The 21st International Conference on
Artificial Reality and Telexistence

ICAT 2011

November 28-30, 2011

Osaka University, Osaka, Japan

ISSN: 1345-1278

THE VIRTUAL REALITY SOCIETY OF JAPAN

Proceedings of
The 21st International Conference on
Artificial Reality and Telexistence

ICAT2011

November 28-30, 2011
Osaka University, Osaka, Japan

Sponsored by:
The Virtual Reality Society of Japan (VRSJ)

Copyright and Disclaimer

ISSN: 1345-1278

The Virtual Reality Society of Japan
Yamakoshi Bld. 301
2-28-3 Hongo, Bunkyo-ku, Tokyo, 113-0033 Japan
TEL: +81-3-5640-8777
FAX: +81-3-5840-8766
office@vrsj.org

Message from the General Chairs and the Program Chairs

We would like to welcome you to the 21st International Conference on Artificial Reality and Telexistence (ICAT), ICAT 2011. ICAT is the oldest international conference on Virtual Reality and Telexistence. We are pleased to host ICAT 2011 in Osaka, one of the ancient capitals in Japan. Since the 7th century, Osaka has been a gathering place and a flourishing economic center. We hope you enjoy your stay in Osaka.

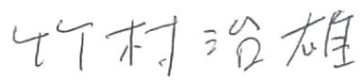
We are very honored to welcome four distinguished keynote speakers: Prof. Woontack Woo from Gwangju Institute of Science and Technology, Prof. Hiroshi Ishiguro from Osaka University and ATR, Prof. Henry Fuchs from University of North Carolina, Chapel Hill, and Prof. Taro Maeda from Osaka University.

This year, the conference received 38 paper submissions. Each submission has been reviewed by at least three experts on the International Program Committee which holds 40 experts from all over the world. Through this rigorous review process, we accepted 20 papers (52.6%). 12 paper submissions were accepted as a poster, and 6 submissions were rejected. We also received 15 poster submissions and 17 demo submissions. Through the review process, 21 posters and 17 demos were accepted.

The conference this year has papers from a wide spectrum of topics. These topics include: multi-modal displays, augmented reality, collaboration and telepresence, and medical systems. We are also delighted to have an organized session on "parasitic humanoid", which includes two keynote talks and four oral presentations. Note that the four papers for these presentations were equally treated as regular paper submissions and are included in the abovementioned statistics.

We gratefully acknowledge the large amount of time and efforts that the organizing committee and the reviewers invested to this conference. It is due to this volunteer efforts that have made ICAT 2011 possible. Now it is your turn, to share ideas, disseminate results, meet old and new friends, and expand our community and push it forward. Enjoy!

General Chair



Haruo Takemura
Osaka University

General Vice Chair

Hirokazu Kato
Nara Institute of Science and Technology

Program Chairs

Kiyoshi Kiyokawa
Osaka University

Torsten Kuhlen
RWTH Aachen University

Dirk Reiners
University of Louisiana

ICAT2011 Organizing Committee

General Chair:	Haruo Takemura (Osaka University)
General Vice Chair:	Hirokazu Kato (Nara Institute of Science and Technology)
Program Co-Chairs:	Kiyoshi Kiyokawa (Osaka University), Torsten Kuhlen (RWTH Aachen University), Dirk Reiners (University of Louisiana)
Publicity Co-Chairs:	Yuichi Itoh (Osaka University), Ehud Sharlin (University of Calgary), Sriram Subramanian (University of Bristol), Daisuke Sakamoto (University of Tokyo)
Local Arrangement Co-Chairs:	Hideyuki Ando (Osaka University), Daisuke Iwai (Osaka University)
Publication Chair:	Masataka Imura (Osaka University)
Financial Chair:	Hirotake Ishii (Kyoto University)
Posters Chair:	Nobuchika Sakata (Osaka University)
Demos Chair:	Masayuki Kanbara (Nara Institute of Science and Technology)

ICAT2011 International Steering Committee

Susumu Tachi (Keio University, Japan)
Michitaka Hirose (University of Tokyo, Japan)
Ming Ouhyoung (National Taiwan University, Taiwan)
Hyun Seung Yang (KAIST, Korea)
Mark Billinghurst (HIT Lab NZ., University of Canterbury, NZ)
Haruo Takemura (Osaka University, Japan)
Zhigeng Pan (Zhejiang University, China)
Tony Brooks (Aalborg University Esbjerg (AAUE), Denmark)
Yasushi Ikei (Tokyo Metropolitan University, Japan)
Hideo Saito (Keio University, Japan)
Sabine Coquillart (INRIA, France)
Yoshifumi Kitamura (Tohoku University)
Bruce H. Thomas (Univ. South Australia, Australia)
Hirokazu Kato (Nara Institute of Science and Technology, Japan)

ICAT2011 Program Committee

Sadagic Amela
Carlos Andujar
Mark Billinghurst
Seungmoon Choi
Henry Duh
Pablo Figueroa
Bernd Froehlich
Watanabe Junji
Hiroyuki Kajimoto
Yoshinari Kameda
Ichiroh Kanaya
Itaru Kitahara
Ernst Kruijff
Tomohiro Kuroda
Marc Erich Latoschik
Rob Lindeman
Mark Livingston
Kazunori Miyata
Guillaume Moreau
Staadt Oliver
Bruno Raffin
David Roberts
Jeha Ryu
Coquillart Sabine
Hideo Saito
Daisuke Sakamoto
Myriam Servières
Fumihisa Shibata
Masanori Sugimoto
Maki Sugimoto
Tsutomu Terada
Bruce Thomas
Masashi Toda
Daniel Wagner
Woontack Woo
Ruigang Yang
Hiroaki Yano
Yanagida Yasuyuki
Gabriel Zachmann
Nelson Zagalo

External Reviewers

Markus Broecker
Tobias Duckworth
Gabjong Han
Scinob Kuroki
In Lee
Stephan Ohl
Gunhyuk Park
Shane Porter
Aaron Toney
James Walsh
Malte Willert

Sponsors' List

ICAT2011 is supported and sponsored by:

The Virtual Reality Society of Japan

and is supported by:

- Tateisi Science and Technology Foundation
- Kayamori Foundation of Informational Science Advancement

Table of Contents

Keynotes

Augmented Reality & DigiLog: Toward Ubiquitous Virtual Reality 2.0 **1**
Woontack Woo

Robots, Humans, and Media **2**
Hiroshi Ishiguro

Toward Improved 3D Telepresence **3**
Henry Fuchs

Immersive Tele-Collaboration with Parasitic Humanoid: How to Assist Behavior Directly in Mutual Telepresence **4**
Taro Maeda, Hideyuki Ando, Hiroyuki Iizuka, Tomoko Yonemura, Daisuke Kondo and Yuki Hashimoto

Papers

Session 1: Emerging Hardware

Fingertip Slip Illusion with an Electrocutaneous Display **10**
Hiroyuki Okabe, Shogo Fukushima, Michi Sato and Hiroyuki Kajimoto

Improvement of Olfactory Display Using Electroosmotic Pumps and a SAW Device for VR Application **15**
Yossiri Ariyakul and Takamichi Nakamoto

Subjective Image Quality Assessment of a Wide-View Head Mounted Projective Display with a Semi-Transparent Retro-Reflective Screen **22**
Duc Nguyen Van, Tomohiro Mashita, Kiyoshi Kiyokawa and Haruo Takemura

Session 2: Outdoor Augmented Reality

An Evaluation of Augmented Reality X-Ray Vision for Outdoor Navigation **28**
Arindam Dey, Graeme Jarvis, Christian Sandor, Ariawan Kusumo Wibowo and Ville-Veikko Mattila

Multiple Camera Augmented Viewport: An Investigation of Camera Position, Visualizations, and the Effects of Sensor Errors and Head Movement **33**
Thuong N. Hoang and Bruce H. Thomas

A User Study on Viewpoint Manipulation Methods for Diorama-Based Interface Utilizing Mobile Device Pose in Outdoor Environment **41**
Masayuki Hayashi, Itaru Kitahara, Yoshinari Kameda and Yuichi Ohta

Session 3: Collaboration / Telepresence

Augmented Fly-Through Using Shared Geographical Data **47**
Sandy Martedi and Hideo Saito

PAC-C3D: A New Software Architectural Model for Designing 3D Collaborative Virtual Environments **53**
Thierry Duval and Cédric Fleury

TouchMe: An Augmented Reality Based Remote Robot Manipulation **61**
Sunao Hashimoto, Akihiko Ishida, Masahiko Inami and Takeo Igarashi

A First Look at a Telepresence System with Room-Sized Real-Time 3D Capture and Large Tracked Display Wall **67**
Andrew Maimone and Henry Fuchs

Session 4: Medical Systems

Augmented Reality Enhanced Image-Guided Surgery System Using CT and Ultrasound Registration for Brain-Shift Estimation **73**
Wei-Chic Huang, Chung-Hung Hsieh, Chung-Hsien Huang, Shin-Tseng Lee, Chieh-Tsai Wu, Yung-Nien Sun, Yu-Te Wu and Jiann-Der Lee

3-Dimensional Visual Navigation for Repetitive Transcranial Magnetic Stimulation Treatment **78**
Yoshihiro Yasumuro, Tatsuya Ogino, Masahiko Fuyuki, Atsushi Nishikawa, Masaki Sekino, Taiga Matsuzaki, Kouichi Hosomi and Youichi Saitoh

Interactive Cutting Method for Physics-Based Electrosurgery Simulation **83**
Yoshihiro Kuroda, Shota Tanaka, Masataka Imura and Osamu Oshiro

Organized Session: Parasitic Humanoid

Controller Reduction for Pseudo-Reference in High-Degree of Freedom Control System **88**
Masaaki Watanabe, Masafumi Okada and Dong Dung Ngyuen

Modeling of Pedestrian's Unintentional Guide Using Vection and Body Sway **94**
Norifumi Watanabe, Hiroaki Mikado and Takashi Omori

A Pedestrian Dynamics Simulator for Wearable Navigation **98**
Kosuke Shinoda, Itsuki Noda and Eimei Oyama

Improvement of Wearable View Sharing System for Skill Training **104**
Yuki Hashimoto, Daisuke Kondo, Tomoko Yonemura, Hiroyuki Iizuka, Hideyuki Ando and Taro Maeda

Session 5: Consistent Augmented Reality

Adaptive Substrate for Enhanced Spatial Augmented Reality Contrast and Resolution **110**
Markus Broecker, Ross T. Smith and Bruce H. Thomas

Occlusion Handling and Image-Based Lighting using Sliced Images in 3D Photo Collections **118**
Frank Nagl, Konrad Kölzer and Paul Grimm

Depth-Assisted Real-Time 3D Object Detection for Augmented Reality **126**
Wonwoo Lee, Nohyoung Park and Woontack Woo

