mm). First, external forces for each strut $F_{e,i}$(t) (i=1~6) are estimated on the tool trajectory of a part of aspherical surface (R=82.8mm). And, the value $F_{v,i}$ given in Equation (2) is defined for each strut. Finally, F_v which equals to the maximum value of $F_{v,i}$ is employed to evaluate the tool trajectory and the workpiece position.

$$F_{v,i} = Max(F_{e,i}(t)) - Min(F_{e,i}(t)) \tag{2}$$

In this study, the effects of the tool posture, the workpiece position and the feed direction on servo motor torque are verified. The definition of the tool posture is shown in Figure 9. The angles α and β are each inclination angles to the feed direction and to the pick feed direction respectively. The evaluated workpiece positions on the machine table are

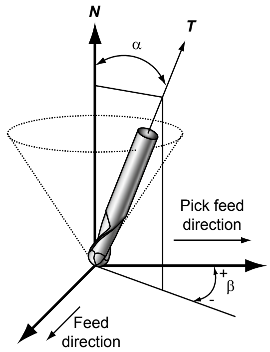


Figure 9: Tool posture.

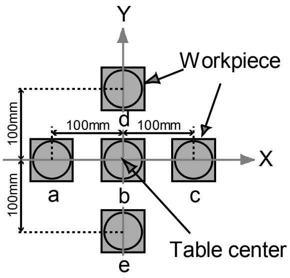


Figure 10: Workpiece position.

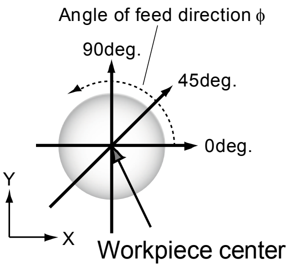


Figure 11: Definition of feed direction.

represented in Figure 10. The definition of the feed direction is shown in Figure 11.

As shown in Figure 12, the interaction between the tool posture and the workpiece position is investigated to reduce the variation of external forces. In this figure, the interactions between α, β and F_v in various workpiece positions are illustrated. The vertical scale is the external forces converted from the estimated servo motor torque. Note that, the result of the workpiece positions (d) and (e) are spared. The tendency of estimated forces especially differs according to the tool posture.

Furthermore, the interaction between α, β and F_v in various feed directions is investigated, and the results are shown in Figure 13. In this case study, the workpiece centre position is fixed at the center of the table (X=0, Y=0). The maximum point and at that α, β varies in every feed direction. It can be seen that the effect of feed direction is also as an important factor as the workpiece position for this parallel kinematic machine tool to reduce the variation of external forces. Based on these results, it is verified that this support system is effective to optimize the tool trajectory and the workpiece position by considering the variation of external forces.

4 INFLUENCE OF MACHINE VIBRATION

Vibration characteristics are also evaluated in various combinations of the tool location and posture in order to support operation planning for a parallel kinematic machine tool. Dynamic characteristics of a parallel kinematic machine tool differ considerably according to the tool location and posture even though the spindle unit exists within working space, and affect the machined results.

An evaluation method of vibrational stiffness is proposed in this study by employing the unit vector of each strut direction obtained from the simulator developed above. The unit vector of each strut direction is calculated as the following equation :

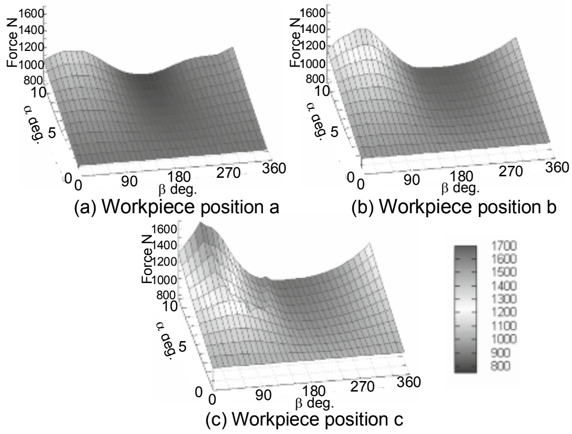


Figure 12: Variation of forces at various workpiece positions.

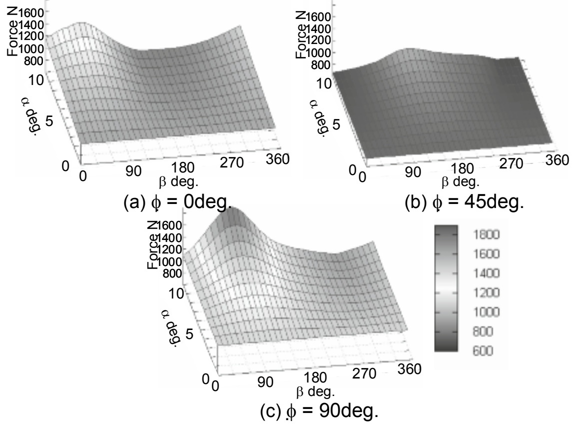


Figure 13: Variation of forces at various feed directions.

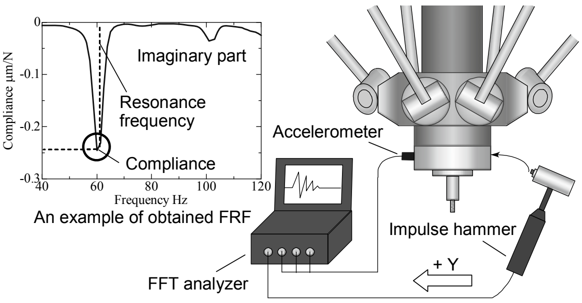


Figure 14: Overview of conducted hammering test.

$$u_i = \frac{P_i - Q_i}{L_i} \tag{3}$$

where P_i is the *i*-th platform joint coordinate, Q_i is the *i*-th base joint coordinate, and L_i is the *i*-th strut length, as shown in Figure 2. The sum of the absolute value of u_i's element is used to estimate dynamic characteristics in each direction (*X, Y, Z*). For example, in case the evaluated value in Y direction is large, it is recognized that the struts directions are close to Y direction and flexural vibration of struts is difficult to occur.

Figure 14 shows the overview of the conducted hammering test. The proposal evaluation method is validated by the hammering tests in various conditions base on the measured compliance and frequency of the primary resonance. The

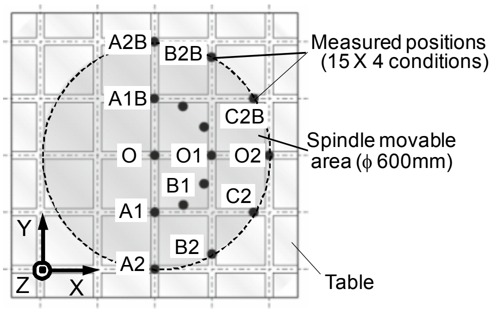
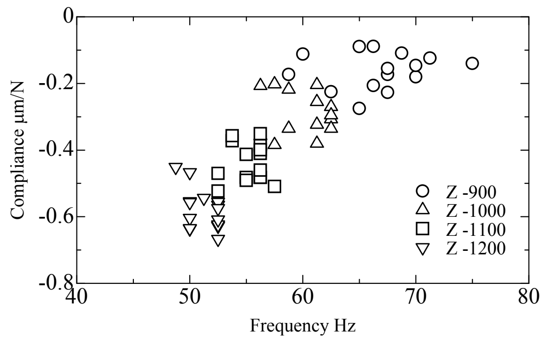


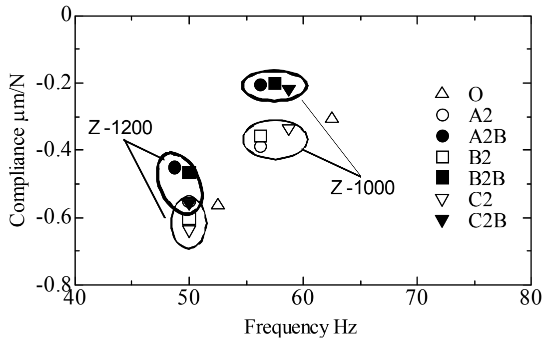
Figure 15: Measured points of hammering test.

Table 2: XY coordinate of measured points.

A2	(0, -300)	B2	(150, -259.8)	C2	(259.8, -150)
A2B	(0, 300)	B2B	(150, 259.8)	C2B	(259.8, 150)
A1	(0, -150)	B1	(75, -129.9)	C1	(129.9, -75)
A1B	(0, 150)	B1B	(75, 129.9)	C1B	(129.9, 75)
O	(0, 0)	O1	(150, 0)	O2	(300, 0)



(a) All tool locations: 60 conditions



(b) Z -1000 and Z -1200

Figure 16: Measurement results (Tool posture is fixed).

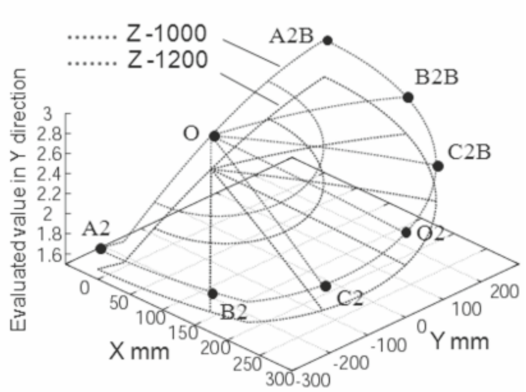
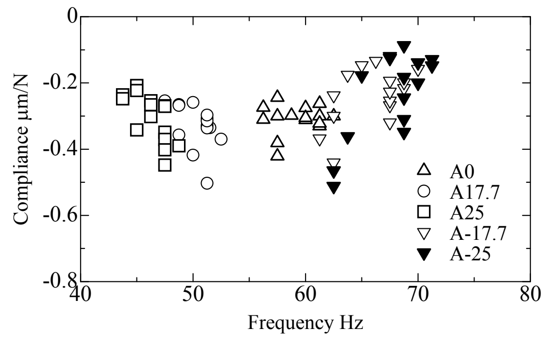
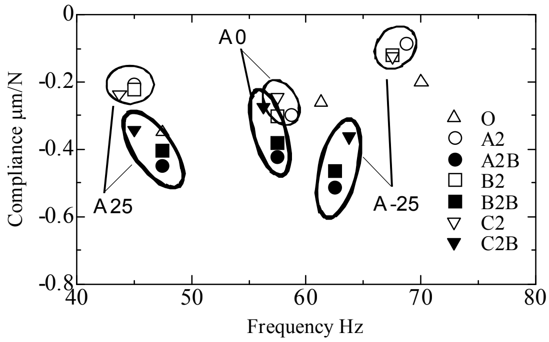


Figure 17: Evaluated value in Y direction.



(a) All tool locations: 75 conditions



(b) A -25, A 0 and A 25

Figure 18: Measurement results (Tool posture is changed).

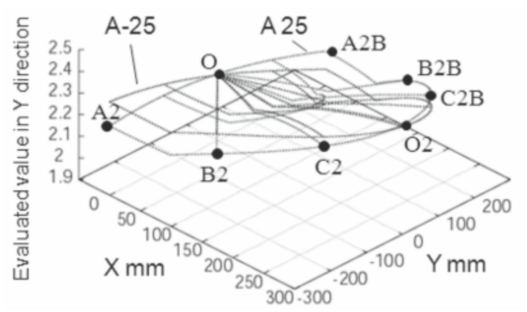


Figure 19: Evaluated value in X direction.

measured points on the table (15points in XY plane) are shown in Figure 15 and summarized in Table 2. Note that, the test is not conducted in the half of the spindle movable area because the structure of the parallel kinematic machine tool is symmetry to the Y axis.

The measurement results in Y direction impulse are shown in Figure 16 in case the spindle unit is kept upright and fixed to Z axis. 4 different Z levels (-900, -1000, -1100, -1200) in each XY coordinate are prepared. From Figure 16(a), it can be seen that vibration characteristics strongly depends on Z levels of the spindle unit. It means that vibrational stiffness becomes better in case the struts length become shorter. Compared to the proposed evaluated value shown in Figure 17, it is recognized that there is a common trend. Higher Z level (-1000) makes high evaluated values and vibration characteristics is improved in +Y area.

The measurement results in X direction impulse are similarly shown in Figure 18. In this experiment, 5 different A values (-25, -17.7, 0, 17.7, 25 deg. around X axis) in each XY coordinate are prepared. The obtained results differ from the last experiment shown in Figure 16 and vibration characteristics is improved in -Y area. Figure 19 shows the proposed evaluated value in this case. It is confirmed that the evaluated values indicate qualitative character, which is similar to the measurement results. Specifically, the better rotational angle of the spindle unit related to the tool posture differs between +Y area and -Y area. Also, it is recognized that effective tool postures can decrease the influence of vibration even at the same tool location.

From the obtained results, the capability of the proposed evaluation method of vibrational stiffness is confirmed in order to reduce the influence of vibration during a cutting operation on a parallel kinematic machine tool. The developed strut motion simulator has prospects for supporting operation planning.

5 SUMMARY

In order to achieve high speed cutting on parallel kinematic machine tools, it becomes important to optimize the tool trajectories and the workpiece position. This paper presents a development of a support system for operation planning of a parallel kinematic machine tool. The conclusions are summarized as follows :

1. In order to establish a methodology to evaluate generated tool trajectories for a parallel kinematic machine tool, a strut motion simulator is developed. The simulator can solve the inverse and the forward kinematic problems. The cutting motion instructed by NC programs can be checked apparently.
2. A new feature is added to the strut motion simulator. External forces, which are gravity, friction and inertia forces, acting each strut are estimated by solving the kinematic model. The tool posture, the feed direction of ball-end milling and the workpiece setup position are evaluated based on the estimated servo motor load during a cutting operation.
3. An evaluation method of vibrational stiffness is proposed by employing the unit vector of each strut direction obtained from the developed strut motion simulator. Vibration characteristics like vibrational stiffness and eigen frequency are evaluated in various combinations of the tool location and posture in order to support operation planning for a parallel kinematic machine tool. The proposal method is validated by the hammering tests in various conditions.

To establish an optimization scheme of cutting parameters such as effective tool trajectories, feed direction and workpiece position which the support system suggests, becomes an issue in the future.

6 ACKNOWLEDGMENTS

This work was supported in part by the Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research No. 17760106 and No. 19760088.

7 REFERENCES

[1] Weck, M., Staimer, D., 2002, Parallel Kinematic Machine Tools – Current State and Future Potentials, Annals of the CIRP, 51/2: 67l-683.

[2] Geldart, M., Webb, P., Larsson, H., Backstrom, M., Gindy, N., Rask, K., 2003, A Direct Comparison of the Machining Performance of a Variax 5 Axis Parallel Kinetic Machining Centre with Conventional 3 and 5 Axis Machine Tools, Int. J. of Machine Tools & Manufacture, 43: 1107-1116.

[3] Terrier, M., Dugas, A., Hascoët, J. Y., 2004, Qualification of Parallel Kinematics Machines in High-Speed Milling on Free Form Surfaces, Int. J. of Machine Tools & Manufacture, 44: 865-877.

[4] Nakamoto, K., Mori, H., Moriwaki, T., 2005, A Fundamental Evaluation Methodology of Tool Trajectory for High Speed Cutting on Parallel Kinematic Machine Tool, Proc. of the 9th Int. Conf. on Mechatronics Technology, ICMT-72 in CD-ROM.

[5] Nakagawa, M., Matsushita, T., Watanabe, S., Kakino, Y., Ihara, Y., 2002, The Improvement of Motion Accuracy of Hexapod-Type Machine Tools and Its Machining Performance, Proc. of 2002 Japan-USA Symp. on Flexible Automation, JS005 in CD-ROM.

[6] Ibaraki, S., Okuda, T., Kakino, Y., Nakagawa, M., Matsushita, T., Ando, T., 2004, Compensation of Gravity- Induced Errors on a Hexapod-Type Parallel Kinematic Machine Tool, JSME Int. J. Series C, 47/1: 160-167.

[7] Okuda, T., Ibaraki, S., Kakino, Y., Nakagawa, M., Matsushita, T., 2004, Disturbance Estimation on a Hexapod- Type Parallel Kinematic Machine Tool by Using a Disturbance Observer, Proc. of 2004 Japan-USA Symp. on Flexible Automation, JL025 in CD-ROM.

Method of Process Parameter Identification and Resist Profile Design for Thin-Film Pattern Formation

Eiji Morinaga[1], Yutaka Matsuura[1], Ryohei Satoh[2], Kouji Nakagawa[3], Reo Usui[1], Yoshiharu Iwata[2], Hidefumi Wakamatsu[1], and Eiji Arai[1]

[1] Division of Materials and Manufacturing Science, Osaka University, Osaka, Japan

[2] Center of Advanced Science and Innovation, Osaka University, Osaka, Japan

[3] Research Center, ASAHI GLASS CO., LTD., Kanagawa, Japan

Abstract

For realization of the next generation thin-film pattern formation technology by the lift-off method, the inversely-tapered resist profile with interstice has been proposed, and its fundamental effectiveness was shown. Furthermore, a method was given, for calculation of thickness distribution of the film pattern formed with the proposed resist profile, and a design method of the proposed resist profile. However, in these methods, there are some problems in identification of process parameters and definition of proper resist profile. This paper proposes a systematic method for identification of the parameters and a design method of proper resist profile.

Keywords:

Parameter identification; Resist profile design; Deposition

1 INTRODUCTION

Thin-film pattern formation technology is required in the electronics field such as displays and semiconductors. Recently, it has been strongly desired to achieve low environmental damage and high cost-performance in the pattern formation process. Therefore, in place of the etching method which is the mainstream, the lift-off method is expected to be effective. The lift-off process has applied to pattern formation in LSI field [1] [2] [3]. However, in the lift-off process, there have been problems caused by inappropriate resist profiles, such as difficulty in resist removal and low pattern accuracy due to formed film on the resist side wall.

For these problems, the inversely-tapered resist profile with interstice has been proposed, and its fundamental effectiveness was shown [4]. Furthermore, a computation method of film thickness distribution formed by the proposed profile was given, and a design method of the proposed profile based on this computation method was suggested. However, this computation method requires identification of process parameters by trial and errors. Furthermore, the design method outputs the optimal resist profile from only the vewpoint of height of the interstice.

This paper proposes a systematic method for identification of the parameters and a design method of proper resist profile from the vewpoint of the height of interstic and taper incline.

2 PROCESS PARAMETER IDENTIFICATION

The inversely-tapered resist profile with interstice shown in Figure 1 is considered, where H is thickness of the resist; $2a$ and $2b'$ are resist pattern widths at top and bottom of the inversely tapered part, respectively; h and w are thickness and depth of the interstice, respectively, l is length from the edge of the interstice to the edge of the electrode.

The film thickness distribution in this resist profile, $T(x,h,\sigma,\kappa)$, can be predicted as follows [4]:

$$T(x,h,\sigma,\kappa) := T_F(x,h,\sigma) + \kappa T_G(x,h,\sigma). \tag{1}$$

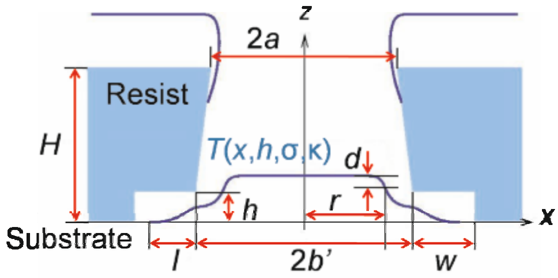


Figure 1: Resist profile and expected film pattern.

About the details of $T_F(x,h,\sigma)$ and $T_G(x,h,\sigma)$ readers should refer to Appendix A. In $T(x,h,\sigma,\kappa)$, $\sigma(\geq 0)$ is the parameter to show breadth of the angular distribution of depositing material onto an arbitrary point on the surface of horizontal substrate without any resist patterns. $\kappa(\geq 0)$ is the parameter to show the ratio of the two processes; the F process, depositing particles reach the substrate directly; the G process, depositing particles reach the substrate via the resist sidewall. These two parameters are the process parameter, and required to be identified by calculations and deposition experiments. However, at present, this identification is carried out by trial and errors.

In this section, an evaluation index of consistency between experimental and computational results will be defined, and then the identification method of σ and κ will be proposed, which is given as an optimization problem based on the index.

First of all, the consistency between experimental and computational results is defined by the following features:

1. the cross sectional area of the thin-film pattern;
2. the film thickness beneath the edge of the interstice;
3. the width of the area where the film is thick;

About the feature 1, the consistency is evaluated by the smallness of the following index:

$$\int_{-(b'+l)}^{b'+l} \left| T_s(x,h,\sigma,\kappa) - T_e(x) \right| dx \ , \tag{2}$$

where $T_s(x,h,\sigma,\kappa)$ and $T_e(x)$ are film thickness distribution obtained by calculation and an experiment, respectively, l is the width of the film under the interstice.

However, it is difficult to express $T_e(x)$ as a mathematical function, and film thickness at only a finite number of positions x_k (k=1,…,N) can be obtained. Therefore, the consistency is evaluated by the following discrete index:

$$\sum_{k=1}^{N} \left| T_s(x_k,h,\sigma,\kappa) - T_e(x_k) \right| \quad \left(-(b'+l) \le x_k \le b'+l \right). \tag{3}$$

About the feature 2, the consistency is evaluated by the smallness of $T_s(\pm b',h,\sigma,\kappa)$-$T_e(\pm b')$. However, minimization of this index may result in σ and κ such that $T_s(\pm b',h,\sigma,\kappa)<T_e(\pm b')$. In this case, the film pattern formed with the resist profile designed with the σ and κ, would be also formed on the resist sidewall. For the reason, $T_s(\pm b',h,\sigma,\kappa)$ is required to be larger than $T_e(\pm b')$. On the other hand, too much largeness of $T_s(\pm b',h,\sigma,\kappa)$-$T_e(\pm b')$ is not good from the viewpoint of the consistency. Therefore, the consistency is evaluated by the following inequality:

$$0 < T_s(\pm b',h,\sigma,\kappa) - T_e(\pm b') \le \varepsilon \ , \tag{4}$$

where ε is a given positive number which is sufficiently small.

About the feature 3, the consistency is evaluated by

$$r_s - r_e \ , \tag{5}$$

where r is defined as the value by which the following equation holds with a given value d:

$$T(0,h,\sigma,\kappa) - T(\pm r,h,\sigma,\kappa) = d \ .$$

By the above arguments, identification of the process parameters is defined as the minimization problem of (3) and (5) under the condition (4). From the point of view of product functions, this minimization problem can be re-defined as follows.

However, this optimization problem is inappropriate from the point of view of mathematical optimization method, since r_s is difficult to describe in analytical form due to the complexity of the function $T(x,h,\sigma,\kappa)$. For small electrical resistance of film pattern electrodes, the consistency from the viewpoint of the feature 1 in the area where film thickness is large, is vey important. For this reason, the indices (3) and (5) are merged into the following index:

$$\sum_{k=1}^{N} \left| T_s(x_k,h,\sigma,\kappa) - T_e(x_k) \right| \quad \left(-r \le x_k \le r \right). \tag{6}$$

In the area where the film thickness $T_s(x,h,\sigma,\kappa)$ is large, $T_s(x,h,\sigma,\kappa)$ can be approximated by $T_F(x,h,\sigma)$, since $T_G(x,h,\sigma)$, which is the film thickness by the G process, tends to be small. Therefore, the index (6) can be re-defined as

$$\sum_{k=1}^{N} \left| T_F(x_k,h,\sigma) - T_e(x_k) \right| \quad \left(-r \le x_k \le r \right), \tag{7}$$

which is a function of only σ.

The condition (4) is also transformed as follows:

$$\frac{T_e(\pm b') - T_F(\pm b',h,\sigma)}{T_G(\pm b',h,\sigma)} < \kappa \le \frac{T_e(\pm b') - T_F(\pm b',h,\sigma) + \varepsilon}{T_G(\pm b',h,\sigma)} \ . \tag{8}$$

This inequality means that κ can be given by this inequality if σ is given by (7). However, the following inequality must be satisfied such that a solution of the inequality (8) exists:

$$T_F(\pm b',h,\sigma) - \left[T_e(\pm b') + \varepsilon \right] \le 0 \ . \tag{9}$$

As a result, the identification problem is defined mathematically as the profile of finding σ which minimizes (7) under the condition (9) and finding κ which satisfies (8).

3 RESIST PROFILE DESIGN

By using the methods suggested in the paper [4] and Section 2, the film thickness distribution formed with a given inversely-tapered resist profile with interstice can be calculated. It is necessary to design, with this method, the resist profile with which the formed film pattern satisfies specifications. In this section, the design method of this resist profile is discussed.

The inversely-tapered resist profile with interstice is determined by the five parameters: H, h, a, b', w. H and b' are dominantly decided by specifications for the thin-filem pattern and demands from front-end processes. For this reason, the design parameters of the resist profile are h, a and w. However, w is decided by the length from the edge of the interstice to the edge of the electrode l, which can be calculated after deciding h and a. Therefore, the resist profile design is considered to be decision of h and a. For this reason, the function of film thickness distribution $T(x,h,\sigma,\kappa)$ described in Section 2, is re-described as $T(x,h,a)$ from the point of view of resist profile design.

The paper [4] proposed a resist profile design method in which the design parameter is only h. This method minimizes h by interative calculation of film thickness beneath the interstice based on a bisectional method, for the purpose of minimization of l.

From the viewpoint of the pattern film electrode, its electrical resistance is also important as well as pattern accuracy. It is necessary to properly design the taper angle of the resist profile since it affects the cross sectional area of the film pattern which control the electrical resistance. However, adequate resist profile can be designed from the viewpoint of the taper angle. This section discusses the ideal resist profile from the viewpoints of both the height of the interstice and the taper-incline, and then the evaluation function for the resist profile is defined. The resist profile design problem is defined as the minimization problem of the function.

It is demanded that the edge of the interstice to the edge of the electrode l is small since l decides pattern accuracy. For this reason h should be small as possible. However, too small h results in formation of the film on the resist sidewall. Therefore it is desirable that h is minimum in the range of $h>T(\pm b',a,h)$, as described in the paper [4]. On the other hand, from the viewpoint of electrical resistance of the film electrode, it is desirable that the area where the film is thick is wide as possible. This is achieved by increasing a as possible. However, it is necessary that $a<b'$ is satisfied for the inversely-tapered resist profile. Therefore, it is desirable that

a is maximum in the range of $a<b'$. The followings are the conditions of the ideal resist profile:

1. Minimize *h* such that $h>T(\pm b',a,h)$ holds.
2. Maximize *a* such that $a<b'$ holds.

The resist profile which satisfies these conditions realizes small electrical resistance of film pattern electrodes and good pattern accuracy.

There is, however, a trade-off between the conditions. Therefore, the following evaluation function is defined:

$$[h-T(\pm b',a,h)]+A(b'-a), \tag{10}$$

where A is a weghting parameter. By minimizng (10) under the conditions $h>T(\pm b',a,h)$ and $a<b'$, the resist profile can be designed well from the points of view of both pattern accuracy and electrical resistance.

4 CASE SYUDY

For evaluation, the proposed methods were applied to the experimental result (H=32.8[μm], h=5.7[μm], a=12.3[μm], b'=14.1[μm]) shown in [4]. 10 points on the substrate x_k (k=1,…,10) were chosen arbitrarily in the area where the film is thick, and then the film thickness over the each points $T_e(x_k)$ was measured. Table 1 shows the values of x_k and $T_e(x_k)$. The specified film thickness $\underline{T}$ and the film thickness beneath the edge of the interstice $T_e(\pm b')$ were also measured, and $\underline{T}$=2.35[μm] and $T_e(\pm b')$=0.552[μm] were obtained. The proposed identification method was applied, by using these values. Since the process parameter identification is a constrained nonlinear optimization problem, the penalty function method was utilized and the following evaluation function was defined:

$$\begin{aligned} &S_1(\sigma) \\ &= \sum_{k=1}^{N} \left|T_F(x_k,h,\sigma)-T_e(x_k)\right| + \rho\{\max[0, T_F(\pm b',h,\sigma)-(T_e(\pm b')+\varepsilon)]\} \\ &(-r \le x_k \le r). \end{aligned} \tag{11}$$

By optimization using the gradient method with ρ=10, the initial value σ=5.0×10^{-2} and ε=0.01[μm], σ=7.73×10^{-2} and $S_1(\sigma)$= 0.417 were obtained. The inequality (8) was

$$1.24\times10^{-5} < \kappa \le 7.19\times10^{-5},$$

and then κ=7.19×10^{-5} was selected. Figure 2 shows the experimental and calculated thickness distribution. The result indicates that this identification is successful.

By using these identified values, H=25.0[μm], and b'=15.0[μm], the resist profile was designed by the proposed design method. The penalty function method was also used. The evaluation function was defined as follows:

$$\begin{aligned} &S_2(a,h) \\ &= [h-T(\pm b',a,h)]+A(b'-a)+\rho[\max(0,a-b')+\max(0,T(\pm b',a,h)-h)] \\ &. \end{aligned} \tag{12}$$

By optimization using the gradient method with ρ=10, A=2.0, the initial values a=10.0[μm] and h=3.0[μm], $S_2(a,h)$= 1.57, a=14.2[μm] and h=1.10[μm] were obtained. By calculation of the film thickness distribution for this resist profile, l=2.45[μm] and $T_s(\pm b)$=1.04 were obtained. w=3.0 was decided based on the value of l. The result of this resist profile design is shown in Figure 3(a), and the result by the existing method (h=5.64×10^{-2}[μm], a=10.9[μm], l=8.23×10^{-2}[μm], $T_s(\pm b)$-=5.61×10^{-2}[μm]) is also shown in Figure 3(b). The calculation results of the film pattern cross sectional area obtained with each designed resist profile, were 82.7[μm^2] (the proposed method), and 65.3[μm^2] (the existing method). By the proposed method, the resist profile is successfully designed well in the sense of both pattern accuracy and electrical resistance.

Table 1:The result of x_k and $T(x_k)$.

x_k	[μm]	$T(x_k)$	[μm]
x_1	1	$T(x_1)$	2.34
x_2	2	$T(x_2)$	2.29
x_3	3	$T(x_3)$	2.29
x_4	4	$T(x_4)$	2.39
x_5	5	$T(x_5)$	2.34
x_6	6	$T(x_6)$	2.26
x_7	7	$T(x_7)$	2.29
x_8	8	$T(x_8)$	2.18
x_9	9	$T(x_9)$	2.08
x_{10}	10	$T(x_{10})$	1.97

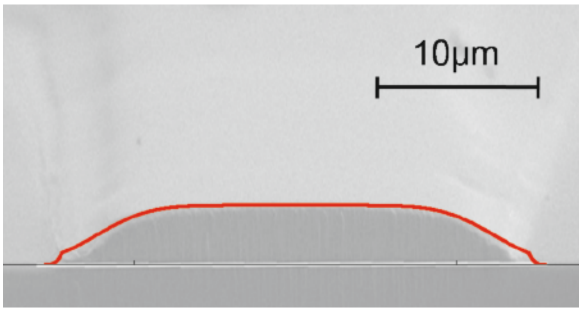
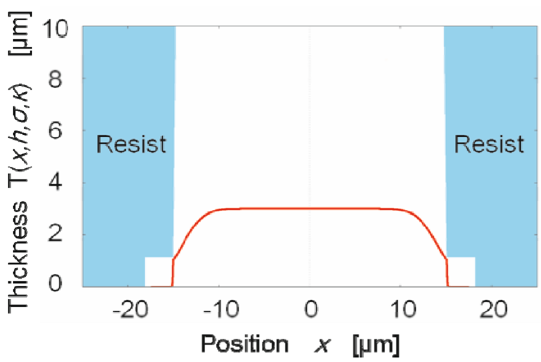
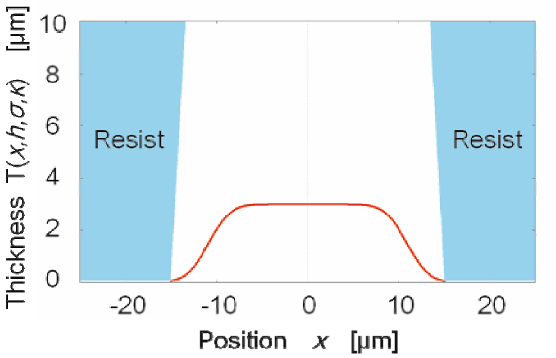


Figure 2: A calculation result and an experimental result.



(a) Result by the proposed method.



(b) Result by the existing method.

Figure 3: Result of the resist design and calculation result.

5 SUMMARY

This paper has proposed two methods which improves the film pattern thickness calculation method suggested in [4] to the film pattern formation by the lift-off process: One is a method for identification of the process parameters without trial and error, and the other is a design method of proper resist profile from the viewpoint of both pattern accuracy and electrical resistance. By the case study, the effectiveness of these proposed methods was verified.

6 REFERENCES

[1] Broers, A. N., Lean, E. G., Hatzakis, M., 1969, 1.75-GHz acoustic-surface-wave transducer fabricated by an electron beam, Applied Physics Letters, 15:98-101.

[2] Sakurai, T. and Serikawa, T., 1979, Lift-off metallization of sputtered Al alloy films, Journal of the Electrochemical Society, 126:1257-1260.

[3] Serikawa, T. and Yachi, T., 1981, Lift-off patterning of sputtered SiO2 films, Journal of the Electrochemical Society, 128:918–919.

[4] Morinaga, E., Satoh, R., Nakagawa, K., Miyagawa, H., Usui, R., Iwata, Y., 2007, Research on next-generation manufacturing method of plasma display panels via lift-off process, Society for Information Display 2007 International Symposium Digest of Technical Papers, 1705-1708.

APPENDIX

A Details of the film thickness distribution equation [4]

Deposition process is modelled by the following two processes; the F process, depositing particles reach the substrate directly; the G process, depositing particles reach the substrate via the resist sidewall. The following assumption is introduced:

Assumption 1. Depositing particles make perfectly inelastic collision at the surface of the substrate, and perfectly elastic collision at the resist sidewall, as shown in Figure 4.

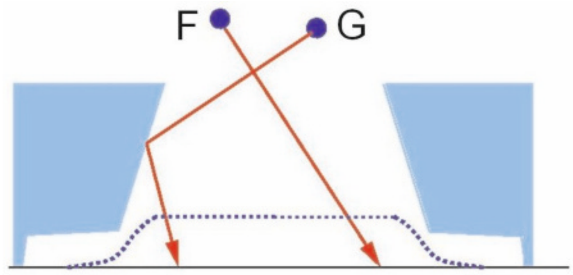


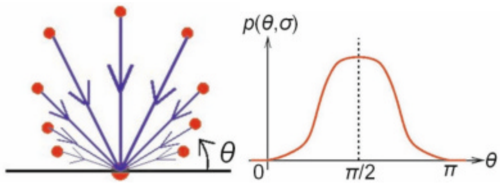
Figure 4: Assumed processes of deposition phenomenon.

First of all, an arbitrary point on the surface of horizontal substrate without any resist patterns is considered. Probability of collision of a particle at the point depends on its incoming angle θ; it is largest for vertical direction, and decreases as the angle approaches horizontal direction, as shown in Figure 5(a). Therefore, as shown in Figure 5(b), the following probability distribution function $p(\theta,\sigma)$ based on the gaussian-like function with variance s is assumed:

$$p(\theta,\sigma) := \begin{cases} \dfrac{1}{Q(\sigma)}\left[q(\theta,\sigma)-q(0,\sigma)\right], \theta \in [0,\pi] \\ 0, \quad otherwise \end{cases} \tag{A1}$$

$$q(\theta,\sigma) := \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}\left(\theta-\frac{\pi}{2}\right)^2\right] \tag{A2}$$

$$Q(\sigma) := \int_0^\pi \left[q(\tau,\sigma)-q(0,\sigma)\right]d\tau \tag{A3}$$



(a) Schematic diagram of angular distribution. (b) Contour of $p(\theta,\sigma)$.

Figure 5: Angular distribution of deposited materials.

As shown in Figure 6, the incoming angle θ with which a particle reach the point in the resist pattern (x,0) is in the range from $\theta_m(x)$ to $\theta_M(x)$. Since the particle reach the point with the possibility $p(\theta,\sigma)$, the amount of the particles which reach the point in unit time through the F process, $f(x,h,\sigma)$, is given by the followings:

$$f(x,h,\sigma) = \int_{\theta_m(x)}^{\theta_M(x)} p(\theta,\sigma)d\theta$$

$$= \frac{1}{2Q(\sigma)}\left[erf\left(\frac{\theta_M(x)-\frac{\pi}{2}}{\sqrt{2}\sigma}\right)-erf\left(\frac{\theta_m(x)-\frac{\pi}{2}}{\sqrt{2}\sigma}\right)\right]$$

$$-\frac{q(0,\sigma)}{Q(\sigma)}\left[\theta_M(x)-\theta_m(x)\right]$$

where erf(x), $\theta_m(x)$, and $\theta_M(x)$ are defined as follows:

$$\operatorname{erf}(z):=\frac{2}{\sqrt{\pi}}\int_0^z e^{-\tau^2}d\tau$$

$$\theta_m(x):=\begin{cases}\operatorname{atan}\dfrac{H}{a-x}, & x\in[-c,a]\\ \pi-\operatorname{atan}\dfrac{H}{x-a}, & x\in(a,b]\\ \pi-\operatorname{atan}\dfrac{h}{x-b'}, & x\in(b,c]\\ 0, & otherwise\end{cases}$$

$$\theta_M(x):=\begin{cases}\operatorname{atan}\dfrac{h}{-b'-x}, & x\in[-c,-b)\\ \operatorname{atan}\dfrac{H}{-a-x}, & x\in[-b,-a)\\ \pi-\operatorname{atan}\dfrac{H}{x+a}, & x\in[-a,c]\\ 0, & otherwise\end{cases}$$

Figure 7 shows schematic diagram of the G process. Incoming paths to the point in the resist pattern (x,0) is determined by geometric relation between the point and the resist side walls. The amount of the particles $g(x,h,\sigma)$ which reach the point in unit time where $g_l(x,h,\sigma)$ and $g_r(x,h,\sigma)$, the amounts in unit time of the particles incoming via the left and right sidewalls, are defined as follows:

$$g(x,h,\sigma)=g_l(x,h,\sigma)+g_r(x,h,\sigma) \tag{A5}$$

$$g_l(x,h,\sigma):=\begin{cases}\dfrac{1}{\cos\chi}\displaystyle\int_{z_l(x)}^{H}p(\theta_l(x,h,z),\sigma)dz, \\ \quad (x\in(-b,c))\cap(z_l(x)<H)\\ \dfrac{1}{\cos\chi}\displaystyle\int_{z_l(x)}^{\bar{z}_l(x,h)}p(\theta_l(x,h,z),\sigma)dz, \\ \quad (x\geq c)\cap\left(z_l(x)<\bar{z}_l(x,h)\right)\\ 0, \quad otherwise\end{cases} \tag{A6}$$

$$g_r(x,h,\sigma):=\begin{cases}\dfrac{1}{\cos\chi}\displaystyle\int_{z_r(x)}^{H}p(\theta_r(x,h,z),\sigma)dz, \\ \quad (x\in(-c,b))\cap(z_r(x)<H)\\ \dfrac{1}{\cos\chi}\displaystyle\int_{z_r(x)}^{\bar{z}_r(x,h)}p(\theta_r(x,h,z),\sigma)dz, \\ \quad (x\leq -c)\cap\left(z_r(x)<\bar{z}_r(x,h)\right)\\ 0, \quad otherwise\end{cases} \tag{A7}$$

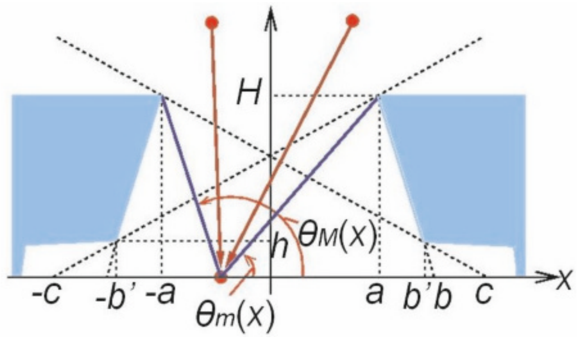


Figure 6: The F process.

$$\theta_l(x,h,z):=\frac{\pi}{2}-2\chi+\operatorname{atan}\frac{(z-h)b-zb'-xh}{zh}$$

$$\theta_r(x,h,z):=\frac{\pi}{2}-2\chi+\operatorname{atan}\frac{(z-h)b-zb'+xh}{zh}$$

$$\bar{z}_l(x,h):=\frac{Hh(x+b)}{H(x-b')+h(b-a)}$$

$$\bar{z}_r(x,h):=\frac{Hh(x-b)}{H(x+b')-h(b-a)}$$

$$z_l(x):=\frac{2HaZ_{l,1}(x)+H(b+x)Z_2}{H(x+b)-h(b-a)}$$

$$z_r(x):=\frac{2HaZ_{r,1}(x)+H(b-x)Z_2}{H(b-x)+2Ha}$$

$$Z_{l,1}(x):=\frac{H(b-a)(x+b)}{H^2+(b-a)^2}$$

$$Z_{r,1}(x):=\frac{H(b-a)(b-x)}{H^2+(b-a)^2}$$

$$Z_2(x):=\frac{H\left(H^2+b^2-a^2\right)}{H^2+(b-a)^2}$$

$$\chi:=\operatorname{atan}\frac{b-b'}{h}$$

Total amount of particles deposited by both of the F and G processes in unit time, $D(x,h,\sigma,\kappa)$, is given with a weighting constant κ>0 by the following equation:

$$D(x,h,\sigma,\kappa):=f(x,h,\sigma)+\kappa g(x,h,\sigma) \tag{A8}$$

Furthermore, if the specified film thickness $\underline{T}$ is smaller enough than H, the following approximation is considered to be valid:

$$T(x,h,\sigma,\kappa)\approx\alpha D(x,h,\sigma,\kappa),\quad \alpha\geq 0 \tag{A9}$$

where α is defined as follows:

$$\alpha:=\frac{\underline{T}}{D(0,h,\sigma)}$$

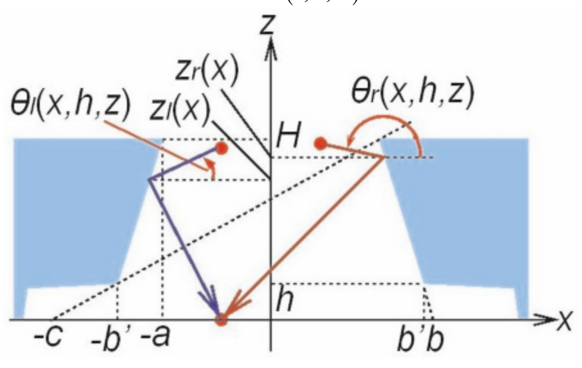


Figure 7: The G process.

The film thickness distribution, $T(x,h,\sigma,\kappa)$, is given by the followings:

$$T(x,h,\sigma,\kappa) := T_F(x,h,\sigma) + \kappa T_G(x,h,\sigma) \tag{A10}$$

where $T_F(x,h,\sigma)$ and $T_G(x,h,\sigma)$ are defined as follows:

$$T_F(x,h,\sigma) := \alpha f(x,h,\sigma)$$

$$T_G(x,h,\sigma) := \alpha g(x,h,\sigma).$$

Visual Inspection of Soldering Joints by Neural Network with Multi-angle View and Principal Component Analysis

Michiya Matsushima, Naohiro Kawai, Hiroyuki Fujie, Kiyokazu Yasuda, and Kozo Fujimoto
Division of Materials and Manufacturing Science, Graduate School of Engineering, Osaka University, Osaka, Japan

Abstract
With the development of microelectronics technology, the demands of the automatic inspection system are ever increasing. The current trends toward miniaturization of components, denser packing of boards, surface mounting technology, and highly automated assembly equipment make the task of inspecting these defects more critical and more difficult for humans. In this paper, we achieved 0% of misjudgment by implementing training category intermediate samples. We also achieved cutting down the processing time and an increase of correct judgments by the improvement of inputs using principle component analysis and multi-angle image.

Keywords:
Visual inspection, Neural networks, Principal component analysis, Intermediate sample, Multi-angle image

1 INTRODUCTION

Nondestructive inspection of solder joints is necessary to keep the quality and reliability of electronics products. Visual inspection of soldering joints is usually done by directly or by glass magnifying by experts. However, there are some problems such as misjudgment caused by the miniaturized and high density assembly. The trend of experts' aging and staff shortage gets worse. In addition, the inspection quality variation caused by the individual variation, physical conditions and fatigue are also some of the problems. So manufacturer demands the automatic visual inspection system in the manufacturing process of electronics devices.

Logical qualification of the solder joints estimation is difficult because the surface sheen makes it difficult to measure the geometry and some judgment criterions of the inspections are qualitative [1]. There are some kinds of information used for automatic nondestructive inspection methods. The two or three dimensional geometric information of the solders [2]-[4], visual information such as intensity of each pixel of optical or X-ray images and so on are utilized for automatic inspection [5]-[11]. The surface of solders should be smooth and sheen and the solders should flow well. Therefore, some image processing methods or learning algorithms are used in some visual inspections. Neural network is one of the learning methods used for visual inspections of solder joints. The neural network is one of the learning system which is possible to categorize complicated pattern information utilizing the repeatability of view pattern by learning good and defect samples [12]-[17]

In this paper, the following three approaches for improving neural network visual inspection system are evaluated.

1. Improvement of input data using principal component analysis (PCA) for high speed learning and inspection.
2. Learning virtual intermediate samples for preventing misjudgment.
3. Utilization of additional information from multi-angle view.

2 NEURAL NETWORK VISUAL INSPECTION SYSTEM

Three-layer neural network which is known to be capable of approximating nonlinear systems with any degree of nonlinearity is applied in our visual inspection system (Figure 1). The neural network visual inspection system consists of the learning phase and the inspection phase.

In the learning phase, the neural network system is given inputs extracted from a good or a defective sample and two outputs of good degree and defect degree of the sample. The system updates the coefficients representing the relationship between the inputs and the outputs by backpropagation algorithm. The training data which are intensities of each pixels or principal components are input and the good degree and defective degree are the outputs. The connection weights are learned so that good degree is 1.0 and defective degree is 0.0 for good samples, and good degree is 0.0 and defective degree is 1.0 for defective samples as shown in Table 1.

The inspection is executed after obtaining converged and well-learned coefficients. The input data of the neural network is the intensity of each pixel of the soldering joint pictures. In our high speed neural network system with PCA, the projective transformed data are the inputs.

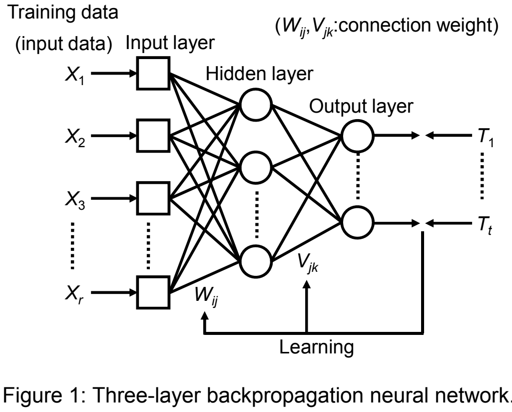


Figure 1: Three-layer backpropagation neural network.

In the inspection phase, there are two thresholds for the judgment. If the good sample degree is higher than a higher threshold and defective sample degree is lower than a lower threshold, the sample is determined to be a good sample. In a contrary case, it is a defective sample. In case that one or both degrees are between two thresholds, the sample is categorized to "undeterminable."

The training values of good degree and defective degree for good samples and defective samples are shown in Table 1 and two thresholds for the judgment are shown in Table 2.

The sigmoid function of the following equation is used as the output function in our neural network.

$$f(x) = \frac{1}{1 + \exp(-\alpha \bullet x)}$$

The parameters of the neural network such as the number of nodes in each layer, the gradience of sigmoid function and the allowable total error in learning are shown in Table 3.

Table 1: Training values.

Samples	Good degree	Defective degree
Good	1.0	0.0
Defective	0.0	1.0

Table 2: Judgment thresholds.

Good degree	Defective degree	Judgment
≥0.8	≤0.2	Good
≤0.2	≥0.8	Defective
else above		Undeterminable

Table 3: The parameters of the neural network.

Camera angle (deg)	90	65	40
Input layer	2496	2288	1872
Hidden layer	100		
Output layer	2		
Gradience of the sigmoid function	1.0		
Allowable total error value	0.01		

3 SAMPLES AND CAMERA SETUPS

The requirements for good solder joint are the following:

1. The solder flows well and forms fillets.
2. The solder surface is clean, shiny and smooth.
3. The fillet ends are thin enough.
4. The form of the solder joints contains no cracks, holes, discontinuity, and the volume of the solder is adequate.

Therefore, the necessary information for inspection is the following:

1. The wetting information.
2. The surface information such as roughness and micro convexoconcave.
3. The geometry information.

Various defectives are detected from this information. In this paper, the solder excess is picked up, for instance, to consider the effect of each approach.

The 120 good samples and 120 defective samples are respectively divided into 2 groups. One is for learning and the other is for inspection. Almost the same numbers of good samples close to defective and defective samples close to good are included in each group. The numbers of samples are shown in Table 4.

The conditions of taking pictures are determined by the angles of camera and light resource. Camera angles are from vertical upward and right angle to the solder surface (The solder surface is assumed that the line from the upper edge of the chip terminal and far edge of the terminal on the board (Figure 2)). The light source angles are selected to avoid the interference to the camera and other devices. These setups are summarized in Table 5.

Table 4: Number of teaching and inspecting samples.

Number of teaching samples	
Good	49
Good (close to defective sample)	11
Solder excess (Close to good sample)	45
Solder excess	15
Number of inspecting samples	
Good	50
Good (close to defective sample)	10
Solder excess (Close to good sample)	45
Solder excess	15

Table 5: The camera and light source setups.

No.	Camera angle (deg)	Light angle (deg)
1	90	65
2	90	40
3	65	90
4	40	90

4 IMPROVEMENT OF THE INPUT DATA FOR FAST LEARNING AND INSPECTION

4.1 Principal component analysis

Generally, the more information of the sample, the more reliable inspection is possible. However, it takes much time if the system has to process a large number of data. Saving the computational effort can be a large merit for selecting the ideal camera setups and additional learning as well as for the efficient inspection. Though we minimize the pictures from 208 x 144-192 into 52 x 36-48, minimizing the picture can cause the loss of information. Therefore, more minimization is critical to the reliable inspection. When two or more images are used for this inspection, the input data goes twice or more and it will take much more time to learn. So we adopt the PCA to degrade the input dimension and cut down the computational effort.

The PCA is a method where the vectors of input data set are projected to the coordination system which makes the distributions of data set large in the axes. The component of the projected axis direction with the maximum distribution is the first principal component. This method enables us to dissolve the correlations of multivariate data and enables the data with lower dimension to represent the feature of the original measured data distribution[8]. The example of PCA applied to the two dimensional data is shown in Figure 3. Projecting the x_1 axis and x_2 axis to the principal component axes of z_1 and z_2 enables us to represent most of the distribution of the data by the z_1 axis only. The contribution rate of each principal component represents the degree of explaining the feature of data set.

4.2 Applying the principal component analysis to the visual inspection

In the learning phase, the principal components calculated from the intensity of teaching sample images are used as the input of the neural network. In the inspection, the principal components of the samples for inspection are calculated with the projective coordinate and used as the input of the neural network formed in the learning phase. To avoid the deterioration of the information, the components are utilized so that the cumulative contribution rate is over 99.99 %. The cumulative contribution rate is the summarized contribution rate from the first (largest contribution rate) principal component.

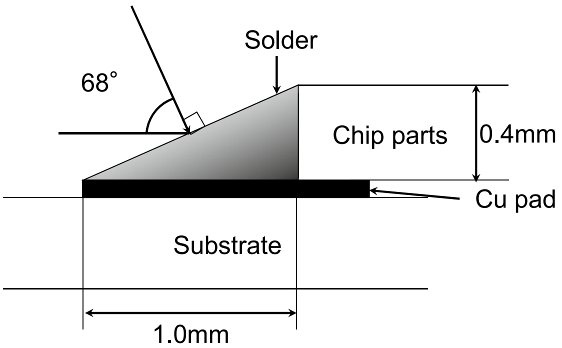


Figure 2: Right angle to the solder surface.

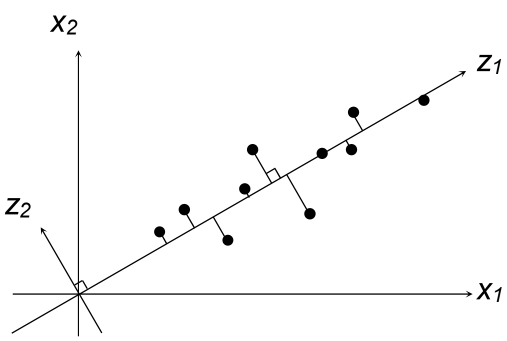


Figure 3: Principal component analysis for two dimensions.

4.3 The effect of applying the principal component analysis on the inspection

The effect of applying the PCA on the inspection results is investigated using the average of the results with the camera and light source setups shown in Table 4. The numbers of inputs for the system with 99.99% cumulative contribution rate and collapsed time for learning and inspection with and without PCA and the inspection results of correct judgment rate, misjudgment rate, and undeterminable sample rate for the inspection with and without PCA with 65 degree camera angle and 90 degree light source are shown in Table 6.

The number of PCA input was decreased to the half of the original inputs. The processing time is also decreased. The total processing time of PCA and neural network with principal component input is about a half of original learning process time. Inspection time for one soldering joint was decreased to about 1/30. The applying PCA enabled to improve the correct determination as well as to decrease the computational effort. The availability is confirmed from this result, the component analysis is applied to the following inspections.

5 PREVENTIVE METHOD OF MISJUDGMENT

5.1 Intermediate samples in category space

When the dependent n-dimensional information is extracted from an image, the data can be considered as a certain point in n dimension Euclid space. The samples in the same category have similar information and located close in distance in the n dimension Euclid space. A set of samples in the same category constructs one category space. In the information space of soldering joint image, the various defective category spaces such as solder excess or lack of

Table 6: Effect of the principal component analysis on the inspection results.

Inputs	Intensity	Principal Component
Learning time of PCA [s]	0	116
Learning time of neural networks [s]	1899	788
Total learning time [s]	1899	904
Inspection time per sample [ms]	2.504	0.082
Correct judgment rate[%]	90.83	91.67
Misjudgment rate [%]	0.84	0.84
Undeterminable rate [%]	8.34	7.5

solder volume, wetting defects, overheating, wicking and so on are considered to exist around the category space of good soldering (Figure 4).

The inspection system must not misjudge defective samples as good. The n dimensional data extracted from teaching solder joint sample images constructs the category spaces around each good and defective sample sets as shown in Figure 5.

The teaching sample does not fulfill all points in each category space so the category space and learned space does not match. The mismatch space causes the misjudgment or incapability of determination (shown in Figure 5). The misjudgment space appears around the boundary zone between leaned good sample space and defective sample space. If the Euclid distance of a good sample and defective sample is short, they are located around the boundary zone. The sample sets of good samples and defective samples whose Euclid distances are small are searched and generate the virtual intermediate sample on the line between the good and defective sample. Learning the intermediate samples as low points of good degrees and defective degrees makes the boundary zone of misjudgment to undeterminable zone (shown in Figure 6).

5.2 Application of intermediate sample in category space

The application method of intermediate samples in category space is the following:

1. Calculate the Euclid distance between the teaching good sample and teaching defective sample.
2. Pick up the set of good and defective samples with the shortest Euclid distance from all teaching samples.
3. Generate the intermediate samples at the internally dividing points of 8:2 and 2:8 between the samples.
4. Pick up the sample set with the shortest Euclid distance except the sample set already used to generate the intermediate samples.
5. Repeat step 3 and 4 several times

The test samples with four camera and light source setups shown in Table 4 are inspected to evaluate the validity of intermediate samples.

Figure 7 shows the results of the inspection with learning the intermediate samples. It shows the decrease of misjudged samples as the effect of learning the intermediate samples although the correct judgments are also decreased.

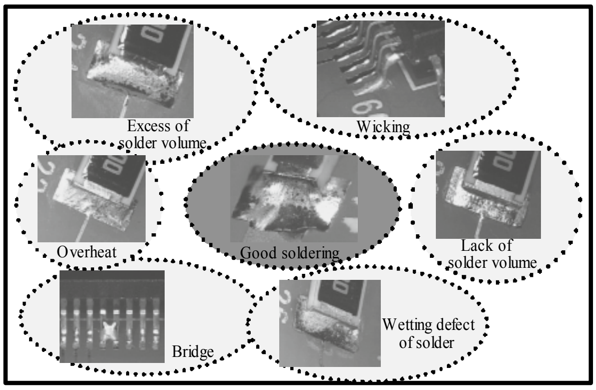


Figure 4: Category space image of good and defective soldering.

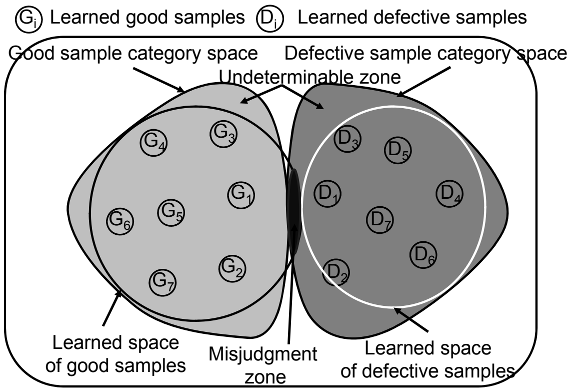
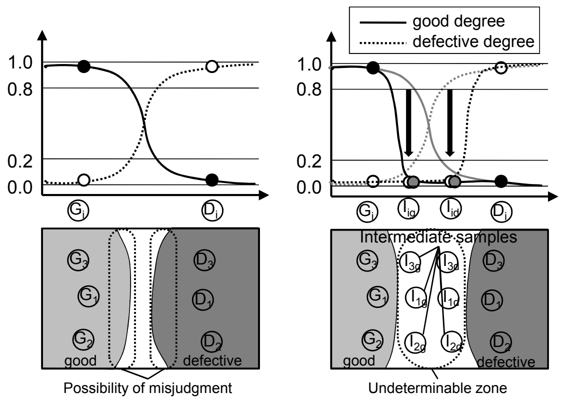


Figure 5: Image of misjudgment and undeterminable zones.



(a) No intermediate sample (b) Learning intermediate samples

Figure 6: Image of intermediate samples effect.

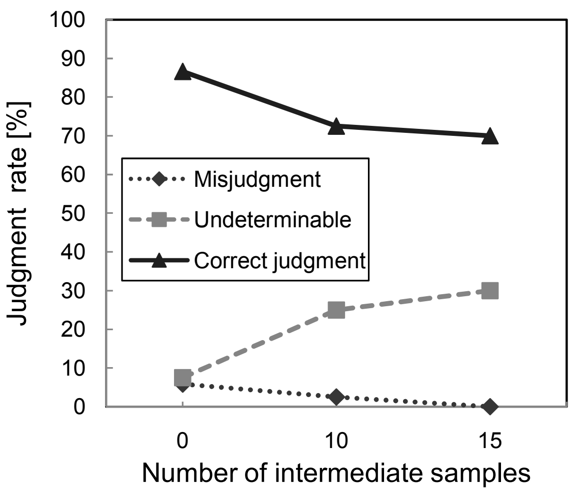


Figure 7: Effect of learning intermediate samples on judgment.

6 THE UTILIZATION OF MULTI-VIEW

6.1 The concept of utilizing multi-view

In some solder joints, it may be difficult to determine whether they are fine or defective if they are only inspected in a certain directions. In some directions, even defective sample sometimes look good. Multi-angle view inspecting system is necessary to inspect the indistinguishable samples. However, the larger the amount of input data becomes, the longer processing time it takes to learn and inspect. If there are any misjudgments, multiple results just confuse the inspection. In order to utilize the multiple images, it is necessary to decrease the input data without thedeterioration of information quality and to prevent the misjudgment.

We applied the PCA to decrease the data amount without the deterioration of information quality. It is possible to decrease the number of undeterminable samples by utilizing multi-angle views obtained from the camera setups which had no misjudgments by learning the intermediate samples.

6.2 Evaluation of the effect of utilizing multi-angle view on undeterminable sample rate

The undeterminable sample rate with the multi-angle view is shown in Figure 8. The undeterminable sample rate was decreased and correct judgment rate was increased by utilizing the multi-angle views obtained from the camera setups which had no misjudgments by learning the intermediate samples.

7 SUMMARY

In the visual inspection system with neural network, the method of fast learning and inspection with the PCA, prevention of the misjudgment with learning intermediate samples, and utilizing the multi-view are evaluated.

The following are the obtained results:

1. The number of the input data was decreased by the PCA without the information loss. The processing time was also cut down and the correct judgment rate was also improved.

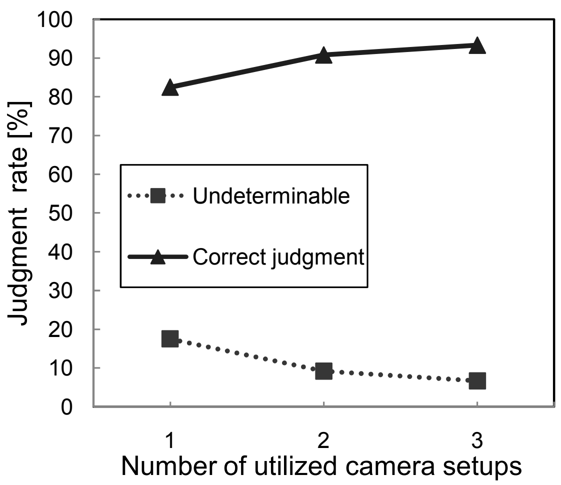


Figure 8: Effect of multi-angle view on the judgment.

2. The misjudgment rate was decreased to 0 by learning the virtual intermediate samples in the category space.
3. The correct judgment rate was improved by utilizing the multi-angle images as input. The rate of the samples unable to determine was decreased.

8 ACKNOWLEDGMENTS

We appreciate Mr. Tateo Nakajima's inspecting the samples as an expert of solder joint inspection.

9 REFERENCES

[1] Robertson, G. I., 1987, Automatic Optical Inspection of Printed Circuit Boards During Assembly, Proceedings of SME Visio'87, 59-67.

[2] Ryu, Y. K., Cho, H. S., 1997, New Optical Measuring System for Solder Joint Inspection, Optics and Lasers in Engineering, 26/6:487-514.

[3] Japan Welding Society Microsoldering Technology Education Committee, 2002,Standard Microsoldering Technology, Nikkan Kogyo Shimbun, Ltd.

[4] Matsuyama, Y., Honda, T., Yamamura, H., Sasazawa, H., Nomoto, M., Ninomiya, T., 1996, Automated Solder Joint Inspection System Using Optical 3-D Image Detection," WACV '96, 116.

[5] Asakura, Y., Takahashi, Y., 2003, Feature Extraction for Fillet Profile of Pb-free Solder and Its Application to Inspection, Trans. JWRI, 32/1:55-58.

[6] Doirona, T., Pinskyb, D. A., Chenc, S., Plottd, M., 1993, An SPC Artifact for Automated Solder Joint Inspection, Measurement, 12/1:1-8.

[7] Oyeleye, O., Lehtihet, E. A., 1999, Automatic Visual Inspection of Surface Mount Solder Joint Defects, International Journal of Production Research, 37/6:1217-1242.

[8] Matsuyama, Y., Honda, T., Yamamura, H., Sasazawa, H., Nomoto, M., Ninomiya, T., 1996, Automated Solder Joint Inspection System Using Optical 3-D Image Detection," WACV '96, 116.

[9] Mahona, J., Harrisb, N., and Vernonc, D., 1989, Automated Visual Inspection of Solder Paste Deposition on Surface Mount Technology PCBs,Computers in Industry, 12/1:31-42.

[10] Mahona, J.1991, Automatic 3-D Inspection of Solder Paste on Surface Mount Printed Circuit Boards, Journal of Materials Processing Technology, 26/2:245-256.

[11] Juha, M., 1986, X-Ray Machine Vision for Circuit Board Inspection, Proceedings of SME Vision'86, 3:41-55.

[12] Roh, Y. J., Park, W. S., Cho, H. S., 2003, Correcting Image Distortion in the X-ray Digital Tomosynthesis System for PCB Solder Joint Inspection, Image and Vision Computing, 21:1063-1075.

[13] Kim, J. H., and Cho, H. S., 1995, Neural Network-based Inspection of Solder Joints using a Circular Illumination, Image and Vision Computing, 13/6:479-490.

[14] Koji, Y., Takatsu, N., Oh, M., 1996, Visual Solder Inspection using Neural Network, Systems and Computers in Japan, 27/1:92-100.

[15] Kim, T. H., Cho, T. H., Moon, Y. S., and Park, S. H., 1999, Visual Inspection System for the Classification of Solder Joints, Pattern Recognition, 32/4:565-575.

[16] Ko, K. W., Cho, H. S., 2000, Solder Joints Inspection using a Neural Network and Fuzzy Rule-Based Classification Method, IEEE Trans. Electronics Packaging Manufacturing, 23/2:93-103.

[17] Misonou, N., Ota, K., Yasuda, K., Fujimoto, K., 2005, Additional learning in Neural Network Visual Inspection System, The 11th Symposium of Micro-joining and Assembly Technology in Electronics, 11:385-390.

[18] Acciani, G., Brunetti, G., and Fornarelli, G., 2006, Application of Neural Networks in Optical Inspection and Classification of Solder Joints in Surface Mount Technology, IEEE Trans. on Industrial Informatics, 2/3: 200-209.

Path Generation and Collision Avoidance of Robot Manipulator for Unknown Moving Obstacle using Real-time Rapidly-exploring Random Trees (RRT) Method

Chikatoyo Nagata[1], Eri Sakamoto[1], Masato Suzuki[1], and Seiji Aoyagi[1]

[1] Dept. of Mechanical Engineering, Kansai University, Osaka, Japan

Abstract

Collision avoidance in a double-arm robot is focused on. One robot arm (Robot2) is taken as a moving obstacle and another arm (Robot1) should move from a start position to the goal one while avoiding the collision with Robot2. A comparatively short path can be searched efficiently in configuration space (C-space), i.e., joint angle space, by using Rapidly-exploring Random Tree (RRT) method. In this study, to cope with unknown obstacle's trajectory, real-time RRT method is proposed. This method assumes that the robot can obtain the information of unknown object movement at every interval.

Keywords:

Robot manipulator; Collision avoidance; RRT search; Path generation

1 INTRODUCTION

A robot has developed mainly for industrial use such as assembly, welding, etc. in a factory, where the environment is arranged for the robot, i.e., the obstacles do not exist. However, in the next generation, a robot is expected to perform complicated tasks, such as assisting a person in the daily home work, nursing her/him in a hospital, etc., where the environment is difficult to arrange for the robot. In such environment, collision avoidance for unknown moving obstacle is necessary.

In this study, collision avoidance in a double-arm robot is focused on. One robot arm (referred to as Robot 2 herein) is taken as a moving obstacle and another arm (referred to as Robot 1) should move from a start position to the goal one while avoiding the collision with Robot 2. A comparatively short path can be searched efficiently in configuration space (C-space), i.e., joint angle space, by using rapidly-exploring random trees (RRT) method. RRT is a contemporary path planner developed in recent years [1] [2].

To cope with a moving obstacle, the path in terms of time, i.e., the trajectory, must be considered. For this purpose, a method is proposed in this study, which regards each step of expanding node of RRT as a unit time.

Under the condition that Robot 1 knows the trajectory of Robot 2 in advance, it is easy for Robot 1 to generate the trajectory by off-line calculation based on the RRT search. On the other hand, if the trajectory of Robot 2 is unknown for Robot 1, the path planning becomes drastically difficult. In home or hospital environment, the trajectory of obstacle is not assured to be known in advance, since the environment is changeable in real time. Thus, to cope with unknown obstacle's trajectory, real-time RRT method is proposed in this study. This method assumes that the robot can obtain the information of unknown object movement at every interval. The robot forecasts future obstacle's trajectory based on the information obtained before and now, and generates the trajectory avoiding the collision with the forecasted trajectory based on the RRT search within the interval. To speed up the RRT search, the probability that the tree expands in the direction to the goal is set to 100% unless the collision is detected.

In this research, the effectiveness of real-time RRT was verified by both a simulation and an experiment. In the simulation, serially linked cylinders were considered as a robot arm. In the experiment, a real-time controller was constructed using ART-Linux. Both in the simulation and the experiment, the proposed real-time RRT method effectively worked for collision avoidance.

2 PATH GENERATION BY USING REAL-TIME RRT

2.1 Basic RRT

A path in terms of time, i.e., a trajectory, in *n*-dimensional configuration space (C-space) is assumed to be expressed by $f = (\theta_1(t), \cdots \theta_n(t))$. There are infinite number of paths, which satisfy the condition that they progress from a given start position and finally reaches a given goal position, while avoiding obstacles. Therefore, RRT is employed, which can efficiently search a path even in a considerably high dimensional C-space.

In RRT, a path from a start configuration $\theta_{\mathrm{init}}(\theta_1(0), \cdots \theta_n(0))$ to a goal configuration $\theta_{\mathrm{goal}}(\theta_1(T), \cdots \theta_n(T))$ (*T*; arrival time) is searched as follows: RRT first samples a point (called as "node"), and samples another point randomly with constant adjacent distance. If the collision is detected, this point is cancelled and another point is sampled again. This process is performed successively until the sampled point finally reaches the goal, where the route connecting the points is called as "tree". This search is fast compared to other searching methods, thus, can cope with comparatively high dimensional space.

Concretely, the basic RRT algorithm is described as follows (see Figure 1): the start position of the robot is assumed to be the node *q_init* in RRT, which is defined as the origin of a tree. When the path is generated, a node *q_rand* is taken at random in the C-space, and it is connected to its nearest node *q_near* among already existing nodes. Then, a node *q_new* is generated on a line of connecting *q_rand* to *q_near*, so that it is apart from *q_near* by constant distance (*e*). Then, *q_new* and *q_near* is connected to each other, making a branch of the tree. The above-mentioned process is successively repeated until the tree reaches *q_goal*, which is the node of given goal position.

In the basic RRT, the direction *q_rand* is randomly taken, that is, there is a possibility *q_rand* is taken in the neighborhood of *q_init*, that is, a branch is extended from a node near the root, even in the ending stage of path searching. Considering this fact, the scheme of basic RRT is rather inefficient in the viewpoint of fast search of a path.

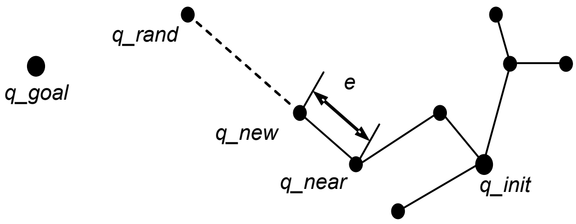


Figure 1: RRT algorithm.

2.2 Improvements is RRT algorithm

2.2.1 Path generation that considers time series

Since the collision avoidance of double-arm robot is considered in this study, it is necessary to detect the collision between two robots dynamically. Assuming that the obstacle movement is uniform and rectilinear, RRT algorithm considering time domain is reported [3]. By contrast, RRT algorithm considering time domain, which can cope with arbitrary obstacle motion, is proposed in this study, in order to generate a path in terms of time, i.e., in order to generate a trajectory.

A proposed algorithm considering time domain is as follows: obstacle's trajectory is assumed to be known in advance. The time at the start position is defined as 0, and the time is incremented by one arbitrary unit, synchronously with an extension step of a branch in the RRT tree. By this process, time series of a path, i.e., a trajectory can be considered.

At every time when a new node is generated, the collision between the node with the obstacle can be checked, since obstacle's trajectory is assumed to be known, as mentioned above. If the collision is detected, the node is cancelled, and another node is re-generated based on the above-mentioned basic RRT algorithm. By successively repeating this process, finally, the trajectory of avoiding the obstacle can be obtained.

2.2.2 Contrivance for achieving real-time RRT

The above-mentioned RRT extends a branch of the tree by taking a node at completely random. This scheme of extending a tree is surely effective for searching a free space in which a robot can move among a comparatively wide unknown space. However, in the view point of fast search of a path which reaches a goal, this scheme of randomly extending a tree is not a smart way. To speed up the RRT search, a method has been reported, in which *q_rand* is set to *q_goal* at a constant rate [4]. This study also follows this method. The feature of this study is to set the rate to 1.0 unless the collision is detected, which means a tree is extended straight to a goal, achieving a fairly fast path generating. And, once the collision is detected, the rate is decreased to 0.5, to efficiently avoid the obstacle. Note that even in this case, the tree is extended to the goal with 50% probability. After it is judged that the danger of collision is removed, the rate is recovered to 1.0. By these processes, the search time is drastically reduced, which is the key point of making real-time RRT search possible.

3 COLLISION DETECTION OF DOUBLE-ARM ROBOT IN THREE DEMENSSIONAL SPACE

Assume that one link be an obstacle, and another link would avoid it. If the distance between two link is more than double the link's radius, as shown in Figure 2 [5]. According to this rule, the collision between two links is detected for all the combination of links in a double-arm robot. As the result, the collision of double-arm robot can be detected in three dimensional space.

The criterion of collision detection is described in detail hereinafter. The equation of link L_M is expressed by using the direction vector $\boldsymbol{m}$. That of the central line of the obstacle is expressed by using $\boldsymbol{m}_h$. The common normal vector of them (***n***), the shortest distance between them (*b*), and the distance from $\boldsymbol{r}_{JM}$ to the intersection of ***n*** and the link (*s*) are obtained as follows [9]:

$$\boldsymbol{n} = \frac{\boldsymbol{m} \times \boldsymbol{m}_h}{\left|\boldsymbol{m} \times \boldsymbol{m}_h\right|}\ , \tag{1}$$

$$b = \left|(\boldsymbol{r}_{JM} - \boldsymbol{r}_h, \boldsymbol{n})\right|\ , \tag{5}$$

$$s = \frac{(\boldsymbol{r}_{JM} - \boldsymbol{r}_h, \boldsymbol{m} \times \boldsymbol{n})}{\left|\boldsymbol{m} \times \boldsymbol{m}_h\right|}\ . \tag{6}$$

Assuming that the radius of the cylinder and the length of the link are *c* and l_M, respectively, then the condition for collision avoidance is $b>2c$ when $0<s<l_M$. When $s>l_M$, it is necessary that $\boldsymbol{r}_{JM+1}$ is apart from $\boldsymbol{m}_h$ by 2*c* at the least, the condition of which is expressed as follows:

$$\{(\boldsymbol{r}_h - \boldsymbol{r}_{JM+1})^2 - (\boldsymbol{r}_h - \boldsymbol{r}_{JM+1}, \boldsymbol{m}_h)^2\} > (2c)^2\ . \tag{7}$$

In the same way, when $s<0$, the condition is as follows:

$$\{(\boldsymbol{r}_h - \boldsymbol{r}_{JM})^2 - (\boldsymbol{r}_h - \boldsymbol{r}_{JM}, \boldsymbol{m}_h)^2\} > (2c)^2\ . \tag{8}$$

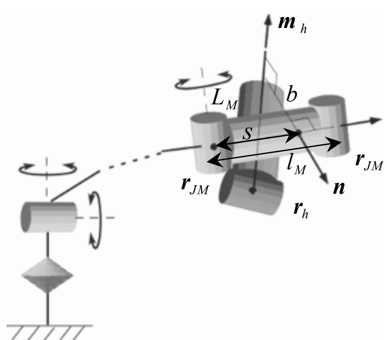


Figure 2: Collision detection.

4 CONCEPT OF REAL-TIME RRT AND SYSTEM COMPOSITION

4.1 Outline of Real-time RRT

In this study, the concept of real-time RRT is proposed. Real-time RRT is a method to generate a trajectory of avoiding a moving obstacle in real-time, being given (or sensing) the data of obstacle's movement (in case of double-arm robot, the data are all the joint angles of the robot arm to be avoided) at every interval. This interval is set to a short period as possible, considering the calculating time of trajcetory based on RRT algorithm. Of course, this interval is usually larger than the unit time of RRT tree extension (see Section 2.2.1).

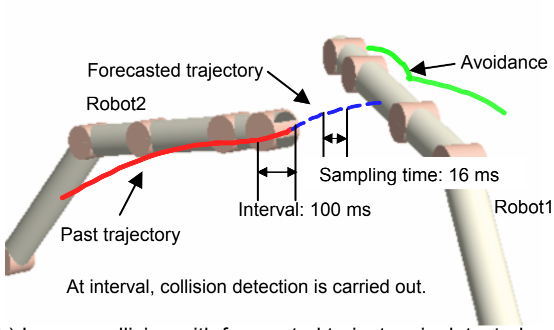
Here, the trajcetory of the obstacle is assumed to be unknown in advance, thus, this trajectory is forecasted in the real-time RRT by extrapolating the past trajectory at above-mentioned every interaval. On the other hand, at every unit time of RRT tree extention, the collision is checked between the trajectory of the robot searched by RRT and that of the obstacle forecasted by extrapolation. The avoidance based on the poposed real-time RRT is schematically shown in Figure 3. As shown in this figure, the interval (sampling time) adopted in the experimental system is 100 ms, as explained later in Section 4.4. On the other hand, sampling time of robot control is 16 ms in the experimental system, which is equivalent to the unit time of RRT tree extension.

4.2 Tree extension in real-time RRT

In real-time RRT, as already explained in Section 2,2,2, *q_rand* is set to *q_goal* at a constant rate of 1.0 unless the collision is detected, which means a tree is extended straight to a goal. Once the collision is detected, the above-mentioned rate is decreased to 0.5, to efficiently avoid the obstacle. In this case, it is possible that *q_rand* is taken in the neighbourhood of tree's root, which is undesirable in the viewpoint of fast search. To address this problem, in the real-time RRT, *q_rand* is taken near the previous *q_new*. Namely, the number of routes of the tree from *q_init* to *q_goal* is only one, and there are no junctions or branches.

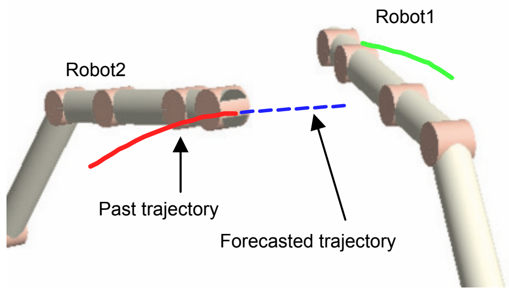
4.3 Obstacle avoidance using registered trajectories

It consumes long calculating time to search a detour path when the robot encounters the collision with obstacle, even the real-time RRT search is employed. To address this problem, in this research, many trajectories for avoiding the obstacle's trajectories are registered in advance, assuming some typical obstacle's trajectories. And, once the collision is detected, the nearest registered trajectory is selected, and a node on this trajectory is selected, which exists in the forward area of the extending direction of the RRT tree, as shown in Figure 4. This node is selected as *q_rand* with probability of 50%, whereas the *q_goal* is selected as *q_rand* with probability of 50%. Then, the *q_new* is generated on the line connecting *q_rand* and the previous *q_new*, so that the distance between *q_rand* and the previous *q_new* be the constant value of *e*. After it is judged that the danger of collision is removed, the probability of the process, in which *q_rand* is set to *q_goal*, is recovered to 100%. By conducting these procedures of using registered trajectories, the searching space is greatly narrowed, so the searching time is much reduced, which makes the real-time path generating possible.

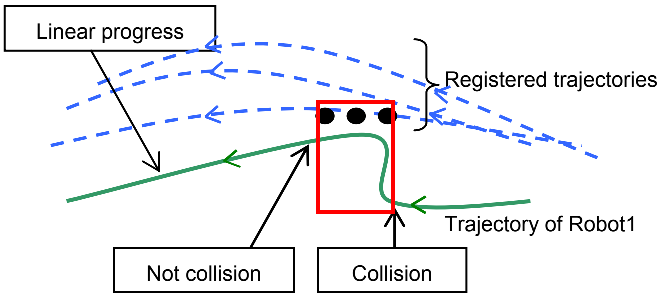


(b) In case collision with forecasted trajectory is detected.

Figure 3: Avoidance based on the proposed real-time RRT.



(a) In case collision with forecasted trajectory is not detected.



A node on the registered trajectory is selected in the neighborhood of the current node. The next node is successively searched with 0.5 probability (0.5 probability to the next node on the registered trajectory, and 0.5 probability to the goal).

Figure 4: Obstacle avoidance using registered trajectories.

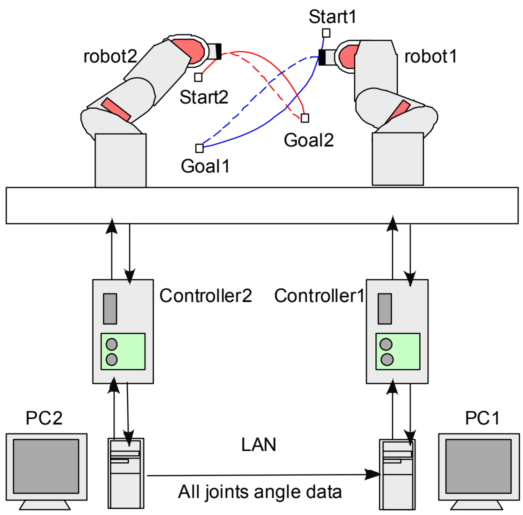


Figure 5: System for real-time RRT and its procedure.

4.4 Procedure of real-time RRT

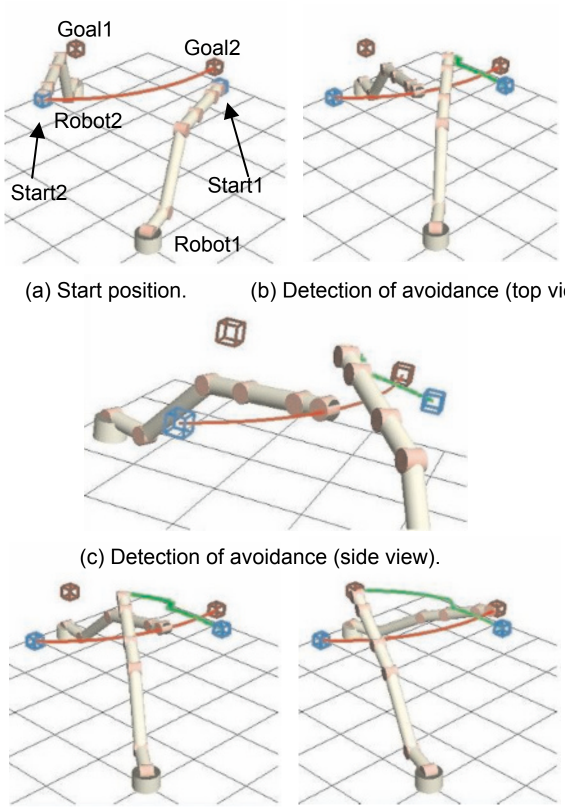
The procedure of real-time RRT conduced in the experiment in this paer is shown in Figure 5, which is as fillows:

1) Robot 2 sends the angle data of all joints to Robot 1 by using local area networks (LAN) at every constant interval of 100 ms.
2) At this interval, Robot 1 forecasts the trajectory of Robot 2 by extrapolating the past trajectory of Robot 2 in the form of spline curve.
3) Robot 1 calculates a trajectory based on RRT, which avoids the forecasted trajectory of Robot 2, while approaching the goal. RRT conducted here is improved one for real-time operation, the detail of which is already explained in Sections 2.2.2, 4.2, and 4.3.
4) The above-mentioned procedures from 1) to 4) are repeated until the Robot 1 reaches the goal position and orientation.

5 VERIFICATION OF PATH GENERATION BY REAL-TIME RRT

5.1 Simulation results of real-time RRT

Simulation results are shown in Figures 6, 7, and 8. Each link was modeled as the cylinder based on the size of a real 6-DOF articulated robot (Kawasaki Heavy Industries, Ltd. type JS-2).



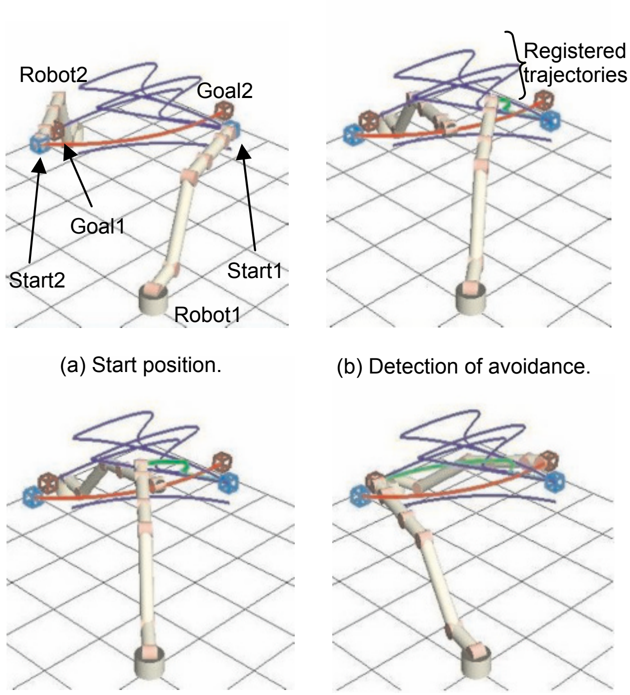
(a) Start position. (b) Detection of avoidance (top view).