Posters

Remote Ikebana with Olfactory and Haptic Media in Virtual Environments **133**
Pingguo Huang, Yutaka Ishibashi, Norishige Fukushima and Shinji Sugawara

QoE Comparison of Competition Avoidance Methods for Management of Shared Object in Networked Real-Time Game with Haptic Media **134**
Yuji Kusunose, Yutaka Ishibashi, Norishige Fukushima and Shinji Sugawara

Development of Inner Strings Haptic Interface SPIDAR-I **135**
Yan Zhu, Tatsuya Koyama, Tatsuro Igarashi, Katsuhito Akahane and Makoto Sato

The Effects of Using a Modified Motorcycle Simulator Training for the Spinal Cord Injury Patients **136**
Siao-Ying Wu, Wen-Hsu Sung, Yun-An Tsai, Henrich Cheng and Jin-Jong Chen

Feasibility Study on the Estimation of Photo Shoot Position and Direction Based on Virtualized Reality Environment Models **137**
Koji Makita, Jun Nishida, Tomoya Ishikawa, Takashi Okuma, Jun Yamashita, Hideaki Kuzuoka and Takeshi Kurata

Obstacle Sensation Augmented by Enhancing Low Frequency Component for Horror Game Sound **138**
Shuyang Zhao, Taku Hachisu, Asuka Ishii, Yuuki Kuniyasu and Hiroyuki Kajimoto

Intra-expo: Augmented Emotion By Superimposing Comic Book Images **139**
Sho Sakurai, Shigeo Yoshida, Takuji Narumi, Tomohiro Tanikawa and Michitaka Hirose

Pan-Tilt Projector Path Planning for Adaptive Resolution Display **140**
Kei Kodama, Daisuke Iwai and Kosuke Sato

Re-PITASu Concept: Touch-Based Interaction Using Range Image Sensor with Image Projected onto Wall Surface **141**
Yuki Uranishi, Goshiro Yamamoto, Hirokazu Kato and Petri Pulli

Localization with Microsoft Kinect using Natural Features and Depth Data **142**
Yuki Takabatake, Yuichi Tamura, Naoya Kashima and Tomohiro Umetani

Real-Time Diminished Reality using Multiple Smartphones **143**
Toshihiro Honda, Takuya Inoue and Hideo Saito

Adaptive Annotation Layout in Projection-Based Mixed Reality by Considering Its Readability **144**
Tatsunori Yabiki, Daisuke Iwai and Kosuke Sato

Model Based Tracking of Rigid Curved Objects using Sparse Polygonal Meshes **145**
Marina Atsumi Oikawa, Goshiro Yamamoto, Makoto Fujisawa, Toshiyuki Amano, Jun Miyazaki and Hirokazu Kato

AR based Co-located Meeting Support System **146**
Igor de Souza Almeida, Jun Miyazaki, Goshiro Yamamoto, Makoto Fujisawa, Toshiyuki Amano and Hirokazu Kato

SURF-Based Line Marker for Augmented Reality **147**
Hiroki Yoshinaga and Yoichi Muraoka

CG Image Generation of Developmental Origami Model of Hypercube **148**
Haruki Chiba, Keimei Kaino, Kuniaki Yajima and Takatoshi Suenaga

Data Adjustment Methods of a Low-Priced Data Glove **149**
Shinichi Hamaguchi, Sanshiro Yamamoto, Kenji Funahashi and Hidenori Kanazawa

Addition of 3D Sound Based on the Position and the Area of an Object in a Silent Video **150**
Miwa Nishimura, Tsuyoshi Kobayashi, Jun Murayama, Yukihiro Hirata, Makoto Sato and Tetsuya Harada

A Web Application for an Interior-Design Simulator using Augmented Reality **151**
Tomoki Tanaka, Takuma Nakabayashi, Keita Kado and Gakuhito Hirasawa

Landscape Simulation in Outdoor Settings using Stereoscopic Augmented Reality **152**
Takuma Nakabayashi, Keita Kado and Gakuhito Hirasawa

AR Whiteboard: Handling Written Contents as Digital Information Using Tools for Whiteboards **153**
Yuta Tsukada, Keita Ushida and Satoshi Tsurumi

Demos

Interactive Cardiovascular Editor Using Echocardiographic Images **154**
Megumi Nakao, Yuji Masuda, Ryo Haraguchi, Ken-Ichi Kurosaki, Koji Kagisaki, Isao Shiraishi, Kazuo Nakazawa and Kotaro Minato

Whirling Interfaces: Smartphones & Tablets as Spinnable Affordances **155**
Michael Cohen, Rasika Ranaweera, Hayato Ito, Shun Endo, Sascha Holesch and Julián Villegas

Collaboration between Heterogeneous 3D Viewers through a PAC-C3D Modeling of the Shared Virtual Environment **156**
Thierry Duval and Cédric Fleury

KINECTing Superheroes in MR Space: Matching Head-Tracking Coordinates and Gesture-Interaction Coordinates **157**
Masaki Oda, Le Van Nghia, Katsuyoshi Tomita, Asako Kimura, Fumihisa Shibata and Hideyuki Tamura

Direct Volume Manipulation for Navigating Liver Resection **158**
Yuya Oda, Megumi Nakao, Keiho Imanishi, Kojiro Taura and Kotaro Minato

Experiencing Shape-COG Illusion in Mixed-Reality Space **159**
Hiroki Omosako, Asako Kimura, Fumihisa Shibata and Hideyuki Tamura

Walk-in-Place Locomotion Interface using Footprint Images **160**
Hidetoshi Kiyofuji, Katsuhide Nagasaki, Jun Murayama and Tetsuya Harada

Skill Transmission by Using Parasitic Humanoid System **161**
Yuki Hashimoto, Daisuke Kondo, Tomoko Yonemura, Hiroyuki Iizuka, Hideyuki Ando and Taro Maeda

Interaction for Remote Collaboration with Tabletop System **162**
Keiji Uemura, Nobuchika Sakata and Shogo Nishida

An Indoor Navigation System using a Wide-View Head Mounted Projective Display with a Semi-Transparent Retro-Reflective Screen **163**
Duc Nguyen Van, Tomohiro Mashita, Kiyoshi Kiyokawa and Haruo Takemura

Owens Luis — A Proposal of a Smart Office Chair in an Ambient Environment **164**
Kiyoshi Kiyokawa, Masahide Hatanaka, Kazufumi Hosoda, Masashi Okada, Hironori Shigeta, Yasunori Ishihara, Fukuhito Ooshita, Hirotsugu Kakugawa, Satoshi Kurihara and Koichi Moriyama

High Dynamic Range 3D Display System with Projector and 3D Color Printer Output **165**
Saeko Shimazu, Daisuke Iwai and Kosuke Sato

Optically Hiding of Information with Polarized Complementary Projection **166**
Mariko Miki, Daisuke Iwai and Kosuke Sato

Interacive Cutting Simulation System of Physics-based Electrosurgery **167**
Yoshihiro Kuroda, Shota Tanaka, Ryosuke Yokohata, Masataka Imura and Osamu Oshiro

STELET Display: Tactile Augmentation with Handheld Tool **168**
Shunsuke Yoshimoto, Naritoshi Matsuzaki, Yoshihiro Kuroda, Masataka Imura and Osamu Oshiro

A Compact Pseudo-Force Glove Using Shape Memory Alloys **169**
Yu Shigeta, Yoshihiro Kuroda, Masataka Imura and Osamu Oshiro

Multi-Viewpoint Interactive Fog Display **170**
Masataka Imura, Asuka Yagi, Yoshihiro Kuroda and Osamu Oshiro

Augmented Reality & DigiLog: Toward Ubiquitous Virtual Reality 2.0

Woontack Woo

Gwangju Institute of Science and Technology

Abstract

In this talk, I will introduce a new concept of "ubiquitous Virtual Reality (UVR)" in the view point of Metaverse and then explain how to realize Virtual Reality in physical space with context-aware Augmented Reality. In UVR-enabled space it is possible to personalize using user's, as well as environmental, context and then selectively share the augmented object with additional (or 3D content as well as text) information according to user's social relationships. I will also explain some core technologies developed in GIST U-VR Lab for last 5 years and demonstrate U-VR applications such as DigiLog Book, Digilog Miniature, CAMAR Tour, etc.

Biography

Prof. Woontack Woo received his BS in Electronics Engineering from Kyungpook National University in 1989 and his MS in Electronics and Electrical Engineering from Pohang University of Science and Technology (POSTECH) in 1991. In 1998, he received his Ph.D. in Electrical Engineering-Systems from University of Southern California (USC), CA, USA. In 1999, as an invited Researcher, he joined Advanced Telecommunications Research (ATR), Kyoto, Japan. Since Feb. 2001, he has been with the Gwangju Institute of Science and Technology (GIST), where he is a Professor in the Dept. of Information and Communications (DIC) and Director of Culture Technology Institute (CTI). The main thrust of his research has been implementing ubiquitous virtual reality in smart space, which include Context-aware Augmented Reality, 3D Vision, HCI, and Culture Technology.

Robots, Humans, and Media

Hiroshi Ishiguro

Osaka University and ATR

Abstract

Studies on interactive robots and androids are not just in robotics but they are also closely coupled in cognitive science and neuroscience. It is a research area for investigating fundamental issues of interface and media technology. This talks introduce the series of androids developed in both Osaka University and ATR and propose a new information medium realized based on the studies.

Biography

Hiroshi Ishiguro (M') received a D.Eng. in systems engineering from the Osaka University, Japan in 1991. He is currently Professor of Department of Systems Innovation in the Graduate School of Engineering Science at Osaka University (2009-) and Group Leader (2011-) of Hiroshi Ishiguro Laboratory at the Advanced Telecommunications Research Institute (ATR). His research interests include distributed sensor systems, interactive robotics, and android science. He has published more than 300 papers in major journals and conferences, such as Robotics Research and IEEE PAMI. On the other hand, he has developed many humanoids and androids, called Robovie, Repliee, Geminoid, Telenoid, and Elfoid. These robots have been reported many times by major media, such as Discovery channel, NHK, and BBC. He has also received the best humanoid award four times in RoboCup. In 2007, Synectics Survey of Contemporary Genius 2007 has selected him as one of the top 100 geniuses alive in the world today.

Toward Improved 3D Telepresence

Henry Fuchs

University of North Carolina, Chapel Hill

Abstract

Telepresence dreams have been inspired for decades by science fiction holodecks, while teleconferencing developments have been inspired by dreams of much-reduced travel budgets and overcrowded classrooms. Meanwhile however, even high-end commercial teleconferencing systems are a far cry from these dreams; they restrict participants to fixed seats, and they inhibit such simple activities as walking around and writing on white boards, and fail to support even simple eye contact with remote participants. This talk with focus on recent developments both in meeting-room based systems, as well as mobile units that allow the remote participant to travel to non-instrumented meeting rooms, to laboratories, factories, hospitals, and clinics. Recent progress in room-based systems include multi-viewer autostereo displays that provide each user with a distinct view, to give an illusion of a window into a remote 3D environment. Recent progress in 3D acquisition of these remote environments include real-time acquisition and reconstruction based on multiple consumer-priced depth-plus-color cameras. To improve mobile telepresence, our UNC team, lead by Professor Greg Welch, has been developing a mobile physical-virtual avatar with a life-size moving human mannequin which takes on the dynamic appearance and mimics the head pose of its remote human "inhabiter." The talk will conclude with speculation about the direction of future progress in these areas, toward systems appropriate for individual offices and homes.