(c) Detection of avoidance (side view).

(d) Searching trajectory. (e) Goal position.

Figure 6: Results of real-time RRT simulation (without using registered trajectories).

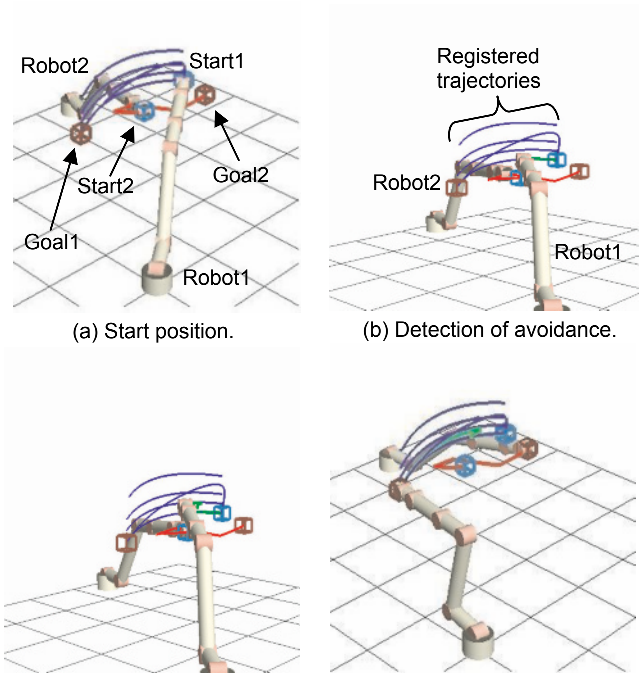
Figure 6 shows the result of searching for trajectory at random by using real-time RRT without using registered trajectories for avoiding obstacle. On the other hand, Figure 7 and Figure 8 show the results with using registered



(a) Start position. (b) Detection of avoidance.

(c) Searching trajectory. (d) Goal position.

Figure 7: Results of real-time RRT simulation (with using registered trajectories, simple trajectory case).



(a) Start position. (b) Detection of avoidance.

(c) Searching trajectory. (e) Goal position.

Figure 8: Results of real-time RRT simulation (with using registered trajectories, complex trajectory case).

trajectories. Here, the trajectory to be avoided is approximately linear in Figure 7, whereas, the trajectory to be avoided is so complex that the forecasting of it seems very difficult for Robot 1 in Figure 8.

In these figures, the trajectory of Robot 2 to be avoided is expressed as red line, the trajectory of Robot 1 of avoiding the moving obstacle, i.e., Robot 2, is expressed as green line, and the trajectories for avoidance which are registered beforehand are expressed as blue lines.

In Figure 6, all the joints of Robot 2 moves linearly in the C-space, and the angles information is sent to Robot 1 via LAN at every interval of 100 ms. All the joints of Robot 1 are moved in C-space on the basis of proposed real-time RRT method. Namely, at every sampling time of 16 ms, the collision detection with the forecasted trajectory is carried out, and if the collision is not detected, all the joints of Robot 1 move linearly in C-space from start angles to goal angles by setting the *q_rand* to *q_goal* unless the collision is detected. Once the collision is detected, the detour trajectory is searched by randomly taking *q_rand* with probability of 50%. Even in this case, the tree is extended to the goal with the other 50% probability, which urges the real-time path generation.

Looking at Figure 6, above-mentioned procedures are successfully performed, thus, the path generation coping with collision avoidance is finely achieved.

In Figure 7, in case that the collision is detected, the range of searching space is narrowed by using one of the registered trajectories. When the collision is detected, the nearest registered trajectory is selected, and a node on this trajectory is selected, as already explained in Section 4.3. This node is selected as *q_rand* with probability of 50%, whereas the *q_goal* is selected as *q_rand* with probability of 50%. After it is judged that the danger of collision is removed, the probability of the process, in which *q_rand* is set to *q_goa*l, is recovered to 100%.

Looking at Figure 7, a nearest registered trajectory is successfully selected, and the total path generation avoiding collision is finely achieve.

In Figure 8, the trajectory of Robot 2 is complex. The robot 2 at first moves in the same direction as Robot 1 does in Cartesian space: then it changes the direction by 180 degree, i.e., the direction approaching Robot 1.

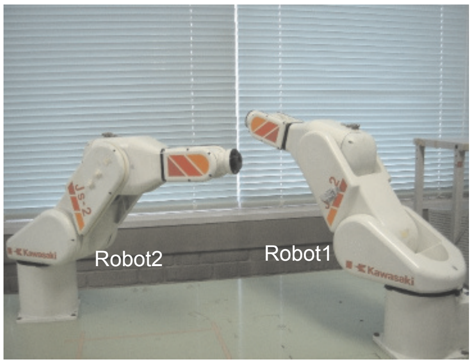
Looking at Figure 8, even under the above-mentioned severe condition, Robot 1 can successfully avoid Robot 2 in real-time in this simulation.

5.2 Experimental result of real-time RRT

Similar experiments were conducted by using two 6-DOF robots (Kawasaki Heavy Industries, Ltd. type JS-2), as shown in Figure 9. The robot controller is equipped with real-time control function, sampling time of which is 16 ms. The control system is constructed by using real-time Linux (ART Linux). On the other hand, the interval of sending the angle data from Robot 2 controller to Robot 1 controller via LAN is set to 100 ms. The system composition is already shown in Figure 5.

Note that, in real applications in the future, in stead of using LAN communication, the configuration of Robot 2 must be measured by some sensor, such as vision sensor, laser range sensor, ultrasonic sensor, and so on. The experiments using a laser range finder is ongoing in the laboratory.

Looking at Figure 9, even in the severe case same as Figure



(a) Start position.

(b) Detection of avoidance.

(c) Searching trajectory.

(d) Goal position.

Figure 9: Results of motion using real-time RRT.

8 in simulation, the real-time path generation while avoiding obstacle is successfully performed in real machines.

5.3 Measurement of processing time

The processing time in simulation in case of basic RRT (called off-line RRT hereinafter) and that in case of real-time RRT were measured, and compared with each other. Also, in case of real-time RRT, the processing time without using registered trajectories and that using them were compared. The results are shown in Table. 1. The CPU is Pentium 4 of 3.4 GHz clock, and memory is 2 GB.

In case of off-line RRT, the range of RRT tree extension is not limited, so the searched space becomes large. Thus, the processing time is rather long.

On the other hand, the processing time is much reduced by approximately 1/5 by proposed real-time RRT. However, there is no effect of using registered trajectories for avoidance, by only looking at these results. The reason for this fact is under investigation. Supposedly, the selected registered trajectory may not be appropriate for this simulation task. In future, it should be investigated how to make the registered trajectories as optimal as possible in order to reduce the processing time.

Table 1: Comparing of processing time [msec].

	Off-line RRT	Real-time RRT (without using registered trajectories)	Real-time RRT (with using registered trajectories)
Fig.7	119.7	24.2	22.8
Fig.8	Can't search	22.2	21.4

6 CONCLUSSIONS

Real-time RRT method for robot's path generation while avoiding collision with unknown moving obstacle is proposed. The robot gets information of obstacle movement at every interval, and forecasts the trajectory of obstacle to be avoided. The robot calculates the trajectory by extending the RRT tree, while the collision detection with the forecasted trajectory is carried out. Once the collision is detected, the detour trajectory is searched by also using the RRT. Here the improvement is provided that the RRT tree always approaches to the goal unless the collision is detected, which reduces the searching time drastically and makes the real-time path generation possible. Also, the method using the trajectories for avoidance, which are registered in advance, is proposed.

At present state, information of obstacle movement is given via LAN communication: however, in future, this information should be obtained by some sensor, such as vision sensor, laser range sensor, ultrasonic sensor, and so on.

REFERENCES

[1] Kuffner J J JrLavalle, S M"RRT-Connect: An Efficient Approach to Single-Query Path Planning", Proc IEEE Int Conf Rob Autom, Vol.2000, No.2, pp.995-1001, 2000

[2] Kazuya Tashiro, Seiji Aoyagi, "Path Planning of Redundant Robot Considering Moving Obstacles", Proc. of 7th SICE System Integration Division Annual Conference, 14-16 Dec. 2006, Sapporo

[3] Hiroto Sakahara, Yasuhiro Masutani, Fumio Miyazaki, "Real Time Motion Planning in Dynamic Environment Containing Moving Obstacles Using Spatiotemporal RRT", T.SICE, Vol.43, No.4, pp.277-284, April 2007

[4] Hiroto Sakahara, "norm-topics-Rapidly-exploring RandomTree", http://robotics.me.es.osakau.ac.jp/~sakahara/topics/rrt.html

[5] Masaharu Takano, "Detailed explanation : Robot Kinematics", Ohmsha, 2004

Path Searching of a Robot Manipulator Using Reinforcement Learning and Self-Organizing Maps

Kenji Hiraoka[1], Seiji Aoyagi[1]

[1] Dept. of Mechanical Engineering, Faculty of Engineering Science, Kansai University, Osaka, Japan

Abstract

In some applications such as search-and-rescue tasks, it is necessary for a robot manipulator to obtain a path, which can adapt to unknown and changeable environment. Reinforcement learning is effective for this purpose, since it is an algorithm for an agent to achieve an objective task by trial and error, where the agent only needs the goal of the task. A path searching method of an articulated robot in three-dimensional Cartesian space is focused on, while avoiding obstacles. In order to reduce the memory resources and the convergence time, a method of restructuring a configuration space by using self-organizing maps (SOM) is proposed.

Keywords:

path searching; reinforcement learning; SOM

1 INTRODUCTION

Reinforcement learning is an algorithm for an agent (a robot manipulator in the present study) to achieve an objective task (moving the robot arm's tip to a desired position in the present study) by trial and error, where the agent only needs the goal of the task, i.e., the way of achieving the task is not necessary to be taught during the transition from a start state to the goal one (this transition is referred to as an "episode" hereinafter in this article) [1]. Therefore, this algorithm is effective for realizing the robot task in case that the environment is unknown or changeable thus the teaching of a robot is rather difficult. In case of an articulated robot, there are infinite paths from a start position to a goal position. It is important to make a searched path as short as possible, while avoiding the collision with unknown obstacles. To this case, applying reinforcement learning should be effective, which is the theme to be dealt with in the present study.

Several researches of applying reinforcement learning to robotics fields have been reported, in which the stand-up motion of a 3-link robot [2], the inchworm locomotion of 2- or 5-link robot [5], the gait of a crawling four-legged robot [3], the path planning of a two-dimensional mobile robot [6], etc. are investigated. In these reports, however, the degrees of freedom (DOF) of the searched space, which are equal to the number of joints, are limited to a rather low order. Also, the resolution of the searched space, which is defined by the division number of movable range of the joint angle, is limited to a coarse value. The reason of the limitation is mainly based on the exponential increase of the number of sate (it is a unit space divided from a total searched space) as the increase of the DOF and the resolution, causing the lack of memory resources and the enormous convergence time in a computer calculation. Under this circumstance, the working space of a robot manipulator or a mobile robot has been generally limited to two-dimensional Cartesian case [2,3,5,6].

To overcome this problem, Actor-Critic algorithm has been employed [2,3,5], in which the action from one state to another state is statistically defined. This algorithm is effective for reducing the number of states compared with Q-learning algorithm [6], in which the action from one state to another is digitally performed. Combining other algorithms with reinforcement learning, such as neural networks [5], genetic algorithm (GA) [4], has been also investigated to optimally divide the searched space to the states. However, in spite of these contrivances, rather long calculating time is still required and the smoothness of the resultant path is still insufficient at present state.

Contrary to other research works described above, in the present paper, a path searching method of an articulated robot in three-dimensional Cartesian space is focused on, while avoiding obstacles. In order to reduce the memory resources and the convergence time, a method of restructuring a configuration space by using SOM in Actor-Critic algorithm (Note: not in conventional Q-learning algorithm) is proposed.

2 PATH SEARCH OF ARTICULATED ROBOT USING REINFORCEMENT LEARNING AND SOM

2.1 Actor-Critic Algorithm of Reinforcement Learning

Reinforcement learning is performed by repeating the transition from a state to another state. In the present paper, Actor-Critic method is used as an algorithm to decide the transition [1]. The concrete procedures are described hereinafter.

The configuration space is divided to many of state $S \in R^n$, where *n* is DOF the robot manipulator. The actor decides the action $a(S)$ randomly based on *n*-dimensional Gaussian distribution, the average and the standard deviation of which are $\boldsymbol{\mu}(S) \in R^n$ and $\sigma(S)$, respectively. Each component of $\boldsymbol{\mu}(S)$ means the average of moving angle, i.e., increment/decrement angle, of the correspondent joint from S to the next state S'. The $\sigma(S)$ means the standard deviation of moving angle, which is common for all joints for

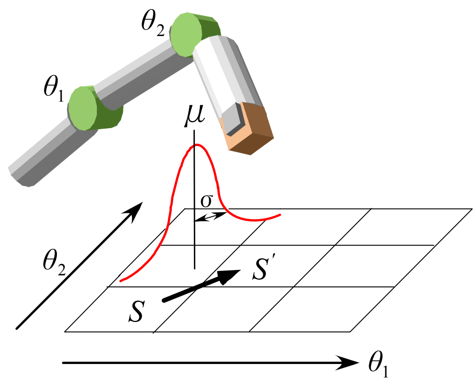


Figure 1: Actor-Critic algorithm in reinforcement learning

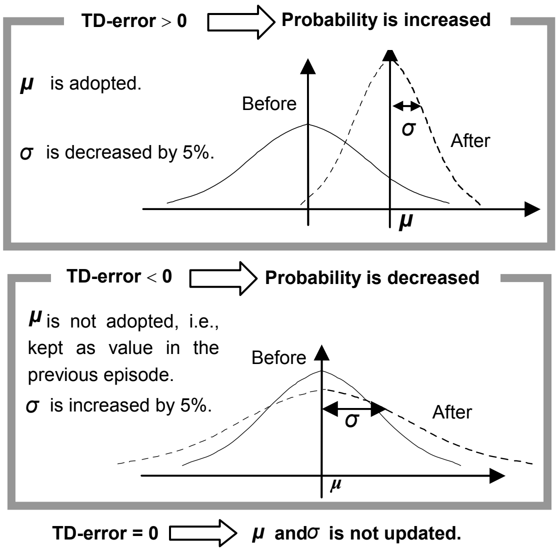


Figure 2: Schematic update rule of μ and σ based on TD-error.

simplicity. Namely, $\mu(S)$ and $\sigma(S)$ are the parameters to decide the action policy $\pi(S)$. The schematic view concerning this procedure is shown in Figure 1 in the two-dimensional case.

Each state S bears a state value $V(S)$, which is updated after the state is transited from S to S' as follows:

$$V(S) \leftarrow V(S) + \alpha\{r + \gamma V(S') - V(S)\}\ , \tag{1}$$

where r is the reward, α is the learning rate, and γ is the discount rate. An action of the actor based on $\pi(S)$ is referred to as a "step" hereinafter. At each step of one episode, the state value is updated, i.e., reinforcement learned, based on Eq. (1). All the state values are set to 0 before the first episode. The reward of r_1 is given when the robot arm's tip reaches finally inside the goal area. Also, the reward of r_2 is given when the required step number for reaching the goal is decreased compared to that in the previous episode. Thus, the r in Eq. (1) is the summation of r_1 and r_2.

The critic deals with the above-mentioned expression $r + \gamma V(s') - V(s)$ as the temporal difference (TD) error [1], and updates $\pi(S)$ at each step according to the following rules: if TD-error>0 holds true, $\mu(S)$ is adopted and $\sigma(S)$ is decreased by 5%. If TD-error<0 holds true, $\mu(S)$ is not adopted, i.e., kept as the value in the previous episode, and $\sigma(S)$ is increased by 5%. If TD-error=0, $\mu(S)$ and $\sigma(S)$ are not updated, i.e., kept as the values in the previous episode (Figure 2).

By iterating the episode many times based on the procedure mentioned above, the shortest path is expected to be obtained finally. Please refer the concrete values of α, γ, r_1, r_2, and those of initial μ and σ to Table 1 in Section 3, which are used in a computer simulation.

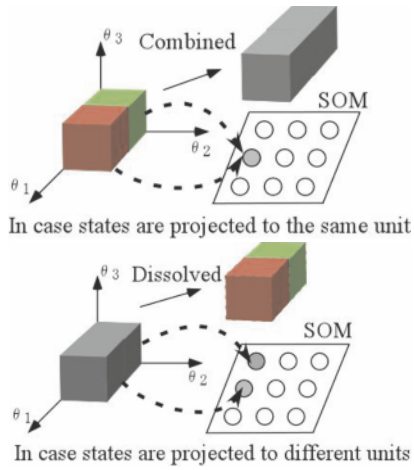


Figure 3: Combination and dissolution of states by self-organizing maps (SOM)

Combinig

$V_{combined} = max(V(S_1), V(S_2))$

$\mu_{combined} = (min(\mu_1(S_1), \mu_1(S_2)), min(\mu_2(S_1), \mu_2(S_2)), \ldots, min(\mu_n(S_1), \mu_n(S_2)))^T$

$\sigma_{combined} = max(\sigma(S_1), \sigma(S_2))$

For speeding up learning process, larger V is adopted.

For achieving fine exploring, smaller μ is adopted.

For avoiding local minimum problem, larger σ is adopted.

Disolving

V, μ, σ of each dissolved state take over ones before the dissolving process, i.e., the values of the state which is combined in a past.

Figure 4: Initialization of V, μ, and σ after combining two states by SOM.

2.2 Self-organizing maps, SOM

Required memory resources and convergence time for reinforcement learning are much influenced by the way to divide the configuration space to the states. In the present paper, SOM algorithm is employed for optimally restructuring the states during the reinforcement learning. In the concrete, the states existing in the area with less influence on the learning result are combined, while those existing in the area with much influence are dissolved, as schematically shown in

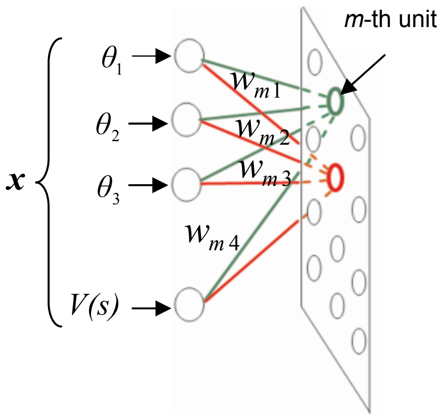


Figure 5: Concept of SOM.

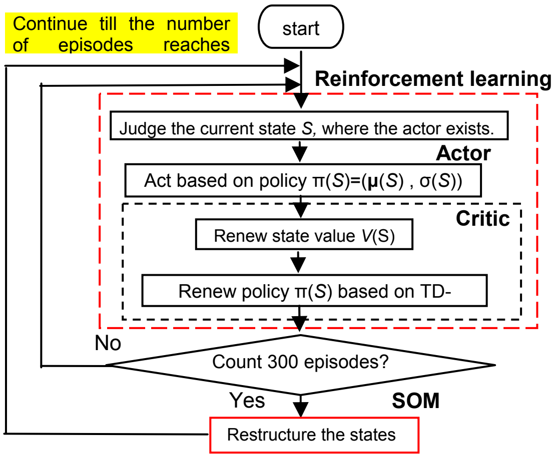


Figure 6 : Flow chart of searching process

Figure 3. The values *V*(*S*) and *π*(*S*) of combined state are initialized by using ones of two states before the combining process, the rule of which is schematically shown in Figure 4. On the other hand, these values of each dissolved state take over ones before the dissolving process, i.e., the values of the state which is combined in a past.

The SOM algorithm in this study is documented hereinafter. The SOM is two-layered neural networks, i.e., it has the input layer and the output layer, forming a projection map between them [1]. The internal weights of the networks are unsupervised learned. Here, the case of searching the path of 3-DOF robot in 3-D space is assumed. The *m*-th unit in the output layer has its weight vector $\boldsymbol{w}_m = (w_{m1}, \cdots, w_{m4})^T$, the dimension of which is as same as that of the input vector $\boldsymbol{x} = (V(s), \theta_1, \theta_2, \theta_3)^T$, as schematically shown in Figure 5. When $\boldsymbol{x}$ is input to the SOM, a "winner" unit *C* is selected among the all units of the output layer, of which $\boldsymbol{w}$ has the nearest Eucid distance to $\boldsymbol{x}$. The $\boldsymbol{w}_m$ in the neighborhood of *C* (including *C*) is updated according to the following expression:

$$\boldsymbol{w}_m(i+1) = \boldsymbol{w}_m(i) + h_{cm}(i)[\boldsymbol{x}(i) - \boldsymbol{w}_m(i)] \tag{2}$$

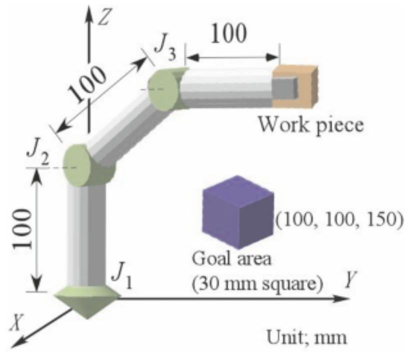


Figure 7 : Model of 3-DOF robot

Table 1 : Parameters used in reinforcement learning and SOM

	Parameter	value
Actor-Critic	Learning rate α	0.1
	Discount rate γ	0.95
	Initial value of average of moving angle μ	0 deg
	Initial value of standard deviation of moving angle σ	6 deg
	Reward for reaching goal r_1	+1
	Reward for decreasing steps r_2	+3
	Number of states	64,000(= 40^3)
SOM	SOM map size	30×30
	Initial value of learning coefficient λ	0.5
	Initial value of dispersion of neighborhood β	5
	Total iteration number	10

$$h_{cm} = \lambda(i)\exp\left(-\frac{\|r_c - r_m\|^2}{2\beta(i)^2}\right) \tag{3}$$

where *i* is the iterative number of the learning, $\lambda(i)$ is the learning coefficient (0< $\lambda(i)$ <1), and $\beta(i)$ indicates the dispersion of the neighborhood of *C*. The $\lambda(i)$ and $\beta(i)$ decrease as the learning progresses. Please refer the concrete values of the map size, initial values of $\lambda(i)$ and $\beta(i)$, etc. to Table 1 in Section 3.

2.3 Proposal of State Restucturing Using SOM

The S having larger $V(S)$ should be an important state for achieving the task, while the S having lower value or the initial value, i.e., the state has not updated at all, should be an unnecessary state. By inputting $\boldsymbol{x} = (V(s), \theta_1, \theta_2, \theta_3)^T$ to the SOM, the states existing in the neighbor with each other and/or having near value of $V(S)$ with each other are clustered together in the output layer. In the computer simulation of the present paper, the movable range of each joint of 3-DOF robot is divided to 40 segments, forming

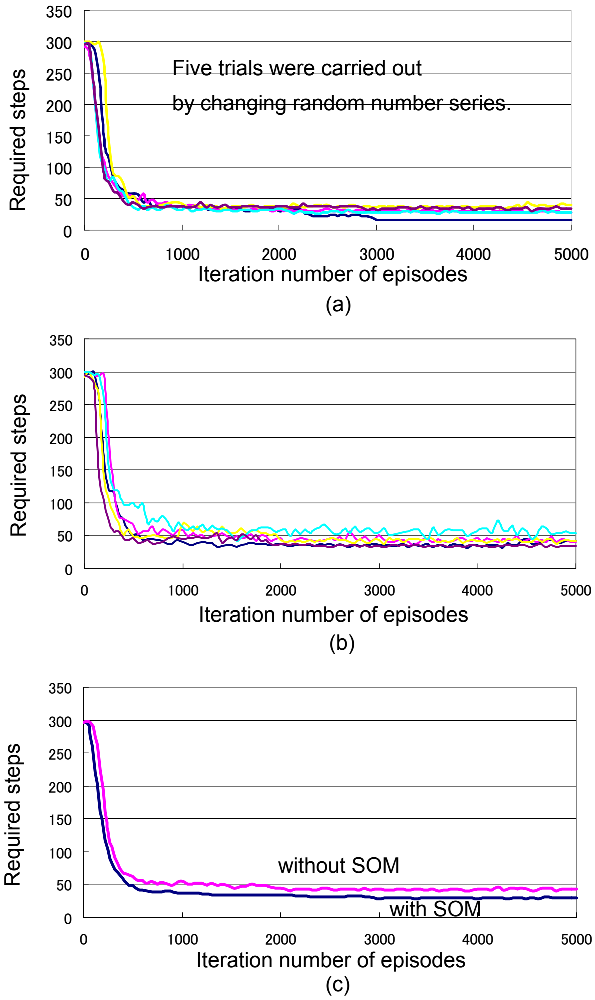


Figure 8 : Relationship between iteration number of episodes and required steps : (a) with SOM; (b) without SOM; (c) averaged results for five trials.

40×40×40=64,000 states in the configuration space. And, the map size of 30×30 is employed in the SOM. In this case, the 64,000 states are restructured to 900 states. In the course of learning process, the SOM is applied at every 300 episodes, details of which are described later in Section 3.

The flow chart of the path searching process using reinforcement learning and SOM is shown in Figure 6.

3 SIMULATION RESULTS

A task is assumed, in which a 3-DOF articulated robot of RPP type (see Figure 7) reaches the goal area (inside a cubic with 30 mm side in Cartesian space) within 300 steps. The concrete values of parameters used in the reinforcement learning and SOM are shown in Table 1.

Computer simulation of searching a path was carried out in two cases with/without applying SOM (it is applied at every 300 episodes in this case). The relationship between the iteration number of episodes (the number of learning) and the required steps is shown in Figure 8. Note that the smaller steps means that the shorter path is achieved. Looking at this

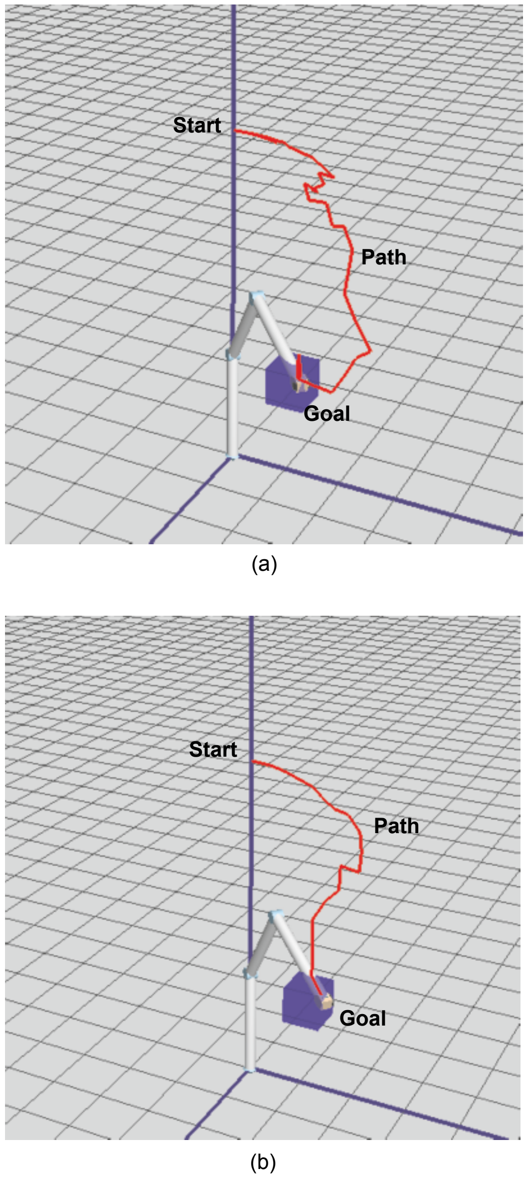


Figure 9 : Resultant path : (a) using only Actor-Critic; (b) using both Actor-Critic and SOM.

figure, the smaller convergence time is achieved by using SOM. Also, the smaller convergence value of the step number, i.e., the shorter path, is realized by using SOM. The resultant path at the 5,000th episode is shown in Figure 9, exhibiting that the smoother and shorter path is realized by using SOM.

4 STATE SPACE DIVIDED BY SOM

The resultant state space at the 5,000th episode, which has been divided by SOM, is shown in Figure 10. In this figure, the area having the same grey scale expresses the same state which was categorized by SOM. For reference, the distribution of state values $V(S)$ in the state space is shown in Figure 11. Comparing these figures with each other, it is confirmed that the areas bearing larger $V(S)$ (see Figure 11)

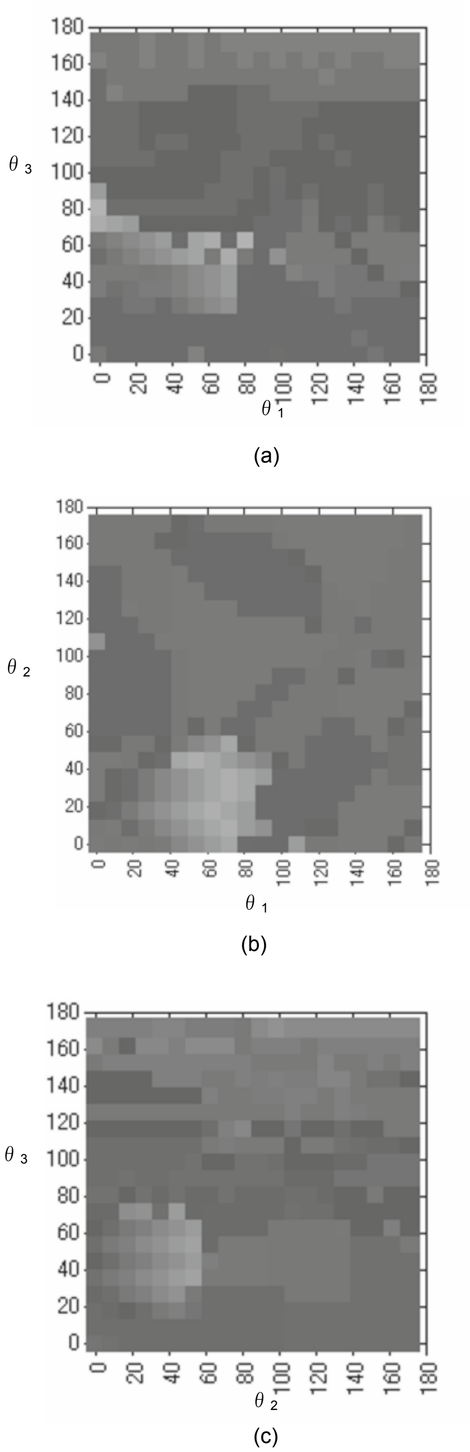


Figure 10 : Resultant state space at 5,000th episode: (a) on θ_1 - θ_3 plane at θ_2 =0 deg; (b) on θ_1 - θ_2 plane at θ_3 =45 deg; (c) on θ_2 - θ_3 plane at θ_1 =54 deg.

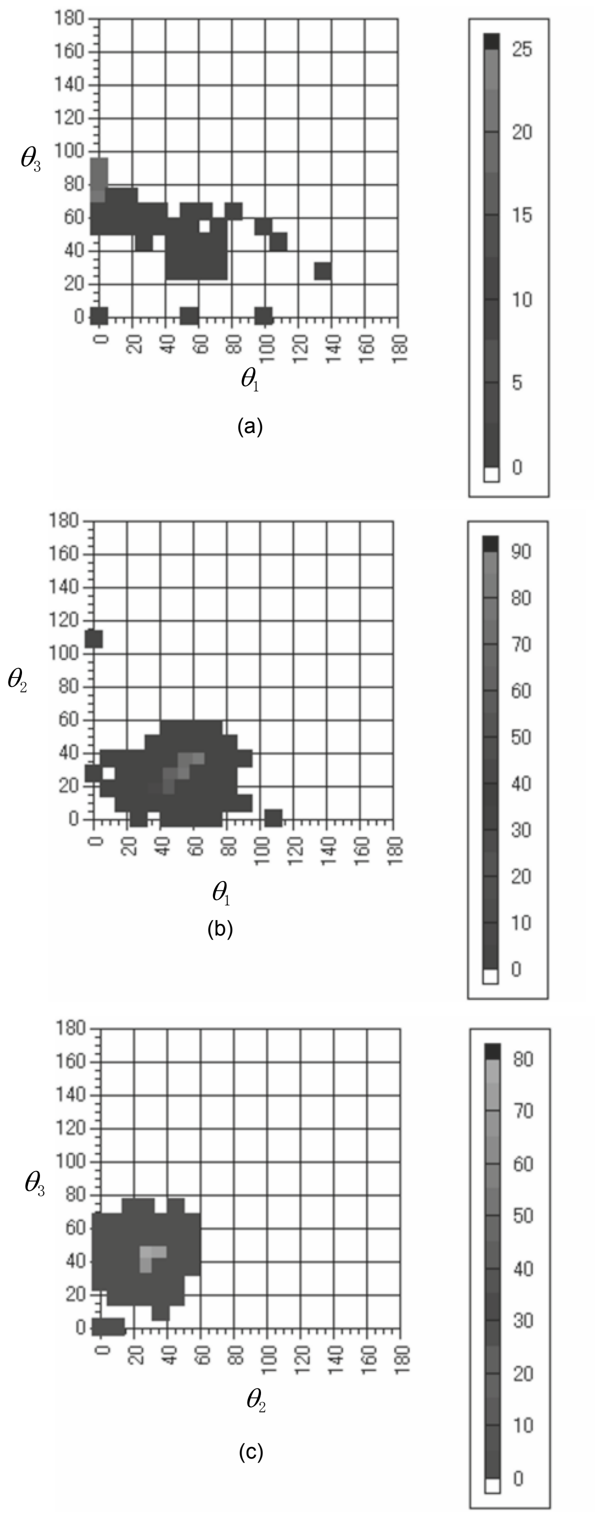


Figure 11 : Distribution of state values $V(S)$.

are divided into many states, i.e., different states (see Figure 10). On the other hand, the areas bearing almost zero in $V(S)$ (see Figure 11) are combined, forming states having comparatively large areas (see Figure 10).

Qualitatively speaking, $V(S)$ of the states existing in the area with much influence on the learning result is frequently updated, and the value of it is increased as the result. Thus, such significant area should be divided into many states by SOM. On the other hand, $V(S)$ of the states existing in the area with less or no influence on the learning result is scarcely updated and kept almost zero. Thus, such meaningless area should be combined by SOM into one state, which has comparatively large area. These presumptions do not contradict the above-mentioned simulation results of the state space and the state values.

5 CONCLUSIONS

The path of an articulated robot is searched in three dimensional Cartesian space by reinforcement learning, while avoiding obstacles. In order to reduce the searched configuration space for achieving the shorter path and the smaller convergence time, three methods are proposed, which are using SOM to restructure the states during the learning.

6 REFERENCES

[1] R. S. Sutton, and A. G. Barto (translated by S. Mikami, and M. Minagawa), "Reinforcement Learning", Morikita Publishing Co., Ltd., Tokyo, 2000.

[2] J. Morimoto, and K. Doya, "Acquisition of Stand-up Behavior by a 3-link-2-joint Robot using Hierarchical Reinforcement Learning", *Journal of the Robotics Society of Japan*, Vol.19, No.5, pp.574-579, 2001.

[3] H. Kimura, T. Yamashita, and S. Kobayashi, Reinforcement Learning of Walking Behavior for a Four-legged robot, *Transactions of the Institute of Electrical Engineers of Japan*, Vol. 122-C, No. 3, pp. 330-337, 2002.

[4] K. Ito, F. Matsuno, and A. Gofuku, A Study of Reinforcement Learning for Redundant Robots –New Framework of Reinforcement Learning that Utilizes Body Image-, *Journal of the Robotics Society of Japan*, Vol. 22, No. 5, pp. 672-689, 2004.

[5] M. Nunobiki, K. Okuda, and S. Maeda, Reinforcement Learning of Multi-Link Robot with Fuzzy ART Neural Networks for State-Space Segmentation, *Journal of the Japan Society for Precision Engineering*, Vol. 71, No. 1, pp. 141-145, 2005.

[6] H. Iwasaki, and N. Sueda, A System of Autonomous State Space Construction with the Self-Organizing Map in Reinforcement Learning, in *Proc. The 19th Annual Conference of the Japanese Society for Artificial Intelligence*, pp. 1-4, 2005.

Improved Positioning Accuracy for a Water Hydraulic Manipulator with State Feedback Controller

Ali Muhammad[1], Jouni Mattila[1], Tapio Virvalo[1], Matti Vilenius[1]

[1] Dept. of Intelligent Hydraulics and Automation, Tampere University of Technology, Tampere, Finland

Abstract

The performance of hydraulic positioning servo systems is very often limited by the poor damping and the variable dynamics. In addition, the positioning accuracy of hydraulic drives suffers from nonlinearities in the valve and the actuator. It is well known that the use of states like velocity and acceleration as feedback signals can significantly improve the dynamic characteristics of such systems. As a result higher position loop gains can be chosen for improved positioning accuracy. Despite remarkable results it has been understood that the state-feedback controllers require considerably good velocity and acceleration feedbacks and researchers have questioned the practical implementation and the robustness of state-feedback controllers when it comes to systems with multiple degrees of freedom such as hydraulic manipulators. In this paper the development of state-feedback controller for a 5-DOF water hydraulic manipulator has been presented. The manipulator joints are only equipped with position encoders, which in general is the minimum instrumentation for an industrial manipulator. The results show that although the controller gains and thus performance is limited by the noise in the feedback signals, it can still provide improvements in tracking and positioning accuracy of the manipulator's end-effecter.

Keywords:

Water Hydraulics, Robotics, Position Control

1 INTRODUCTION

Hydraulic position servo systems regularly exhibit poor damping and variable dynamics, limiting the closed loop positioning gains. In addition, the positioning accuracy of hydraulic drives suffers from several nonlinearities which exist in the valve and the actuator. The most dominant of these nonlinearities are the valve offset and hysteresis, the friction in the load and the actuator and the leakage both in the valve and the actuator. Despite their drawbacks, hydraulic servo systems still compete with their electrical counterparts in demanding industrial applications because of their high power to size ratio and reliability [1].

The most popular controller used in hydraulic position servo systems is the P-controller. Though robust, it gives a very sluggish response when tuned using robustness criteria [2]. Another drawback of the P-controller is that it provides a very limited possibility to influence the system dynamics. As a result, a lot of research has been done to investigate and to develop better solutions.

In [3], the traditional PI-controller has been modified to implement a nonlinear PI-controller. However, the testing of the controller was limited only to a single joint of a hydraulic manipulator. On the other hand, Virvalo has suggested avoiding the integral term in controllers for hydraulic position servo systems [4]. It has been shown that because of the nonlinearities in the valve, the use of the integrator can lead to a hunting behaviour. Moreover, Virvalo has compared the performance of the P-controller, the filtered P-controller (P-controller with the first order lag) and the state-feedback controller with varying loads [5]. The results show that the performance of the state-feedback controller depends on the load variation. Linjama [2] has analyzed the robustness of the same controllers with a focus on the parameter variations and sampling time. He concluded that a robustly designed state-feedback controller can lead to unstable system when practically realized as time discrete. In [6], Torben tests a set of linear controllers on a two joint manipulator driven by linear hydraulic drives. Bonchis has documented the performance of a very wide range of controllers for hydraulic drives in his work [7]. But once again all the tests were performed on a single joint of a mining manipulator.

Since hydraulic drives exhibit nonlinear and variable characteristics, many researchers have investigated solutions using nonlinear and adaptive controllers. The use of linearised pressure dynamics as a feedback has been studied in [8]. The system was found to be robust against the load and parameter variations, but position tracking results were not remarkable. In the same study, Torben also tests adaptive controllers on a two joint manipulator driven by linear hydraulic drives. In this case, despite attaining better accuracy, the position tracking exhibit chattering due to the noise in the velocity feedback signal. In [9], Tochizawa conducts experiments with a two joint manipulator driven by rotary hydraulic actuators. The controllers have been compared under the condition of varying load on the manipulator. The adaptive controller was found much more complicated to implement but it did not guarantee any more accuracy than the classical controller.

The above paragraphs provide just a glimpse of the developments in the control of hydraulic position servo systems. However, majority of the research has been done with oil hydraulic actuators, and more importantly, with systems composed of one or two actuators. Considering the nonlinearities of hydraulic components, the situation gets worse with water hydraulic servo systems. The choice of the water hydraulic components is much more limited and the quality of the water hydraulic servo valves is not as high as that of the oil hydraulic servo valves. The leakage is normally higher due to low viscosity, and so is the friction in the actuator. Despite, the characteristic advantages of hydraulics together with water as the pressure medium (fire and environmentally safe, chemically neutral, not activated, not affected by radiation) are highlighted in critical applications such as remote handling operations in the International Thermonuclear Experimental Reactor (ITER) [10].

With these developments in water hydraulic applications, it is clear that an investigation in the control techniques is required to improve the position control of these systems. The positioning response is important not only to move such water hydraulic manipulators from point to point in space, but also to employ the force control techniques, such as the position-based impedance control (PBIC), where the performance of the scheme heavily depends on the performance of the internal position control loop [11].

In the following section, the background and the choice for state-feedback controller for the study will be discussed briefly. The experimental setup of 5-DOF (Degrees Of Freedom) water hydraulic manipulator has been presented in section 3. In section 4 the development of state-feedback controller for the first joint of the manipulator as an example has been discussed. Position tracking measurements and accuracy results of the manipulator's end-effecter in Cartesian space will be presented in section 5. In the last section, the conclusions will be drawn on the basis of results and some future goals will be described.

2 CHOICE OF STATE-FEEDBACK CONTROLLER

A dexterous manipulator composed of water hydraulic actuators is a requirement for carrying out remote handling operations in the diverter region of ITER. The application is demanding in terms of compactness, reliability, dexterity and accuracy. The manipulator should be operable remotely and include the master-slave scheme. The positioning response and accuracy are important not only for point-to-point motion of the manipulator in space but also for the employment of the force control schemes, such as position based impedance control (PBIC) or hybrid position and force control [11]. Since hydraulic components (both water and oil) exhibit poor damping limiting position loop gain, the performance can only be improved by using control techniques that can influence the dynamics of the system.

It is well known that the use of states such as velocity and acceleration as feedback can improve the dynamics of fluid power (hydraulic) servo systems. As a result, higher closed loop gains can be used for improved dynamic response and reduced steady state error. Controllers employing such technique are widely termed as state-feedback controllers. The characteristics of the state-feedback controller can be useful in hydraulic manipulators, where the dynamic response of each drive propagates further, to strongly affect the end-effecter positioning response.

The design of the state-feedback controller is based on the principal of selecting new pole locations to achieve the desired dynamics. Figure 1 shows the state-feedback controller with ideal feedback signals.

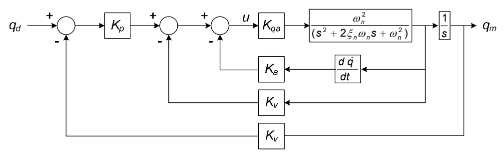


Figure 1: State-feedback controller with ideal feedbacks

Mathematically the modified linear model of the system can be written as in equation 1, which is the transfer function between the control signal of the servo valve and the position of the actuator. Where K_{qa} is the velocity gain, ω_n is the natural frequency and ξ_n is the natural damping of the system. K_p, K_v and K_a are the position, velocity and acceleration feedback gains respectively.

$$\frac{Q(s)}{U(s)} = \frac{K_p K_{qa}\omega_n^2}{s^3 + (2\xi_n\omega_n + K_a K_{qa}\omega_n^2)s^2 + (\omega_n^2 + K_v K_{qa}\omega_n^2)s + K_p K_{qa}\omega_n^2} \quad (1)$$

Assuming ideal feedback of the states (velocity and acceleration), the poles of the above system can be selected as desired, by modifying the feedback gains. In practice this means using both velocity and acceleration sensors, which will result in additional cost and complicated instrumentation, especially in the systems with several hydraulic actuators. If the velocity sensor will be used alone, the quality of the acceleration signal will depend on the resolution of the velocity sensor, and if the acceleration sensor will be used alone, the integration will result in an offset in the velocity signal.

In practice, velocity and acceleration signals are obtained by numerical differentiation of the position signal. This results in a poor quality velocity and acceleration signals. The reason is the quantization noise, which is inversely proportional to the resolution of encoder/resolver and directly proportional to the sampling frequency. Increasing the resolution of the encoder/resolver may improve the performance, but this will result in increased instrumentation cost. Also, it does not provide the complete answer, because the quantization noise may still appear at small velocities at which these manipulators needed to be derived. At the same time modern servo systems require sampling frequencies approaching few kilo hertz. In Figure 2, the realization of the state-feedback controller is shown where only position feedback signal is available.

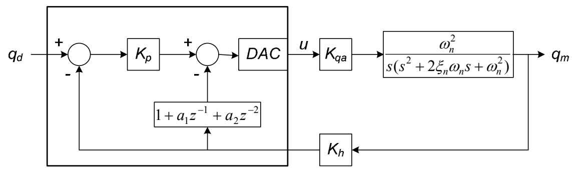


Figure 2 : Realization of state-feedback controller

As discussed in last section, lots of work has been carried out to study the application of state-feedback controllers on fluidic servo systems. However, in most cases the results have been presented using a single hydraulic drive and the control input signal is a large step. On the other hand a hydraulic manipulator is composed of multiple hydraulic drives connected in series to form a non-rigid structure. In addition a large step is never a realistic input for manipulators of larger size as majority of hydraulic manipulators are. Most of the time these manipulators are required to follow smooth position profiles in space. These profiles are designed according to the dynamic limitations of the manipulator and the task at hand. In most hydraulic manipulators, the driving velocity is limited because of large manipulator size and supply pressure. These lower velocities and accelerations result in a much higher quantization noise and thus strongly restrict the servo loop gains or result in vibrations of small amplitude. Because of the manipulator's chain like structure, these small vibrations can propagate from joint to joint and lead to undesired behavior of the manipulator's end-effecter.

To fully utilize the capabilities of the state-feedback controller, velocity and acceleration signals of a relatively good quality are required. This requirement reduces its practical application in several cases [2]. In [12] Mäkinen evaluates the effects of the feedback signal quality on the position servo control of a pneumatic drive. The best results were achieved by using an observer to calculate the velocity and the acceleration feedback signals. Mäkinen also shows that the use of a filter results either in too much delay or in a failure to eliminate the quantization noise completely. In [13], Virvalo concludes that a high resolution of the position encoder is required to implement the state-feedback controller for the hydraulic drives. The study showed that good results can be obtained when an n-sample estimator is used to calculate the velocity and acceleration signals. Implementation of an n-sample estimator is shown in equation 2.

$$\dot{q}(k) = \frac{q(k) - q(k-n)}{n \cdot T_s} \tag{2}$$

3 EXPERIMENTAL SETUP

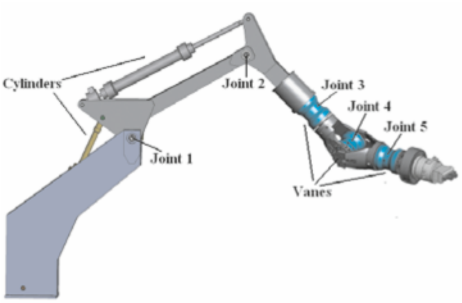


Figure 3: 5-DOF water hydraulic manipulator

Experiments were performed using the water hydraulic manipulator shown in Figure 3. The manipulator consists of a 2-link planar arm with shoulder and elbow rotational joints, powered by water hydraulic cylinder actuators. The end of the planar arm is fitted with a robotic wrist, with three rotational joints powered by water hydraulic vane actuators. The dimensions of the cylinders for the first and the second joint are 63/32-350 mm and 50/28-350 mm respectively. The specific volume of vanes is 2.8×10^{-5} m^3/rad with a motion range of 270°. The cylinders and vanes are driven by flow control servo valves with a flow rate of 6 l/min and 4 l/min respectively at a nominal pressure of 3.5 MPa (35 bar) per control notch, respectively. The joint angles are measured using pulse encoders with a resolution of 5000×25 pulses/revolution. The sampling frequency of the system is 1 kHz. A supply pressure of 20 MPa (200 bar) is used.

4 DEVELOPMENT OF CONTROLLER

During the experimental study the state-feedback controller was developed for each joint of the manipulator. However, due to limited space only the example of the first joint (joint 1 as shown in Figure 3) is presented here. The example is selected because of the wide variations in the dynamics of the link which not only depends on its own position but also the posture of the rest of the manipulator's links.

The dynamic parameters of the joint can be estimated by driving it in the open loop, while keeping the rest of the joints in different positions. The nominal value of natural frequency ω_n was determined to be 35 rad/s. The velocity gain K_{qa} of the system when the cylinder is derived outwards is measured approximately to be 3.5 deg/s/V. It is difficult to estimate the natural damping of a hydraulic servo system but a good approximation can be 0.15.

The initial tuning of state-feedback controller has been widely discussed for example in [2] and [7]. Initial tuning parameters can be found using the following set of equations 3, 4 and 5.

$$K_p = \frac{(1...2).K_{cr}}{K_{qa}K_H} \tag{3}$$

$$K_v = \frac{(\omega / \omega_n)^2 - 1}{K_{qa}} \tag{4}$$

$$K_a = \frac{2\xi_n((\omega / \omega_n).(\xi / \xi_n) - 1)}{K_{qa}\omega_n} \tag{5}$$

Where, $K_{cr} = 2\xi_n\omega_n$ is the open loop critical gain, $\omega = (0.9…1.5)\omega_n$ is new selected frequency of the system, $\xi = 0.3…0.7$ is the new selected damping and K_H is the position feedback gain.

Using the above set of equations when $\omega = 1.1 \times \omega_n = 38.5$ rad/s, $\xi = 0.7$ and $K_{cr} = 10$ 1/s, the calculated gains were found to be $K_p = 166$ V/rad, $K_v = 3.5$ V/rad/s and $K_a = 0.6$ V/rad/s^2. The simulated step response of the system with these tuning parameters is shown in Figure 4. The system remains stable without any overshoot.

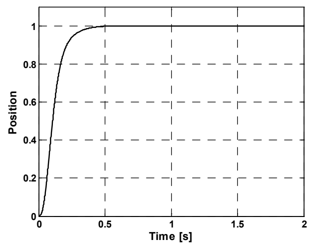


Figure 4: Simulated step response

The above simulation makes use of the ideal velocity and acceleration feedback signals. In modern practice the control systems are implemented using digital computers with time discrete equivalent of continuous time models. As mentioned earlier the velocity and acceleration feedbacks are obtained by the numerical differentiation or some variation of it (n-sample estimators) of position feedback signal. As feedback signals obtained by direct differentiation are too noisy, the n-sample estimators were opted as previous studies show their usefulness. The velocity feedback signal is obtained with a value of $n = 8$ and acceleration signal is obtained with a value of $n = 4$. These n-sample estimators though result in a reduction of quantization noise but at the same time introduces delay in the feedback signal. In this case velocity feedback signal is delayed by 8 samples and acceleration feedback signal is delayed by 12 samples. The idea is to find a balance between the delay and quantization noise as both factors results in the reduction of possible feedback gains. Figure 5 shows the obtained velocity and acceleration feedback signals for the position profile of Figure 7. The phenomenon of quantization noise is especially visible in case of acceleration feedback signal.

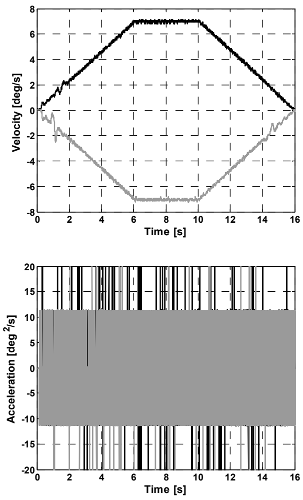


Figure 5: Velocity and acceleration feedback signals

Hence, with the tuning parameters obtained during the design phase the manipulator joint repeatedly went unstable for a step input. To achieve a stable response the gains needed to be retuned and lowered. The new values of the gains were found to be as K_p = 116 V/rad, K_v = 2.8 V/rad/s and K_a = 0.0524 V/rad/s^2. This finding confirms the concerns rose in [2] about the practical implementation of state-feedback controller. With these tunings the position step response from -15^o to 11^o and 37^o to 11^o is shown in Figure 6. The valve saturation is fixed at ±20% as the movement of the joint is too fast if valve is not saturated. The zoomed plot of step response shows some decaying vibrations of very low amplitude, however the response is stable and the steady state accuracy is better than 0.05^o.

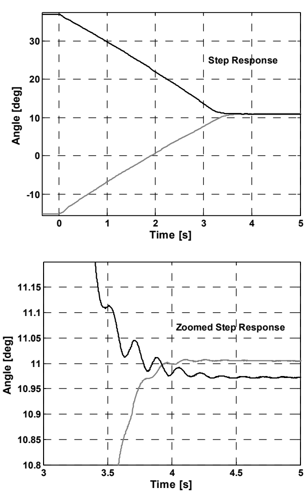


Figure 6: Step response of the joint 1

As mentioned earlier that a large step input is seldom a realistic input for a manipulator. Hence, a smooth position profile is the next input tested. With the same tuning during the profile tracking, the motion of the joint though remained stable however, minor vibrations were observed during profile tracking. In case of a manipulator these small vibrations can lead to instability and undesired behavior at the end-effecter. The reason of vibrations during profile tracking is the quantization noise. Hence a smaller value of gains had to be chosen. After some tuning of gains in such a way that no vibrations were observed during the profile tracking the tuned values of gains were further reduced to $K_p = 110$ V/rad, $K_v = 2.0$ V/rad/s and $K_a = 0.05$ V/rad/s^2.

Figure 7 shows the position tracking response from -9^o to 80^o and 80^o to -9^o. The input trajectory has an approximate velocity of 7^o/s. The tracking response indicates the controller follows the input without any serious vibration. The tracking error remains less then 1.0^o all times and steady state accuracy remains better than 0.05^o.

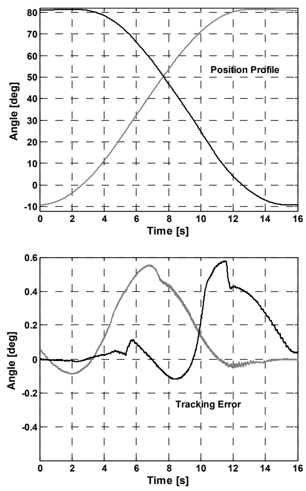


Figure 7: Tracking response of joint 1

From the above experiments and discussion it is quite clear that quantization noise in the velocity and acceleration feedback signals results in the instability during the practical implementation of state-feedback controller. Further reduction in the feedback gains is required when the joint is to be driven with smooth position profiles of reduced velocity. In fact the behavior was found consistent during the development of controllers for the rest of the joints of the manipulator.

5 CARTESIAN MEASUREMENTS

Following the same steps as described in prvious section the state-feedback controller was designed for each joint of the 5-DOF water hydraulic manipulator. The manipulator is driven across in its workspace using smooth position trajectories. The tracking response and the error for four different profiles are shown in Figure 8 and Figure 9. The tracking error remains smaller than 5 mm for linear motion and smaller than 0.1^o for revolute motion. The notable deviation at the start of the profile is believed to be the limited pressure to move all the joints simultaneously. Even though the feedback gains were limited due to the delay and quantization noise in the velocity and acceleration signals, still much higher position gains as compared to P-controller were possible to achieve an improved tracking and positioning accuracy at the end-effecter of the manipulator. To be mentioned here the movement of the joints and thus the workspace of the manipulator were slightly limited, as the manipulator went unstable at the workspace boundary. The velocity and acceleration feedback gains needed to be decreased much further to keep the manipulator stable in entire workspace and thus the no noticeable improvements were achieved in that case.

6 CONCLUSIONS

In this study the implementation of the state-feedback controller for water hydraulic manipulator was analyzed. The results show that the performance of the controller depends on the feedback signal quality. The presence of quantization noise and delay results in the reduced feedback gains, so a care must be taken during the implementation. However, even with these reduced gains a considerable improvement in positioning and tracking accuracy of the manipulator was achieved as compared to a traditional P-controller. The manipulator was found stable in the majority of work space. The stability and accuracy of the manipulator can also be guaranteed in entire workspace if the gains are scheduled as a function of manipulator's posture. However this is beyond the scope of current study and is a future goal.

The future work will include the further enhancement of controller by obtaining the velocity and acceleration feedbacks by using the observer. This may result in improved accuracy and robustness. The performance of the controller needs to be tested with varying conditions of load. Once the dynamic performance and positioning accuracy are satisfactory, force control needs to be included to make the manipulator practically applicable for remote handling operations.

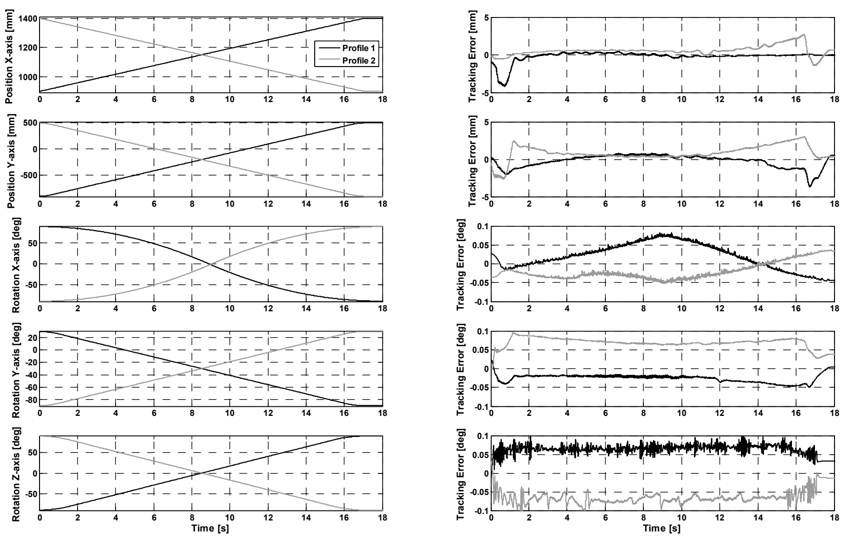


Figure 8: Position response and tracking error of the manipulator

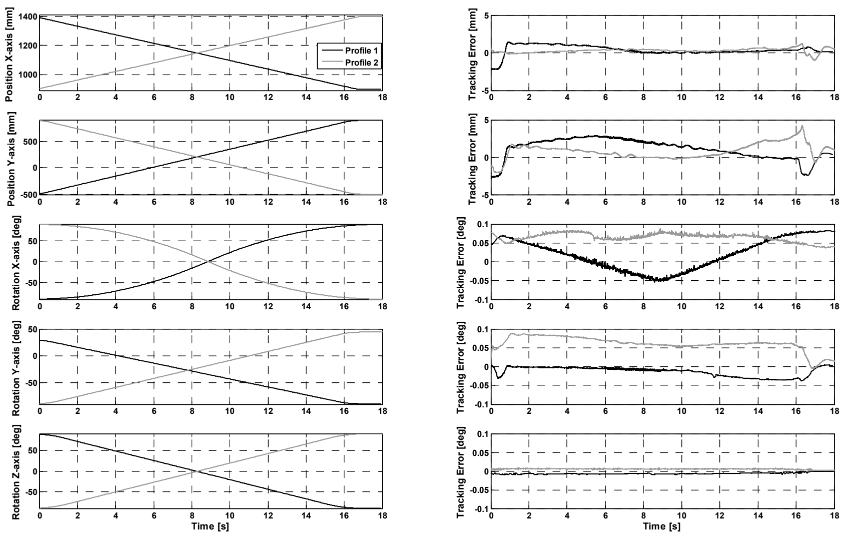


Figure 9: Position response and tracking error of the manipulator

7 ACKNOWLEDGMENTS

This work, supported by the European Communities under the contract of Association between EURATOM/TEKES, was carried out within the framework of the European Fusion Development Agreement. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

8 REFERENCES

[1] Measson Y., David O., Louveau F., Friconneau J.P., “Technology and control for hydraulic manipulators”, Symposium On Fusion Technology (SOFT 2002), Helsinki, Finland, 2002.

[2] Linjama M., Virvalo T., “Robustness of some low-order controllers in an electrohydraulic position servo drive”, Bath Workshop on Power Transmission and Motion Control (PTMC 2001), Bath, United Kingdom, 2001.

[3] Sepehri N., Khayyat A.A., Heinrichs B., “Development of nonlinear PI controller for accurate positioning of an industrial hydraulic manipulator”, Journal of Mechatronics, Vol. 7, No. 8, pp. 683-700, 1997.

[4] Virvalo T., “PI and PID-controllers in a hydraulic position servo system - What it is all about”, Proceedings of the Fifth International Conference on Fluid Power Transmission and Control (ICFP 2001) April 3-5, 2001, Hangzhou, China, 2001.

[5] Virvalo T., Linjama M., “Experimental study of robustness of some controllers in a hydraulic position servo drive”, Hydraulika a Pneumatika, June 5-6, 2001, Ostrava, Czech Republic, 2001.

[6] Andersen T.O, Hansen M.R., Pedersen. H.C., Conrad F., “Comparison of linear controllers for a hydraulic servo system”, Proceedings of the 6^{th} JFPS International, Symposium on Fluid Power, Tsukuba, November 7-10, 2005.

[7] Bonchis A., Corke P.I., Rye D.C., “Experimental evaluation of position control methods for hydraulic systems”, IEEE Transactions on Control Systems Technology, Vol. 10, No. 6, November 2002.

[8] Andersen T.O, Hansen M.R., Pedersen. H.C., Conrad F., “Feedback linearization applied on a hydraulic servo system”, Proceedings of the 6^{th} JFPS International, Symposium on Fluid Power, Tsukuba, November 7-10, 2005.

[9] Tochizawa M., Edge K.A., “A comparison of some control strategies for a hydraulic manipulator”, Proceedings of the American Control Conference, San Diego, California, June 1999.

[10] Siuko, M., Pitkäaho, M., Raneda, A., J. Poutanen, J., Tammisto, J., Palmer, J. and Vilenius, M., “Water hydraulic actuators for ITER maintenance devices”, Fusion Engineering and Design 69 (2003) pp. 141-145.

[11] Tafazoli S., Salcudean S.E., Hashtrudi-Zaad K., Lawrence P.D., “Impedance control of a teleoperated excavator”, IEEE Transactions on Control Systems Technology, Vol. 10, No.3, May 2002.

[12] Mäkinen E. Virvalo T., “Influence of Position Feedback Signal Quality on Behaviour of a State Controlled Pneumatic Position Servo”, Proceedings of the ICMA 02, The Fourth International Conference on Machine Automation human-Friendly Reliable Mechatronics, Tampere, Finland, September 11-13, 2002.

[13] Virvalo T., Linjama M., Mattila J., “Comparing different controllers of an electrohydraulic position servo”, Proceedings of the 4th International Conference on Fluid Power Transmission and Control, '97 ICFP, September 9-11, Hangzhou, China, 1997

Trajectory Tracking Feedback Control of Robot Manipulators with Coupled Dynamics of Force and Position

Hiroaki Ichii[1], Sadao Kawamura[2]

[1] Dept. of Control Engineering, Nara Nationl College of Technology, Nara, Japan

[2] Dept. of Robotisc, Ritsumeikan University, Shiga, Japan

Abstract

When a robot manipulator contacts an object, a Coulomb friction force occurs in the motion direction on the surface. Therefore, it seems that the previously proposed control methods can not obtain good control performance with trajectory tracking. This paper proposes a new trajectory tracking control method for cases that there is a Coulomb friction force between a robot and an object. It is mathematically proven that trajectory tracking errors between actual motions and desired motions can be effectively reduced as position and velocity feedback gains increase. The trajectory tracking performance of the proposed feedback control method is demonstrated by some simulation results.

Keywords:
Tracking; Coulomb Friction; Decoupling Control

1 INTRODUCTION

If a robot manipulator contacts the surface of an object as seen in Fig.1, a constraint force is generated between the end effecter of the robot and the surface. Many papers on stability and constraint forces (or hybrid position/force control) have been published so far [1][2][3]. However, the characteristics of Coulomb friction between the end effecter and the constraint surface have not been investigated enough so far even though there are the papers on characteristics of static friction. In general, the constrain force occurs in the normal direction to the surface and the end effecter of the robot manipulator moves in the orthogonal directions to the normal direction. In the previous researches, they focused on orthogonality or decoupled dynamics between position directions and force directions. In other words, Coulomb friction was ignored because the Coulomb friction force determined by the contact force exists in position directions. However, since there is Coulomb friction between the robot and the object in practice, the characteristics of Coulomb friction should be considered.

We investigated the characteristics of Coulomb friction and pointed out that there is another singular point with Coulomb friction in the reference [4]. In the reference, the reason why the singular points occur was explained and the qualitative characteristics on the singular points were revealed through numerical examples. Moreover, we proposed a new control method to overcome the problems of Coulomb friction force. However, the stability of the robot motion controlled by the proposed method has not been proven in the reference [4].

In this paper, we prove the stability of the robot motion. At first, the robot dynamics with Coulomb friction is described and the coupling problem of the constraint force is explained through some simulation results. Next, a Lyapunov function candidate is proposed. In the stability analysis, it is mathematically proven that the trajectory tracking error between the actual motion and the desired motion can be effectively reduced as the position and velocity feedback gains increase. Even though the motion control direction is coupled with the constraint force direction, the trajectory tracking with high accuracy is guaranteed for non-linear dynamics including Coulomb friction. Finally the trajectory tracking performance of the proposed feedback control method is demonstrated by some simulation results.

2 MODELING OF A ROBOT AND COULOMB FRICTION

In order to simplify the explanation, we consider a three DOF robot as shown in Fig.1. The end effector is geometrically constrained by the surface

$$\phi(\boldsymbol{p}) = 0 \tag{1}$$

where $\boldsymbol{p}(3\times 1)$ is a position of the end effecter in the Cartesian coordinate system Σ_b. Hereafter we assume that the constraint surface is sufficiently smooth. The constraint force $f(1\times 1)$ is generated to the normal direction of the constraint surface. Generally, when a robot contacts a surface, the constraint force becomes positive, where the inside direction is set positive as seen in Fig.1. However, in order to simply explain the following parts, here we consider that the constraint force f takes both positive and negative.

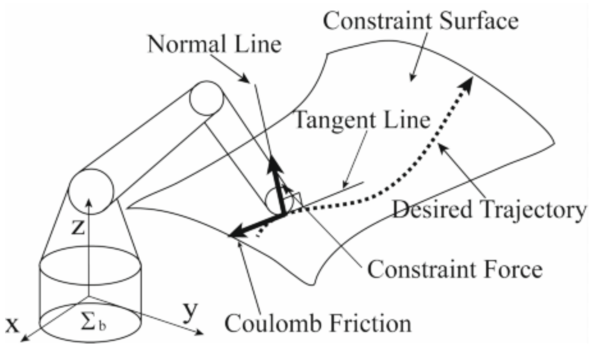


Figure 1: A robot arm and A Constraint Surface.

we may set the sufficiently large feedforward input. The magnitude of the Coulomb friction force generated to the motion direction is determined by the friction coefficient $\mu(1\times 1)$ and the constraint force f, and the direction is set to the reverse of the motion direction.

In general, the robot moves along a desired trajectory on the surface and may stop at some points on the surface.

Therefore, static friction must be considered. However, since the main purpose of this paper is to investigate the coupling problems of constraint force f when the robot moves, we will not treat static friction in the followings. Then the model of the Coulomb friction force f_c is given by

$$f_c = a(\dot{\boldsymbol{p}})\mu f \tag{2}$$

where, $a(\dot{\boldsymbol{p}})(3\times 1)$ is shown by

$$a(\dot{\boldsymbol{p}}) = \begin{cases} [\dot{x}\,\dot{y}\,\dot{z}]^T / \|\dot{\boldsymbol{p}}\| & \|\dot{\boldsymbol{p}}\| \geq \varepsilon \\ [\dot{x}\,\dot{y}\,\dot{z}]^T / \varepsilon & \|\dot{\boldsymbol{p}}\| < \varepsilon \end{cases} \tag{3}$$

Here, each of $\dot{x}$, $\dot{y}$ and $\dot{z}$ is an element of the velocity vector $\dot{\boldsymbol{p}}$.

The dynamics of the robot is shown by

$$\boldsymbol{R}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \left(\frac{1}{2}\dot{\boldsymbol{R}}(\boldsymbol{q}) + \boldsymbol{S}(\boldsymbol{q},\dot{\boldsymbol{q}}) + \boldsymbol{B}\right)\dot{\boldsymbol{q}} + \boldsymbol{g}(\boldsymbol{q}) = \boldsymbol{u}(t) + \boldsymbol{d}f - \boldsymbol{J}^T\boldsymbol{a}(\dot{\boldsymbol{p}})\mu f\,. \tag{4}$$

Where $\boldsymbol{q}(3\times 1)$: a joint angle vector, $\boldsymbol{R}(\boldsymbol{q})(3\times 3)$: an inertia matrix, $(1/2\dot{\boldsymbol{R}}(\boldsymbol{q}) + \boldsymbol{S}(\boldsymbol{q},\dot{\boldsymbol{q}}))\dot{\boldsymbol{q}}(3\times 1)$: terms of centrifugal force and Coriolis force, $\boldsymbol{B}(3\times 3)$: a diagonal matrix whose elements are coefficients of viscosity of the joint, $\boldsymbol{g}(\boldsymbol{q})(3\times 1)$: a gravity term, $\boldsymbol{u}(3\times 1)$: a control input, $\boldsymbol{J}(3\times 3)$: a Jacobian matrix, and $\boldsymbol{d}(3\times 1)$: a normal direction unit force vector in the joint angle coordinates given by

$$\boldsymbol{d} = \boldsymbol{J}^T\left(\frac{\partial\phi}{\partial\boldsymbol{p}}\right)^T \Bigg/ \left\|\frac{\partial\phi}{\partial\boldsymbol{p}}\right\| \tag{5}$$

where $\partial\phi/\partial\boldsymbol{p}(1\times 3)$: a normal vector of the surface.

3 CONVENTIONAL CONTROL METHODS AND COUPLING PROBLEMS

3.1 A conventional control method

Here, the input $\boldsymbol{u}(t)$ is chosen as

$$\boldsymbol{u}(t) = -\boldsymbol{K}_p\Delta\boldsymbol{q} - \boldsymbol{K}_d\Delta\dot{\boldsymbol{q}} - \boldsymbol{d}f_d(t) + \boldsymbol{g}(\boldsymbol{q}_d(t)) \tag{6}$$

where $\Delta\boldsymbol{q}(t)(3\times 1)$: a joint angle error ($\Delta q(t) = \boldsymbol{q}(t) - \boldsymbol{q}_d(t)$), $\boldsymbol{q}_d(t)(3\times 1)$: a desired trajectory, $f_d(t)(1\times 1)$: a feedforward input. $\boldsymbol{K}_p = \mathbf{diag.}\{k_{p1}, k_{p2}, k_{p3}\}(3\times 3)$: a position feedback gain matrix, and $\boldsymbol{K}_d = \mathbf{diag.}\{k_{d1}, k_{d2}, k_{d3}\}(3\times 3)$: a velocity feedback gain matrix.

Here, by substituting Eq. (6) to Eq. (4), we have

$$\boldsymbol{R}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \left(\frac{1}{2}\dot{\boldsymbol{R}}(\boldsymbol{q}) + \boldsymbol{S}(\boldsymbol{q},\dot{\boldsymbol{q}}) + \boldsymbol{B}\right)\dot{\boldsymbol{q}} = -\boldsymbol{K}_p\Delta\boldsymbol{q} - \boldsymbol{K}_d\Delta\dot{\boldsymbol{q}} - \boldsymbol{d}f_d + \boldsymbol{g}(\boldsymbol{q}_d) - \boldsymbol{g}(\boldsymbol{q}) + \left\{\boldsymbol{d} - \boldsymbol{J}^T\boldsymbol{a}(\dot{\boldsymbol{p}})\mu\right\}f \tag{7}$$

Next, we multiply Eq. (7) by $\boldsymbol{d}^T\boldsymbol{R}(\boldsymbol{q})^{-1}$ from the left side of Eq. (7). Using of the relation of $\boldsymbol{d}^T\ddot{\boldsymbol{q}} = -\dot{\boldsymbol{d}}^T\dot{\boldsymbol{q}}$, we obtain

$$\eta f = -\dot{\boldsymbol{d}}^T\dot{\boldsymbol{q}} + \boldsymbol{d}^T\boldsymbol{R}(\boldsymbol{q})^{-1}\left\{\left(\frac{1}{2}\dot{\boldsymbol{R}}(\boldsymbol{q}) + \boldsymbol{S}(\boldsymbol{q},\dot{\boldsymbol{q}}) + \boldsymbol{B}\right)\dot{\boldsymbol{q}} - \boldsymbol{g}(\boldsymbol{q}_d) + \boldsymbol{g}(\boldsymbol{q}) + \boldsymbol{K}_p\Delta\boldsymbol{q} + \boldsymbol{K}_d\Delta\boldsymbol{q} + \boldsymbol{d}f_d\right\} \tag{8}$$

where, η is given by

$$\eta = \boldsymbol{d}^T\boldsymbol{R}(\boldsymbol{q})^{-1}\left\{\boldsymbol{d} - \boldsymbol{J}^T\boldsymbol{a}(\dot{\boldsymbol{p}})\mu\right\}. \tag{9}$$

3.2 Coupling problems on Coulomb Friction

We treat the case that the robot can trace the desired trajectory as seen in Fig. 1. In practice, the robot cannot perfectly realize the desired trajectory. In this case, unfortunately it is possible to generate the circulation of the error increase. Fig. 2. shows the mechanism of the circulation. Namely, since the constraint force appears in the motion direction, it works to increase the position and velocity errors. The increase of the errors may enlarge the constraint force.

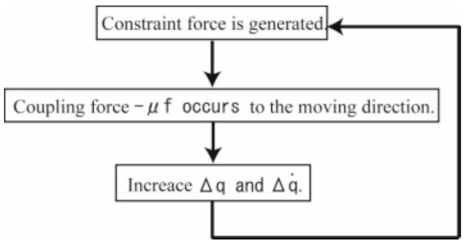


Figure 2: Cycle of Coupling Constraint Force.

3.3 A two links model

In order to investigate the circulation problem on Coulomb friction, the simulation model with two links is utilized as seen Fig.3. Here, to simplify the explanation, the gravitational term is not included. The control input $\boldsymbol{u}(t)$ is given by

$$\boldsymbol{u}(t) = -\boldsymbol{K}_p\Delta\boldsymbol{q} - \boldsymbol{K}_d\Delta\dot{\boldsymbol{q}} - \boldsymbol{d}f_d \tag{10}$$

where $f_d(1\times 1)$: a feedforward input, $\boldsymbol{K}_p(2\times 2)$: a position feedback gain matrix, $\boldsymbol{K}_d(2\times 2)$: a velocity feedback gain matrix.

Here, we consider that the tip of the robot is constrained by a line as shown Fig. 3. Each link has same length 0.5[m]. The center of mass of each link is set $l_{g1} = l_{g2} = 0.25$ [m]. The mass of the link 1 is fixed with 15.0 [kg], and the mass of the link 2 is fixed with 10.0 [kg]. The cross section of the link is assumed to be a square and the density is set constant. The moment of inertia is fixed with the mass of a link. The friction coefficient μ of the constraint surface is set 0.3.

The robot moves along the desired trajectory from y_0 to y_1 in Cartesian coordinates shown by

$$\begin{cases} x(t) = x_0 \\ y(t) = 2(y_0 - y_1)\left(\dfrac{t}{T}\right)^3 - 3(y_0 - y_1)\left(\dfrac{t}{T}\right)^2 + y_0 \end{cases} \tag{11}$$

where, $x_0 = 0.5$ [m], $y_0 = 0.95\times\sqrt{3.0}/2.0$ [m], $y_1 = -0.95\times\sqrt{3.0}/2.0$ [m], and $T = 5.0$ [sec] : a terminal time.

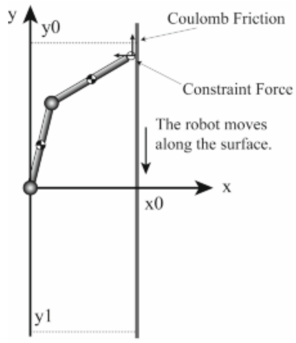


Figure 3: A robot Arm with Two Links and A Constraint Line.

3.4 Simulation results

We calculated the constraint force f, the angle error $\Delta \boldsymbol{q}_1$ and the angular velocity error $\Delta \dot{\boldsymbol{q}}_1$ when the robot is controlled by the feedback control input shown by an Eq. (10). The desired trajectory is given by Eq. (11). Here, we set the feedforward input f_d as follows:

$$f_d = \begin{cases} 50.0\sin\left(\pi t - \dfrac{\pi}{2}\right) + 50.0 & (t < 2.0) \\ 0.0 & (t \geq 2.0) \end{cases} \quad (12)$$

In this simulation, we set the position feedback gain matrix $\boldsymbol{K}_p = \mathrm{diag.}\{800 \quad 800\}$, and the velocity feedback gain matrix $\boldsymbol{K}_d = \mathrm{diag.}\{800 \quad 800\}$. The angle error $\Delta \boldsymbol{q}_1$, the angular velocity error $\Delta \dot{\boldsymbol{q}}_1$, and the constraint force f are shown in Fig. 4, in Fig. 5, in Fig. 6, respectively.

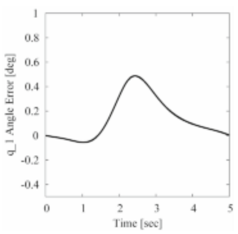


Figure 4: Error $\Delta \boldsymbol{q}_1$.

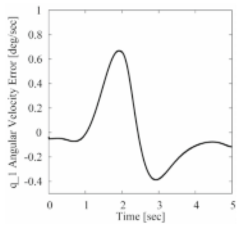


Figure 5: Error $\Delta \dot{\boldsymbol{q}}_1$.

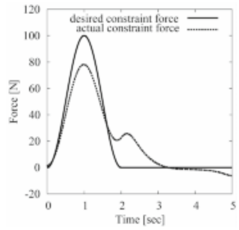


Figure 6: Constraint Force.

There is the peak of the actual force at 1.0 [sec] because the feedforward input f_d is added. As the result, the angle and the angular errors increase around 2.0[sec]. Next, the constraint force has a considerably large peak after 2.0[sec] because of increase of the angle and the angular velocity errors. It means the circulation problem explained in subsection 3.2.

4 A PROPOSED CONTROL METHOD

4.1 A decoupling control method on Coulomb friction

The cause of the feedback errors/constraint force coupling is that the feedback errors appear in the right hand side of Eq. (8). Thus, we eliminate those errors from Eq. (8). For this purpose, we propose a new control method represented by

$$u = -k_p \boldsymbol{R}(\boldsymbol{q})\boldsymbol{\Psi}\Delta\boldsymbol{q} - k_d \boldsymbol{R}(\boldsymbol{q})\Psi\Delta\dot{\boldsymbol{q}} - \boldsymbol{d}f_d \quad (13)$$

where a projection matrix $\boldsymbol{\Psi}(3 \times 3)$ is given by

$$\boldsymbol{\Psi} = \boldsymbol{I} - \frac{\boldsymbol{d}\boldsymbol{d}^T}{\|\boldsymbol{d}\|^2} . \quad (14)$$

4.2 Conditions

We assume that the Jacobian matrix is nonsingular and

$$\eta \geq \varepsilon_0 \quad (15)$$

where ε_0 : a positive constant. In the reference [4], we defined Coulomb friction singular points as $\eta = 0$. The condition given in Eq.(15) means the nonsingularity of Coulomb friction singular points. This condition can be satisfied if there is no link with extremely small (or large) inertia in comparison with other links of the robot. The details are shown in the reference [4].

Next, we assume

$$\|\Delta\boldsymbol{q}\| < \gamma \quad (16)$$

where γ is a positive constant. If we set γ sufficiently small, then we obtain the following equation

$$\|\boldsymbol{\Psi}_f \Delta\boldsymbol{q}\| \leq \varepsilon_a \|\Delta\boldsymbol{q}\| \qquad (0 \leq \varepsilon_a < 1) \quad (17)$$

where a projection matrix $\boldsymbol{\Psi}_f$ is given by

$$\boldsymbol{\Psi}_f = \frac{\boldsymbol{d}^T\boldsymbol{d}}{\|\boldsymbol{d}\|^2} . \quad (18)$$

Eq.(17) holds if γ is sufficiently small because the component of the angle error in the force control direction becomes small as the angle error converge to zero. Here, we utilized the smoothness of the constraint surface to obtain Eq.(17).

From the Eq.(17) $\|\Delta\boldsymbol{q}\|$ is satisfied the following inequality

$$\|\Delta\boldsymbol{q}\| \leq \|\boldsymbol{\Psi}\Delta\boldsymbol{q}\| + \|\boldsymbol{\Psi}_f\Delta\boldsymbol{q}\| \leq \|\boldsymbol{\Psi}\Delta\boldsymbol{q}\| + \varepsilon_a\|\Delta\boldsymbol{q}\| . \quad (19)$$

Therefore, we obtain

$$\|\boldsymbol{\Psi}\Delta\boldsymbol{q}\| \geq (1 - \varepsilon_a)\|\Delta\boldsymbol{q}\| . \quad (20)$$

Moreover, we assume

$$\left\|\boldsymbol{\Psi}_f \Delta\dot{\boldsymbol{q}}\right\| \le \varepsilon_b(\dot{\boldsymbol{q}}_d)\left\|\Delta\boldsymbol{q}\right\| \qquad (0 \le \varepsilon_b(\dot{\boldsymbol{q}}_d) < 1)\ , \tag{21}$$

To understand the meaning of Eq(21), the following equations are important. At first, using the relation of $\boldsymbol{I} = \boldsymbol{\Psi} + \boldsymbol{\Psi}_f$, we obtain

$$\Delta\dot{\boldsymbol{q}} = \boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}} + \boldsymbol{\Psi}_f \Delta\dot{\boldsymbol{q}}\ . \tag{22}$$

On the other hand, we calculate the error of the joint angluar velocity $\Delta\dot{\boldsymbol{q}}$ as

$$\Delta\dot{\boldsymbol{q}} = \boldsymbol{\Psi}\Delta\boldsymbol{q} + (\boldsymbol{\Psi} - \boldsymbol{\Psi}_d)\dot{\boldsymbol{q}}_d\ . \tag{23}$$

Then from Eq.(22) and Eq.(23) we have

$$\boldsymbol{\Psi}_f \Delta\dot{\boldsymbol{q}} = (\boldsymbol{\Psi} - \boldsymbol{\Psi}_d)\dot{\boldsymbol{q}}_d\ . \tag{24}$$

From Eq.(24) we obtain Eq.(21).