Biography

Henry Fuchs (PhD, Utah 1975) is the Federico Gil Distinguished Professor of Computer Science and Adjunct Professor of Biomedical Engineering at the University of North Carolina at Chapel Hill. He is a member of the National Academy of Engineering, a fellow of the American Academy of Arts and Sciences, and a fellow of the ACM. He is recipient of the 1992 ACM SIGGRAPH Achievement Award, the 1992 Academic Award of the National Computer Graphics Association, and the 1997 Satava Award of the Medicine Meets Virtual Reality Conference. Fuchs has been active in computer graphics since the 1970s, coauthoring over 170 publications on rendering algorithms (BSP Trees), real-time hardware (PixelPlanes and PixelFlow), virtual environments, tele-immersion systems and medical applications. He is a member of the Steering Committee of the International Symposium on Mixed and Augmented Reality (ISMAR), External Advisory Board of Harvard's Neuroimage Analysis Center, Editorial Advisory Board of Computer & Graphics Journal, and the Editorial Advisory Board of International Journal of Virtual Reality. He has also served as a member of the National Research Council's Computer Science and Telecommunications Board and DARPA's Information Science and Technology Study Group (ISAT). He is co-founder and Special Technical Advisor of InnerOptic Technology, Inc.

Immersive tele-collaboration with Parasitic Humanoid: How to assist behavior directly in mutual telepresence

Taro MAEDA*1,*2, Hideyuki ANDO*1,*2, Hiroyuki IIZUKA*1,*2, Tomoko YONEMURA*1,*2,

Daisuke KONDO*1,*2, and Yuki Hashimoto*1,*2

*1. Graduate School of Information Science & Technology, Osaka University, *2. JST CREST

ABSTRACT

Telexistence allows you to control a robot intuitively without the need to learn special skills by exploiting the complete physical association between a controller and a robot. In this technology, sensory and motor information based on the robot's and user's embodiments is translated to digital data, which can be recorded, and transmitted to each other to share their experiences. The technology can be available not only for interactions between a robot and a human but also between humans. In realizing this system, how to share the motions and sensations of humans and how to adjust or revise their motions become important. In this paper, we propose an approach for sharing first-person perspectives. We have developed a view-sharing system to realize such interactions using the Parasitic Humanoid (PH). PH is a wearable robotic human interface for sampling, recording, and replaying the sensory and behavioral experiences of the wearer from the first person perspective. Connecting PH wearers can realize skill transmission, telecollaboration, and sharing of experiences between humans.

KEYWORDS: Embodiment, Motion Induction, Human Hack, Ability Extension.

INDEX TERMS: K.6.1 [Management of Computing and Information Systems]: Project and People Management—Life Cycle; K.7.m [The Computing Profession]: Miscellaneous—Ethics

1 INTRODUCTION

Telexistence is a technology that enables a human operator to feel that he or she exists where a robot exists, rather than where they actually exist when controlling the robot [1,2]. The feeling of presence makes it easier for operators to recognize the surrounding environment and to achieve a task with the robot because they feel as if they are actually working there instead of the robot. For telexistence technology, effectively communicating sensations and motions between an operator and a robot in real-time is important so that the operator feels present. The technology applies not only to robot-human interaction but also to human-human interaction between two distant people [3]. A human works as if he exists at the distant place by "possessing" a partner even if he does not actually exist there. In the same way as robot-human interaction, sensations, perceptions, and motions must be shared between humans. By capturing human behaviors, we elicit key factors to establish effective interactions that can be applied to robot-human interactions.

In the situation of ordinary telexistence in Fig.1(a), the purpose of the person A is to assist the robot B in the scene B. In the system shown in Fig1.(b), the purpose of the person A is also to assist the person B in the scene B. It is a kind of mutual human to human telepresence. It is an ideal method for directly assisting behavior.

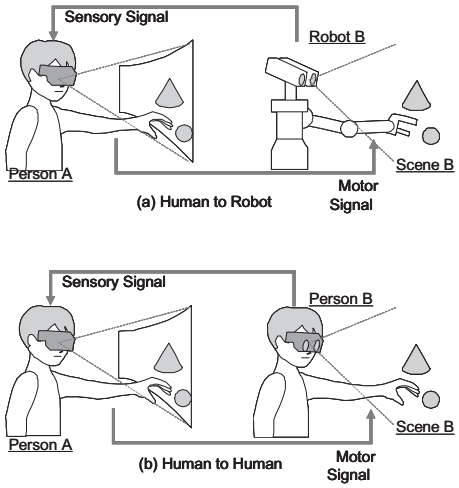


Figure 1. Mutual telexistence for tele-collaboration

To realize the system in Fig. 1(b), the person B should act the role of the robot B in Fig.1 (a). There are a few problems to be solved.

1. How to intercept Person B's sense and motion information.
2. How to synchronize Person B's behaviour to Person A's behaviour.
3. How to keep the sense of reality in the scene B for Persons A and B.

We propose solutions for these problems as follows.

The solution for 1 is Parasitic Humanoid technology.

The solution for 2 is the method of motion induction.

The solution for 3 is the view sharing system with image stabilization technique.

The view sharing system is a system for mutual telexistence. It transmits the information of senses and behaviors between Parasitic Humanoids optimized to make the wearers share their experiences in their first person perspectives.

2 MUTUAL TELEPRESENCE FOR IMMERSIVE TELE-COLLABORATION

In conventional studies to establish human-human collaborative interactions between two distant people, the feeling of presence is

2-1, Yamada-oka, Suita, Osaka 565-0871, JAPAN
{t_maeda, hide, iizuka, yonemura, kondo, y.hashimoto}@ist.osaka-u.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

not seriously considered for the following two reasons. First, most studies only pay attention to how visual information should be presented or shared without considering the viewpoints. For example, to simplify collaboration and increase effectiveness, Kuzuoka et al. [4] and Fussel et al. [5] developed similar shared-view video support systems where a worker's view is captured by a head-mounted camera, but the captured images are displayed to an instructor on a monitor on a desk. The instructor can see the worker's view, but the perspectives are not consistent.

The inconsistent perspectives require the user (instructor) to consciously interpret from the displayed images such situations as the worker's movements and the spatial relation among objects and the worker. On the other hand, a consistent perspective induces an immersive experience and the feeling of presence, and users can unconsciously understand the surrounding environments.

If an immersive environment has been established to display images, it can produce such illusions of self-sense as self-ownership [6], body-swapping [7], and pseudo-haptics [8] by artificially manipulating visual and tactile sensations. In the rubber hand illusion, a rubber hand's incongruent directions suppress the induction of the rubber hand illusions [6], which suggest that it can be seen as the user's body from different perspectives. This observation shows that, for users, perspective is crucial for inducing the illusion of feeling their own bodies. Another reason is that previously studied collaborations have been collaborative work between workers and instructors. The worker tries to achieve a goal with an instructor's help, meaning that the worker always needs to understand the instructor's behaviors as instructions. For example, the mutual view sharing system proposed by Nishikawa et al. provides consistent perspectives, but the worker and the instructor perform different tasks, such as achieving a task and instructing behavior to teach the worker procedures [9].

Consequently, the feeling of presence is never exploited in conventional ways. In the present work, we consider collaboration where the worker and instructor are simultaneously performing the same action in an immersive environment.

Mentally rehearsing movements is an effective way to acquire skills [10]. Even though mental rehearsing is purely imaginary, we hypothesize that if an immersive environment can be established between two humans and a non-skilled person can see a skilled person's movements as if there were her/his own movements, the skill can be effectively transmitted. Therefore, we share first-person perspectives between humans in which a non-skilled person easily recalls the body images or movements of skilled persons to acquire a skill or to achieve a task in real-time with help from skilled persons.

Below we tested our hypothesis. In the next section, we introduce a system developed for sharing first-person perspectives. Then we examine simple collaborated behavior in interaction between two subjects to clarify the fundamental effects of our view-sharing system. We exploited the collaboration findings and applied our view-sharing system to a concrete juggling skill and evaluated skill transmission performance. Finally, the results of our view-sharing and skill transmission are discussed.

3 Parasitic Humanoid: Wearable Robotics as Behavioral Interface

What is the information which assists an individual behavior? To answer the question, it is necessary to consider how a human perceives the world.. The world is filled not by information but phenomena. Only the act of measurement can convert the phenomena to information. Measurement is defined as the evaluation of a physical phenomenon according to some method and scale. For the information of human perception, the scale is defined by the scale of the human body. In human information technology, it is important to recognize the significance of scaling and measuring in relation to human embodiment. Most wearable computers today derive their usage from concepts in desktop computing such as data-browsing, key-typing, device-control, and operating graphical user interfaces. If wearable computers are anticipated to be fitted continuously as clothing, we must re-examine their use from the viewpoint of behavioral information. In this paper, we consider the role of wearable computers as a behavioral interface.

The Parasitic Humanoid (PH) is a wearable robot for modeling nonverbal human behavior [1]. This anthropomorphic robot senses the behavior of the wearer and has internal models to learn the process of human sensory motor integration, thereafter it begins to predict the next behavior of the wearer using the learned models. When the reliability of the prediction is sufficient, the PH outputs the difference from the real behavior as a request for motion to the wearer. Through this symbiotic interaction, the internal model and the process of human sensory motor integration approximate each other asymptotically. This process is available to transmit modalities such as senses of sight, hearing, touch, force and balance with human embodiment. This synergistic multimodal communication between distant people wearing PH can realize experience-sharing, skill transmission, and human behavior supports.

3.1 The Usage of Anthropomorphic Robots

Wearable computing and wearable robotics have separate histories. Most recent research on wearable robotics is motivated by interest in powered assist devices [2]. These devices are typically too heavy and consume too much energy for mobile use. On the other hand, research in mobile wearable robotics [3] does not take advantage of the embodiment of wearable devices.

Consider the usage of anthropomorphic robots as an interface for human behavior. This is perhaps the only pragmatic usage, because the anthropomorphic shape is usually disadvantaged compared to optimized designs for other purposes. One successful example is the Telexistence system [11]. However, such a robot is too complicated and expensive to be an interface of human behavior, and therefore too socially unacceptable to be considered as a pragmatic solution, except for some specific purpose, such as teleoperation of robots in a hazardous environment (although commercial systems are quickly driving down the cost of such devices). A more serious concern regards the safety for the users under common circumstances in modern life. There will be situations of disorder in which the control system of the robot continues to move although it has to stop to avert a collision. A solution in this situation is a lightweight and low power design such that a surveyor can easily prevent undesirable motion. However, this strategy makes it difficult for the robots to support themselves.

Wearable technologies supply a solution to the problem. Wearable sensory devices can construct a wearable humanoid without muscles and skeletons, if they are of the proper type and in sufficient number (Figure 1). This robot may be too weak to move by itself, and can not assist the wearer with mechanical power. However, it is safe and light for the wearer, and can assist his or her behavior with multi-modal stimulation, when the worn robot is continuously capturing, modeling and predicting the behavior of the wearer.

3.2 Parasitic Humanoid

We refer to such a wearable robot as Parasitic Humanoid (PH). PH is a wearable robot for modeling nonverbal human behavior.

This anthropomorphic wearable robot senses the behavior of the wearer and has internal models to learn the process of human sensory motor integration, thereafter it begins to predict the next behavior of the wearer using the learned models. When the reliability on the prediction is sufficient, the PH outputs the difference from the real behavior as a request for motion to the wearer. When the correction is not adequate, the wearer does not follow it. In this case, the PH corrects its internal model by the difference and continues learning. When the correction is adequate, the wearer follows the correction and corrects his internal behavioral model for himself. In this case, the PH does not correct its internal model and raises the estimation of the reliability to the model. Through this symbiotic interaction, the internal model and the process of human sensory motor integration approximate each other asymptotically. As a result, the PH acts as a symbiotic subject for information of the environment. The relationship is similar to oneness between horse and rider, although the role of the wearer corresponds to the horse, not following like a sheep. The symbiotic relationship between these partners acts as a high performance organism.

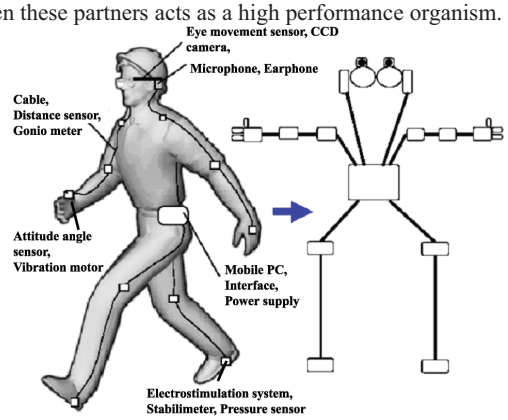


Figure 2. Wearable sensory devices construct a wearable humanoid without muscle

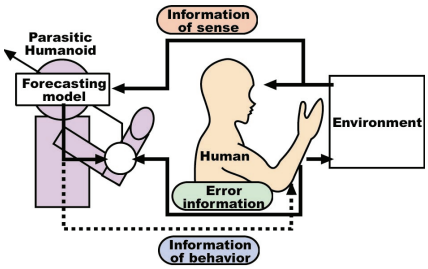


Figure 3. The symbiotic relationship between the wearer and the Parasitic Humanoid.

3.3 Capturing and Retrieving Sense and Behavior

The most direct usage of PH as a behavioral interface is capturing and retrieving behavior. It is not only for scientific research to analyze human behaviors, but also to enhance daily life. When you play golf, you may want to capture and retrieve your best shot. Or, you may download the data of the swing from the PH of Tiger Woods. You may exchange data for dance steps like name cards in a ballroom. You may share today's behavioural experiences with a distant person like telephone communication. Behavioral interfaces will make up a new style of communication.

Video see-through head-mounted displays are a common and useful method to record, transmit and reproduce visual information in human sight. When using a video see-through head-mounted display in telexistence or augmented reality, it is important to arrange the camera eye and human eye in conjugate in optics. However, earlier studies do not attempt to realize this. They align only the optical axis because of the difficulty of implementation. In this method, we may misidentify the distance to the object because of the difference of the angle of vergence. Moreover, the user's action may be obstructed because of the difference of motion parallax. In PH, we designed a video see-through head-mounted display in which camera eye and human eye are arranged in conjugate in optics. We investigate the influence given to depth perception and work performance when there is agreement or disagreement in viewpoint by using this system. This device is especially useful for skill transmission between PH wearers.

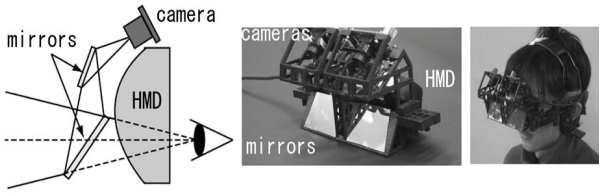


Figure 4. Zero Eye Offset and Orthoscopic Video See-Through Head-Mounted Display (VST-HMD)

4 How to Induce Human Motion: Behavioral Interface Devices using Illusion in Sensory Motor Process

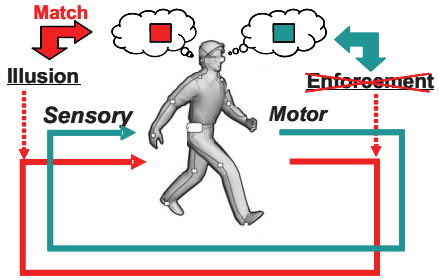


Figure 5. The concept to induce human behavior: Not by physically force but by sensory

In this system, when the wearer and the PH target the same behaviour, their targeted motions should be realized in a very similar manner. If the wearer knew the correct motion targeted by PH, he would achieve the motion naturally. What is the way to induce the motion in the wearer as if he knew the motion targeted by PH?

In the only neat way to induce human motion, the behavior might not be enforced physically but induced with sensory illusion. Human motion and sensory perception are not independent. These are integrated in a continuous relationship in neural processing as a loop connected between human embodiment and the outer world (Figure 5). In the first state, the subject has the motion plan A, and has the perception A. When outer force changes the motion from A to B physically, it also causes the change of the perception from A to B. The subject may become aware of the conflict between the motion plan A and the perception B. In contrast, when an illusion changes the perception from A to B, the subject may change the motion plan from A to B voluntarily. The latter case might be a better way to induce human behavior, avoiding a conflict to the free will of the subject.

We propose sensory motor interface devices using illusion for behavioral assist on Parasitic Humanoid system. These should

realize the synchronization of the wearer's motion with the target motion collaboratively. When the distant PHs share the sensory and motion data with telecommunication, the wearers should also share their behavioral experience like the mutual telexistence in Figure 1(b). In this situation, the collaborative persons would trace each other's behavior. It is an advantage

5 TELECOMMUNICATION NETWORK WITH PARASITIC HUMANOID: SYNERGISTIC MULTIMODAL COMMUNICATION IN COLLABORATIVE MULTIUSER WITH PARASITIC HUMANOID

PH is also an electro mechanical device. Therefore the sensory motor information of wearer measured by PH is available to be recorded, transmitted and reproduced as electrical information. When many PHs and their wearers connect to an information network with sensory motor information, a novel synergetic multimodal communication might be realized providing the oneness not only between PH and the wearer but also among collaborative users. Our aim to transmit modalities such as senses of sight, hearing, touch, force and balance is to realize experience-sharing, skill transmission, and human behavior supports between distant people. This system may create a situation such as if an expert exists right in front of injured people, instead of the cooperator who is actually there, through multimodality transmission by PH network. The skills of the doctor are communicated to the cooperator and their experiences are shared with each other. Such multimodal communication is non-verbal, immediate and intuitive. It might break the limits of learning skills through verbal communication.

5.1 VIEW-SHARING SYSTEM

Here, we assume the situation such as the scene of an accident as an example (Figure 6). Our concept is that the bystander performs the emergency treatment according to the expert's direction transmitted from another place. As shown in Fig.1, two users in different places wear the video-see-through HMD. The user on the left (a) is a doctor who can direct the emergency treatment, and the other (b) is a bystander who is a layperson, and executes treatment according to the expert's directions. We name user (a) "expert", and user (b) "bystander". The expert, who has knowledge of techniques for emergency treatment, is not on site, and uses the view sharing system to understand the condition of the injured person and the environment. The expert then directs the bystander. The bystander achieves cardio pulmonary resuscitation performing, for example, cardiac massage. The bystander can receive the method of cardiac massage; the posture, the position to place the hands, timing, etc., by sharing the view with expert. For the expert user, in order to provide the bystander user with precise direction, sharing their first person view is important. Because sharing their first person view helps the expert to understand the situation at the bystander's location with accuracy, such as the spatial relationship of the bystander and injured person, and furthermore notice what within arms' reach is required.

5.2 Sensory-Motor Collaboration by View Sharing System

As the prototype of the synergistic communication between collaborative users, we have a developed view-sharing system to share the first-person perspectives between remote two people [12]. The system consists of a head-mounted display and cameras, which realize a video-see-through system. The users wearing the video-see-through HMDs can see his own and the partner's view and also can send his own view to the partner. The sharing system has been applied to a skill transmission and learning task.

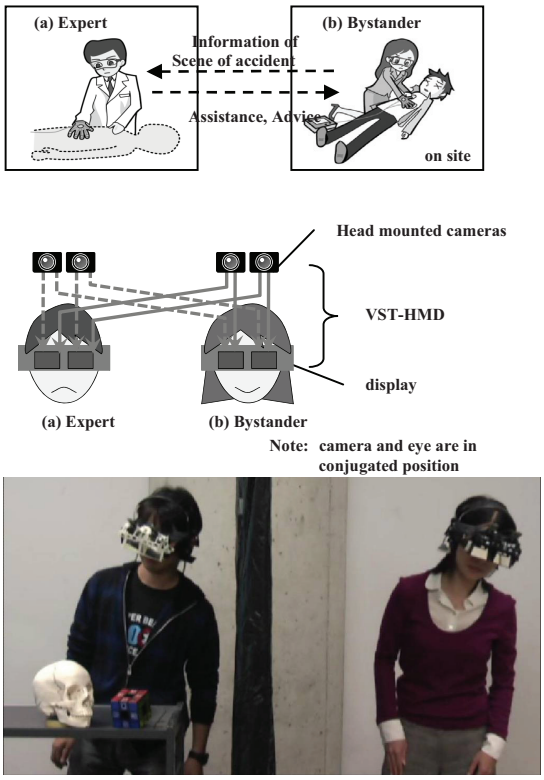


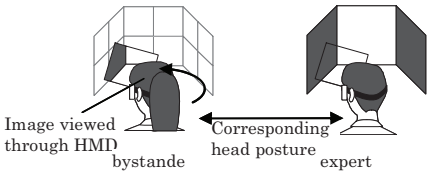
Figure 6. View sharing system.