4.3 Error dynamics

Using the proposed decoupling control method, Dynamics of a robot manipulator can be written by

$$\boldsymbol{R}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \left(\frac{1}{2}\dot{\boldsymbol{R}}(\boldsymbol{q}) + \boldsymbol{S}(\boldsymbol{q},\dot{\boldsymbol{q}}) + \boldsymbol{B}\right)\dot{\boldsymbol{q}} + \boldsymbol{g}(\boldsymbol{q})$$
$$= -\boldsymbol{K}_p\boldsymbol{R}\boldsymbol{\Psi}\Delta\boldsymbol{q} - \boldsymbol{K}_d\boldsymbol{R}\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}} - \left\{\boldsymbol{d} - \boldsymbol{J}^T\boldsymbol{a}(\dot{\boldsymbol{p}})\mu\right\}f - \boldsymbol{d}f_d + \boldsymbol{g}(\boldsymbol{q}_d)\ . \tag{25}$$

Since

$$\dot{\boldsymbol{q}} = \boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}} + \boldsymbol{\Psi}\dot{\boldsymbol{q}}_d\ , \tag{26}$$

we obtain

$$\ddot{\boldsymbol{q}} = \dot{\boldsymbol{\Psi}}\Delta\dot{\boldsymbol{q}} + \boldsymbol{\Psi}\Delta\ddot{\boldsymbol{q}} + \dot{\boldsymbol{\Psi}}\dot{\boldsymbol{q}}_d + \boldsymbol{\Psi}\ddot{\boldsymbol{q}}_d\ . \tag{27}$$

From Eqs.(26) and (27), the error dynamics of the robot manipulator is rewritten into

$$\boldsymbol{R}\left\{\dot{\boldsymbol{\Psi}}\Delta\dot{\boldsymbol{q}} + \boldsymbol{\Psi}\Delta\ddot{\boldsymbol{q}} + \dot{\boldsymbol{\Psi}}\dot{\boldsymbol{q}}_d + \boldsymbol{\Psi}\ddot{\boldsymbol{q}}_d\right\}$$
$$+ \left\{\frac{1}{2}\dot{\boldsymbol{R}} + \boldsymbol{S}(\boldsymbol{q},\dot{\boldsymbol{q}}) + \boldsymbol{B}\right\}\left\{\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}} + \boldsymbol{\Psi}\dot{\boldsymbol{q}}_d\right\}$$
$$= -k_p\boldsymbol{R}\boldsymbol{\Psi}\Delta\boldsymbol{q} - k_d\boldsymbol{R}\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}} - \left\{\boldsymbol{d} - \boldsymbol{J}^T\boldsymbol{a}(\dot{\boldsymbol{p}})\mu\right\}f$$
$$-\boldsymbol{d}f_d - \left(\boldsymbol{g}(\boldsymbol{q}) - \boldsymbol{g}(\boldsymbol{q}_d)\right). \tag{28}$$

Next we multiply Eq.(28) by $\boldsymbol{d}^T\boldsymbol{R}^{-1}$ from the left side of an Eq.(28). Using of the relation of $\boldsymbol{d}^T\boldsymbol{\Psi} = 0$, we obtain the following equation which dose not involve feedback gain k_p , k_d .

$$\boldsymbol{d}^T\left\{\dot{\boldsymbol{\Psi}}\Delta\dot{\boldsymbol{q}} + \dot{\boldsymbol{\Psi}}\dot{\boldsymbol{q}}_d\right\} + \boldsymbol{d}^T\boldsymbol{R}^{-1}\left\{\frac{1}{2}\dot{\boldsymbol{R}} + \boldsymbol{S}(\boldsymbol{q},\dot{\boldsymbol{q}}) + \boldsymbol{B}\right\}\left\{\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}} + \boldsymbol{\Psi}\dot{\boldsymbol{q}}_d\right\}$$
$$= -\boldsymbol{d}^T\boldsymbol{R}^{-1}\left\{\boldsymbol{d} - \boldsymbol{J}^T\boldsymbol{a}(\dot{\boldsymbol{p}})\mu\right\}f - \boldsymbol{d}^T\boldsymbol{R}^{-1}\boldsymbol{d}f_d$$
$$- \boldsymbol{d}^T\boldsymbol{R}^{-1}\left(\boldsymbol{g}(\boldsymbol{q}) - \boldsymbol{g}(\boldsymbol{q}_d)\right) \tag{29}$$

Here, from Eq.(29), the constraint force f is shown by

$$f = \frac{1}{\eta}\Big[-\boldsymbol{d}^T\left\{\dot{\boldsymbol{\Psi}}\Delta\dot{\boldsymbol{q}} + \dot{\boldsymbol{\Psi}}\dot{\boldsymbol{q}}_d\right\}$$
$$-\boldsymbol{d}^T\boldsymbol{R}^{-1}\left\{\frac{1}{2}\dot{\boldsymbol{R}} + \boldsymbol{S}(\boldsymbol{q},\dot{\boldsymbol{q}}) + \boldsymbol{B}\right\}\left\{\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}} + \boldsymbol{\Psi}\dot{\boldsymbol{q}}_d\right\}$$
$$-\boldsymbol{d}^T\boldsymbol{R}^{-1}\left\{\boldsymbol{d}f_d + \left(\boldsymbol{g}(\boldsymbol{q}) - \boldsymbol{g}(\boldsymbol{q}_d)\right)\right\}\Big]. \tag{30}$$

Finally by substituting Eq.(30) into Eq.(28) the dynamics without the constraint force f is represented by

$$\boldsymbol{R}\left\{\dot{\boldsymbol{\Psi}}\Delta\dot{\boldsymbol{q}} + \boldsymbol{\Psi}\Delta\ddot{\boldsymbol{q}}\right\} = -k_p\boldsymbol{R}\boldsymbol{\Psi}\Delta\boldsymbol{q} - k_d\boldsymbol{R}\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}$$
$$-W\left[\left\{\frac{1}{2}\dot{\boldsymbol{R}} + \boldsymbol{S}(\boldsymbol{q},\dot{\boldsymbol{q}}) + \boldsymbol{B}\right\}\left\{\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}} + \boldsymbol{\Psi}\dot{\boldsymbol{q}}_d\right\}\right.$$
$$\left.+\boldsymbol{d}f_d + \boldsymbol{g}(\boldsymbol{q}) - \boldsymbol{g}(\boldsymbol{q}_d)\right] - \boldsymbol{R}\left(\dot{\boldsymbol{\Psi}}\dot{\boldsymbol{q}}_d + \boldsymbol{\Psi}\ddot{\boldsymbol{q}}_d\right)$$
$$+\left\{\boldsymbol{d} - \boldsymbol{J}^T\boldsymbol{a}(\dot{\boldsymbol{p}})\mu\right\}\frac{1}{\eta}\boldsymbol{d}^T\left(\dot{\boldsymbol{\Psi}}\Delta\dot{\boldsymbol{q}} + \dot{\boldsymbol{\Psi}}\dot{\boldsymbol{q}}_d\right) \tag{31}$$

where

$$\boldsymbol{W} = \boldsymbol{I} - \left\{\boldsymbol{d} - \boldsymbol{J}^T\boldsymbol{a}(\dot{\boldsymbol{p}})\mu\right\}\frac{\boldsymbol{d}^T\boldsymbol{R}^{-1}}{\eta}\ . \tag{32}$$

4.4 Stability analysis

Here we define a Lyapunov function candidate as

$$V(\Delta\boldsymbol{x}) = \frac{1}{2}\Delta\boldsymbol{x}^T\boldsymbol{N}\Delta\boldsymbol{x} \tag{33}$$

where $\boldsymbol{\Delta x} = [\boldsymbol{\Delta q}^T, \Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}]^T$ and the matrix $\boldsymbol{N}$ is defined as

$$\boldsymbol{N} = \begin{bmatrix}(k_p + \alpha k_d)\boldsymbol{R}_0 & \alpha\boldsymbol{R}\\ \alpha\boldsymbol{R} & \boldsymbol{R}\end{bmatrix}. \tag{34}$$

In Eq.(34), the matrix $\boldsymbol{R}_0$ is a positive definite matrix and should be set an estimated inertia matrix $\boldsymbol{R}(\boldsymbol{q})$. All elements of $\boldsymbol{R}_0$ are constants. We set α an appropriate positive constant to satisfy

$$V \ge 0 \tag{35}$$

for any $\boldsymbol{\Delta x}$.

Next we calculate the derivative of V as

$$\dot{V} = (\Delta\ddot{\boldsymbol{q}}^T\boldsymbol{\Psi} + \Delta\dot{\boldsymbol{q}}^T\dot{\boldsymbol{\Psi}})\boldsymbol{R}\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}} + \alpha(\Delta\ddot{\boldsymbol{q}}^T\boldsymbol{\Psi} + \Delta\dot{\boldsymbol{q}}^T\dot{\boldsymbol{\Psi}})\boldsymbol{R}\boldsymbol{\Psi}\Delta\boldsymbol{q}$$
$$+ (k_p + \alpha k_d)\Delta\dot{\boldsymbol{q}}^T\boldsymbol{R}_0\Delta\boldsymbol{q} + \frac{1}{2}\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}\dot{\boldsymbol{R}}\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}$$
$$+ \alpha\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}\dot{\boldsymbol{R}}\boldsymbol{\Psi}\Delta\boldsymbol{q} + \alpha\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}\boldsymbol{R}\dot{\boldsymbol{\Psi}}\Delta\boldsymbol{q}$$
$$+ \alpha\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}\boldsymbol{R}\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\ . \tag{36}$$

From the assumption given by Eq.(16), we obtain the following equations

$$\left\|\Delta\dot{\boldsymbol{q}}\right\| \le \left\|\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\right\| + c_0\ , \qquad \left\|\dot{\boldsymbol{R}}\right\| \le c_1\left\|\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\right\| + c_2\ ,$$
$$\left\|\dot{\boldsymbol{\Psi}}\right\| \le c_3\left\|\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\right\| + c_4\ , \quad \left\|\boldsymbol{S}(\boldsymbol{q},\dot{\boldsymbol{q}})\right\| \le c_5\left\|\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\right\| + c_6 \tag{37}$$

where $c_0 \sim c_6$ are appropriate positive constants.

From Eq.(37) and Eq.(31), the first term of the right hand side of Eq.(36) can be written into

$$(\Delta\ddot{\boldsymbol{q}}^T\boldsymbol{\Psi} + \Delta\dot{\boldsymbol{q}}^T\dot{\boldsymbol{\Psi}})\boldsymbol{R}\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}} \le$$
$$- k_p\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}\boldsymbol{R}\boldsymbol{\Psi}\Delta\boldsymbol{q} - k_d\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}\boldsymbol{R}\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}$$

$$+c_7\|\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\|^3+c_8\|\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\|^2+c_9\|\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\|, \tag{38}$$

and the second term of the right hand side of Eq.(36) can be written into

$$\begin{aligned}\alpha(\Delta\ddot{\boldsymbol{q}}^T\boldsymbol{\Psi}+\Delta\dot{\boldsymbol{q}}^T\dot{\boldsymbol{\Psi}})\boldsymbol{R}\boldsymbol{\Psi}\Delta\boldsymbol{q}\le\\ &-\alpha k_p\Delta\boldsymbol{q}^T\boldsymbol{\Psi}\boldsymbol{R}\boldsymbol{\Psi}\Delta\boldsymbol{q}-\alpha k_d\Delta\boldsymbol{q}^T\boldsymbol{\Psi}\boldsymbol{R}\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\\ &+\alpha(c_{10}\|\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\|^2+c_{11}\|\Delta\boldsymbol{q}\|^2+c_{12}\|\Delta\boldsymbol{q}\|)\end{aligned} \tag{39}$$

where $c_7\sim c_{12}$ are appropriate positive constants.

Here, to drive Eq.(38) and Eq.(39) we used the following relation

$$\left\|\boldsymbol{d}-\boldsymbol{J}^T\boldsymbol{a}(\dot{\boldsymbol{p}})\mu\right\|\le c_{13} \tag{40}$$

where c_{13} is an appropriate positive constant.

Therefore, from the first term to the third term of the right hand side of Eq.(36), we obtain the following equation

$$\begin{aligned}&(\Delta\ddot{\boldsymbol{q}}^T\boldsymbol{\Psi}+\Delta\dot{\boldsymbol{q}}^T\dot{\boldsymbol{\Psi}})\boldsymbol{R}\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}+\alpha(\Delta\ddot{\boldsymbol{q}}^T\boldsymbol{\Psi}+\Delta\dot{\boldsymbol{q}}^T\dot{\boldsymbol{\Psi}})\boldsymbol{R}\boldsymbol{\Psi}\Delta\boldsymbol{q}\\ &\quad+(k_p+\alpha k_d)\Delta\dot{\boldsymbol{q}}^T\boldsymbol{R}_0\Delta\boldsymbol{q}\\ &\quad\le(k_p+\alpha k_d)\Delta\dot{\boldsymbol{q}}^T\boldsymbol{R}_0\Delta\boldsymbol{q}\\ &\quad-\alpha k_p\Delta\boldsymbol{q}^T\boldsymbol{\Psi}\boldsymbol{R}\boldsymbol{\Psi}\Delta\boldsymbol{q}-k_d\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}\boldsymbol{R}\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\\ &\quad-k_p\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}\boldsymbol{R}\boldsymbol{\Psi}\Delta\boldsymbol{q}-\alpha k_d\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}\boldsymbol{R}\boldsymbol{\Psi}\Delta\boldsymbol{q}\\ &\quad+(c_7+\alpha c_{10})\|\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\|^3+c_8\|\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\|^2+c_9\|\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\|\\ &\quad+\alpha c_{11}\|\Delta q\|^2+\alpha c_{12}\|\Delta q\|.\end{aligned} \tag{41}$$

Here, using of the relation $\boldsymbol{I}=\boldsymbol{\Psi}_f+\boldsymbol{\Psi}$ the first term of the right hand side of Eq.(41) can be written into

$$\begin{aligned}(k_p+\alpha k_d)\Delta\dot{\boldsymbol{q}}^T\boldsymbol{R}_0\Delta\boldsymbol{q}&=(k_p+\alpha k_d)\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}\boldsymbol{R}_0\boldsymbol{\Psi}\Delta\boldsymbol{q}\\ &+(k_p+\alpha k_d)\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}_f\boldsymbol{R}_0\Delta\boldsymbol{q}\\ &+(k_p+\alpha k_d)\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}\boldsymbol{R}_0\boldsymbol{\Psi}_f\Delta\boldsymbol{q}.\end{aligned} \tag{42}$$

Moreover, it is important to note that the following relation holds if $\boldsymbol{R}_0$ is close to $\boldsymbol{R}(q)$.

$$\|\boldsymbol{R}_0-\boldsymbol{R}\|\le\varepsilon_c\lambda_{\max}[\boldsymbol{R}_0]\qquad(0<\varepsilon_c<1) \tag{43}$$

Using Eqs.(17), (21) and (42), the part of the right hand side of Eq.(41) from the first term to the fifth term can be shown

$$\begin{aligned}&-\alpha k_p\Delta\boldsymbol{q}^T\boldsymbol{\Psi}\boldsymbol{R}\boldsymbol{\Psi}\Delta\boldsymbol{q}-k_d\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}\boldsymbol{R}\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\\ &\quad-(k_p+\alpha k_d)\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}\boldsymbol{R}\boldsymbol{\Psi}\Delta\boldsymbol{q}+(k_p+\alpha k_d)\Delta\dot{\boldsymbol{q}}^T\boldsymbol{R}_0\Delta q\\ &\le(k_p+\alpha k_d)\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}(\boldsymbol{R}_0-\boldsymbol{R})\boldsymbol{\Psi}\Delta\boldsymbol{q}\\ &\quad+(k_p+\alpha k_d)\varepsilon_b\lambda_{\max}[\boldsymbol{R}_0]\|\Delta\boldsymbol{q}\|^2\\ &\quad+(k_p+\alpha k_d)\varepsilon_a\lambda_{\max}[\boldsymbol{R}_0]\|\Delta\boldsymbol{q}\|\|\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\|\\ &\quad-\alpha k_p\lambda_{\min}[\boldsymbol{R}]\|\boldsymbol{\Psi}\Delta q\|^2-k_d\lambda_{\min}[\boldsymbol{R}]\|\boldsymbol{\Psi}\Delta\dot{q}\|^2\\ &\le(k_p+\alpha k_d)\lambda_{\max}[\boldsymbol{R}_0]\left\{\varepsilon_b\|\Delta\boldsymbol{q}\|^2+(\varepsilon_a+\varepsilon_c)\|\Delta\boldsymbol{q}\|\|\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\|\right\}\\ &\quad-\alpha k_p\lambda_{\min}[\boldsymbol{R}]\|\boldsymbol{\Psi}\Delta q\|^2-k_d\lambda_{\min}[\boldsymbol{R}]\|\boldsymbol{\Psi}\Delta\dot{q}\|^2\\ &\le-\left\{\alpha k_p(1-\varepsilon_a)^2\lambda_{\min}[\boldsymbol{R}]\right.\\ &\quad\left.-(k_p+\alpha k_d)\left(\varepsilon_b+\frac{\varepsilon_a+\varepsilon_c}{2}\right)\lambda_{\max}[\boldsymbol{R}_0]\right\}\|\Delta\boldsymbol{q}\|^2\\ &\quad-\left\{k_d\lambda_{\min}[\boldsymbol{R}]-(k_p+\alpha k_d)\left(\frac{\varepsilon_a+\varepsilon_c}{2}\right)\lambda_{\max}[\boldsymbol{R}_0]\right\}\|\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\|^2.\end{aligned} \tag{44}$$

The part of the right hand side of Eq.(36) from the fourth term to the seventh term can be shown

$$\begin{aligned}&\frac{1}{2}\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}\dot{\boldsymbol{R}}\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}+\alpha\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}\dot{\boldsymbol{R}}\boldsymbol{\Psi}\Delta\boldsymbol{q}\\ &\quad+\alpha\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}\boldsymbol{R}\dot{\boldsymbol{\Psi}}\Delta\boldsymbol{q}+\alpha\Delta\dot{\boldsymbol{q}}^T\boldsymbol{\Psi}\boldsymbol{R}\boldsymbol{\Psi}\Delta\dot{\boldsymbol{q}}\\ &\quad\le c_{13}\|\Psi\Delta\dot{\boldsymbol{q}}\|^3+c_{14}\|\Psi\Delta\dot{\boldsymbol{q}}\|^2\\ &\quad+c_{15}\|\Delta q\|^3+c_{16}\|\Delta\boldsymbol{q}\|^2\end{aligned} \tag{45}$$

where, $c_{13}\sim c_{16}$ are appropriate positive constants.

Consequently, from Eq.(36),(41),(42),(44) and (45) we obtain

$$\dot{V}\le\zeta_3\|\Delta\boldsymbol{x}\|^3-\zeta_2\|\Delta\boldsymbol{x}\|^2+\zeta_1\|\Delta\boldsymbol{x}\| \tag{46}$$

where ζ_1, ζ_2 and ζ_3 are positive constants. The reason why ζ_2 can be set positive will be explained in the following manner.

Here, we set $k_p=\rho$, $k_d=a\rho$ where ρ is a positive constant and the character a means the ration between gains.

The coefficient ζ_2 is determined by

$$\zeta_2=\mathbf{min}[a_1,a_2]. \tag{47}$$

The values of a_1 and a_2 are represented by

$$\begin{aligned}a_1&=\rho\left[\alpha(1-\varepsilon_a)^2\lambda_{\min}[\boldsymbol{R}]\right.\\ &\quad\left.-(1+\alpha a)\left(\varepsilon_b+\frac{\varepsilon_a+\varepsilon_c}{2}\right)\lambda_{\max}[\boldsymbol{R}_0]\right]-c_{17}\\ a_2&=\rho\left[\alpha\lambda_{\min}[\boldsymbol{R}]-(1+\alpha a)\left(\frac{\varepsilon_a+\varepsilon_c}{2}\right)\lambda_{\max}[\boldsymbol{R}_0]\right]-c_{18}\end{aligned} \tag{48}$$

where c_{17}, c_{18} are positive constants.

In Eq.(48),if ε_a, ε_b, and ε_c are sufficiently small, we obtain the following inequalities.

$$\alpha(1-\varepsilon_a)^2\lambda_{\min}[\boldsymbol{R}]-(1+\alpha a)\left(\varepsilon_b+\frac{\varepsilon_a+\varepsilon_c}{2}\right)\lambda_{\max}[\boldsymbol{R}_0]>0$$

(49)

$$\alpha\lambda_{\min}[\boldsymbol{R}]-(1+\alpha a)\left(\frac{\varepsilon_a+\varepsilon_c}{2}\right)\lambda_{\max}[\boldsymbol{R}_0]>0 \tag{50}$$

Therefore if we set ρ a sufficiently large value, then ζ_2 can become a positive constant with a sufficiently large value .

The right hand side of Eq.(46) becomes a third order function with $\|\Delta\boldsymbol{x}\|$ and the roots of the function can be represented by zero, r_1 and r_2 as follows:

$$r_1=\frac{\zeta_2-\sqrt{\zeta_2^2-4\zeta_3\zeta_1}}{2\zeta_3},\quad r_2=\frac{\zeta_2+\sqrt{\zeta_2^2-4\zeta_3\zeta_1}}{2\zeta_3}. \tag{51}$$

As explained, ζ_2 can become large as the feedback gain ρ increases. It means that the r_1 becomes small and r_2 becomes large. Since the time derivative of the Lyapunov function given by Eq.(36) is negative in the range from r_1 to r_2 , the decrease of r_1 and the increase of r_2 mean increase of the stable range of $\|\Delta\boldsymbol{x}\|$. In particular, the decrease of r_1 directly realize the improvement of trajectory tracking performance. More rigorously speaking, we obtain

$$\|\Delta x(t)\|\le\sqrt{\frac{\lambda_{\max}N}{\lambda_{\min}N}}\;r_1 \tag{52}$$

for any time t. The details to derive Eq.(52) is shown in [5] and [6]. Moreover, it is proven that the constraint force represented by Eq.(30) is bounded because the angle $\boldsymbol{q}$ and angular velocity $\dot{\boldsymbol{q}}$ are bounded.

4.5 Simulation results

The simulation results on the proposed control method are shown in Fig. 7, Fig.8, and Fig.9. In the proposed control method, we could increase the feedback gains. In the conventional control method, however, the feedback gains could not be set more than 800 with the same sampling time.

Fig. 7 and Fig.8 shows the position errors and the velocity errors of the joint 1 on the conventional control method and the proposed control method, respectively. As seen in Fig. 7 and Fig. 8, it is confirmed that the position and the velocity errors on the proposed control method become smaller than those errors on the conventional control method. On the other hand in Fig.9, it is observed that the proposed control method could eliminate the second peak force. Consequently, it is expected that the proposed control method is useful for trajectory tracking of robot manipulators with Coulomb friction.

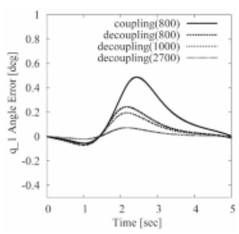


Figure 7: Error of $\Delta\boldsymbol{q}_1$.

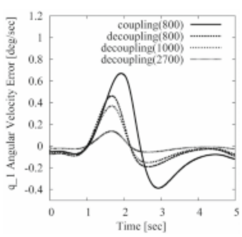


Figure 8: Error of $\Delta\dot{\boldsymbol{q}}_1$.

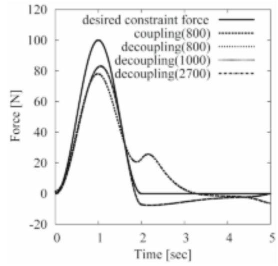


Figure 9: Constraint Force.

5 SUMMARY

In this paper, we pointed out that the circulation problem on Coulomb friction occurs when position and/or velocity feedback loops are utilized. To solve the circulation problem, we have proposed a new control method for trajectory tracking of robot manipulators with coupled dynamics of force and position. We have proved mathematically that the trajectory tracking performance is improved by increasing on the feedback gains. Moreover, the effectiveness of the proposed control method have been demonstrated through some simulation results.

In the simulation, we utilized the exact inertia matrix $\boldsymbol{R}(\boldsymbol{q})$. However, it should be consider that the estimated inertia matrix contains errors. Therefore, it is necessary to investigate the robustness of the proposed control method with the estimated inertia matrix from both of theoretical and practical view points. It is one of the future works with the proposed control method.

[1] McClamorch N.H. and Wang D . , "Feedback Stabilization and Tracking of Constrainted Robots", IEEE Transaction Automatic Control,33,419-426 1988.

[2] Mills J.K. and Goldenberg A.A. , "Force and Position Control of Manipulators during Constrained Motion Tasks", IEEE Transaction on Robotics and Automation, 5, 30-46, 1989.

[3] Suguru Arimoto, "Contorl Theory of Non-linear Mechanical Systems", Oxford Science Publications 1996.

[4] Hiroaki Ichii and Sadao Kawamura, "Coulomb Friction Singular Points of Robot Dynamics with Mechanical Constraints", SICE Annual Conference 2008.

[5] D.M. Dawson, Z. Qu, F.L. Lewis and J.F. Dorsey ,"Robust control for the tracking of the robot motion", INT. J. CONTROL, Vol. 52, No. 3, pp. 581-895, 1990.

[6] Sadao Kawamura, Jinwoo Jun, Katsuya Kanaoka, Hiroaki Ichii, "A Cascaded Feedback Control Scheme for Trajectory Tracking of Robot Manipulator Systems with Actuator Dynamics", IEEE/RSJ International Conference on Intelligent Robots and Systems 2006 (IROS2006).

Experimental Study of Lubricant Depletion in Laser-assisted Magnetic Recording

Norio Tagawa[1], Hisamitsu Yano[2], and Atsunobu Mori[1]

[1] Dept. of Mechanical Engineering, High Technology Research Center, Kansai University, 3-3-35, Yamate-cho, Suita, Osaka 564-8680, Japan

[2] Kansai University, presently Sharp Corporation

Abstract

Laser–assisted magnetic recording (LAMR) is one of the novel technologies that are employed for achieving magnetic recording densities higher than 1 Tb/in^2. In this study, a fundamental research on the lubricant depletion caused due to laser heating in LAMR systems was conducted. That is, the behavior of the lubricant films under laser heating was investigated experimentally. The obtained results revealed that lubricant depletion occurred and the lubricant could not be replenished. In addition, the effects of laser power, laser irradiation duration, and lubricant film thickness on the lubricant film depletion were studied and the lubricant depletion mechanisms were discussed.

Keywords:

Laser–assisted Magnetic Recording Systems; Lubricant Depletion; Perfluoropolyether Lubricant Film

1 INTRODUCTION

The storage density of magnetic hard disk drives has been increasing very rapidly and hard disk drives with a magnetic recording density of approximately 200 Gb/in^2 have found practical applications. The head-disk interface (HDI) spacing has been decreased to as low as 10 nm in order to achieve such ultra-high magnetic recording density. In current hard disk drive systems, the carbon overcoat is coated with a lubricant film that has a thickness of around 1– 2 nm. In order to realize magnetic recording densities of more than 1 Tb/in^2, the HDI spacing is expected to become less than 2–3 nm. Thus far, various novel technologies have been studied and developed in order to achieve a magnetic recording density of more than 1 Tb/in^2. Laser–assisted magnetic recording (LAMR) is one such novel technology. However, with regard to this novel technology, it has been suggested that there exists a critical HDI issue associated with the use of ultra-thin liquid lubricant films on the disk surface. This problem is attributed to the fact that these films are heated to high temperatures by using laser beams in order to reduce their magnetic coercivity. Thus far, theoretical and experimental researches have been carried out in order to understand the mechanism of the lubricant depletion caused due to laser heating in LAMR systems [1]–[4]. However, the existing knowledge regarding the abovementioned issue is still limited because of the complicated behavior of lubricant films. Therefore, in this study, we have conducted a fundamental research of the lubricant depletion caused due to laser heating in LAMR systems. In other words, the behavior of the lubricant films under laser heating was investigated experimentally. The obtained results revealed that lubricant depletion occurred and the lubricant could not be replenished. In addition, the effects of laser power, laser irradiation duration, and lubricant film thickness on the lubricant film depletion were studied and the lubricant depletion mechanisms were discussed.

2 EXPERIMENTAL

2.1 Test disk and test lubricant

The disks used in this study were 2.5 in. glass substrate disks with a Ta thin film layer of thickness 50 nm and a carbon overcoat of thickness 4.5 nm on one side of the disk. The average surface roughness was about 0.5 nm, as measured by atomic force microscopy (AFM). The disks were coated with an approximately 2.26-nm-thick lubricant film by the dip-coating process. Conventional Zdol2000 was used as the test lubricant.

2.2 Experimental setup

A schematic of the experimental setup, comprising an air bearing spindle, slider mount fixture, slider load/unload mechanism, and optical pickup, is shown in Fig. 1. A laser beam from the optical pickup unit was irradiated through the one side of the disk surface on which the lubricant film as well as various other films was not deposited, and the Ta thin-film layer was heated locally to temperatures of up to or greater than approximately 200 ºC by using a conventional focusing servo control used for optical disk actuators. The specifications of the optical pickup related to the laser beam are listed in Table 1 In this experiment, the test disk was rotated, whereas the slider was not used.

Table 1: Specifications of the optical pickup.

Spot diameter	0.9μm
Wavelength	660nm

Table 2: Laser irradiation test conditions.

Lubricant	Zdol2000
Lubricant thickness [Å]	12.4, 22.6
Linear velocity [m/s]	8.00
Laser power [mW]	1,2,3,4,5
Laser irradiation duration [s]	1,5,10

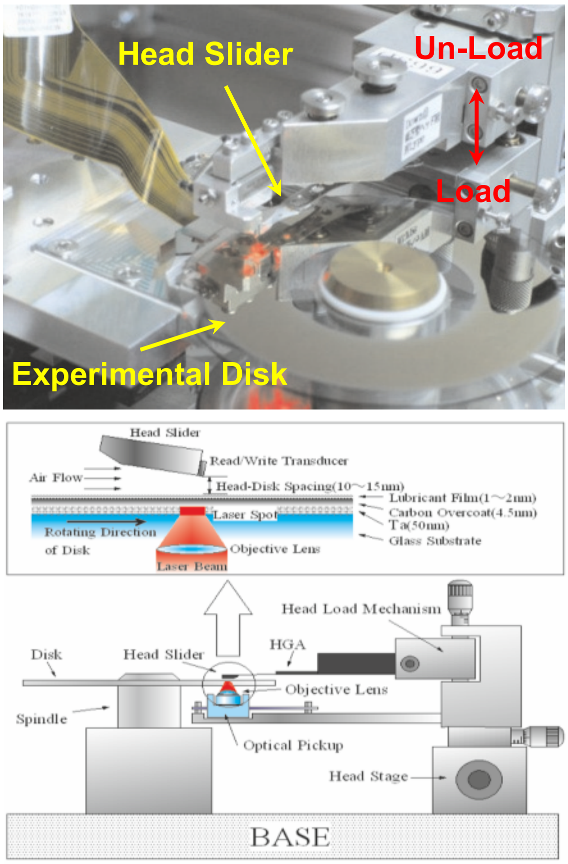


Figure 1: Schematic of the experimental setup for the laser-assisted magnetic recording.

2.3 Experimental conditions

The laser irradiation test conditions of this experiment are listed in Table 2. The disk rotational velocity was 8 m/s and the laser irradiations were conducted by varying the laser power as well as the laser irradiation duration. After each laser irradiation, the changes in the thickness of the lubricant film were investigated, using a surface reflectance analyzer (SRA).

3 EXPERIMENTAL RESULTS AND DISCUSSION

Figures 2 and 3 show the SRA image of the laser irradiated track and the lubricant film thickness profile, respectively, after the laser irradiation. From these figures, it can be seen that the lubricant depletion occurred and the lubricant film profile changed. The lubricant depletion depth and width were about 5.4 Å and 11 μm, respectively, for a laser power of 5 mW and laser irradiation duration of 10 s. Another interesting observation is that a raised ridge is formed that surrounds the lubricant-depleted region. This may indicate that the lubricant depletion is driven by the thermocapillary stress induced by the disk surface temperature gradient resulting from the non-uniformity of the laser spot intensity distribution [1]. This raised ridge was, however, vanished due to the lubricant spreading after around one day. On the other hand, the depleted lubricant could not be replenished in this case. The effects of laser power on the lubricant depletion depth and width are shown in Figs. 4 and 5, respectively. In this experiment, it is observed that the lubricant depletion does not occur when the laser power is less than 3 mW. In other words, it is found that the laser power should be less than 3 mW to prevent the lubricant depletion of Zdol2000 due to laser heating. Furthermore, it can be seen that the lubricant depletion depth increases very remarkably when the laser power is more than 4 mW. With regard to the lubricant depletion width, it is about 8~12 μm. Accordingly, the depletion width is found to be ten times greater than that of the laser beam spot. The effects of laser irradiation duration on the lubricant depletion depth and width are shown in Figs. 6 and 7, respectively. Based on these results, it is suggested that the lubricant depletion proceeds significantly rapidly in the laser spot region within a very short time frame because the lubricant depletion proceeds very rapidly within 1 s. After 1 s, the lubricant depletion depth appears to be proportional to the laser irradiation duration for the time range considered in our experiment. The lubricant depletion width increases very rapidly within 1 s, which is very similar to the increase in the lubricant depletion depth. After 1 s, it becomes almost saturated and constant, independent of the irradiation duration. Therefore, it is suggested that the effect of laser power on the lubricant depletion due to laser heating is greater than that of the laser irradiation duration in laser-assisted magnetic recording. We also investigated the lubricant depletion in the case of the 1.24-nm-thick lubricant film. Figure 8 shows the lubricant film thickness profile after the laser irradiation. It can be observed that the lubricant depletion depth and width are approximately 0.1 Å and 100 μm, respectively and that the raised ridge is not formed. It is found that the depletion width is approximately one hundred times greater than that of the laser beam spot. In addition, it is confirmed that the lubricant depletion is replenished in this case. These experimental results are considerably different from those obtained in the case of thick lubricant films and these results suggest that the lubricant film depletion in the case of thin lubricant films is driven dominantly by the mobile lubricant evaporation due to the high temperature. Therefore, it is considered that the lubricant depletion due to laser heating depends largely on the lubricant film thickness.

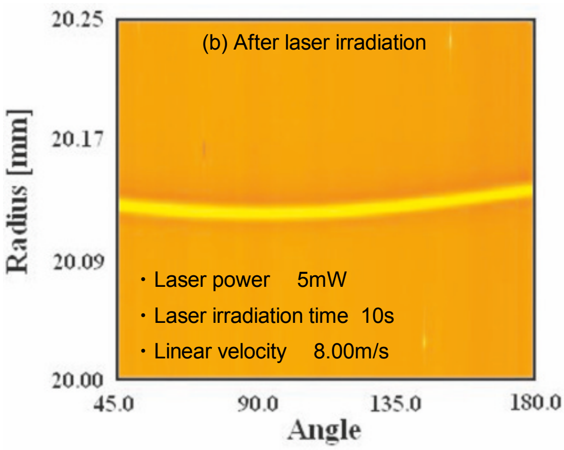


Figure 2: SRA images of laser irradiated track.

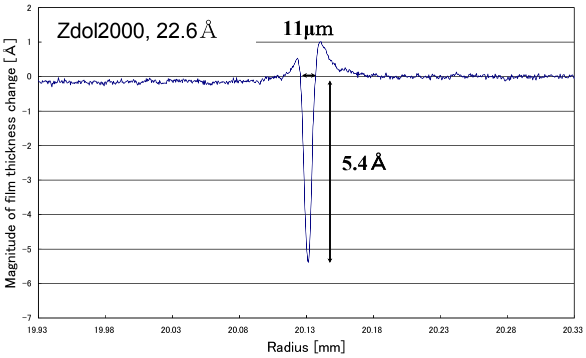


Figure 3: Lubricant film thickness profile after laser irradiation.

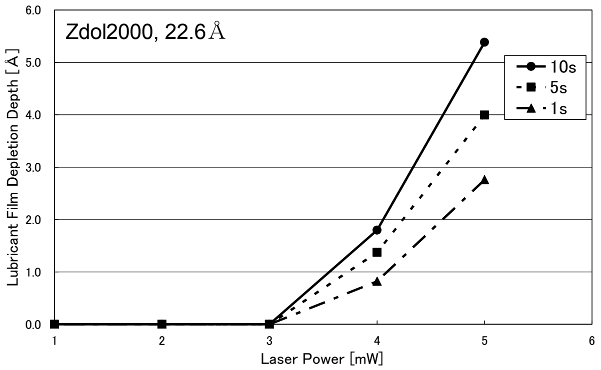


Figure 4: Depletion depth vs. laser power.

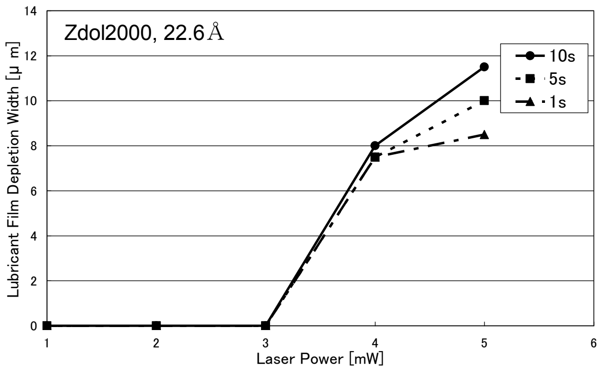


Figure 5: Depletion width vs. laser power.

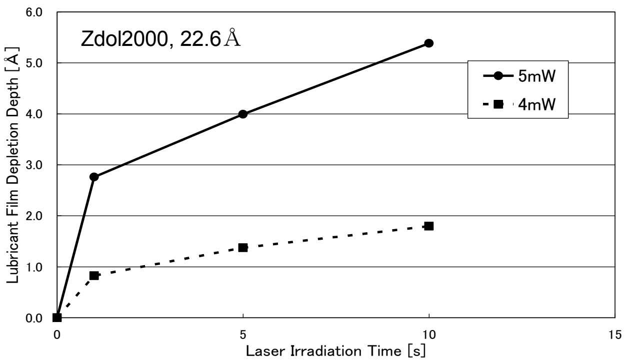


Figure 6: Depletion depth vs. laser irradiation time.

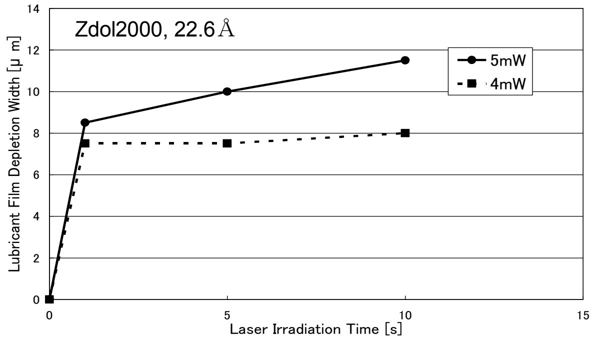


Figure 7: Depletion width vs. laser irradiation time.

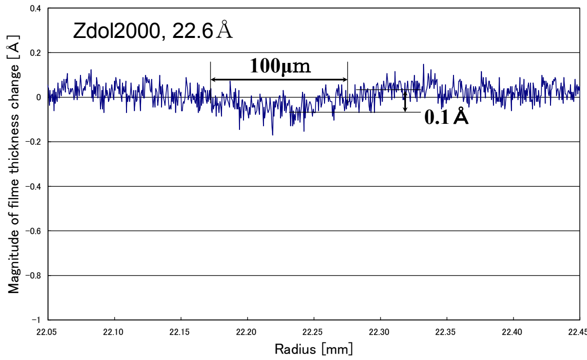


Figure 8: Lubricant film thickness profile after laser irradiation.

4 SUMMARY

We have conducted a fundamental research on the lubricant depletion caused due to laser heating in LAMR systems. In this study, we focused on the thermal stability of the ultra-thin liquid lubricant films in laser-assisted magnetic recording. In other words, the behavior of the lubricant films under laser heating was investigated experimentally. The obtained results revealed that lubricant depletion occurred and the lubricant could not be replenished. In addition, the effects of laser power, laser irradiation duration, and lubricant film thickness on the lubricant film depletion were studied and the lubricant depletion mechanisms were found to be driven by not only the lubricant evaporation but also the thermocapillary stress induced by the disk surface temperature gradient resulting from the non-uniformity of the laser spot intensity distribution. In addition, it was concluded that the lubricant depletion due to laser heating depends largely on the lubricant film thickness.

5 ACKNOWLEDGMENT

This work was supported in part by a project of the High Technology Research Center of Kansai University, titled “The development of innovative nanotechnology for the Tb/in^2 class ultrahigh-recording-density information storage systems for supporting a ubiquitous society” and by a grant from the Storage Research Consortium (SRC).

6 REFERENCES

[1] L. Wu., 2006, “ Lubricant Depletion due to Laser Heating in Heat-Assisted Magnetic Recording,” Presented at Symposium on Frontier of Magnetic Hard Disk Drive Technology and Tribology, June 23-24, Santa Clara, California, USA.

[2] J. Zhang, R. Ji, J. W. Xu, J. K. P. Ng, B. X. Xu, S. B. Hu, H. X. Yuan, and S. N. Piramanayagam., “Lubrication for Heat-Assisted Magnetic Recording Media,” IEEE Trans, Magn., Vol. 42, No. 10, pp. 2546-2548, Oct. 2006.

[3] N. Tagawa, H. Yano, and A. Mori., 2007, “Study of Head-Disk Interface in Laser-Assisted Magnetic Recording,” Program and Extended Abstracts of 17th Annual ASME Conference on Information Storage and Processing Systems, pp. 171-173, June 18-19, Santa Clara, California, USA.

[4] L. Wu., “Modeling and simulation of the lubricant depletion process induced by laser heating in heat-assisted magnetic recording system.” Nanotechnology, Vol. 18, pp. 1-8, 2007.

Light Transmittance Memory Effect of Ferroelectric Materials Induced by Electrical Imprint Field

Toshinori Ohashi, Hiroshi Hosaka and Takeshi Morita

Dept. of Human Environmental Studies, Graduate School of Frontier Sciences,
The University of Tokyo, Kashiwa, Japan

Abstract

In this paper, the possibility of the light transmittance memory effect was examined. Recently, we have pointed out that an imprint electrical field originates the memory effect of various properties of ferroelectric materials, such as strain and permittivity. In this paper, using a lead lanthanum zirconate titanate, it was confirmed that the light transmittance had two stable values depending on the polarization state in the absence of an electrical field in the presence of the imprint electrical field. The control of the polarization direction with the pulsed voltage successfully demonstrated the light transmittance memory effects.

Keywords:
PLZT; imprint electrical field; light transmittance; memory effect; ferroelectric material

1 INTRODUCTION

In these days, optical switches have been paid attention for broadband communication. For example, optical switches in the multiplexing do not convert optical signals into electrical signals. Therefore advanced optical switches are indispensable for larger capacity of transmission. Optical switches are mainly classified into three groups such as mechanical switches, Micro Electro Mechanical Systems (MEMS) optical switches and waveguide switches. Waveguide switches utilize functional materials whose refractive index depends on applied heat, light, or electrical field. One of those materials is lead lanthanum zirconate titanate (PLZT).

PLZT is a transparent ferroelectric material that has various nonlinear electro-optic effects, such as variable birefringence and variable light scattering [1]-[4] (Figure 1) which was applied for optical scanners, optical switches and optical shutters. [5]-[8] Conventionally, those devices need a continuous external electrical field so as to remain a certain optical value and the polarization direction of ferroelectric materials is not reversed. Those optical devices consume much electrical power and need complicated drivers to apply an electrical field continuously.