However, there are two problems in the system that may make the performance of the system for sharing senses and motions less effective. One is that the displayed view swings regardless of the person's own motion because the view is captured on the basis of the partner's motion. The discrepancy between motion and visual information causes a breakdown of the user's embodiment, which makes it difficult to maintain proper spatial perception such asthe positional relationship between an object and the world, a recognition of posture with which the partner sees the view, and the relationship between his own and the other world (Figure 7). This problem can be overcome by giving both the sender's motion and visual information to the receiver to maintain the coherence between them in the receiver's situation. By realizing an interface device that supports the receiver to follow the sender's motion and that displays the flow of images in response to the receiver's actual motion, the embodiment or coherency between motion and visual information can be maintained and the proper spatial perception can be sent to the receiver.

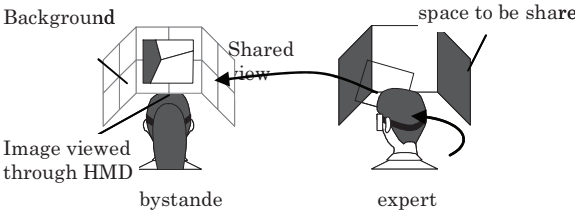
Another problem is a narrow view angle. Conventional video-see-through HMDs are not capable of sufficient view angles compared with the naked eye. It is shown that wider view angles improve the immersiveness of the view, and abilities of spatial perception and searching in a space [13][14]. A new design for video-see-through HMD with wider view angles resolves the lack of the visual information in the view-sharing system.

We propose a spatial stabilizer of flow of views on a video-see-through HMD with a wide view angle to overcome these two problems. The effect of the spatial stabilizer and view angles are tested and evaluated by the performance of communication of spatial perception from one to another. As shown in Figures 6 and 7, two users wear the video-see-through HMDs at the different places. The left user (a) is an expert who can direct the emergency treatment, and another (b) bystander is a layperson who actually executes it under the instruction of expert's direction. In order to

follow the other user's head motions, a tracking support using guide markers is introduced. In Figure 8, two markers are in front of the user, (A) is fixed in front of the user, indicating the posture of user's own head. Another one (B) is the target marker, and it shows the partner's current head posture. Therefore, the positional relationship between center marker (A) and target marker (B) reflects the positional relationship between the user and the partner. The user has to follow the target marker by moving their head.



(a) The expert moved and the image is send to the bystander's eyes.



(b) The bystander moves following the expert, and the bystander's orientation corresponds to expert's. It means the expert's direction is transmitted to the bystander.

Figure 7. Overview of view sharing and motion tracking.

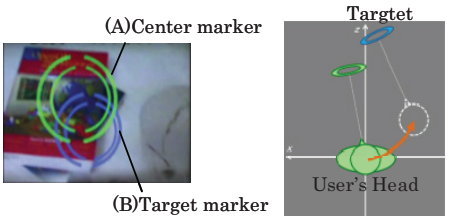


Figure 8. Positional relationships of guide markers.

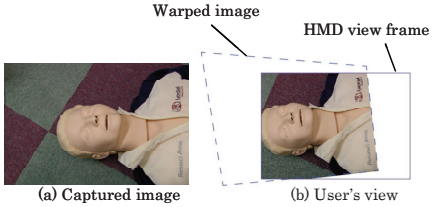


Figure 9. Method of image stabilization.

When both markers are aligned, the user and partner's motion are successfully corresponding. It is not easy for the user to track the marker in the partner's view, because when the users are moving, the image provided to the user does not correspond to their head motion. Ordinary, one's view of the background and its optical flow is an important cue of one's orientation during head motion. In this system, a background that does not correspond to the user's head motion makes the user feel disoriented or that the scene wobbles. As a result, the user's head movement is disturbed. The shaky images due to inconsistent motions between two must be eliminated.

5.3 Image Stabilization

If the user can track the other's vision with precision such that their view corresponds to the user's head motion, the user can recognize the other user's view as their own background naturally. However, perfect correspondence is difficult. The background image that does not correspond the user's head motion gives the user a disturbance during the task. The shaky images due to inconsistent motions between the two must be eliminated. In this section, the method for eliminating and stabilizing the background will be introduced. The stabilization technique is sometimes used in the field of tele-robotics [15] where the camera image attached on the mobile robot needs to be stabilized. We also used a stabilization technique in our system. The function, named "image stabilizer", transforms the image to generate the nearest image to the correct one. The function slides, warps and rotates the partner's camera image according to the relative position and orientation between the user and the partner in real time. The concept image is shown in Figure 9. We investigated the effect of our proposed stabilizer function and view angles in the spatial perception task sent by the partner. From the results of our experiments, it was shown that the view angle is crucial to perceive the positional relationships between objects in the partner's image flow and that the stabilizer function could compensate for the gap of ability of spatial perception between narrow and wide view angles (Figure 10)

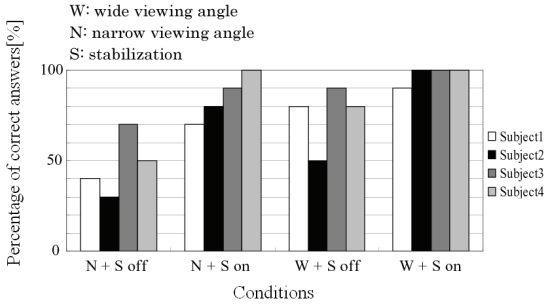


Figure 10. Performance of spatial cognition.

6 Motor Skill Transmission in View-Sharing System

Mentally rehearsing movements is an effective way to acquire skills [16]. Even though mental rehearsing is purely imaginary, we hypothesize that if an immersive environment can be established between two humans and a non-skilled person can see a skilled person's movements as if they were her/his own movements, the skill can be effectively transmitted. Therefore, we share first-person perspectives between humans in which a non-skilled person easily recalls the body images or movements of skilled persons to acquire a skill or to achieve a task in real-time with help from skilled persons. We tested our hypothesis.

For example, we exploited the collaboration findings and applied our view-sharing system to a concrete juggling skill and evaluated the skill transmission performance. Here, the skill transmission immediately improved a non-skilled person's performance with online PH assistance by an expert. In addition, skill transmission is different from skill acquisition. If the skill was acquired, the number of failures should also be decreased in the unassisted trial in Figure 11.

Figure 12 shows the theremin performance which is defined as the RMSE of tone pitches between the master and the beginner. The horizontal axis shows the trial number and the labels over the bar shows the condition of each trial (I : Individual condition, S: Side-by-side condition, V: View sharing condition). The

performances of the same conditions are connected with the lines. At the first trial, the beginner has to play the theremin alone without any knowledge after listening to the target song once. The performance of the first trial shows the original skill that the subject has. After that, the subject learns how to move their hand from the master in the side-by-side condition. The performances improve over the original ones. However, as shown in the figure, the subject can show the best performance with the view-sharing system although the difference becomes smaller through repeated learning since the individual skill of the subjects improves at the same time. However, it is obvious that the skill of the master is transmitted to the beginner through the view-sharing system. At the 3rd and 5th trial, where the condition is view-sharing, the beginner can already play the song as well as he plays at the end of the experiment. This result can be seen in another pair as well.

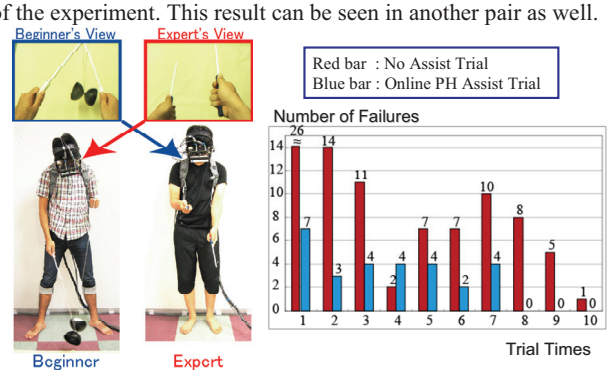


Figure 11. Performance of juggling skill transmission in Online PH Assist.

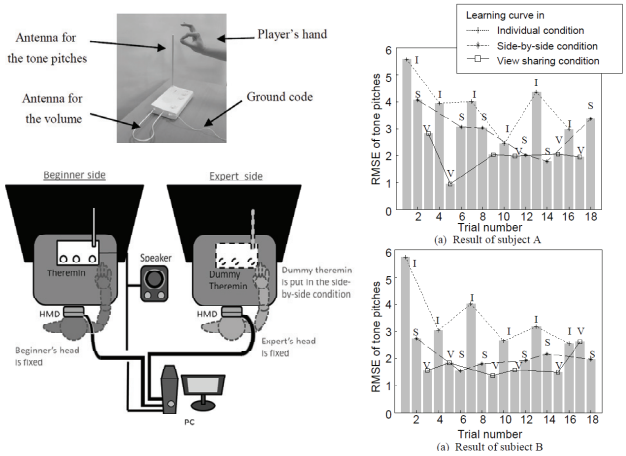


Figure 12. Performance of playing theremin skill transmission in Online PH Assist.

Our results indicate that the skill is transmitted better to the beginner with the view-sharing system than with side-by-side method. What we focus on in this paper is the online skill transmission which means that the master skill is continuously transmitted to a non-skilled person in a short time span. Actually, the beginner could not play the song well soon after the view-sharing condition in which he played well. This means that skill acquisition cannot be expected here. Another aspect of the system is the offline skill transmission which indicates the improvement of learning skills over a long time span as a skill acquisition. Our view-sharing system can be regarded as a learning method for this case. These two aspects are not exclusive but complementary. Actually, the skill acquisition would happen in our experiment as well. The view-sharing system would work better in the skill acquisition in the same manner of the online skill transmission but other experiments are required to clarify the effectiveness.

7 Conclusion

In this paper, we proposed an approach for sharing first-person perspectives to establish collaboration and to transmit a skill from one to another. We developed a view-sharing system to realize such interaction using the Parasitic Humanoid (PH). PH is a wearable robotic human interface for sampling, modeling, and assisting nonverbal human behavior. This anthropomorphic robot can sample, record, and replay the sensory and behavioral experiences of the wearer from the first person perspective of himself. PH can also transmit and share the experience to the other PH wearer. It is a novel technique of telepresence from human to human that can realize telecollabolation and skill transmission between PH wearers.

8 Acknowledgements

This research was supported by JST, CREST.

References

[1] T. Maeda, H. Ando, M. Sugimoto, J. Watanabe, T. Miki: Wearable Robotics as a Behavioral Interface -The Study of the Parasitic Humanoid, Proc of 6th International Symposium on Wearable Computers, pp.145-151 (2002)

[2] S. Jacobsen: Wearable Energetically Autonomous Robots, DARPA Exoskeletons for Human Performance Kick Off Meeting, 2001

[3] W. W. Mayol, B. Tordoff and D. W. Murray: Wearable Visual Robots, International Symposium on Wearable Computing, 2000.

[4] H. Kuzuoka, T. Kosuge, and M. Tanaka, "GestureCam: A Video Communication System for Sympathetic Remote Collaboration," *Proceedings of the 1994 ACM conference on Computer Supported Cooperative Work*, pp. 35-43, 1994.

[5] S. R. Fussell, L. D. Setlock, and R. E. Kraut, "Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vol. 5, No. 1, 513-520, 2003

[6] M. Tsakiris and P. Haggard, "The Rubber Hand Illusion Revisited: Visuotactile Integration and Self-attribution," *Journal of Experimental Psychology, Human Perception and Performance*, Vol. 31, pp. 80-91, 2005.

[7] M. Tsakiris and P. Haggard, "The Rubber Hand Illusion Revisited: Visuotactile Integration and Self-attribution," *Journal of Experimental Psychology, Human Perception and Performance*, Vol. 31, pp. 80-91, 2005.

[8] H. H. Ehrsson, "The Experimental Induction of Out-of-Body Experiences," *Science*, Vol. 317, No. 5841, pp. 1048, 2007.

[9] A. Lécuyer, J. M. Burkhardt, and L. Etienne, "Feeling Bumps and Holes without a Haptic Interface: the Perception of Pseudo-Haptic Textures," *ACM Conference in Human Factors in Computing Systems (ACM SIGCHI'04)*, 2004.

[10] S. Blakeslee and M. Blakeslee, Body Has a Mind of Its Own: How Body Maps in Your Brain Help You Do (Almost) Anything Better, Random House USA, 2007.

[11] S.Tachi, H.Arai, T.Maeda : Tele-Existence Simulator with Artificial Reality(1) - Design and Evaluation of a Binocular Visual Display Using Solid Models, IEEE International Workshop on Intelligent Robot and Systems (IROS'88), Oct. 1988, Tokyo,Japan

[12] Hiroki Kawasaki, Hiroyuki Iizuk, Shin Okamoto, Hideyuki Ando, Maeda Taro, Collaboration and Skill Transmission by First-Person Perspective View Sharing System, 19th IEEE International Symposium on Robot and Human Interactive Communication, September 2010. (in press)

[13] Arthur, K. W., Effects of Field of View on Performance with Head-mounted Displays, ISBN:0-599-73372-1, University of North Carolina at Chapel Hill Doctoral Thesis, 2000.

[14] Tao Ni, Doug A. Bowman, Jian Chen, Increased display size and resolution improve task performance in Information-Rich Virtual Environments, Proceedings of Graphics Interface 2006, pp.139-146, 2006.

[15] N.Shiroma, J.Kobayashi, E.Oyama, Compact image stabilization system for small-sized humanoid, 2008 IEEE International Conference on Robotics and Biomimetic, pp. 149-154, Feb. 2009.

[16] S. Blakeslee and M. Blakeslee, Body Has a Mind of Its Own: How Body Maps in Your Brain Help You Do (Almost) Anything Better, Random House USA, 2007.

Fingertip Slip Illusion with an Electrocutaneous Display

Hiroyuki Okabe[1)] Shogo Fukushima[1), 2)] Michi Sato[1), 2)] Hiroyuki Kajimoto[1), 3)]

1) The University of Electro-Communications

2) JSPS Research Fellow

3) Japan Science and Technology Agency

ABSTRACT

Development of an intuitive pointing device is one of the most significant issues at present in the graphical user interface (GUI) field. Current pointing devices are categorized into two types: the force-based type and the position-based type. The force-based devices require a small input area, but require non-intuitive force-position translation. In contrast, position-based devices are easy to manipulate because of the relatively intuitive position-position translation, but require a large input area. To address the trade-offs between the two device types, we propose a pointing-stick-based input device that uses a newly discovered illusory slip sensation. The slip illusion is induced by presenting a tactile flow generated by an electro-tactile stimulus to the fingertip, while also applying a shear force at the fingertip. This enables us to operate the force-based type of pointer as intuitively as the position-based type. In this paper, we investigate the conditions for occurrence of the illusion, focusing on the shear force at the fingertip, the velocity of the tactile flow, and the directional dependence between the shear force and the tactile flow.

KEYWORDS: electrocutaneous display, haptic illusion, pointing device, slip sensation.

INDEX TERMS: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Input devices and strategies; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems —Artificial, augmented, and virtual realities

1 INTRODUCTION

Pointing devices for graphical user interfaces can be categorized into two types. One type of pointing device, such as a mouse or a touchpad, translates the position of the user's finger to the position of a pointer. The other type of pointing device, such as a joystick or a pointing stick, translates the force of the finger to the speed of the pointer.

There is a well-known trade-off between the two device types. In the position-based input devices, position-to-position translation is quite intuitive, but the device requires a relatively large input area for the finger motion. In contrast, the force-based input devices require a small installation area, but the force-to-position translation is somewhat non-intuitive. For example, it can be difficult to draw even a simple circle.

To resolve the trade-off between intuitive operation and area requirements, Ikeda et al. presented a small touch pad with a fingerprint sensor [1]. This touch pad translates the motion of a fingerprint to the position of a pointer, so that the device requires only a small installation area. It achieved positional type input with small finger movements, but at the same time this small movement may actually make the operation harder.

The other solution is to use haptic illusion. Some researchers have added tactile feedback to a pointing stick [2][3]. Tsuchiya et al. proposed Vib-touch, which added vibration to a force-based pointing device. By controlling the vibrational pattern in accordance with the applied force, the device can reproduce a variety of tactile sensations, such as force, viscosity and elasticity. However, we have found a new haptic illusion of "slip," which occurs when tactile flow is presented at the fingertip, while a shearing force is also applied to the fingertip (Figure 1). By using this illusion, we aim to resolve the trade-off.

In this paper, we investigated the occurrence conditions for the illusion, focusing on the shear force at the fingertip and the velocity of the tactile flow, and the directional dependency between the shear force and the tactile flow.

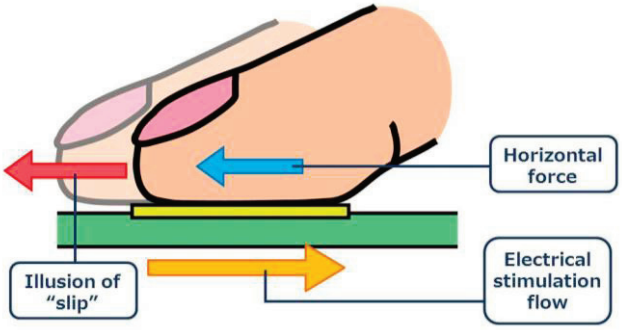


Figure 1. Conceptual image of the slip illusion.

2 SLIP ILLUSION

2.1 Analysis and proposal

To present a subjective feeling of finger motion while the finger does not actually move, we observed the components of haptic sensation related to finger motion. They are divided into two categories. One is the cutaneous sensation, which is because of shear force, frictional vibration, or motion of a pattern on the skin. The other is proprioceptive sensation, which is because of the force and the joint angle. There are therefore roughly five types of cue for finger motion (cutaneous sensation caused by shear force, vibration, and pattern motion; and proprioceptive sensation caused by force and joint angle).

1) 1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, Japan
{h.okabe, shogo, michi, kajimoto}@kaji-lab.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Our problem here is that we want to present a subjective feeling of finger motion while the finger does not actually move. As we are focusing on the force-based type input device, shear force does exist, and so two of the five cues (cutaneous sensation by shear force and proprioceptive sensation by force) exist. What is essentially lacking is proprioceptive sensation from the joint angle. We must supplement the existing sensation by enhancing the other cues.

One possibility is the addition of vibration as a cue for the subjective feeling of motion, which was done by Tsuchiya et al [2][3]. It is a simple method, but users cannot grasp how far the finger moves accurately, and it requires additional cues such as vision.

Our proposal is to add another cutaneous cue, using the motion of a pattern on the skin. A similar method was proposed by Pasquero and Hayward [4]. They showed that cutaneous motion of a pattern presented on a finger helps in scrolling and task selection. As their main concern was the task completion time, the subjective feeling was not reported. In our case, we focus on inducing an illusory finger motion, which has a larger potential application area.

2.2 Preliminary observations

To represent the cutaneous motion of a pattern, we used an electrocutaneous display (Figure 2, and described in detail in the next section) [5], and a commercially available pin type mechanical tactile display (KGS Corp., DotView DV-2). A single line was presented and moved while participants (the authors) put their finger on the display and generated a shear force.

In this preliminary trial, we found two interesting phenomena. The first was that when we used an electrocutaneous display, a "constant slip" sensation occurred, as we expected. However, when we used a mechanical tactile display, this illusory sensation was not reported. We believe that the flatness of the electrocutaneous display surface helped the illusion, while the non-flat surface of the mechanical tactile display might have hindered the illusion.

The other phenomenon was that when shear force was not applied, the illusion did not occur. The shear force or deformation of the skin by the shear force therefore seems to be a main contributing factor to the illusion.

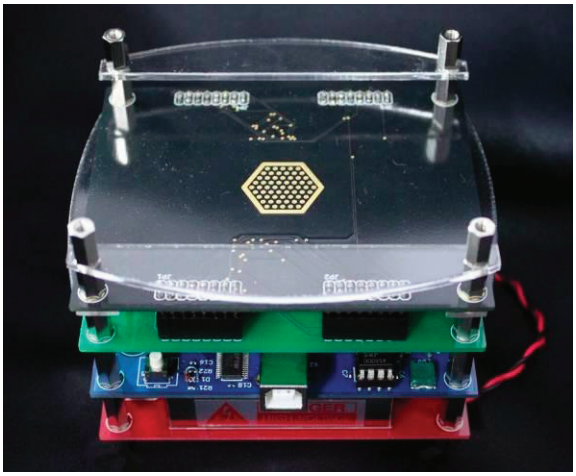


Figure 2. Electrocutaneous display [5]

3 Experiment 1: Shear force measurement during tactile flow presentation

We conducted an experiment to verify that the proposed method can present a pseudo-slip sensation. A moving pattern with variable velocity was presented, and the participants were asked to exert a shear force until they subjectively felt "slip". Our expectation was that the shear force should decrease because of the slip illusion.

3.1 Experimental system

3.1.1 Electrocutaneous display

We used an electrocutaneous display, as shown in Figure 2. The display has 61 electrodes arranged hexagonally. The distance between each electrode is 2.0 mm, and the electrode diameter is 1.0 mm. The pulse amplitude used was 0.0–3.0 mA, the pulse width was 0.05 ms, and the pulse frequency was 60 pps (pulses per sec).

A moving line pattern with variable speed was presented, as shown in Figure 3. During the experiment, the subjects controlled the pulse amplitude freely, so that the elicited sensation remained clear.