Recently, our group has focused on a memory effect utilizing an imprint electrical field. An imprint electrical field is observed mainly in ferroelectric thin films and has been conventionally researched on in order to be removed. [9] We have pointed out that the various properties of ferroelectric materials, such as strain, permittivity and light transmittance, can obtain memory effects in the presence of the imprint electrical field. The imprint electrical field can be intentionally induced by applying a high voltage electrical field to ferroelectric materials at a high temperature for a long time. A shape memory piezoelectric actuator and refractive index memory optical switch had been demonstrated by this technique. [10] - [12]

In this study, the memory effect on light transmittance is proposed using PLZT as an optical shutter or an optical switch. This optical memory realizes a pulse operation, and enables the maintenance of the on or off state in the absence of an external electrical field; thus, the energy consumption would be decreased, and a simple operation becomes possible; then matrix switches would be fabricated easily (Figure 2).

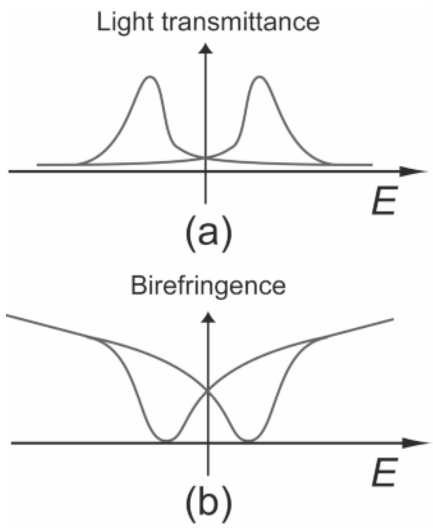


Figure 1: Electro-optic effects of PLZT. (a) is for variable light scattering and (b) is for varialble birefringence.

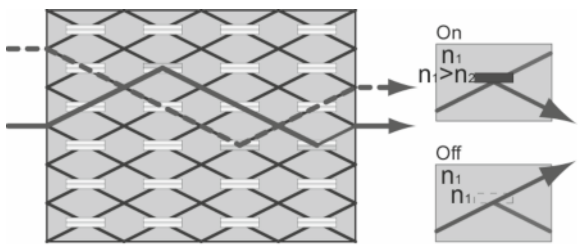


Figure 2: Matrix switch.

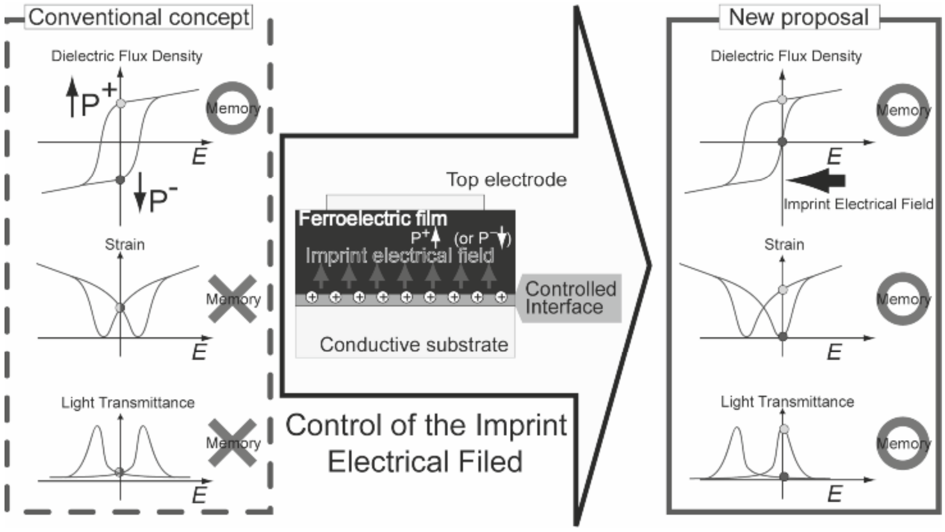


Figure 3: Principle of memory effect induced by controlling impinrt electrical field.

2 PRINCIPLE

Ferroelectric material has a spontaneous polarization direction. The polarization direction is reversed when the external electrical field becomes larger than the coercive field. The relationship between an electrical field and the various properties of ferroelectric materials such as strain, permittivity is explained as a butterfly-shaped curve.

An imprint electrical field is sometimes observed in research on the ferroelectric thin films for ferroelectric random access memory (FeRAM) devices. [13] The imprint electrical field is sometimes induced in the process of forming ferroelectric thin films especially epitaxial thin films. In the presence of the imprint electrical field, the butterfly-shaped curve of the various properties shits to horizontal axis; then the various properties have memory effect that there are two different values depending on the polarization direction in the absence of an external electrical field, as shown in Figure 3. The residual strain or the lattice defect at the boundary surface must be related to this electrical imprint field; however, the conclusive mechanism of this imprint electrical field is unclear until now. The imprint electrical field can be intentionally induced by applying a high voltage electrical field to ferroelectric materials at a high temperature for a long time. This imprint electrical field is a serious problem in FeRAM research and has been researched on in order to be removed; on the contrary, it is the principle phenomenon in this study.

Applying an electrical field to PLZT its light transmittance changes because of valuable a light scattering effect. The principle of this effect is based on the refractive index discontinuities at domain and grain boundaries. Variation of ferroelectric domain size, density of strain-relieving 71 ^{o} and 109 ^{o} domains and proportion of rhombohedral to tetragonal grains and domains as functions of polarization induce the valuable light scattering effect. [1] The relationship between the light transmittance and the intensity of the electrical field is explained as a butterfly-shaped curve, as shown in Figure 1. The light transmittance of PLZT becomes maximum when the electrical field matches a coercive electrical field. At the initial state, the butterfly-shaped curve of the light transmittance is symmetric against electrical field; thus, there is no memory effect even with polarization reversal. However, in the presence of an imprint electrical field, the relationship between the light transmittance and the applied electrical field shifts to the horizontal axis, as shown in Figure 3; then, two stable light transmittances can be maintained depending on the polarization direction in the absence of an external electrical field.

An optical shutter using the PLZT in the presence of an imprint electrical field has two stable light transmittances in the absence of an external electrical field, and each value can be selected by controlling the polarization direction by adjusting pulsed voltages. Such pulse voltages can be obtained easily using transformer from low voltage, it means that a large amplifier is not required.

3 EXPERIMENTS AND RESULTS

3.1 Fabrication of an optical shutter

An optical shutter that uses a PLZT plate (Zr : Ti = 65 : 35, La = 8.19 $^{o}/_{o}$, 5 x 1.5 x 0.5 mm^{3}) was fabricated (Figure 4). PLZT has the Pockels (primary) or Kerr (secondary) electro-optic

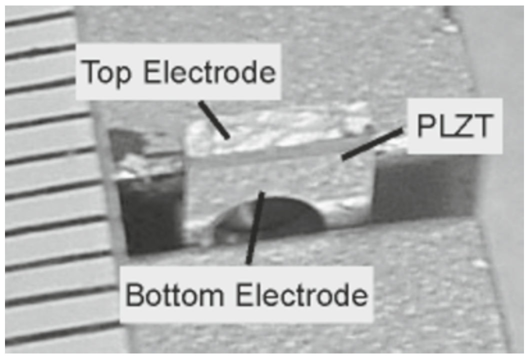


Figure 4: Photo of optical shutter.

effect, depending on the chemical composition. In this experiment, PLZT which had the Pockels effect was used because the light transmittance characteristics relates to the butterfly-shaped curve against external electrical field. On one side of PLZT, a brass block (5 x 1.5 x 30 mm^3) was glued as a bottom electrode with a conductive adhesive (DOTITE 705A), and on the other side, a rectangular top electrode was pasted with the same adhesive. After fabricating both electrodes, their both side surfaces (5 x 0.5 mm^2) were polished to mirror finish using a precision polishing machine (Musashino Denshi MA-150) so that the laser beam can penetrate, whose spot size was 0.48mm.

3.2 Electrical imprint control

The imprint electrical field was sometimes induced in the process of forming ferroelectric thin films especially epitaxial thin films. It might be related to be residual strain due to lattice mismatching. However, the bulk of PLZT does not have an imprint electrical field in the initial condition. In order to induce an imprint electrical field, a 2.5 kV/mm electrical field was applied in the thickness direction for 10 h at 120 ^oC in silicone oil. The electrical field was supplied from an amplifier (NF HVA4321) and the silicone oil was heated by an electric hot plate (IKA RH digital KT/C). The electrical field supply was turned off when the temperature of the silicone oil cooled down enough. In the previous research about a shape memory piezoelectric actuator, the intensity of the imprint electrical field could be controlled depending on those parameters, such as the temperature, the intensity of the electrical field and the voltage supplied time. [11]

3.3 Measurement of light transmittance

Before and after the electrical imprint field treatment, the change of the light transmittance of PLZT was measured. The laser (633 nm, 0.8 mW, 0.48 mm spot size, Edmund 61318-I) went through PLZT, and the electrical voltage was supplied from a function generator (NF WF1974) through a high-speed amplifier (NF 4010, M-2601). A photo diode (Edmund 54522-I) was set at about 300 mm distance from PLZT to detect the intensity of the penetrated laser. The experimental setup is shown in Figure 5. When a triangular electrical field (0.9 kVpp, 0.1 Hz) was applied, the relationship between the light transmittance and the electrical field was measured as shown in Figure 6 (before the imprint electrical field treatment (a) and after the imprint electrical field treatment (b)). Before the electrical imprint field treatment, the relationship between the light transmittance and the applied voltage was symmetric. In the presence of the imprint electrical field, the PLZT obtained the memory effect. It was confirmed that transmittance characteristics had two stable light transmittances in the absence of an external electrical field due to internal electrical field.

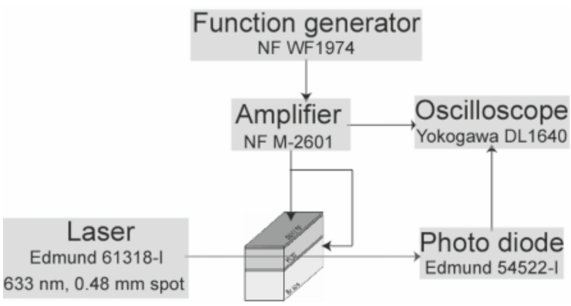


Figure 5: Experimental setup

Using this sample, the light transmittance of PLZT was controlled with the pulse voltages, as shown in Figure 7. The applied pulse voltage was ± 450 V, and the pulse width was 86 ms. The intensity of the pulsed voltage was the same

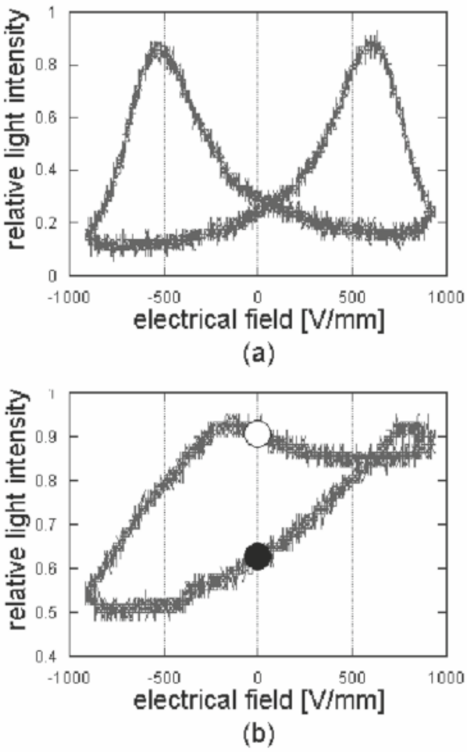


Figure 6: The relationship between the light transmittance and applied electrical field of PLZT without an imprint electrical field (a) and with an imprint electrical field (b).

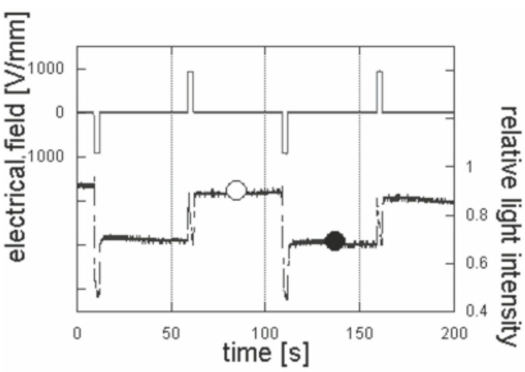


Figure 7: The light transmittance controlled with the pulsed voltages.

intensity of the triangular electrical field for measuring the butterfly-shaped curve. Without an external electrical field, the light transmittance maintained each stable value and was controlled depending on the polarization direction. The light transmittance memory effect was successfully demonstrated. The amount of the memory effect which was the difference between the minor light transmittance and the major light transmittance corresponded to the amount presumed from the butterfly-shaped curve as shown in Figure 6. The amount of the memory effect could be maintained for several days at least. The relationship between the amount of elapsed time and the amount of the memory effect is ongoing research.

4 CONCLUSIONS

The principle of the light transmittance memory induced by an electrical imprint was proposed and a PLZT optical shutter was fabricated. It was found that an electrical imprint can be controlled with a relatively high electrical field at a high temperature. Furthermore, the light transmittance can be controlled by adjusting the pulsed voltage in the presence of an imprint electrical field. It maintained a certain value depending on the polarization direction. In our next work, the detailed mechanism of an imprint electrical field will be investigated.

Acknowledgments

This research was partially supported by the Ministry of Education, Culture, Sports, Science, and Technology through a Grant-in-Aid for Scientific Research on Priority Areas, No. 438, “Next Generation Actuators Leading Breakthroughs”, as well as by the Murata Science Foundation, the Futaba Electrons Memorial Foundation, Japan Chemical Innovation Institute and Support Center for Advanced Telecommunications Technology Research (SCAT).

References

[1] Land, C.-E., 1974, Ferroelectrics, 7:45-51.

[2] Maldonado, J.-R., Meitzler, A.-H., 1971, Proc. IEEE, 59:368.

[3] Smith, W.-D., Land, C.-E., 1972, Appl. Phys. Lett., 20:169.

[4] Land, C.-E., Smith, W.-D., 1973, Appl. Phys. Lett., 23:57.

[5] Kosaka, T., Fukunaga, D., Uchiyama, K., Shiosaki, T., 2007, Proc. ISAF2007, 29Ps-A-6.

[6] Hikita, K., Tanaka, Y., 1989, Ferroelectrics, 94:73-80.

[7] Nashimoto, K., Nakamura, S., Morikawa, T., Moriyama, H., Watanabe, M., Osakabe, E., 1999, Jpn. J. Appl. Phys., 38:5641-5645.

[8] Shames, P.-E., Sun, P.-C., Fainman, Y., 1998, Appl. Opt., 37:3717-3725.

[9] Grossmann, M., Lohse, O., Bolten, D., Boettger, U., Schneller, T.,2002, J. Appl. Phys., 92:2680-2687.

[10] Morita, T., Kadota, Y., Hosaka, H., 2007, Appl. Phys. Lett., 90:082909.

[11] Kadota, Y., Hosaka, H., Morita, T., 2008, Jpn. J. Appl. Phys., 47:217-219.

[12] Ohashi, T., Hosaka, H., Morita, T., 2008, Jpn. J. Appl. Phys., 47:3985-3987.

[13] Morita, T., Cho, Y., 2006, Jpn. J. Appl. Phys., 45:4489-4492.

A Minimization of Environmental Burden of High-Speed Milling

Nelfa Desmira[1], Hirohisa NARITA[1], Hideo FUJIMOTO[1]

[1] Graduate School of Engineering, Nagoya Institute of Technology, Japan

Abstract

An environmental burden analyzer of machine tool operation has been developed. This system can evaluate cutting conditions from the view points of total environmental burden due to electric consumption, cutting tool status, coolant, lubrication oil and metal chip of machine tool by considering the emission intensities and other parameters related to the evaluation factor of machine tool operations. Global Warming Potential (GWP) of 100 years impact was considered and equivalent CO_2 emissions (based on emission intensities?) were evaluated. The environmental burden analysis of high speed milling due to cutting speed was determined by using the aforementioned system and is presented in this paper. We figured out that high speed milling is effective in reducing CO_2 emission.

Keywords:

Environmental Burden, Machine Tool, High-speed Milling

1 INTRODUCTION

Sustainable manufacturing has become a global issue due to increasing awareness that machining process creates an environmental burden. This effort has received significant interest from (already a part of governments) governments, industries and consumers. Until now, there are many tools and methods that have been developed in order to minimize the effect of manufacturing process to environmental burden. Those tools included Life Cycle Assessment (LCA), Life Cycle Impact Assessment (LCIA), Environmental Impact Assessment (EIA), and Design for Environment (DfE) [9]. The newest developed life cycle impact assessment is Eco-Indicator 99, a successor of Eco-Indicator 95 [6]. This system assesses the impact of emission to human being and ecosystems. The impact indicators include global warming potential (GWP) and ozone depletion potential (ODP).

In a previous research [10], we have developed an environmental burden analyzer based on LCA policy. This system was applied to machine tool operations such as milling, turning and grinding.

This paper presents the results and analysis on environmental burden of high-speed milling. Moreover, the optimal cutting condition for high-speed milling in order to decrease CO_2 emission is determined.

2 ENVIRONMENTAL BURDEN ANALYZER

2.1 System Overview

The environmental burden analyzer as depicted by Figure 1 consists of information about the workpiece, the cutting tool and conditions and the NC program as inputs [9]. All parameters related to the machining operation and processes like coolant and lubrication oil were estimated.

Electric consumption, cutting tool status, coolant quantity, lubrication oil and metal chip quantity of the machine tool were calculated by using the environmental burden analyzer formula. The database consists of emission intensities, electric power and consumption of machine tool components and other parameters related to the evaluation factor.

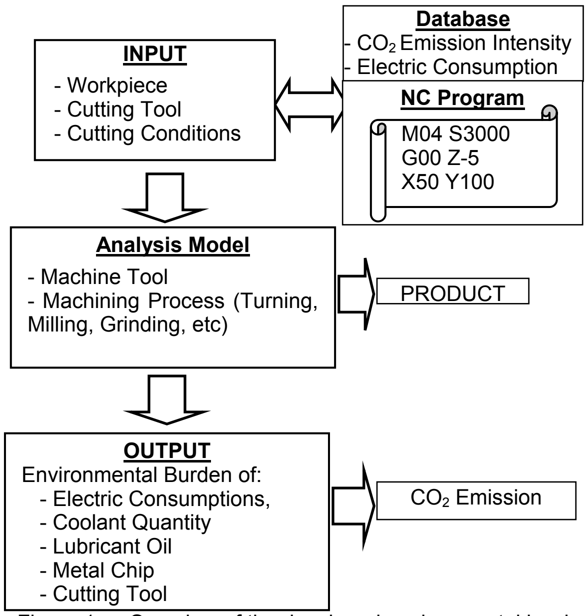


Figure 1: Overview of the developed environmental burden analyzer for machine tool operations

The environmental burden analyzer focuses on global warming and is evaluated based on embodied energy and emission intensity data for Japan [8]. Global warming potential (GWP) is a time dependent index used to compare the specific green house gases relative to CO_2. All gases included in Kyoto Protocol are weighted in their GWP over a 100-year time horizon [2]. In this study, values of GWP for 100 years as listed in Table 1 were considered and the equivalent CO_2 emissions were evaluated as the environmental burden.

Table 1 Characterization factors of GWP [2]

Green House Gas	CO_2	CH_4	N_2O
Global Warming Potential	1	25	298

2.2 Algorithm of Environmental Burden

Total environmental burden (equivalent of CO2 emission) is the sum of electric consumption of machine tool components, coolant quantity, lubricant oil quantity, cutting tool status and metal chip quantity. It is calculated using the following equation:

$$P_e = E_e + C_e + LO_e + \sum_{i=1}^{N} T_{e_i} + CH_e \quad (1)$$

where:

Pe : Environmental Burden of machining operation [kg-CO_2]
Ee : Environmental Burden of machine tool component [kg-CO_2]
Ce : Environmental Burden of coolant [kg-CO_2]
LOe : Environmental Burden of coolant [kg-CO_2]
Te : Environmental Burden of cutting tool [kg-CO_2]
CHe : Environmental Burden of metal chip [kg-CO_2]
N : Number of tool in NC- machining process

Environmental Burden of Electric Consumption (Ee)

The electric consumption is calculated from running time, but if servo and spindle motor are varied dynamically, it is calculated according to the machining process time.

The environmental burden of machine tool component (Ee) is calculated as follows:

$$E_e = k \times \begin{pmatrix} SME + SPE + SCE + CME + CPE + TCE1 \\ + TCE2 + ATCE + MGE + VAE \end{pmatrix} \quad (2)$$

where:

k : CO_2 emission intensity of electricity [kg-CO_2/kWh]
SME : electric consumption of servo motor [kWh]
SPE : electric consumption of spindle motor [kWh]
SCE : electric consumption of cooling system spindle [kWh]
CME : electric consumption of compressor [kWh]
CPE : electric consumption of coolant pump [kWh]
TCE1 : electric consumption of lift up chip conveyor [kWh]
TCE2: electric consumption of chip conveyor in machine tool [kWh]
ATCE: electric consumption of ATC [kWh]
MGE : electric consumption of tool magazine motor [kWh]
VAE : vampire energy (stand-by) of machine tool [kWh]

Environmental Burden of Coolant (Ce)

Coolant stored in a tank and during the machining process is supplied to the cutting area by using a coolant pump. Used coolant is separated from the metal chip before it is re-stored and reused. This system is applied until a new coolant is added. Meanwhile, the cutting fluid is given to the metal chip during the addition of the new coolant. Water is applied in order to minimize vaporization at regular periods.

The environmental burden of the coolant is described as follows:

$$C_e = \frac{CUT}{CL} \times \begin{Bmatrix} (CPe + CDe) \times (CC + AC) \\ + WAe \times (WAQ + AWAQ) \end{Bmatrix} \quad (3)$$

where:

CUT : coolant usage time in NC program [s]
CL : mean interval of coolant update [s]
CPe : environmental burden of cutting fluid production [Kg-CO_2/L]
CDe : environmental burden of cutting fluid disposal [Kg-CO_2/L]
CC : initial coolant quantity [L]
AC : additional supplement quantity of coolant [L]
WAe : environmental burden of water distribution [Kg-CO_2/L]
WAQ : initial quantity of water [L]
AWAQ : additional supplement quantity of water [L]

Environmental Burden of Lubricant Oil (LOe)

Since lubricant oil is used by both spindle and slide way, two equations were formulated. Oil-air lubrication is assumed for spindle while oil lubrication for slide way. Oil-air lubrication and oil lubrication are supplied regularly by a pump to the spindle and slide way parts respectively. In oil-air lubrication case, 0.01 cc ~ 0.06 cc is infused into one oil pipe by the pump.

Equation (4) describes the total environmental burden of lubrication oil and equations (5) and (6) are described by the environmental burden of lubrication oil from spindle and slide way respectively.

$$LOe = Se + Le \quad (4)$$

$$Se = \frac{SRT}{SI} \times SV \times (SPe + SDe) \quad (5)$$

$$Le = \frac{LUT}{LI} \times LV \times (LPe + LDe) \quad (6)$$

Where:

Se :environmental burden of spindle lubrication oil [kg-CO_2]
Le :environmental burden of slide way lubrication oil [kgCO_2]
SRT : spindle runtime in NC program [s]
SV : discharge rate lubricant oil of spindle [L]
SI : mean interval between discharge [s]
SPe :environmental burden of spindle lubrication oil production [kg-CO_2/L]
SDe : environment burden of spindle lubrication oil disposal [kg-CO_2/L]
LUT : slide way runtime NC program [s]
LI : mean interval between supplies [s]
LV : lubricant oil quantity supplied to slide way [L]
LPe : environmental burden of slide way lubricant oil production [kg-CO_2/L]
LDe : environmental burden of slide way lubricant oil disposal [kg-CO_2/L]

Environmental Burden of Cutting Tool (Te)

Cutting tool is calculated from tool life and environmental burden of cutting tool is computed by comparing machining time with tool life. This relationship is described by the following equation:

$$T_e = \frac{MT}{TL \times (RGN + 1)} \times \{(TPe + TDe) \times TW + RGN \times RGE\} \quad (7)$$

where:

MT : machining time [s]

TL : tool life [s]

TPe : environmental burden of cutting tool production [kg-CO_2/L]

TDe : environmental burden of cutting tool disposal [kg-CO_2/L]

TW : tool weight [kg]

RGN : total number of regrinding process

RGE : environmental burden of regrinding [kg-CO_2]

Environmental Burden of Metal Chip (CHe)

Metal chips are collected and recycled in one electric furnace after being separated from the coolant. This process generates environmental burden. However input energy seems to be different for some metal type, but the electric consumption rate is represented in [kWh/t], thus, environmental burden is also calculated from the metal chip weight in this study as described as follows:

$$CH_e = (WPV - PV) \times MD \times WDe \quad (8)$$

Where:

WPV : workpiece volume [cm^3]

PV : product volume [cm^3]

MD : material density of work piece [kg/cm^3]

WDe : environmental burden of metal chip processing [kg-CO_2/kg]

2.3 System Configuration

In this research, an environmental burden analyzer by using the aforementioned algorithm has been developed for high-speed milling. However, this system is applied for machining process in general such as turning, grinding, drilling, etc. Figure 2 shows the example of environmental burden result.

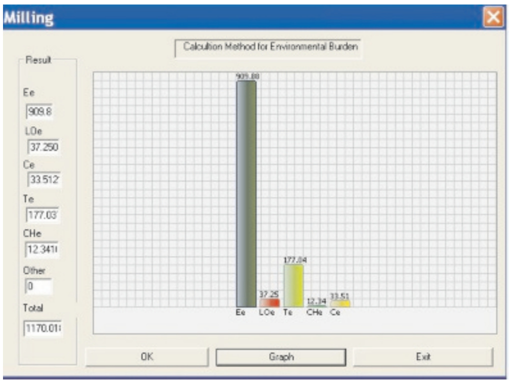


Figure 2: An example of environmental burden analyzer for machine tool operation.

3 ENVIRONMENTALBURDEN OF HIGH-SPEED MILLING

High-speed milling is characterized by a high rotation speed of spindle [17] around five to ten times faster than the speed of conventional milling. Cutting speed of high-speed milling depends on the type of workpiece material machined and type of operation applied [14]. In other words, applied cutting speed for high-speed milling is diverse for various type of materials i.e. 800m/min is high-speed milling for some material, but conventional for other materials.

There are five determinant factors in order to determine total of environmental burden of machine tools [10]. Those five factors mentioned above are electric consumptions of machine tool (Ee), coolant quantity (Ce), lubricant oil quantity (LOe), metal chip quantity (CHe), and cutting tool status (Te). Amount of each factor has two possibilities to decrease and increase. 2^5 probabilities are proposed. However, since the number of metal chip quantity is constant, then the probability changed to 2^4. From these probabilities, three patterns of environmental burden of machine tools due to high-speed milling as shown in Figures 3, 4 and 5 are proposed. In this concept, feed/tooth is remained constant.

Environmental burden of electric consumption of machine tool, coolant quantity and lubricant oil quantity are determined from machining time. Each number of those factors is decreased as machining time is decreased. In other words, high-speed milling is able to reduce machining time, thus effectively reducing the environmental burden or CO_2 emission. However, CO_2 emission generated by coolant quantity and lubricant oil quantity is smaller compared to the CO_2 emission generated by electric consumption.

On the other hand, the environmental burden of cutting tool is managed from the view point of tool life. This factor is calculated by comparing machining time with tool life. Tool life denoted by contact of tool edge and workpiece is defined as the required time of the cutting tool to operate before it reaches its end life limit [1]. Tool life is determined from tool wear. It is difficult to explain the relationship between tool wear of high-speed milling so far. This has become one of the reasons why we proposed the patterns of CO_2 emission due to high-speed milling. More details about this pattern will be explained briefly.

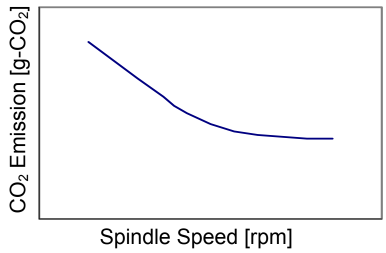


Figure 3: Minimum patterns of CO_2 emissions Due to high-speed milling

Figure 3 called minimum pattern; shows that total of CO_2 emission of machining tool is inversely proportional to the spindle speed and reaches its optimum point at the highest spindle speed. This pattern will occur because CO_2 emission generated by electric consumption is decreased due to decreasing machining time. Actually, CO_2 emission generated from cutting tool is increased, but its increase is not as high as the decrease in CO_2 emission by electric consumptions. In this pattern, increasing the cutting speed decreases the rate of tool wear and the rate of tool life is increased as well as.

Figure 4 shows a parabolic pattern where environmental burden strikes the optimum point at the stated spindle speed [11]. While the lowest and the highest spindle speed generates more CO_2 emission. This occurs because at the lowest spindle speed, CO_2 emission generated by electric consumptions is major, since it's required machining time is longer. While at the highest spindle speed, the CO_2 generated from cutting tool quantity is mainly higher than the other factors, due to decreasing tool life. Moreover, there is more heat transfer produced by the cutting tool to the workpiece during the machining process, and the necessity of coolant and lubricant oil to mitigate the heat transfer becomes more substantial. These conditions will then result to more CO_2 emission. However, there is an optimal point where the CO_2 emission generated is minimal.

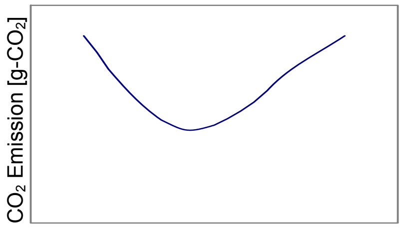


Figure 4: A parabolic pattern of CO_2 emission Due to high-speed milling

In the contrary, maximum pattern as depicted in Figure 5, shows that the total of CO_2 emission reaches its optimum point at the lowest spindle speed and vice versa.

In this pattern, electric consumption generated few of CO_2 emission, thus, CO_2 emission of cutting tool increase markedly. This is mainly caused by increasing of tool wear due to the high-speed milling and affected the increasing of CO_2 emission of cutting tool. Unfortunately, since the machining tool capable of effectively reducing electric consumption is not yet available, this pattern is not discussed further.

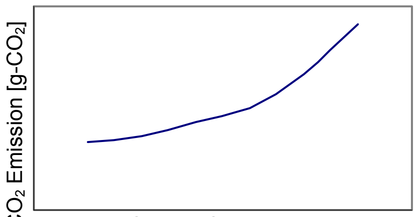


Figure 5: A maximum pattern of CO_2 emission due to high-speed milling

Cutting length and material work piece are also considered has an influence to produce CO_2 emission. It has a relation with tool life and machining time, and affects the electric consumptions, coolant quantity, lubricant oil quantity, metal chip quantity and cutting tool status respectively.

4 CASE STUDY

4.1 Experimental Data

Experiments have been conducted by using machining centre, a cutting tool is carbide square coating with diameter 10mm, flute number is 6, teeth number is 2, helix angle is 30 degree, workpiece is SKD11 (60HRC) [3]. In these experiments, cutting speeds were varied from 20 ~ 200m/min and feed/tooth was kept constant. The axial depth is 0.1 mm, the radial depth is 10 mm, the feed rate is 0.1 mm/rev and cutting length is 60m. Airflow is using instead of coolant. These data provided tool wear information and using those tools wears data, environmental burden of high-speed milling is determined.

4.2 Parameter Setting

The emission intensity using in this analyzer are summarized in Table 2. These factors are taken from some environmental report, technical report, home page and industrial table [4,5,6,7,12.13.15,16]. Production and disposal process factors are considered, but other factors such as transportation etc are ignored.

Table 2 CO_2 emission intensities [10]

Electricity [kg-CO_2/kWh]	0.381
Cutting fluid production [kg-CO_2/L]	0.9776
Cutting fluid disposal [kg-CO_2/L]	0.0029
Dilution liquid (water) [kg-CO_2/L]	0.189
Spindle and slide way lubricant oil production [kg-CO_2/L]	0.469
Spindle and slide way lubricant oil disposal [kg-CO_2/L]	0.0029
Cutting tool production [kg-CO_2/kg]	33.7478
Cutting tool disposal [kg-CO_2/kg]	0.01346
Regrinding [kg-CO_2/number]	0.0184
Metal Chip Processing [kg-CO_2/kg]	0.0634

Electric power and consumption of machine tool components are described in Table 3.

Table 3 Electric powers and consumption [10]

NC controller [kW]	0.16
Cooling system of spindle [kW]	0.45
Compressor [kW]	1
Coolant Pump [kW]	0.25
Lift up conveyor [kW]	0.1
Chip conveyor in machine tool [kW]	0.6
ATC [Wh]	0.08
Tool Magazine [Wh] (1 round)	0.087
Vampire (stand-by) energy [kW]	0.64

Other factors required for calculating environmental burden are shown in Table 4.

Table 4 Other parameters related to the evaluation factor 10]

Initial cutting fluid quantity [L]	8.75
Additional supplement of cutting fluid [L]	4.3

Total dilution fluid quantity [L]	175
Additional supplement of dilution fluid [L]	82.25
Mean interval between replacements of coolant pump [Month]	5
Discharge rate of spindle lubrication oil [mL]	0.03
Mean interval between discharge for spindle lubrication [s]	480
Lubricant oil supplied to slide way [mL]	228
Mean interval between supplies [hour]	2000
Total number of regrinding process	2
Material density of cutting tool [g/cm^3]	11.9
Material density of work piece [g/cm^3]	7.1
Coolant tank capacity of machine tool [L]	175

4.3 Results

The experimental result, as depicted in Figure 6, shows that environmental burden reached a maximum value when at the lowest spindle speed on high-speed milling and vice versa. This result fits with the minimum pattern of environmental burden mentioned above.

From Figure 6, by using mathematic (Least Square Method), a function is determined as follows:

$$y = 106346x^{-0.7127} \tag{9}$$

where:

y = Environmental burden equivalent of CO_2 emission

x = Spindle speed

By using this simple equation, it is possible to determine appropriate cutting conditions in order to minimize CO_2 emission.

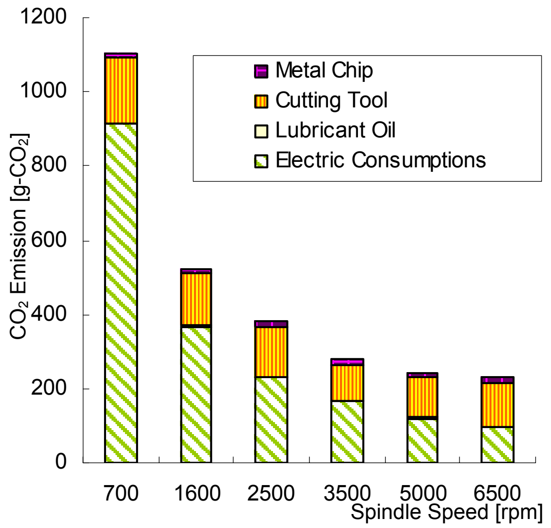


Figure 6: CO_2 emissions due to high-speed milling

By using a minimization program, we find the optimum point of spindle speed (y axis) which produces the minimum value of CO_2 emission (x axis).

From the result, minimum location is located at spindle speed 6365 rpm while CO_2 emission generated is 206.89 g-CO_2. Figure 7 is shown the minimum point of spindle speed and minimum CO_2 emission more detailed.

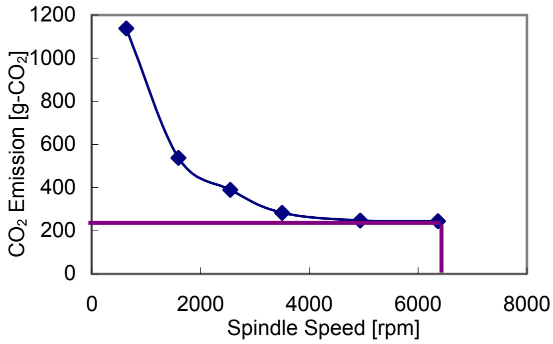


Figure 7: An example to decide the cutting condition

Spindle speed is decided 6365 rpm, by using equation (9), CO_2 emission generated is 206.89g-CO_2. From this information, the appropriate cutting condition can be decided. In this case, we can define cutting condition as follows:

$$V_c = \frac{\pi \cdot d \cdot n}{1000} \tag{10}$$

$$V_c = \frac{3.14 \cdot 10 \cdot 6365}{1000} = 199.681 m/\min$$

$$V_f = \frac{f_z \cdot z \cdot 1000 \cdot V_c}{\pi \cdot d} \tag{11}$$

$$V_f = \frac{0.1 \cdot 2 \cdot 1000 \cdot 199.681}{3.14 \cdot 10} = 1271.85 mm/\min$$

Finally, the cutting speed is decided 199.681m/min and feed velocity is 1274.85mm/min. In other words, by using this cutting condition, the

4.4 Analysis of high-speed milling

CO_2 emissions are high in cases with spindle speeds lower than 3000 rpm and cutting speeds lower than 80 m/min. Therefore, the conventional milling operations emit more CO_2 because they have longer machining time and higher electric consumption. Moreover, coolant and lubricant oil are needed more during the machining process and tool life and tool wear increase as well. In contrary, when the spindle speed is higher than 3000 rpm or cutting speed higher than 110 m/min, CO_2 emission decreases and reaches its optimum point at spindle speed 5000-6000 rpm or cutting speed 155-200 m/min. In other words, high-speed milling decreased the CO_2 emission because of the reduction of the electric consumption due to the shortening of machining time. Since the rate of tool wear of high-speed milling compared to conventional milling is decreased, the rate of tool life is also increased and the CO_2 emission generated from cutting tool is decreased. Besides these, coolant and lubricant oil are also decreased. From this analysis, it is found that high-speed milling is effective in reducing CO_2 emission. Therefore, the first pattern from the three patterns of CO_2 emission due to high-speed milling stated above is proved.

5 CONCLUSION

In this research, we have developed environmental burden system for machine tools focused on global warming and equivalent of CO_2 emission is evaluated as the environmental burden. The effect of high-speed milling to reduce CO_2 emission is analyzed based on tool wear information from some experiments by various cutting speeds. We have showed possibilities of CO_2 emission variability due to the high-speed milling. There are three patterns available for environmental burden of high-speed milling. An equation is determined in order to decide appropriate cutting conditions to minimize CO_2 emission. Moreover, high-speed milling has its limitation range of spindle speed to minimize the CO_2 emission generated (optimal high-speed milling).

References

[1] Caldeirani Filho J., Diniz A. E., 2002, Influence of Cutting Conditions on Tool Life, Tool Wear and Surface Finish in the Face Milling Process, Journal of the Brazilian Society of Mechanical Sciences, 24/1:10-14.

[2] IPCC Fourth Assessment Report (AR4), 2007, Changes in Atmospheric Constituents and in Radiative Forcing

[3] Imaizumi, H., Ouseji, 2007, The state-of-the-art trend of the cutting tool, Japanese Society of Die and Mould Technology (in Japanese):1-6.

[4] Ministry of Economy, Trade, and Industry, 2004, About Declaration of Evaluation Report related to Budget Request in Fiscal Year 2002 (Policy Evaluation), Section 6 Construction of Industrial Facilities (Regional Economic Industry U.

[5] Environmental Report of Tokyo Waterworks, 2002 (in Japanese.

[6] Ministry of Housing Spatial Planning and the Environment, 2000, Eco-Indicator 99 Manual for Designers.

[7] Mizukami, H., Yamaguchi, R., Nakayama, T., and Maki, T., 2002, Off-Gas Treatment Technology of ECOARC, NKK Technical Report (in Japanese), No 176:1-5.

[8] Nansai, K., Moriguchi, Y., and Tahono, S., 2002, Embodied Energy and Emission Intensity Data for Japan Using Input-Output Tables (3EID) –Inventory Data for LCA-, Center for Global Environment Research, National Institute of Environmental Studies, Japan.

[9] Narita, H., Kawamura, H., Norihisa, T., Chen, Lian-Yi., Fujimoto, H., , and Hasebe, Takao., 2006, Development of Prediction System for Environmental Burden for Machine Tool Operation (1st Report, Proposal of Calculation Method for Environmental Burden), The Japan Society of Mechanical Engineer, Vol. 49, No. 4:1188-1195.

[10] Narita, H., Nelfa Desmira., Fujimoto, H., 2008, Environmental Burden Analysis for Machining Operation using LCA Method, The 41st Conference on Manufacturing System (CIRP), 65-68.

[11] Narita, H., Fujimoto. H., 2007, Environmental Burden Analysis Due to High-Speed Milling, Proceedings of the 19th International Conference on Production Research (ICPR19), 081.pdf, CD-ROM.

[12] New Energy and Industrial Technology Development Organization (NEDO), 2003, Development of Environment-Conscious Technology Laboratory, First Section Meeting of “Development of Energy Saving Wastewater Treatment Technique”, Program Original Register of Development of Energy Saving Wastewater Treatment Technique.

[13] SETAC, 1993, Guidelines for Life Cycle Assessment: A code of Practice.

[14] Schmitz T.; Davies M.; Dutterer B.; Ziegert J., 2001, The application of high-speed CNC machining to prototype production

[15] The Society of Non-Traditional Technology, 1995, Survey Report (Additional Volume) of Basic Survey Research for Construction of Environmental Burden Evaluation System – Material Inventory Data – (in Japanese).

[16] Tokyo Electronic Power Company Homepage, 2003, The Earth, People and Energy TECPO Sustainability Report.

[17] W. Bouzid Sai, 2005, An Investigation of Tool Wear in High-Speed Turning of AISI 4340 Steel, International Journal Manufacturing Technology, 26:330-334

Fast Estimation Method of Workpiece Shape in NC Machining Process for Prediction of Instantaneous Cutting Force

Jun'ichi Kaneko[1], Koji Teramoto[2], Kenichiro Horio[1] and Yoshimi Takeuchi[3]
[1] Dept. of Mechanical Engineering, Graduate School of Engineering, Saitama University, Saitama, Japan
[2] Dept. of Mechanical Systems Engineering, Muroran Institute of technology, Hokkaido, Japan
[3]Dept. of Mechanical Engineering, Graduate School of Engineering, Osaka University, Osaka, Japan

Abstract
This study deals with geometric calculation method to predict instantaneous cutting force in NC machining process. In this paper, a new algorithm to introduce performance of new graphics hardware (GPU) is proposed. In the developed algorithm, GPU in rendering process estimates distance from viewpoint to tool swept volumes, initial workpiece shape and locus surface of cutting edge and detects the nearest surface from reference plane on each view line. By referring color on rendered image, GPU can detect a part of cutting edge removing workpiece in each cutter location and rotate angle in short time.

Keywords:
Geometric Calculation; Cutting Force Prediction; GPGPU

1 INTRODUCTION

This study deals a new geometrical calculation method of workpiece shape in order to realize fast prediction of instantaneous cutting force in NC machining process with complicated tool moving steps.

In order to predict instantaneous cutting force, it is required to estimate cutting depth over the whole part of cutting edges in micron unit. Such cutting depth is known as undeformed chip thickness. Because the undeformed chip thickness is determined by geometric relation among cutter location of endmill, tool rotation angle and workpiece shape changed during each cutter location, prediction of accurate instantaneous cutting force is required to estimate local shape around cutting tool at every position where cutting tool locates. Especially, in case products are designed with complicated free-form surface and to be machined by ball endmill with small diameter, workpiece surface during NC machining is complicated curved surface consisted of large number of cusps generated by hundreds of thousands of cutter locations.

In conventional studies, geometric estimation of workpiece shape during machining process is achieved by either solid model kernels developed for CAD software or discrete geometric models proposed for CAM simulation, Z-map [1], Voxel, G-Buffer [2]. However, in case of actual products, applying such conventional geometric models often cause serious problem in validity. For example, the solid models usually require logical operation between workpiece volume and trajectory of cutting tool volume for each tool moving, and each operation takes much computational time. On the other hand, the discrete geometric models require computer memory capacity in proportion to both resolution of description and scale of workpiece. So, it is quite difficult to handle with the convention geometric models by the current architecture of general computers, when machined workpiece has size of several metric unit.

To solve these problems, we propose a new calculation algorithm based on GPU (Graphics hardware) computing technology in order to estimate actual cutting depth on each part of cutting edge at arbitrary moment. GPU is computational device designed to draw 3-dimentional computer graphics with high frame rate. Then, high performance for scientific calculation in general purpose is expected. For example, the latest GPU can transform coordinate hundreds of millions of times for one second and detect the nearest polygon from viewpoint millions of times for one second.

The developed prototype system is designed for 3-axis control machining. And it can estimate the cutting depth accurate enough to predict instantaneous cutting force at arbitrary moment in NC program consists of hundreds of thousand steps within several ten milliseconds. This performance means that prediction of instantaneous cutting force requires only a same time to draw shape of workpiece in machining.

2 ESTIMATION OF ACTUAL CUTTING DEPTH BASED ON IDEA OF TOOL SWEPT VOLUMES

In NC machining process, cutting tool is moved usually according to an order described in NC program. Therefore, a sequence about removal of workpiece volume is easy to estimate in geometric simulation. So, in this study, we propose a new method to check for existence of cutting between each part of cutting edge and machined workpiece volume. The proposed method is based on the idea of tool swept volume [3]. By introducing this idea, the existence check of cutting can be performed by logical product conjunctions about inclusion relations between each tool swept volume and a part of cutting edge.

Figure 1 shows the idea of tool swept volume. In this study, we define a volume has occupied by cutting tool within ith step as tool swept volume TSVi. As illustrated, the tool swept volume is defined by shape of cutting tool, the start point and the end point of the tool moving step in NC program. Furthermore, it is easy to estimate inclusion relation between a point and single tool swept volume, because the tool swept volume has usually simple shape.

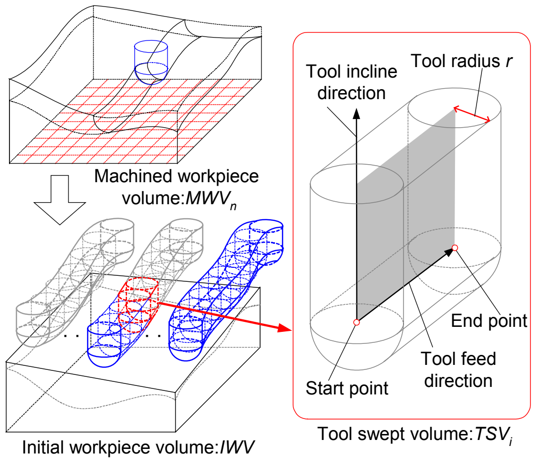


Figure 1: Tool swept volume in NC machining process.

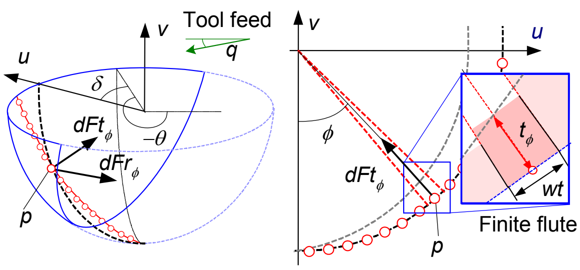


Figure 2: Ideal cutting depth of finite flute on curved cutting edge of ball endmill.

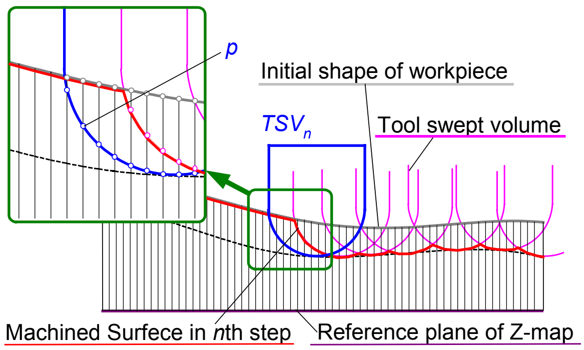


Figure 3: Height comparison in Z-map method.

Therefore, we define volume of machined workpiece at the moment when ith tool moving step is finished as MWV_i and volume of initial workpiece before machining as IWV. By using these variables, we can describe MWV_n by following equation.

$$MWV_n = MWV_{n-1} \cap \overline{TSV_n} = IWV \cap \left(\bigcap_{i=1}^{n} \overline{TSV_i} \right) \quad (1)$$

Here, we introduce variable p as a point on cutting edge at arbitrary cutter location and tool rotation angle. When p satisfies all following conditions, we can assume that a finite flute on cutting edge corresponding to p is removing workpiece volume and generates new workpiece surface in nth tool moving step.

1. p exists inside of IWV.
2. p exists outside of envelope surface include MWV_1 to MWV_{n-1}.

When we define Boolean function of existence e as following,

- When p exists inside of volume V, $e(p, V)=1$.
- When p does not exist inside of volume V, $e(p, V)=0$.

we can estimate whether the finite flute corresponding to p exists inside of MWV_n with Equation 2.

$$e(p, MWV_n) = e(p, IWV) \times \sum_{i=1}^{n} (1 - e(p, TSV_i)) \quad (2)$$

In conventional study, ideal cutting depth on each cutting edge can be estimated by following.

$$t_\phi = r + fb - \sqrt{f^2(b^2 - 1) + r^2} \quad (3)$$

$$b = sin(\delta - \theta) sin\phi cos q + cos\phi sin q \quad (4)$$

,where f is feed rate per one cutting edge and r is tool radius. θ is tool rotation angle. ϕ is the angle of a line between p and center of cutting edge against -v direction. δ is twist angle of cutting edge at p from bottom and t_ϕ is ideal cutting depth at p. The relation of these variables is illustrated in Figure 2.

In these equations, it was assumed that workpiece has flat surface and shape of workpiece machined by former steps of tool moving can be ignored. So, we introduce new assumptions that f is quite small and influence of edge on the undeformed chip can be ignored. Then, we regard at_ϕ as actual cutting depth at p for machined workpiece shape. at_ϕ can be described as Equation 5.

$$at_\phi = e(p, MWV_n) \times t_\phi \quad (5)$$

By using t_ϕ and $e(p_\phi, MWV_n)$, it is enable to predict instantaneous cutting force at arbitrary cutter location and tool rotation angle.

So, in the flowing sections, we propose a new calculation method to realize fast existence evaluation $e(p, MWV_n)$ by using geometric calculation with GPU. The proposed method is focused on 3 axis control machining for fabrication of mold with ball endmill.

3 NEW ESTIMATION METHOD OF EXISTENCE EVALUATION BY USING GPU DRAWING FUNCTIONS

3.1 Existence evaluation from height information about envelop surface of tool swept volumes and initial workpiece shape

In 3 axis control machining, direction of cutting tool axis is fixed in machining process. So, we introduce Z-map representation method in order to estimate workpiece shape with discrete information of height corresponding to each lattice point on reference plane of Z-map. As illustrated in Figure 3, workpiece shape machined by nth step can be regarded as a set of point, which is the nearest intersection among initial workpiece volume and tool swept volume of the n unit.

So, we introduce the idea of discrete representation of Z-map to the Boolean function of existence e. If machined workpice shape can be described by Z-map representation, we can judge the results of function *e* in equation 2 whether following conditions are satisfied.

1. *p* is located near the reference plane from the upper surface of initial workpiece volume.
2. A vertical line crossing *p* intersects initial workpiece volume.
3. *p* is located near the reference plane than all tool swept volumes from 1 to *n*.

In this study, we propose a method to judge these conditions from rendered images by GPU. GPU has ability to detect the nearest surface from viewpoint and render an image with color the nearest surface has on each view line. So, in the following section, we explain new rendering algorithm of GPU to judge whether the conditions are satisfied.

3.2 Height comparison from rendered color images of each shape by GPU

As mentioned above, we introduce GPU ability to existence evaluation. In the proposed algorithm, a set of two color images is rendered at each cutter location. And, results of existence evaluation for all finite flute and all rotation angels are estimated from each set of two color images. In following section, we explain about drawing settings for two color images and estimation of existence evaluation.

Rendering process of two color images for one cutter location

First, we define a view filed with orthographic projection and view field coordinates system ***Xv*** in order to render depth information as illustrated in Figure 4. The direction of Z axis on ***Xv*** matches is the central axis of endmill. And, each view line is arranged along the Z axis. The origin of ***Xv*** is an intersection between Z axis and the reference plane of Z-map. In Figure 4, *r* is radius of cutting tool and z_t is height of cutter location from the reference plane of Z-map.

Then, we set shape information of render targets to GPU. The targets are consisted of initial workpice shape, shape of tool swept volume and cutter locus. The initial workpice shape is defined by a set of approximated polygons by CAD software. Shape of tool swept volume is consisted by a set of cylindrical shape and some parts of sphere. These shapes are defined separately for each tool moving steps. Cutter locus is correspond to trajectory of one cutting edge, which is hemisphere turned to the direction of tool feed. And, these targets have different color in order to distinguish each other.

As Figure 4 shows, the cutter locus is always set center of the view field. On the other hand, the initial workpiece shape and shape of tool swept volumes are placed on workpiece coordinates system and did not shift together with the view field. Because cutting tool is usually smaller than workpiece, the view field captures only a part of initial workpiece shape and shape of tool swept volumes.

In the following process, two images at the same view filed for different targets are rendered. They are required to judge whether the three geometric conditions in section 3.1 are satisfied. The targets of the first image **A** are all shapes and it is correspond to the results of 1’ and 3’. The target of the second image **B** is only initial workpiece shape and it is correspond to the result of 2’. These images are shown in buffer memory of GPU as Figure 5 illustrated.

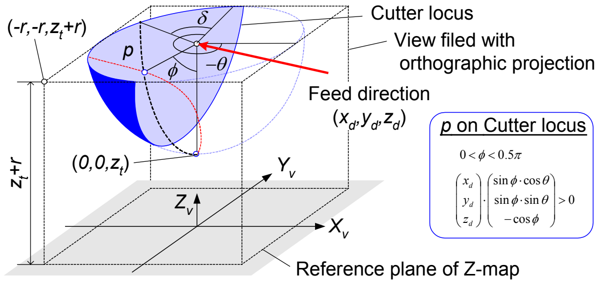


Figure 4: Cutter locus on view field coordinates system.

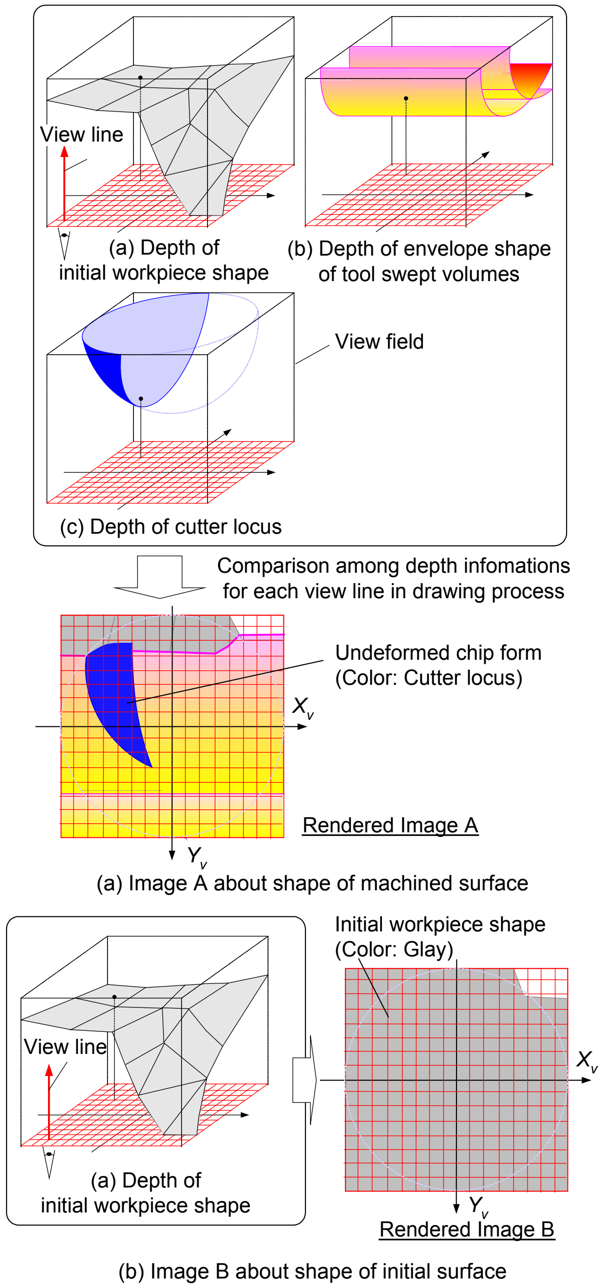


Figure 5: Image rendering for detection of the nearest intersection.

Detect the nearest intersection from rendered color image

On rendered image A and B, the finite flutes of one cutting edge lie on red curved line as shown in Figure 6. The 2D coordinates (x_p, y_p), which are correspond to an intersection between the reference plane and a view line passing *p* can be calculated by following equations.

$$x_p = r \cdot \sin\phi \cdot \cos(-\theta + \delta) \tag{6}$$

$$y_p = r \cdot \sin\phi \cdot \sin(-\theta + \delta) \tag{7}$$

Because color of each part of image A indicates the nearest shape from reference plane of Z-map, the judgement about the three conditions is possible by referring color information on (x_p, y_p).

For example, when (x_p, y_p) on image A is included in the area colored by cutter locus, *p* on the cutting edge satisfies both condition 1 and condition 3. On other hand, *p* on the cutting edge satisfies condition 2, when (x_p, y_p) on image B is included in the area colored by the initial workpiece volume. Then, these results mean that the finite flute of *p* locates inside of machined workpiece MWV_n and a part of cutting edge corresponding to *p* is removing workpiece volume.

In the proposed method, the whole part of cutter locus at the same cutter location is always included in one view field. So, we estimate the existence evaluation $e(p, MWV_n)$ in any rotation angle θ and cutter position ϕ from only two rendered images per one cutter location. In the developed prototype system, which is explained following, GPU renders the two images for each cutter location at start point of tool moving step. Then, the system estimates three components of the instantaneous cutting force 120 times by referring the rendered two images.

4 EVALUATION OF THE PROPOSED METHOD

In order to evaluate the proposed method and algorithms, we develop a prototype simulation system and conduct an experiment using NC program on the supposition of finish machining for mold. In the prototype system, we introduce OpenGL graphics library and GLSL (OpenGL Shading Language) extended library in order to realize accurate calculation of depth information for each pixel on rendered images.

4.1 Process flow of developed prototype system

Figure 7 illustrates process flow of the developed system. In the process flow, area A is corresponding to a part, which is repeated for each cutter location. And, area B shows a part, which is performed by GPU rendering functions. In this process flow, the part requires most geometric calculation is comparison of the depth between all polygon and all view line at each cutter location. Then, such part is included completely in the rendering process shown in area B. So, it is expected that the geometric calculation in cutting force prediction can be achieved as quick as preview process of 3D model in CAD software system performed by GPU rendering function.

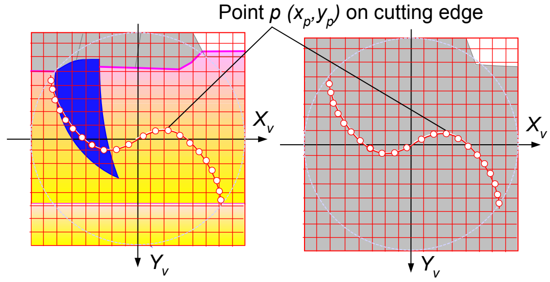


Figure 6: Point *p* on rendered image and relation between *p* and the area colored by cutter locus.

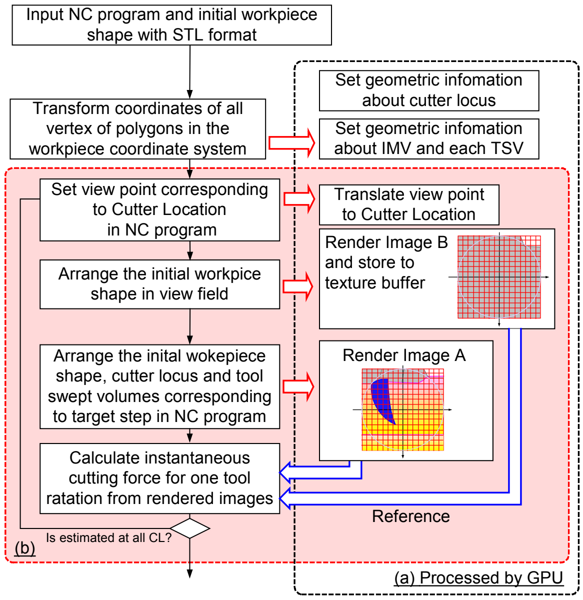
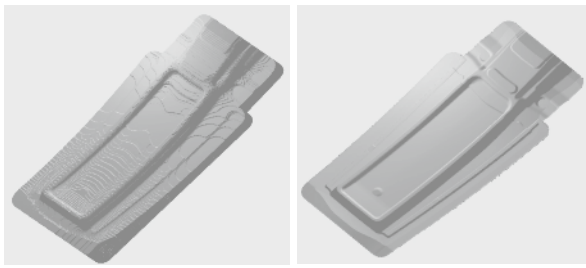


Figure 7: Process follow of image rendering for detection of the nearest intersection.

4.2 Initial workpiece shape and NC program

NC program for the evaluation was generated by commercial CAM systems. The NC program consists of 478,896 tool moving steps. And, six kinds of ball endmill are changed and used in sequence. Figure 8 shows initial work piece shape before finish machining and finished workpiece shape. The initial workpiece shape is equivalent to the shape when rough machining process with large square endmill finished. The initial workpiece shape is approximated as a set of 53,012 polygons in the developed system.



(a) Initial workpice

(b) Machined workpiece by whole steps of the NC program

Figure 8: Workpiece shape for evaluation.

4.3 Rendering images with OpenGL and GLSL

In the developed system, we set color of drawing target as following.

- Initial workpiece shape: Gray that it becomes dark depending on the distance between the reference plane of Z-map and the intersection.
- Shape of tool swept volume: Yellow to red that it depends on position of intersection in gutter shape.
- Cutter locus: Blue.

Figure 9 illustrates the change of workpiece shape in machining by GPU functions. The view fields in these drawing are fixed on the reference plane of Z-map. And, the width of view fields is set to the width of initial workpiece in *X* direction. In the same way, the height is set to the width of initial workpiece in *Y* direction.

The developed system can set the arbitrary moment of rendering target. In the process flow illustrated by Figure 7, we can set freely a range of tool swept volumes that are arranged into the view field. So, by ignoring tool swept volumes after the step of rendering target at range (b) in Figure 7, estimation of *p* existence at each cutter location is achieved easily.

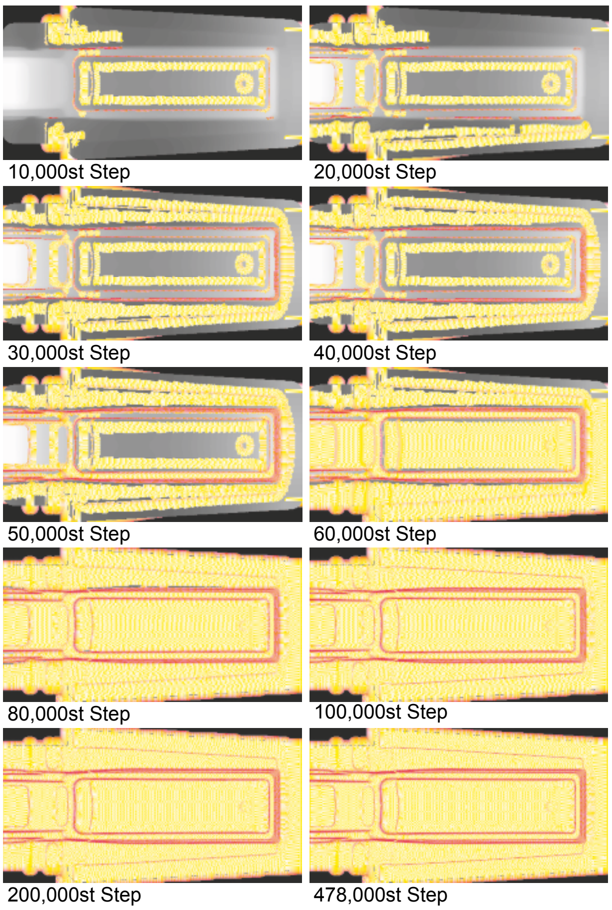


Figure 9: Rendered images of whole shape of workpiece volume in machining process halfway.

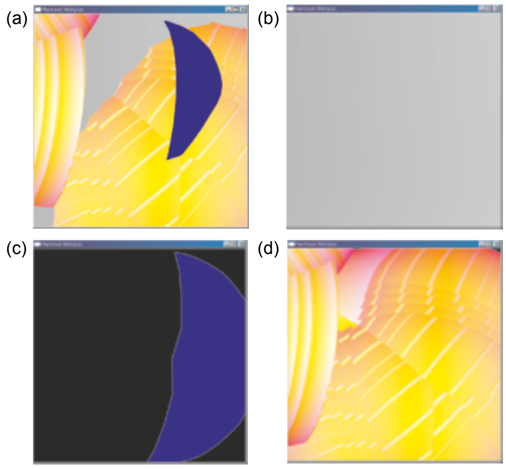


Figure 10: Rendered images from the viewpoint corresponding to cutter location of 1,300th step.

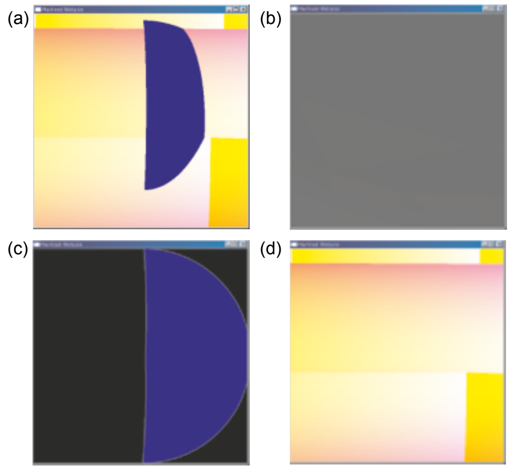
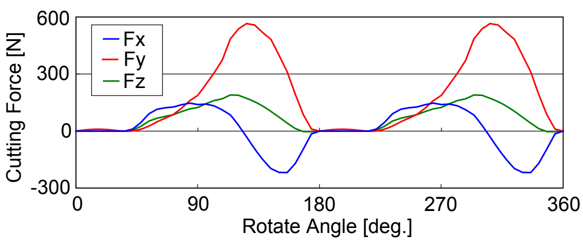
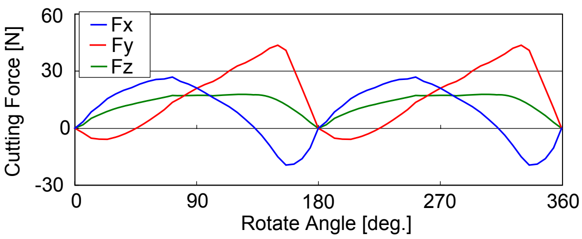


Figure 11: Rendered images from the viewpoint corresponding to cutter location of 200,000th step.



(a) In cutter location of 1,300th step.



(b) In cutter location of 200,000th step.

Figure 12: Calculated instantaneous cutting force from rendered image A and B.

4.4 System performance

Figure 10 and Figure 11 show sample of rendered images at another step. Figure 10 corresponds to 1,300th step in NC program. Figure 11 corresponds to 200,000th step. 4 images in each figure illustrate following results.

(a) Image A

(b) Image B

(c) Rendered image when only the cutter locus is called

(d) Rendered image when only the tool swept volumes corresponding tool moving steps before the target step are called

In the developed system, Image A and Image B are set to consist of 262,144 pixels, which is 512 points in height and 512 points in width.

When we compare (a) with (c), it is clear that the developed system can specify the part which cutting edge participates in workpiece removal from the results of rendering. On the other hand, from Figure 10 (b), Figure 10 (d) and Figure 10 (a), we can find that Image A shows the nearest shape of machined workpiece surface from the reference plane of Z-map.

In the evaluation, we introduce a cutting force model [2] [4] [5] to calculate the instantaneous cutting force from the actual cutting depth. In each ball endmill, a cutting edge is regarded as a set of 90 finite flutes. Figure 12 shows sample of calculated instantaneous cutting force. Figure 12 (a) is corresponding to 1,300th step in NC program as Figure 10. And, Figure 12 (b) is corresponding to 1,300th step in NC program as Figure 11.

In order to evaluate performance of the developed system, we execute the system on the Windows PC with following specification as listed in Table 1.

Table 1: Specification of computer.

CPU	Name	Core2Duo Processor E8400
	Clock	3.0 GHz
	Memory	2048 MB
GPU	Name	GeForce8800GTX
	Clock	680 MHz (GPU Core)
	Memory	768 MB

As a result, the developed system can predict the 120 points of instantaneous cutting force in one tool rotation within about dozens milliseconds per arbitrary cutter location. The average of required time for one tool rotation is about 40 milliseconds. This result means that the proposed method in this paper can realize prediction of the instantaneous cutting force with sufficient performance.

5 CONCLUSION

We conclude this paper as follows.

1. We investigate problems in the conventional estimation method of machined workpiece shape for prediction of instantaneous cutting force.
2. We propose a new geometric calculation method to estimate undeformed chip thickness at arbitrary cutter location and tool rotation angle.
3. We propose a new calculation method in order to introduce the performance of graphics hardware to evaluation about workpice removal in a finite flute on cutting edge.
4. We developed the prototype system and conduct the experimental simulation. The results show that our developed system can apply to the NC program designed for finish machining of large mold. And, the system can estimate the instantaneous cutting force at arbitrary cutter location within dozens milliseconds.

6 REFERENCES

[1] Takeuchi, Y., Sakamoto, M., Abe, Y., Orita, R., 1989, Annal of the CIRP, 38/1:429-432.

[2] Takata, S., Tsai, M., Inui, M., Sata, T., 1989, Annals of the CIRP, 38/1:417-420.

[3] Wang, W., Wang, K., 1986, IEEE Computer Graphics and Applications, 6/12:8-17.

[4] Kaneko, J., Teramoto, K., Onosato, 2002, Proceedings of ICMT2002, 260-265.

[5] Kaneko, J., Teramoto, Horio, K., Takeuchi, Y., 2006, Mechatronics for Safety, Security and Dependability in a New Era, 153-156.

Development of 2-dimensional contouring by endless wire saw

Yoshitaka Morimoto[1]

[1] Dept. of Mechanical Engineering, Faculty of Engineering, Kanazawa Institute of Technology, 3-1 Yatsukaho, Hakusan, Ishikawa 924-0838, Japan

Abstract

A new contouring machining device by wire saw has been developed. The endless wire saw is feedback controlled to contour the target trajectory by detecting the position of cutting point with the PSD sensor unit or LIS unit. The controlled variables are obtained previously by calculating the target point on the work coordinate system to avoid the wire deflection during machining by the cutting forces. The total positioning resolution of the developed device is realized within 2μm by using the 2-axis stepping motor driven stage.

The endless wire is welded by the laser-beam welding equipment so that the diameter of the welded part does not exceed the original one. The tensile strength of the welded wire is enough strong that it can be used as the practical contouring tool.

Various shapes are machined to evaluate the machining ability. The trajectory error of machined cow's backbone (80HV) settles in 100μm by the proposed position feedback. The experimental procedure demonstrates that this device is useful to obtain the 2-dimentional contouring profile.

Keywords:

Wire saw; Contouring; PSD; Micro cutting; Positioning control

1 INTRODUCTION

Cutting machine with diamond wire saw is widely used at the semiconductor industry. The main usage is off course cutting of Si wafer [1], [2]. The productivity of cutting performance is remarkably improved and the accurate straight cutting is realized.

There exists another usage for a diamond wire saw. That is surgery medical treatments [3]-[5]. In this field, the apparatus for this operation is not always straight cutting only. The operation is required to cut the surgical site with the complicated trajectory not to injure the important site.

We look at the diameter of diamond wire saw that is enough fine to realizing the complicated trajectory by machining. In this case, the manual operation is not suitable for the complicated machining. There is one more serious problem when the fine diamond wire is used for this kind of machining. This leads only to the wire deflection by cutting force. As the machined profile depends on the trajectory of a diamond wire saw, the automated cutting machine to realize this cutting process has to equip the feedback control of the cutting position considering the wire deflection. In this case, the detecting method of the wire deflection is the key factor.

A new 2-dimentional computer controlled cutting machine that applied two kinds of wire deflection detecting methods has been developed. One method is that the wire guide and PSD sensor are used to detect the wire deflection indirectly. The other one is that two linear image sensors are used for detecting the wire deflection directly.

The detecting methods, the positioning performance and the cutting ability of the developed system are reported. The cutting results obtained by the developed cutting machine are evaluated in this report.

2 DEVELOPED CUTTING MACHINE

The developed cutting machine is controlled by position feedback. The position detection of cutting point is so difficult because the cutting point is covered with the cutting tips, cutting fluid. Two kinds of detecting method are tried to obtain the cutting point that depends on the deflection of wire.

In this section, the control system of the developed cutting machine is introduced and the detecting methods are described.

2.1 Driving mechanism and positioning system

Two dimensional stages driven by ball-screw driving mechanism and two stepping motors are used. The schematic diagram is shown in Figure1. The wire saw is positioned to the designed point with the positioning command prepared previously.

The feedback control is applied to this machine by using the position detecting sensor. In this case, point to point control is executed by this system.

Endless wire saw is set between three pulleys to minimize the machine size. The wire is connected by the laser-beam

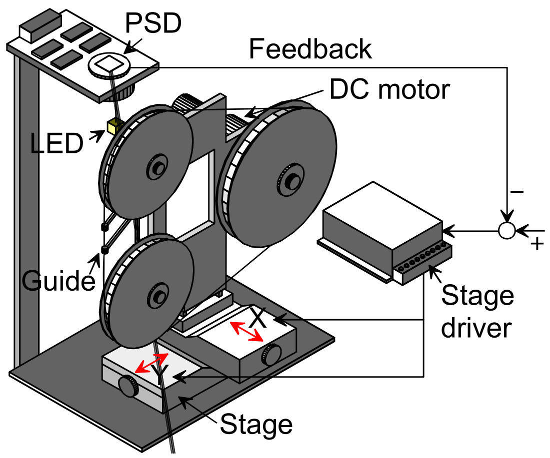


Figure1: Schematic diagram of developed cutting machine

welding. Connected point is not always the same diameter as the one of original wire. The tensile strength test is executed to determine the welding condition previously to establish the moderate performance.

2.2 Cutting point detection method

The wire has so fine diameter of 0.2mm that it cannot avoid the deflection by the cutting force. This leads to worse trajectory during machining. Two types of detecting methods are applied in order to solve this problem.

2.2.1 Position sensitive detector type

Position sensitive detector (PSD) is widely used for the electric device industry and optics device industry. The principle of the detecting method is that the infrared red emitted from the linkage set to the wire is received by the two dimensional detector device in which the analog signal proportional to the position is output from the amplifier.

The linkage guide for detecting wire cutting position is attached to just above the cutting point. The two dimensional coordinates of the wire position as shown in Figure 2 and Figure 3 can be calculated by the following method. The default geometric parameters are measured and determined previously. The position signals from the PSD are used to calculate the wire position by next equations.

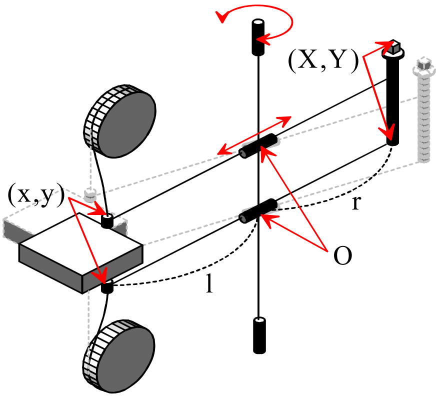


Figure2: Detecting mechanism for position and wire deflection

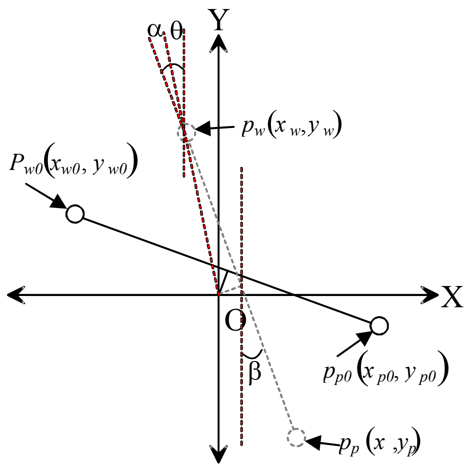


Figure3: Geometric alignment of detecting mechanism

$$x_w = -x_{w0} + r\cos\beta - l_{1p}\sin\beta \tag{1}$$

$$y_w = -y_{w0} + r\sin\beta + l_{1p}\cos\beta \tag{2}$$

Where,

$$\overline{OP}_p = \sqrt{(x_{p0}+x_p)^2 + (y_{p0}+y_p)^2} \tag{3}$$

$$\theta_p = \tan^{-1}\frac{x_p + x_{p0}}{y_p + y_{p0}}, \qquad y_{p0} > 0 \tag{4}$$

x_w : current wire position of x direction,

y_w : current wire position of y direction,

x_{w0}: original wire position of x direction,

y_{w0}: original wire position of y direction.

When the wire attaches to the workpiece surface, the machining starts and the wire deflection also begins simultaneously. Our method realizes to calculate the wire position easily. In this case, although the deflection inside the work piece is not considered strictly, the appropriate position of wire can be recognized. The detecting method itself is enough accurate, although the welded point of wire sometimes causes the vibration of linkage.

2.2.2 Linear image sensor (LIS) type

Another trial to detect the wire position directory by using LIS is executed. The linear image sensor (LIS) is ordinary used for image processing. The resolution depends on the array size of device. We chose the minimum resolution of 0.02mm that is enough fine for the wire diameter of 0.2mm. Two LIS

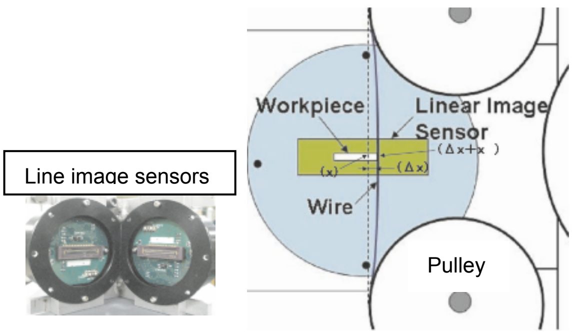


Figure4: Line image sensors and schematic diagram of developed cutting machine

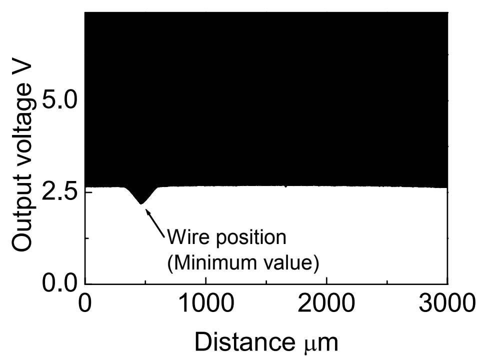


FIgure5: Detecting result of wire position by LIS

are used to detect two dimensional deflection of wire position as shown as Figure 4. The shadow of the wire is reflected on the LIS, the output signal shows the minimum voltage as shown in Figure 5. Although the lighting direction and the light source itself are very severe, once the setup is completed, the steady signal is obtained.

As the detecting length of LIS is proportional to the measurement, the wire position is calculated by the following equations.

$$x = X + \Delta x \quad (5)$$

$$y = Y + \Delta y \quad (6)$$

Where,

x : wire position of x direction,

y: wire position of y direction,

X : stage position of x direction,

Y : stage position of y direction,

Δx : wire deflection of x direction,

Δy : wire deflection of y direction.

2.3 Calculation of cutting force based on deflection measurement

Two kinds of methods to detect the sire position can be used not only to calculate the wire deflection but also to calculate the two dimensional cutting forces. When the preload is set to the wire from the cutting direction, the deflection of wire is measured and the calibration between the load and deflection is executed. Once we obtain the calibration factor between these, the cutting force can be calculated from the next equation.

$$F_x = k_x \Delta x \quad (7)$$

$$F_y = k_y \Delta y \quad (8)$$

Where,

F_x: cutting force of x direction,

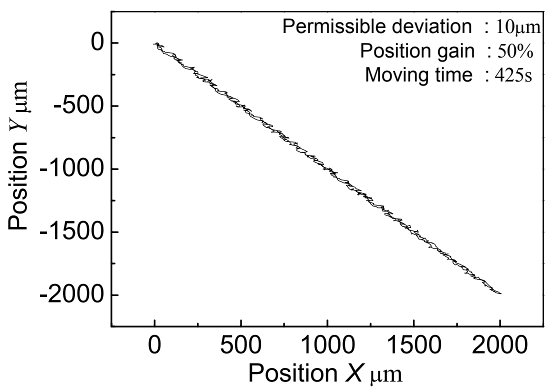
F_y: cutting force of y direction,

k_x, k_y: Calibration factor for cutting forces.

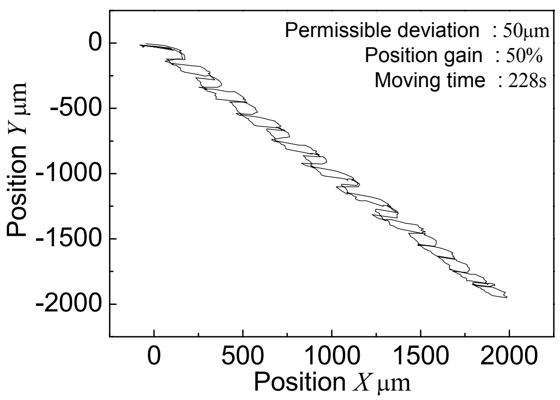
3 EVALUATION OF POSITIONING PERFORMANCE

3.1 Setting of permissive deviation of positioning

The positioning control and the trajectory control are executed by the bang-bang control because of the adaption of stepping motor's driving mechanism. Then, the permissive deviation is inquired to positioning the wire to the target position. Figure 6 (a) and (b) show the deference between trajectories with the two permissive deviation values during 2 dimensional



(a) Permissible deviation 10μm



(b) Permissible deviation 50μm,

Figure 6: Controlled results by change of permissible deviation

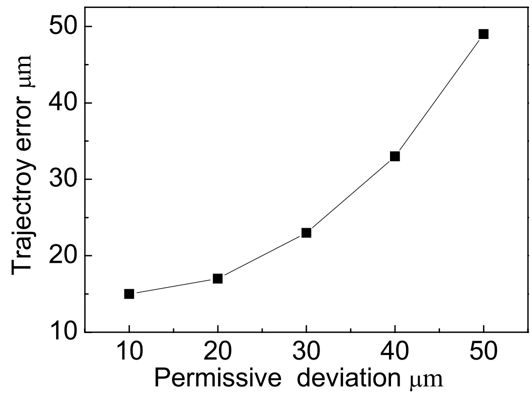


Figure 7: Relation between permissible deviation and trajectory error

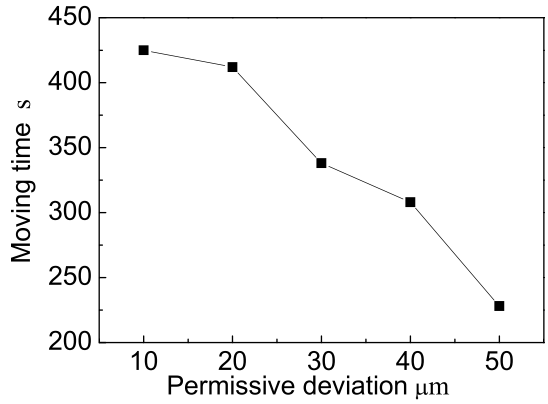


Figure 8: Relation between Permissible deviation and moveing time

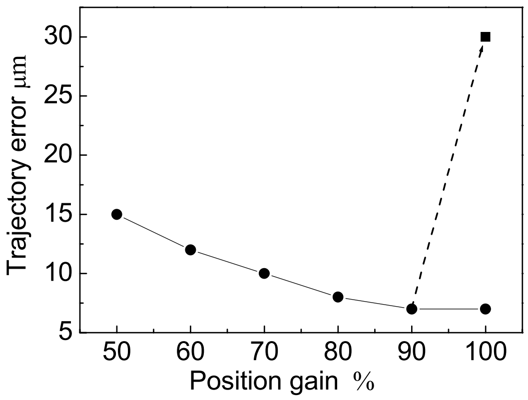


Figure 9: Relation between position gain and trajectory error

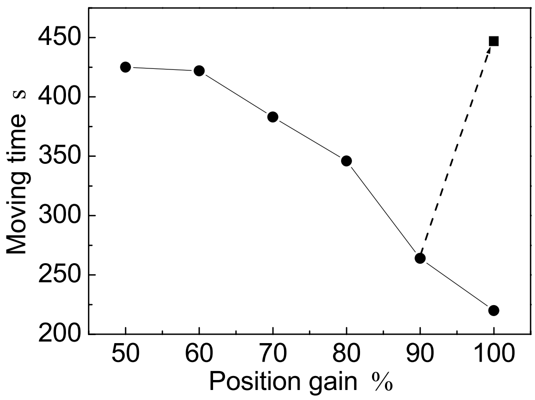


Figure 10: Relation between position gain and trajectory error

straight interpolation trajectories. In this case, air cutting is applied to evaluate just the trajectory accuracy.