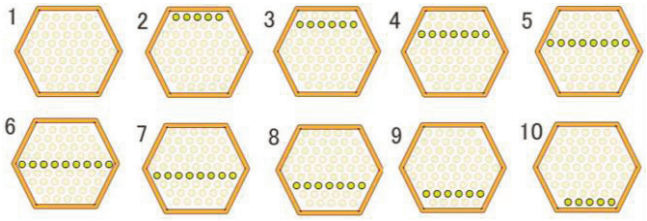


Figure 3. Electrical stimulation flow (1 to 10)

3.1.2 Measurement of shear force

The electrocutaneous display was placed inside a hard case with four pressure sensors (NITTA Corp., FlexiForce), with one placed on each side wall. When the participants applied a shear force, it was measured using the sensors (Figure 4). In this experiment, we used one pressure sensor attached to the front side.

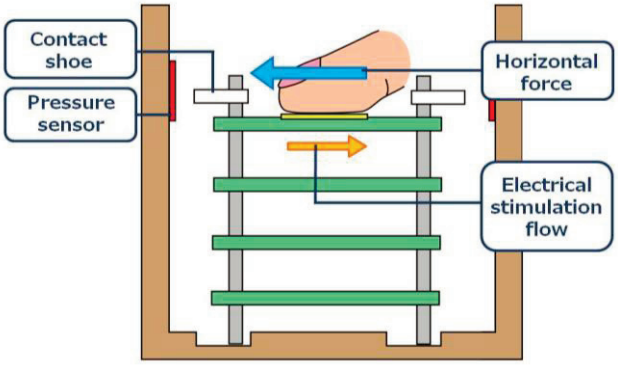


Figure 4. Measurement of shear force

3.2 Experimental conditions

The participants closed their eyes during the measurements. We prepared four velocity conditions (0, 10, 30, and 50 mm/sec). The 0 mm/sec condition required no electrical stimulation. The participants answered 5 times for each condition, and thus the total number of trials was 20 (=4×5). Three participants (2 males, and 1 female, aged 22–23 years old) took part in the experiment.

3.3 Experimental procedure

First, the participants placed their right index fingers on the electrodes. While the display presented electrical stimulation flow to the finger, the participants gradually distorted their fingers forward until they perceived the slip sensation. When the participants perceived the slip sensation, they pressed a button, and the shear force at that moment was recorded. Figure 5 shows the appearance of the experiment.

Figure 5. Appearance of experiment 1

3.4 Results

Figure 6 shows the experimental results. The horizontal axis gives the velocity of the electrical stimulation; and the vertical axis shows the normalized shear force. We subtracted the force at 0 mm/s from all experimental data. For example, under the 10 mm/sec condition, participants perceived a slip sensation with 3.14 N smaller force than in the 0 mm/sec condition. Error bars indicate the standard deviations.

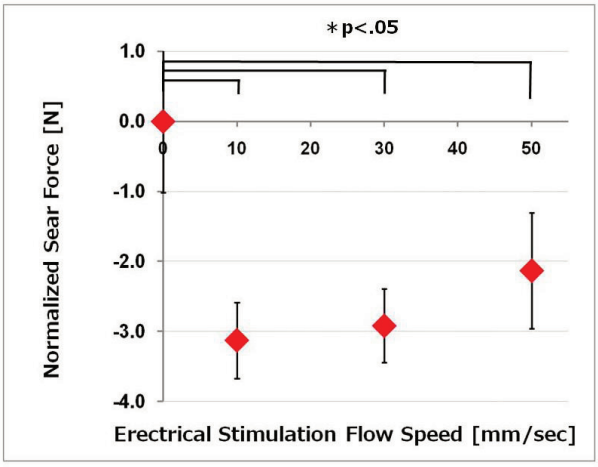


Figure 6. Experimental results

3.5 Discussion

From the results of analysis of variance, the effects of the flow velocity were found to be significant (ANOVA, $F(3, 6)=10.43$; $p<0.01$). For all of the 10, 30 and 50 mm/sec conditions, the required shear force was lower when compared to the 0 mm/sec condition. This result indicates that the participants perceived the illusory slip sensation when they were presented with both a moving electrical stimulation to the fingertip and finger distortion. All participants also reported that they felt continuous motion of their fingers, although their fingers were actually not moving.

4 Experiment 2: Shear force measurement with constant vertical force

The results of experiment 1 imply that a motion speed of around 10 mm/sec is optimal, while speeds of less than 10 mm/sec were not observed. In the next experiment, we observed the details around and below 10 mm/sec.

Also, in experiment 1, we did not control the vertical force of the finger, which in fact greatly affects the slip conditions. Because it is possible that the shear force was reduced simply because the vertical force was attenuated, the next experiment was carried out under constant vertical force conditions.

4.1 Experimental system

To feed back the vertical force information to the participants, we placed the experimental system on an electronic force balance (Figure 7).

Figure 7. Experimental system

4.2 Experimental conditions

Six participants (4 males and 2 females, aged 21–28 years old) took part in the experiment. We presented a moving line as in experiment 1, with 11 different velocities (from 0, 2, 4 … 20 mm/sec). As before, no electrical stimulation flow was applied for the 0 mm/sec condition. For each speed, the shear force measurement was carried out three times. The total number of trials was 33(=11×3).

4.3 Experimental procedure

First, the participants pushed their right index finger on to the display while watching the value of the electronic force balance. They were asked to maintain a vertical force of between 1 N and 3 N during the measurements. Next, the electrocutaneous display presented the tactile flow to the finger, and the participants gradually distorted their fingers forward until they perceived the slip sensation. When they felt the slip sensation on their fingers, they pressed a button, and the shear force was measured. The participants repeated this trial 3 times for each speed condition. If the vertical force was not kept between 1 N and 3 N, the subject restarted the trial.

4.4 Results

Figure 8 shows the experimental results. The vertical axis, horizontal axis, and error bars are the same as those in Figure 6.

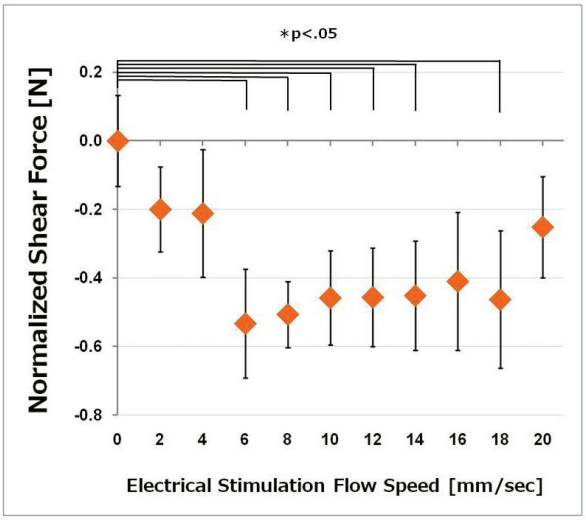


Figure 8. Experimental result 1

4.5 Discussion

The experimental results showed that the shear force was attenuated. Results of analysis of variance showed that the velocity of the tactile flow was significant (ANOVA, $F(11, 55)=1.94$; $p<.05$). The t-test revealed that the results for 0 mm/sec vs. 6,8,10,12,14,18 mm/sec of electrical stimulation flow were significant (independent t-test; $p<0.05$), and the minimum point for the shear force is 6 mm/sec. Therefore, by presenting both a moving tactile flow to the fingertip and the shear force, the required force to perceive the slip sensation was clearly attenuated.

5 Experiment 3: Directional Dependency

When considering the application of the illusion to the pointing device, the slip sensation should occur in any direction. In this experiment, we tested whether the slip sensation was evoked in four different directions (up, down, left, right) (Figure 9).

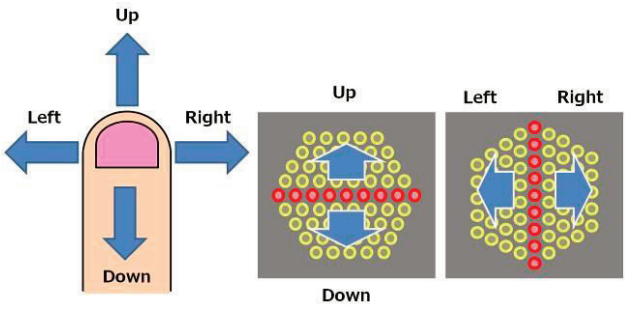


Figure 9. Direction of shearing force (left) and tactile flow by electrical stimulation (Right)

5.1 Experimental device

The experimental device was the same as that in experiment 2. To measure the four-way shear forces, we used four force sensors attached to the inner walls of the case.

5.2 Experimental condition

An electrical stimulation flow at 10 mm/sec was presented with five directional conditions (up, down, left, right, and nothing, i.e. no electrical stimulation). Four directional conditions for the shear force were given (up, down, left, right). For each combination, the measurement was carried out three times. The total number of trials was 60 (=5×4×3). Six participants (4 males and 2 females, aged 21–25 years old) participated in the experiment. Figure 10 shows the appearance of experiment 2.

Figure 10. Appearance of experiment 2

5.3 Experimental procedure

The procedure for the experiment was the same as that in experiment 2. After three measurements using the same directional conditions (direction of the flow and direction of the shear force), the participants changed the direction of the shear force. Then, the direction of the tactile flow was changed and the measurements were conducted again in the same way.

5.4 Results

The experimental results are shown in Table 1. The values of each cell show the normalized shear force, in which the test values were reduced by subtraction of the values of the 0 mm/sec case. The force units used are newtons. A negative value (blue cell) means that a lower shear force was required to perceive the slip sensation than with no electrical stimulation flow, indicating that the slip illusion was induced.

Table 1. Experimental results 2

Normalized Shear Force [N]		Direction of Shear Force			
		Up	Down	Left	Right
Direction of Electrical Stimulation	Down	-0.50	-0.24	0.05	0.12
	Up	-0.32	0.27	0.56	-0.43
	Right	0.27	0.24	-0.60	-0.94
	Left	0.03	0.52	-0.67	-0.74

5.5 Discussion

Variance analysis in experiment 3 indicated that interaction between the fingertip shear force and the electrical stimulation flow was significant (ANOVA, $F(12, 60)=6.96$; $p<0.01$).

Looking back at experiments 1 and 2, the slip sensation occurred when the finger was distorted "upwards" and the electrical stimulation flow was presented as "downwards (i.e., opposite direction to the finger shear force)". We therefore expected that the direction of the finger shear force and the electrical stimulation flow must be in opposite directions.

However, the results indicated that when the direction of shear force was "downward", the direction of the electrical stimulation flow was better in the "downward (same direction)" case than in the "upward (opposite direction)" case. However, when the direction of the shear force was left or right, the direction of the electrical stimulation may be either right or left.

These results may indicate that the tactile flow direction was sometimes misinterpreted as being in the opposite direction when combined with the shear force. Further study is necessary for this experiment.

6 Conclusions and Future Work

In this paper, we observed a new tactile illusion of slip, where a combination of electrical stimulation flow and finger shear force was applied. The sensation is continuous while the finger remains stationary.

We conducted two experiments to validate the illusion. The results clearly showed that the user felt the slip sensation with less shear force than under the actual slip conditions, suggesting the existence of a slip illusion.