The trajectory accuracy of Figure 6 (a) is superior to the one of Figure 6 (b) obviously. The machining time of Figure 6 (a) is linger than the one of Figure (b). There is a bifurcation point between the permissive deviation and the machining time.

Then, the relation between the permissive deviation and the trajectory accuracy is evaluated as shown in Figure 7. The more permissive deviation increases, the worce trajectory error degenerates.

Next, the moving time of this trajectory is evaluated because it has a close relation to the machining time. Figure 8 shows the relation between the permissive deviation and the moving time. The moving time inreases with the magnificaion of the permissive deviation.

On the other hand, the position gain is another important parameter for the machining accuracy. The relation between the position gain and the trajectory error is evaluated as shown in Figure 9. The trajectroy error decreases with the increase of position gain. There is a critical point of position gain as the trajectory error suddenly degenerated in this figure.

Almost same tendency is obtained in the case of relation between the position gain and the moving time as shown in Figure 10.

Table 1: Cutting conditions

Wire diameter [mm]	ϕ0.220
Average grain size [μm]	30～40
Wire speed [m/s]	0.795
Workpiece, Hardness	Cow bones, HV80

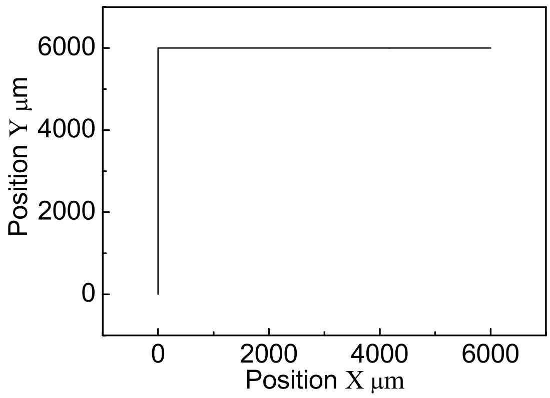


Figure 11: Disired trajectory

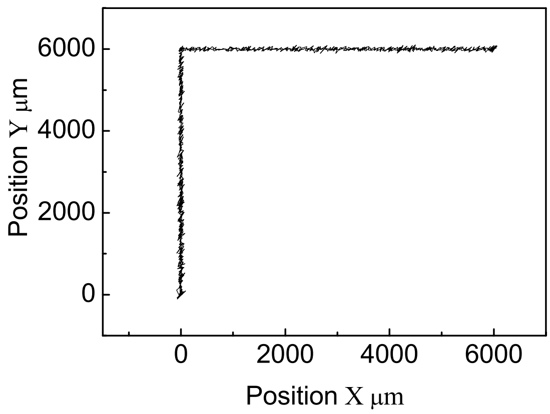


Figure 12: Machined result

Therefore, moderate parameters' combinations are applied to cover the machining time and the trajectory error. The posiotin gain of 90% and the permissive deviation of 20μm are chosen from the ezperiments.

Each parameter is set as same value for both two kinds of detecting method because of the almost same tendencies.

4 MACHINING RESULT

4.1 Machining result of cow born

Our developed cutting machine can be applied as the born cutter for human bone and artificial bone in the future. Then, the cow's backbone is chosen as the target workpiece because its hardness is the hardest one than any other born (HV80).

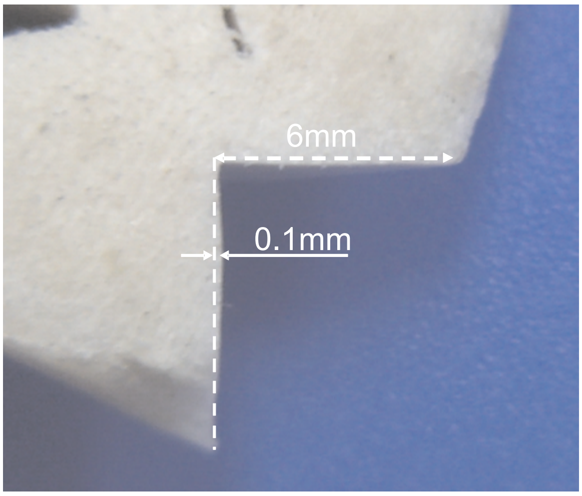


Figure 13: Machined result (profile)

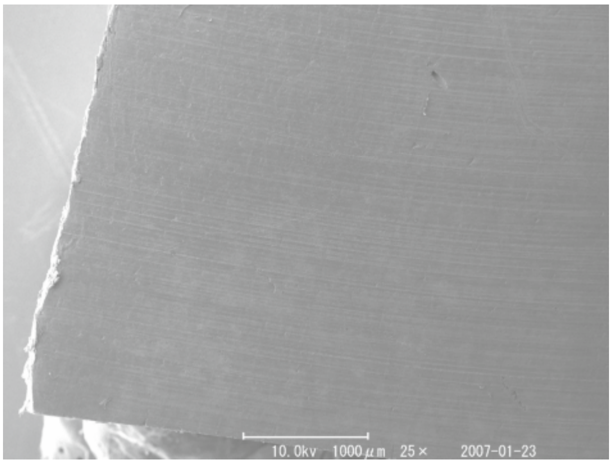


Figure 14: SEM image of macined surface

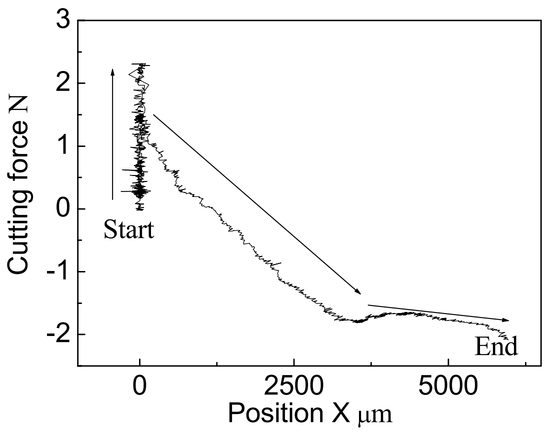


Figure 15: Calculated cutting force (X direction)

Figure 11 shows the target trajectory by using the straight interpolation cutting. Each independent axis motion is

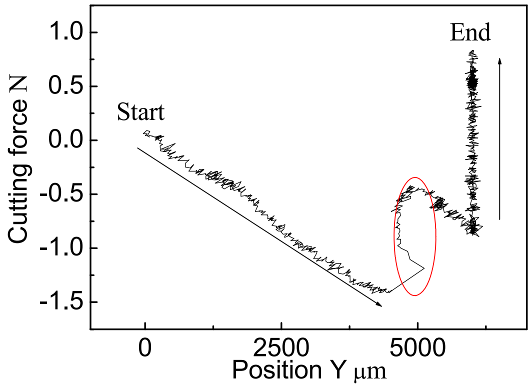


Figure 16: Calculated cutting force (Y direction)

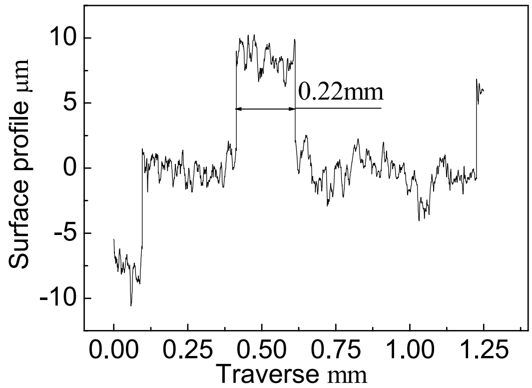


Figure 17: Surface profile of machined bone

controlled by the feedback control as described previously. The cutting conditions are set as shown in Table 1.
Figure 12 shows the measurement results of wire position that is equivalent to the machined trajectory. The diamond grains electrodeposited on the steel wire (JIS G3522) are used for the cutting tool. Although some vibration is observed during motion, the trajectory itself shows enough accurate for the wire saw machining.
Figure 13(a) shows the shot image of the cow backbone after machining by this operation. This figure agrees with the profile of Figure 12. The machined trajectory is accurate enough. The small curvature of 0.1mm is observed. The reason of this result is that the sensing position is just on the surface at the start point. The thickness of workpiece changes its value during machining. Then, the sensor cannot cover the exact deflection of wire.
Figure 14 shows the SEM image of the machined surface of the workpiece. The wire trajectory is influenced by the characteristic cartilage fibril. Although the periodic tongue is observed clearly, the machined surface shows smooth and steady one. These results show that this machine is available for the contouring machine of surgical operation.

4.2 Cutting force calculation

The wire deflection can be calculated from the deference between the wire position and the stage position by equations

(1), (2), (5), (6). The wire deflection is proportional to the load as described in eq. (7), (8). Therefore, the cutting forces during machining can be obtained.
Figure 15 shows the calculation result of the x directional cutting force. The cutting force changes from 0N to 2.4N at the beginning of start point (around X = 0). The reason of this result is that the wire is pulled to the X direction by the cartilage fibril during Y directional cutting. After the wire starts to move to X direction, the cutting force varies from 1N to -2N and settles in constant value around the end. The cutting force depends on the cartilage fibril apparently from this result.
Figure 16 shows the Y directional cutting force. The cutting force decreases from 0N to -1.5N till Y = 4.2mm. Then the cutting force varies suddenly. Machined profile is not affected by this temporary change of cutting force. Therefore, the bone tissue would be changed at this point. The wire cuts across the cartilage fibril. Then, the wire is disturbed around this area. Another disturbance of the cutting force occurs just around the corner because of the change of direction effects the wire deflection. It is obvious that the cutting machine controls the wire to the correct position and its trajectory agrees with the desired one accurately.

These results shows that this contouring machine is available for the practical use not only surgical operation but also other contouring machining.

4.3 Characteristic of machined surface

Figure 17 shows the surface profile of the machined cow bone. The periodic asperity can be seen. The pitch is 0.22mm that is almost equal to the wire diameter of 0.22mm. Once the wire positions to a point where it stays at the same position because of the control method, the wire continue to cut at the same place until settling into the permissible deviation. That is the reason why this surface is obtained.

5 SUMMARY

The new 2-dimensional contouring machine by endless wire saw has been developed by adapting the wire position feedback. The main results obtained are as follows.

(1) Endless wire is positioned by the 2-dimentional stages by the detecting the wire deflection.

(2) Cutting force is calculated by the wire deflection.

(3) Machined trajectory shows enough accurate and the machined profile is agreed with the designed profile.

(4) The machined surface is affected to the cartilage fibril.

(5) Further investigation on the tool life is required to improve the productivity.

6 ACKNOWLEDGMENTS

The author sincere thanks to Professor ICHIDA who gave me a chance to study. Dr. FUNADA who welded the wire by the laser-beam welding method and all who contributed to preparing the instructions.

7 REFERENCES

[1] Suwabe Hitoshi, Shirasu Hideo et al, Development of Electrodeposited Endless Wire Tool of Diamond Grains : Welding Method and Evaluation of Welding Point, JSPE, 59, 6,1993, pp939-944

[2] Ishikawa Kenichi, Slicing Technology of Hard and Brittle Materials: Present and Future, JSPE, 60, 2, 1994, pp.163-167.

[3] Scientific information "T-SAW" SPINE SURGERY, Koshiya Co., ltd.

[4] Kole HK, Surgical operations on the alveolar ridge to correct occlusal abnormalities, Oral Surg 12, 1959,pp 277-282.

[5] Bell WH, Condit CC, Surgical-orthodontic correction of adult bimaxillary protrusion, Journal of Oral Surg, 28, 1970, pp578-583.

[6] Noriaki Ando, Peter Korondi, and Hideki Hashimoto : Development of Micromanipulator and Haptic Interface for Networked Micro- manipulation, IEEE/ASME trans. of Mechatronics, 6, 4, 2001, pp.417-422.

Application of Cusp Surface Analysis to Jumping Phenomenon

Yasufumi Kume

[1] Department of Mechanical Engineering, School of Science and Enginering, Kinki University, Higashiosaka, Osaka, Japan

Abstract

This paper describes the system of three-dimensional representation for Cusp Surface Analysis and the application of this system to jumping phenomenon. Cusp surface is the quantitative cusp catastrophe model in our previous papers. cusp surface analysis proposed by Cobb L.. bases on cusp model in catastrophe theory by Thom R. The psychological data used experimental data of visual recognition for ambiguous figures are analyzed and characteristics of data are discussed. Therefore, effectiveness of cusp surface analysis for psychological data is verified.

Keywords:

Cusp Surface Analysis; Jumping Phenomenon; Visual Recognition

1. INTRODUCTION

. Catastrophe theory is a controversial new way of thinking about change-change in course of events, change in an object's shape, change in a system's behavior, change in ideas themselves. Its name suggests disaster. But it deals with changes as quiet as the dancing of sunlight on the bottom of a poll of water and as subtle as transition form waking to sleep. Mathematical principles are ideally suited to analyze –because they were created to analyze – smooth, continuous, quantitative change: the smoothly curving paths of planets around the sun, the continuously varying pressure of a gas as it is heated and cooled, the quantitative increase of a hormone level in the bloodstream. But there is another kind of change , too, that is less suited to mathematical analysis. The cusp catastrophe occurs in system whose behavior depends on two control factors. Its graph is three-dimensional, a curved surface with a pleat. Every point of the surface represents an equilibrium state. All the points on the underside of the pleat are unstable maxima. All the points along the fold line, which forms the lip on the pleat, are semi-stable points of points of inflection. All the rest of the points are stable minima. For certain combinations of values of the control factors, there are two possible stable states, one on the upper surface of pleat and one on the lower surface beneath the pleat. The behavior of the system under these conditions is called "bimodal" meaning that the same conditions permit either of two stable states.

Cusp surface analysis [4,5] is based on cusp model in catastrophe theory [3]. Figure 1 shows transition of research for cusp model. In this cusp surface analysis, psychological experimental data by visual recognition for ambiguous figures is investigated about characteristics of data . Then, effectiveness of cusp surface analysis for psychological data is clarified. The three dimensional geometric feature of stochastic cusp catastrophe model in the jumping phenomenon of psychological data is described. The effectiveness of this analysis is verified. The geometric feature of stochastic cusp catastrophe model in the discontinuous phenomenon of psychological data governed by irregular nature was clarified including the human factors, and the model was quantified by three-dimensional representation for cusp surface analysis. The result obtained by this paper is summarized as follows. It was shown by stochastic cusp catastrophe model proposed by this paper from visual perception and clear degree of ambiguous

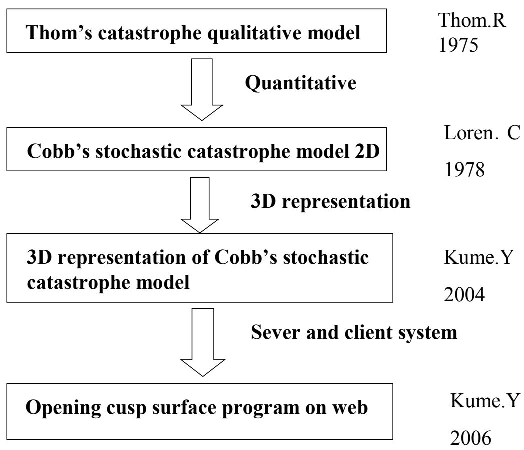


Figure 1 Transition of research for cusp model

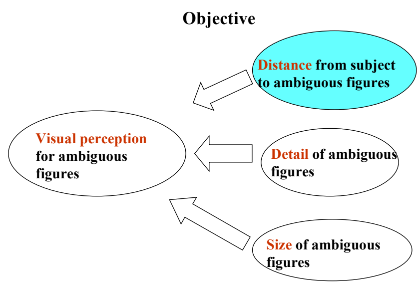


Figure 2 Effect of degree of distinct for visual perception

figures [6,7,8] can be presume. In this paper, the system of three-dimensional representation for Cusp Surface Analysis is described and effectiveness of cusp surface analysis for psychological data is verified. Figure 2 shows the elements of visual perception for ambiguous figure.

In this paper, catastrophe phenomenon is abrupt jumping generated in continuous phenomenon. Human visual perception depends on objects. Effect of distance from ambiguous figures to subject, the size of ambiguous figures and the detail of ambiguous figures on visual perception is clarified by cusp surface analysis. In ambiguous figures, the relationship between distance and visual perception, the relationship between the size of ambiguous figures, and the relationship between the detail of ambiguous figures and visual perception generation of catastrophe phenomenon is clarified. Figure 2 shows these relationships.

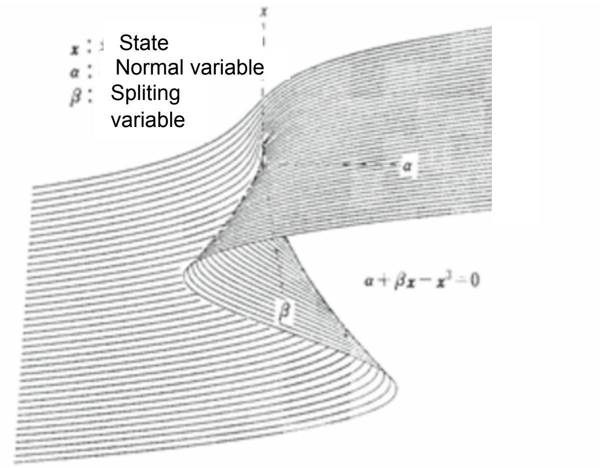


Figure 3 Thom's cusp catastrophe model[1]

2. CATASTROPHE THEORY AND CUSP SURFACE ANALYSIS

In cusp surface analysis[4,5], it is assumed that the variable expressed with X_1, X_2, ・・・, X_v exists in relation to parameter $A(\underline{X}),B(\underline{X}),C(\underline{X})$. The formula is changeable into the

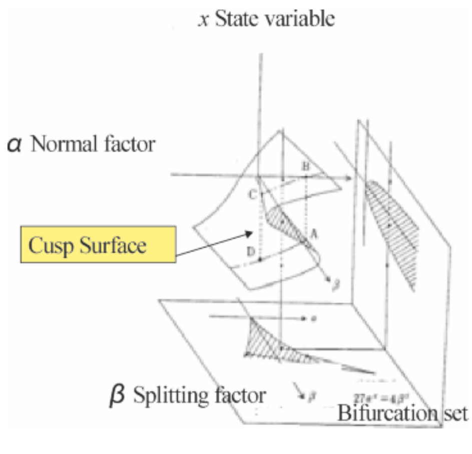


Figure 4 3 D cusp model

Cusp catastrophe model is used. Catastrophe theory is proposed by Rene Thom in French mathematician as mathematical method described development of configuration in nature [3]. Figure 3 and figure 4 show 3D cusp catastrophe model and cusp surface. In cusp surface analysis, it is assumed that the independent variable expressed with X_1, X_2,・・・, X_v exists in relation to parameter $A(\underline{X}),B(\underline{X}),C(\underline{X})$. Then formula is changeable into the form of the following formula.

$$
\begin{aligned}
0 &= A(\underline{X}) + B(\underline{X})[Y - C(\underline{X})] - D[Y - C(\underline{X})]^3 \\
A(\underline{X}) &= A_0 + A_1X_1 + A_2X_2 + \ldots\ldots + A_vX_v \\
B(\underline{X}) &= B_0 + B_1X_1 + B_2X_2 + \ldots\ldots + B_vX_v \\
C(\underline{X}) &= C_0 + C_1X_1 + C_2X_2 + \ldots\ldots + C_vX_v
\end{aligned} \quad (1)
$$

,where D is a constant, and $\underline{X}$ is independent variable vector, three control factors are scalar values and predicted value Y is decided, then v is number of independent variables. A ($\underline{X}$) is unsymmetrical factor, B ($\underline{X}$) is bifurcation factor and C ($\underline{X}$) is linear factor. In parameter estimating, most likelihood method is used. Statistical testing which is satisfied criterion of catastrophe model is not single. Three criterions are proposed by Cobb. This prediction equation is a cubic polynomial in $\boldsymbol{Y}$ which means that for each values of $\boldsymbol{X}$ there are either one or three predicted values of $\boldsymbol{Y}$. For simplicity, it is assumed in these in these figures that C = 0 and D = 1. According to terminology generally used in the literature on catastrophe theory, A and B closely related to the so-called normal and splitting factors, respectively (the term " normal " is used simply because this factor is perpendicular, i.e. normal, to the splitting factor in the canonical representation of the control space). The terminology is recommended for two reasons: (a) to prevent confusion with the normal distribution, and (b) to emphasize the fact that the statistical model is not as flexible as the topological model. In three-dimensional representation for cusp surface analysis, it became possible to treat quantitatively psychological data. Likelihood for cusp model is compared with likelihood of linear model. When ratio of logarithm likelihood is larger than χ^2 value by means of χ^2 testing, the condition of criterion is satisfied. In the case of factor D=0, equation (1) is transformed as follows.

$$
Y = C(\underline{X}) - \frac{A(\underline{X})}{B(\underline{X})} \quad (2)
$$

In equation (1), it take place that unsymmetrical factor A ($\underline{X}$)=A_0, and bifurcation factor B ($\underline{X}$)=B_0. A ($\underline{X}$) is unsymmetrical factor, B ($\underline{X}$) is bifurcation factor and C ($\underline{X}$) is linear factor. In parameter estimating, most likelihood method is used. Statistical testing which is satisfied criterion of catastrophe model is not single. Then three criterions are proposed by Cobb. Likelihood for cusp model is compared with likelihood of linear model. When ratio of logarithm likelihood is larger than χ^2 value by means of χ^2 testing , the condition of criterion is satisfied. In the case of factor D=0, equation (1) is transformed as shown in equation (2). In equation (1), as the value of Y has one value to the value of X, this model does not form catastrophe model. Also, it take place that unsymmetrical factor A ($\underline{X}$)=A_0, and bifurcation factor B ($\underline{X}$)=B_0. Then, when $D \neq 0$ and one at least for A($\underline{X}$), B($\underline{X}$)and($\underline{X}$) is not zero, this criterion is satisfied. When data point in estimated bimodal range do not include in the ratio of constant, bimodal type does not form. Then, when data point existed 10% of bimodal range at least, condition of criterion is satisfied. When these are satisfied condition of criterion, cusp catastrophe model is ormed. In this paper, the three dimensional geometric feature of stochastic cusp catastrophe model in jumping phenomenon of physiological data is described. The effectiveness of this analysis is s verified in the discontinuous phenomenon of physiological data governed by irregular nature was clarified including the human factors. The model was quantified by three-dimensional representation for cusp surface analysis. In this paper, A, B and C will be called the asymmetry, bifurcation, and linear factors, respectively. Cusp Surface Analysis offers three separate tests to assist the user in ev1aluating the overall acceptability of the cusp catastrophe model. To confirm a cusp model all three tests should be passed. The cusp catastrophe model may be said to describe the relationship between a dependent variables if all of these three conditions hold:

1) The chi-square test shows that the likelihood of the cusp model is significantly higher than that of the linear model.

2) The coefficient for cubic term and at least one of the coefficients of the asymmetry

and bifurcation factors are significantly different from zero.

3) At least 10% of the data points fall in the bimodal zone of the estimated model.

3. CUSP SURFACE ANALYSIS SYSTEM

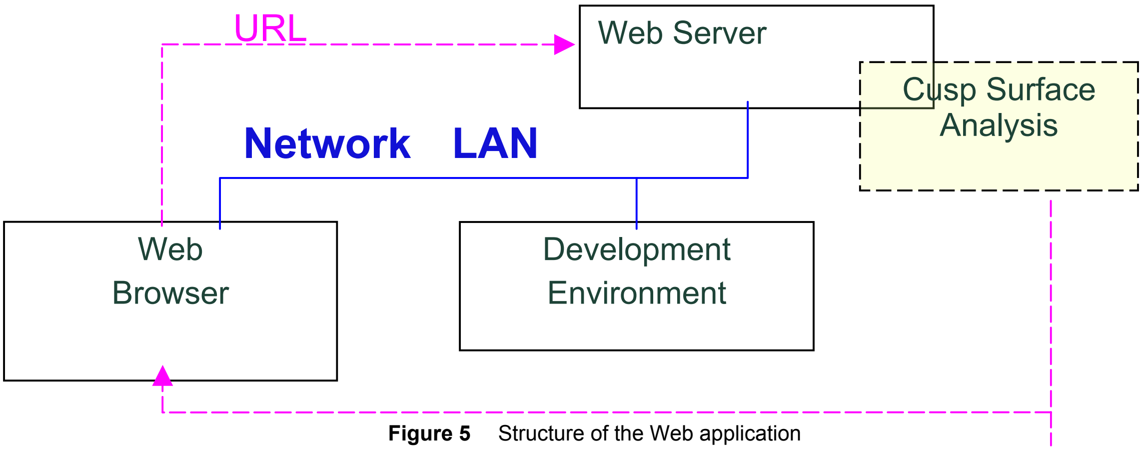


Figure 5 Structure of the Web application

Figure 5 above shows interchange of information between Web application server and the client. Web browser indicates the Web application to client in **Figure 5.** A system developed cusp surface analysis program is deployed on Web server. The cusp surface analysis program is opened and performed by the URL in a Web page. When registered URL on Web server is appointed, The cusp surface analysis program is corresponding to it and carrying out as HTML in a Web browser via a Web server(* Deploy means installing of Web application to Web server collecting necessary files and compressing them to one file for Web application and in other words, arranging packaged the War file for Web application).

The cusp surface analysis system is performing calculation of the cusp surface analysis on a Web server by Linux, Apache, Tomcat, and JAVA servlet. Eclipse3.2 as tool of the developer is used in order to build cusp surface analysis system on Web. Apache is not only used as Web server software, but Tomcat software is also used for processing JAVA servlet and JAVA Server Pages(JSP). Tomcat is able to perform independently as Web server. However, we make it plug-in with Apache for improving capacity and security of processing as well as for more stable system. The client can simply use cusp surface analysis system through Web browser without regarding any complicated environment. Interchanging information on Web server is performed by HTTP protocol. The client demands using cusp surface analysis system with HTTP. Web server responses the result with HTTP demanded by client. Client, that is, web browser shows the result with HTML transformed by HTTP. It is cralified that the client can save performed data to both Web server and local directory. And opening saved data from both Web server and local directory is also available.

4. APPLICATION OF CUSP SURFACE AALYSIS TO JUMPING PHENOMENON

Twenty-one kinds of figures from A to U (A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U) that change from Man's face to woman figure linearly are used and experimental data are obtained. Figure6 is picked out A, F, K, Q, U from used figures. K is Fisher's Figure and A is the figure seen man's face and U is the figure seen woman's figure.

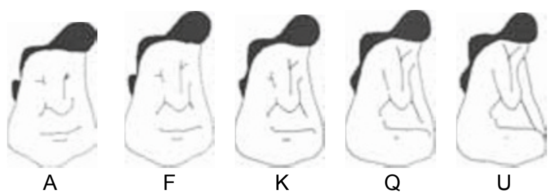


Figure 6 ambiguous figures

These sizes of figures (42.0 length, 29.5 width) are used. Subjects look at these figures from 2m to 30m at 2m intervals in fifteen ways. Independent variables X_1,X_2 are the kind of figure, the distance between figures and subject respectively. Then, X_1 quantities as A=1, B=2 , ▪▪▪, U=21. Dependent variable Y is human visual recognition and to quantify the judgment of man or woman, man's face is 0.0, woman' figure 1.0 and yes or no is 0.5. Used experimental data is shown in Table. Where combination of X_1 and X_2 is at random and experiments is performed 126 times. Using this data, cusp surface analysis is performed. Figure 7 shows view of experiments for smbiguous figures,

5. Judgment of Visual Recognition for Ambiguous Figure

Next, it is investigated by questionnaire that ambiguous figures are judged on man's face or woman's figure by putting stress on any part of ambiguous figure shown in Figure 8 After subject look at panel of drawing Fisher's figure K as shown in Figure 6 at distance 2m from the figure, he or she answer which part put stress in this figure. Size of panel is used width 29.5cm and length 42.0cm. Result of questionnaire is obtained. From this result, it is clarified that a person that regard ambiguous figure as man's face selects the part of black hair encode by thick line as shown in Figure 8. Then, it seems that there are many people who consider who look at hair in figure man's face.

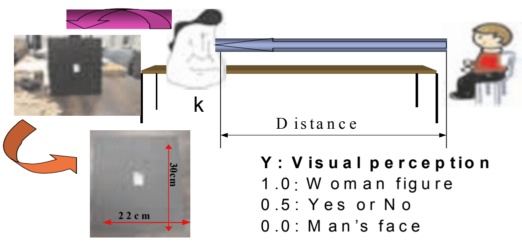


Figure 7 View of experimemts for distance to ambiguos figure

6. ANALYTICAL RESULT

In the result of estimating parameter from data of experimental results, it is as follows.

$$0 = A(\underline{X}) + B(\underline{X})[Y - C(\underline{X})] - 12.705[Y - C(\underline{X})]^3 \tag{3}$$

$$\begin{aligned} A(\underline{X}) &= -0.104 + 0.159X_1 + 0.835X_2 \\ B(\underline{X}) &= 13.818 - 2.506X_1 - 0.601X_2 \\ C(\underline{X}) &= 0.09 - 0.029X_1 - 0.019X_2 \end{aligned} \tag{4}$$

Table 1: Experimental results

Case	X_1	X_2	Y	Case	X_1	X_2	Y	Case	X_1	X_2	Y
1	3	2.0	0.0	**22**	2	2.0	0.0	**43**	1	4.0	0.0
2	15	30.0	0.0	**23**	10	30.0	1.0	**44**	11	24.0	0.5
3	7	8.0	0.0	**24**	19	18.0	1.0	**45**	18	12.0	1.0
4	8	16.0	0.0	**25**	21	8.0	1.0	**46**	8	30.0	0.5
5	1	24.0	0.0	**26**	5	16.0	0.0	**47**	3	28.0	0.0
6	9	28.0	0.0	**27**	13	10.0	1.0	**48**	15	22.0	1.0
7	11	4.0	0.0	**28**	15	24.0	1.0	**49**	10	6.0	0.0
8	12	14.0	0.0	**29**	1	26.0	0.0	**50**	12	14.0	1.0
9	18	22.0	1.0	**30**	14	4.0	1.0	**51**	13	10.0	1.0
10	5	6.0	0.0	**31**	4	22.0	0.0	**52**	7	2.0	0.0
11	10	18.0	0.0	**32**	8	12.0	0.5	**53**	6	18.0	1.0
12	21	20.0	1.0	**33**	20	14.0	1.0	**54**	20	16.0	1.0
13	20	26.0	1.0	**34**	16	20.0	1.0	**55**	14	20.0	1.0
14	14	12.0	0.0	**35**	11	28.0	1.0	**56**	9	26.0	0.5
15	2	10.0	0.0	**36**	17	6.0	1.0	**57**	16	8.0	1.0
16	17	12.0	1.0	**37**	12	26.0	1.0	**58**	5	26.0	1.0
17	4	28.0	0.0	**38**	9	12.0	0.0	**59**	4	20.0	0.0
18	13	2.0	0.0	**39**	3	18.0	0.0	**60**	21	4.0	1.0
19	16	18.0	0.0	**40**	6	16.0	0.0	**61**	17	6.0	0.0
20	19	6.0	1.0	**41**	7	22.0	0.0	**62**	2	22.0	1.0
21	6	8.0	0.0	**42**	18	14.0	1.0	**63**	19	8.0	0.5
Case	X_1	X_2	Y	Case	X_1	X_2	Y	Case	X_1	X_2	Y
64	1	2.0	0.0	**85**	1	6.0	0.0	**106**	4	2.0	0.0
65	13	30.0	0.5	**86**	9	30.0	0.0	**107**	16	28.0	0.5
66	11	16.0	1.0	**87**	5	10.0	1.0	**108**	11	26.0	0.0
67	18	24.0	1.0	**88**	14	14.0	0.0	**109**	10	4.0	0.0
68	4	26.0	0.5	**89**	13	4.0	1.0	**110**	12	30.0	0.5
69	21	6.0	1.0	**90**	7	16.0	0.0	**111**	3	6.0	0.0
70	16	28.0	0.5	**91**	4	20.0	0.0	**112**	19	14.0	1.0
71	3	22.0	0.5	**92**	18	2.0	1.0	**113**	17	16.0	1.0
72	10	12.0	0.0	**93**	12	24.0	0.5	**114**	20	10.0	1.0
73	12	18.0	1.0	**94**	10	26.0	0.5	**115**	5	12.0	0.0
74	5	14.0	0.0	**95**	8	28.0	0.5	**116**	9	22.0	0.5
75	20	10.0	1.0	**96**	20	22.0	0.0	**117**	18	18.0	0.5
76	15	20.0	1.0	**97**	16	18.0	0.0	**118**	13	20.0	0.0
77	6	8.0	0.0	**98**	6	12.0	1.0	**119**	2	24.0	0.0
78	17	4.0	1.0	**99**	17	8.0	1.0	**120**	14	8.0	0.5
79	2	4.0	0.0	**100**	21	26.0	0.5	**121**	6	12.0	0.0
80	7	30.0	0.5	**101**	15	8.0	1.0	**122**	1	20.0	0.0
81	9	16.0	0.0	**102**	2	14.0	0.0	**123**	8	10.0	0.0
82	8	2.0	0.0	**103**	3	30.0	0.5	**124**	21	22.0	0.5
83	14	14.0	1.0	**104**	11	12.0	1.0	**125**	7	2.0	0.0
84	19	8.0	1.0	**105**	19	6.0	1.0	**126**	15	4.0	0.0

From this estimating result, cusp model is drawn. It is shown in Figure 9. In this figure, distance X_2 is constant from 1 to 7 (2, 10, 20, 30, 60, 65 m) and cross section of visual recognition is shown. From this estimating result , cusp model can be drawn. From 1 to 7 (2, 10, 20, 30, 60, 65 m) in Figure 9 distance X_2, kind of figure X_1 and cross section of visual recognition, Y is shown. Figure 9 is cross section plane that cut off at Y-X_2 plane and probability density function. Also, in Figure 9, upper and lower lines are mode and central line is antimode. As the data used at this time is X_2=2～30m range, figure at X_2=60～70m are expected data. Figure 9(a) is cross section of Figure 8 cut off at Y-X_2 plane and Figure 9(b) is probability density function . Also, in Figure 9(a) and Figure 9(b), upper and lower lines show mode and central line shows antimode. In data used at this time, as X_2 is rannge from 2 to 30m, X_2 is expected figure of range from 60 to70m. From Figure 9 there is no antimode at X_2=65. Also, growing up X_2, curve converge Y=0.5. This figure shows that anyone cannot recognize the figure growing up distance. Next, when the kind of figure X_1 is constant, it is investigated. When the kind of figure X_1 is constant, cross section of relationship between distance of figure X_2 and visual recognition is shown in from□to□(A, K, U) in Figure 10. Figure 10(a) is cross section of Figure 9 cut off at Y-X_1 plane in parallel and Figure 9(b) is its probability density function. Also, in Figure 10, upper and lower lines show mode and central line shows antimode. When recognition for figure X_1=1(A) is most near to man's face, the probability of recognition for man's face is high and when recognition for figure X_1=21(U) is most near to woman's figure, the probability of recognition for woman's figure is high are understood by probability density function. Also, in ambiguous figure X_1=11(K), it seems that probability of man is comparable with probability of woman. But in this experiment, the probability of recognition for woman's figure is high. When recognition for figure X_1=1(A) is most near to man's face, the probability of recognition for man's face is high and when recognition for figure X_1=21(U) is most near to woman's figure, the probability of recognition for woman's figure is high are understood by probability density function. Also, in Fisher's figure X_1=11(K), it seems that probability of man is comparable with probability of woman. But in this experiment, the probability of recognition for woman's figure is high.

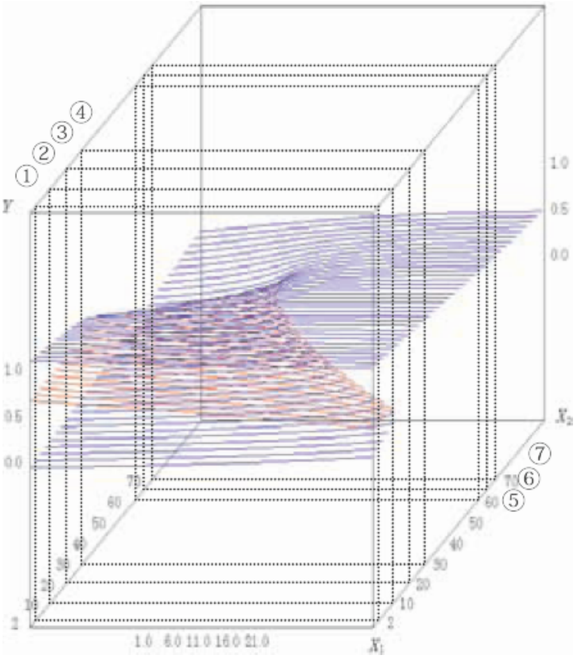
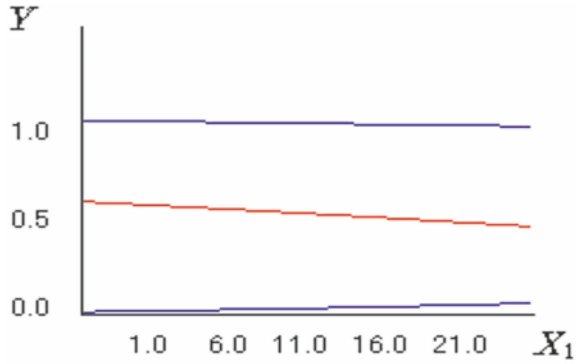
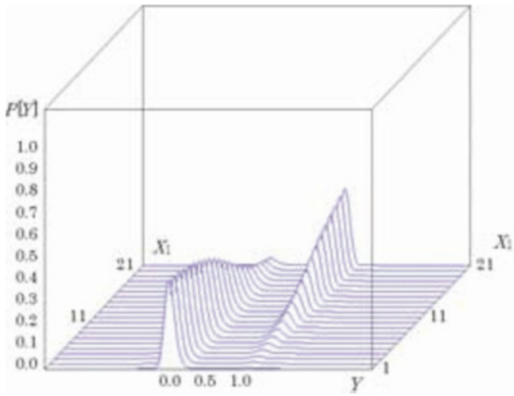


Figure 8: A cusp surface analysis result and a section of Ambiguous figures

Growing up distance between subject and figure, pointed out figure cannot be recognized clearly. In X_1=11(ambiguous figure, K), It seems that man's face and woman's figure is recognized at equal probability. But probability that is recognized to woman's figure is high. It seems that any person recognize man's face at part of Hair of Fisher's figure. As psychological data is included, it is disperse that subject look at which part of ambiguous figure. From this questionnaire, it is understood that it has to consider it. In the case of probability density function X_1=1(man's face, A), probability that recognize as man's face is high. In the case of X_1=21 (woman's figure, U), probability that recognize as woman's figure is high. In X_1=11 (Fisher's figure, K), man's face and woman's figure are recognized at equal probability by data used this time. Is regard as man's face part of hair in Fisher's figure. As psychological data is included human factor, ambiguous figure occur subject.

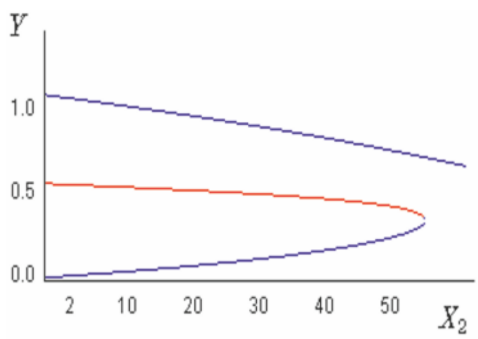


(a) Y-X_1 cross section of Figure 3（X_2=2)

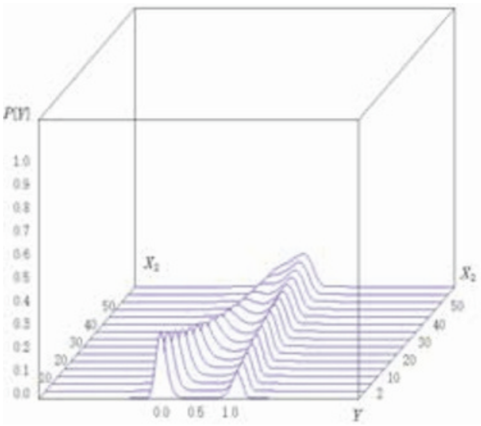


(b) Probability Density Function (Y)

Figure 9: Y-X_1 cross section of Fig.7 and PDF(X_2=2)



(a) Y-X_1 cross section of Figure 7（X_1=11)



(b) Probability Density Function (*Y*)

Figure 10: *Y*-X_1 cross section of Fig.2 and PDF(X_1=11)

7. Conclusion

In data of ambiguous figure, catastrophe phenomenon takes place and a person that look at the part of hair in figure regard as Man's face. It seems that it is clarified that the judgment of ambiguous figure is changed by distance from figure. That is lengthening distance to figure, recognition of figure becomes difficult. But there is case of cusp model or case of linear model. This reason is that is tangled by shape of material and it is fall down on the way.

8. References

[1]Kume,Y.,2004,Three-Dimensional Representation for Cusp Surface Analysis - Application of Cusp Surface Analysis to Analysis of TEEMA's Questionnaire - Proc.of 8th International Conference on Mnufacturing & Management,1105-1112.

[2]Kume,Y.;Okada,K, 2005, Management of Creative Process Using Cusp Surface Analysis Proc. of 11th HCII 2005,CD edition.

[3] Thom,R.,1975, Structural Stability and Morphogenesis.

[4] Loren, C., 1978, Stochastic Catastrophe Models and Multimodal Distributions, Behav. Sci., 23, 360-374.

[5] Loren, C., 1988, Cusp Surface analysis User's Guide.

[6] Fisher, G.H., 1967, Preparation of Ambiguous Stimulus Materials, Percept. Psychopys., 2, 421-432.

[7] Attneave, F., 1971, Multistability in Perception, Sci, Am., 225, 62-71.

[8] Poston, T. ; Stewart, I., 1978, Nonlinear Modeling of Multistable Perception, Behav. Sci., 23, 318-334.

.

Authors' Index