We also conducted an experiment to see if the illusion occurs in other directions. The results remain unclear, but seem to indicate that the tactile flow direction was sometimes misinterpreted as being in the opposite direction when combined with the shear force. Further study is necessary to explore this point.

Our future work includes an in-depth study of the last experiment, and fabrication of the pointing device and its evaluation.

References

[1] Atsutoshi Ikeda, Yuichi Kurita, Jun Ueda and Tsukasa Ogasawara, Development of a Small Pointing Device Utilizing the Fingerprint Deformation at the Time of an Incipient Slip, Information Processing Society of Japan, vol. 45, no. 7, pp. 1769–1778, 2004 (in Japanese).

[2] Sho Tsuchiya, Masashi Konyo, Hiroshi Yamada, Takahiro Yamauchi, Shogo Okamoto and Satoshi Tadokoro, Vib-Touch: Virtual Active Touch Interface for Handheld Devices, 18th IEEE International Symposium on Robot and Human Interactive Communication, pp. 12–17, 2009.

[3] Sho Tsuchiya, Masashi Konyo, Hiroshi Yamada, Takahiro Yamauchi, Shogo Okamoto and Satoshi Tadokoro, Virtual Active Touch II: Vibrotactile Representation of Friction and a New Approach to Surface Shape Display, Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3184–3189, 2009.

[4] Jerome Pasquero and Vincent Hayward, Tactile Feedback Can Assist Vision During Mobile Interactions, Proc. 2011 Annual Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 2011.

[5] Hiroyuki Kajimoto, Electro-tactile Display with Real-time Impedance Feedback, EuroHaptics 2010, pp. 285–291, 2010.

Improvement of Olfactory Display Using Electroosmotic Pumps and a SAW Device for VR Application

Yossiri Ariyakul and Takamichi Nakamoto

Graduate school of science and engineering, Tokyo Institute of Technology

ABSTRACT

Previously, air blowing or heating techniques has been used for presentation of the scents. As a consequence, the scents of low-volatile substances are difficult to be released at an acceptable speed. Also, the effect from heat to the nearby environment cannot be avoided. In this research, we proposed an olfactory display using miniaturized electroosmotic pumps and a surface acoustic wave device utilizing SAW atomization technique. In this manner, the olfactory display is miniature in size, works soundlessly, and is able to produce even the scents of low-volatile substances without heat radiation to the adjacent environment. In addition, we optimized the experimental parameters to solve the problem of reproducibility when the number of odor components increase. As a result, by using an odor sensing system we could confirm that the developed device possesses ability to control odor intensity and to blend multiple of odor components. Owing to these benefits it can be expected to be utilized in a variety of virtual reality applications.

KEYWORDS: Olfactory display, low-volatile scent, electroosmotic pump, SAW streaming, QCM gas sensor.

INDEX TERMS: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities;

1 INTRODUCTION

During the past decades, extensive research efforts have been dedicated to the development of virtual reality for entertainment [1], engineering [2], education [3], and medical applications [4]. The virtual reality experience can be made more realistic with a three dimensional stereoscopic vision and audio, or other sensory domains using the experience of tactile and thermal expressions [5], wind blowing, and the emission of aromas [6]. Unlike other senses, as the olfaction system is connected directly to the Paleomammalian brain, it has an intense connection with memory and induces an emotional connection to us. Therefore, the introduction of olfactory information using olfactory display could have a particular impact on the potential of the virtual reality.

The first attempt to add the sense of smell into virtual reality started in 1960s when the Sensorama [7], a crude multimedia movie machine, was developed. This work can be regarded as a predecessor of modern multimedia system. Even though there were a lot of limitations in technology at that time, it can present the odor according to the 3D scene on the display with wind blowing and chair vibration. Since then, there have been discussions whether movies need odor presentation. However, as human perceive the world through five senses, the lack of olfaction in the virtual world may cause users to feel like wearing a space suit without exposure to the air in the real world [8].

To include an olfactory sense into virtual reality, nowadays olfactory displays based on several concepts have been extensively developed by several groups. For example, our group has developed olfactory displays using mass flow controllers [9] and solenoid valves [10]. In the former, odors are mixed with a flow ratio determined by the mass flow controller and in the latter, the high speed solenoid valves which only two states of ON and OFF are used to adjusted the strength of the odor by varying the ON/OFF frequency of the valves. The developed devices equipped with up to 31 different odor components are able to mix the odors at arbitrary recipes, and then deliver it to the user's nose through a tube attached to a headset. Our group also developed the olfactory display using inkjet devices in which the fine droplets of liquid-phase odor samples are spouted using the inkjets on to the heater and evaporated to generate odor [11]. As the liquid droplets are enforced to be vaporized, this device is suitable for presenting the odor of low-volatile substances.

The olfactory display based on inkjet system has also been developed by Okada [12]. As this device presents odor for a short time, it can avoid the adaptation problem that users tend to be unable to detect smell when they inhale it continuously for a period of time. In addition, the olfactory display using functional high polymers coupled with a Peltier device, above which fragrance is encapsulated in a hydrogel chip, has been developed by Kim [13]. Besides, the wearable olfactory display with the capability to present scent information spatially based on the user's position has also developed by Hirose [14]. And, the scent projector with the capability to deliver odor in vortex ring shape to a specific area up to 1.5m has also developed by Yanagida [15].

Taking the current situation into account, olfactory displays based on several techniques have been so far studied and they are useful in certain situations. However, our understanding is that the olfactory display for general usage in virtual reality should possess the 4 following capabilities.

(1) It should be able to present any kinds of smells vividly without the lack of reproducible release over multi cycles. Moreover, the odor presentation should be made at sufficient speed. In other words, the presentation should be able to be started and stopped as fast as possible.
(2) It should be able to adjust the intensity and amount of the emitting odor precisely and in real time according to the user's needs.
(3) As primary smells have not been so far discovered unlike primary colors, it is profitable if the olfactory display can blend multiple odor components for presenting various smells based on a limited number of odor components.
(4) As home electronic apparatuses, such as television, home-theatre, or video game, are getting thinner, smaller, and more delicate, the olfactory display should be also small

2-12-1, Ookayama, Meguro-ku, Tokyo, 152-8552, JAPAN
E-mail: nakamoto@mn.ee.titech.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

and thin enough. Besides, it should work soundlessly and do not produce heat to the nearby circumstance.

Indeed, no existing device at the moment could fulfill these requirements. For example, the mass flow controllers and solenoid valves based olfactory display has difficulty to present scents of low-volatile substances, as their principles rely on the natural evaporation characteristic of the liquid-phase odor sample to generate smell, and also the noise associated with the valve switching and air blowing cannot be avoided. For the inkjet based olfactory display developed by our group, it requires skill to handle and the odor blending becomes more difficult when the number of odor components increased as it lacks of self-priming capability unlike an EO pump. Moreover, the heat radiated from the heater affects nearby environment. For the olfactory displays developed by other groups, there is no reported regarding the capability to blend smells, and devices are still bulky. For these reason, innovative techniques are needed obviously.

As the primary requirement of the development of an olfactory display is an accurately adjustable odor presenting capability, in this research we employ a kind of miniaturized liquid pump called EO (Electroosmotic) pump to drive the liquid-phase odor sample to another device as the flow rate can be controlled linearly by adjusting the driving voltage with high reproducibility. To produce the odor, we make use of a SAW (Surface Acoustic Wave) device to atomize the liquid-phase odor sample emited from the EO pump into fine particles which are suddenly vaporized in to the air. Therefore, the response time of the odor emission is fast. And as the liquid is atomized forcibly, even the scents of low-volatile substances can be also rapidly generated. Both EO pump and SAW device work soundlessly, do not produce heat to the nearby surrounding, and are miniature. Therefore, the olfactory display based on this concept can be expected to be in a small size enough to be applied to many systems. In addition, by using multiple EO pumps filled with different odor components the ability to blend odors can be expected.

In the previous paper [16], we have successfully employed a basic olfactory display developed based on this new atomization technique to present smells and we were able to confirm its reproducibility. In the present paper, to realize the capability to control the odor intensity and to blend smells, a lot of parameters, such as the rearrangement of the SAW device, the optimization of the SAW parameters, and the improvement of EO pumps driving method, were optimized. Consequently, we can solve the problem of reproducibility when the number of odor components increases and the device also became thinner. Moreover, the odor sensing system used to evaluate the developed device was also improved to reduce the noise occurring in the experiment.

2 SYSTEM OVERVIEW

2.1 Electroosmotic pump

In recent years, micro-fluidic devices and their applications have received a lot of attention due to the rapid growing progress in the field of micro-fluid systems. Micro-pump is one of the most important micro-fluidic components as it can provide the precise flow rate for the micro-channel. EO pump used in this study is a kind of Micro-pump that can produce the flow rate of liquid by utilizing electroosmotic phenomenon [17]. It can drive ethanol or de-ionized water with high stability quietly due to no mechanically moving part. Besides, as the relationship between the flow rate and electric driving voltage is linear according to the electroosmotic phenomenon, the flow rate can be controlled easily by adjusting the applied voltage.

Because of these advantages, it was employed here to supply liquid droplet onto the SAW substrate to atomize sample liquid to generate smell. The EO pumps used in this study were made of Polypropylene with 11.5mm in length and 6mm in diameter. The reservoir (100μL) is equipped on the top of the device as shown in Figure 1.

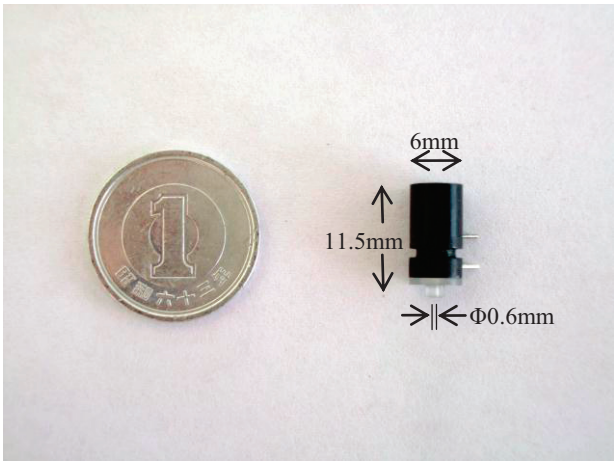


Figure 1. An EO pump used in this study.

2.2 SAW device

SAW device basically consists of an input transducer to convert electrical signal to surface acoustic wave, which then travels along the solid surface to the output transducer where it is reconverted to electrical signal. With this characteristic, it is typically used as a RF filter in communication field. However, when the liquid is dropped onto its substrate as shown in Figure 2, the SAW radiates a longitudinal wave into liquid causing various liquid motions, such as vibration, flow, soaring, and atomization depends on the amplitude of the electric voltage applied to the SAW device. This phenomenon is called SAW streaming [18].

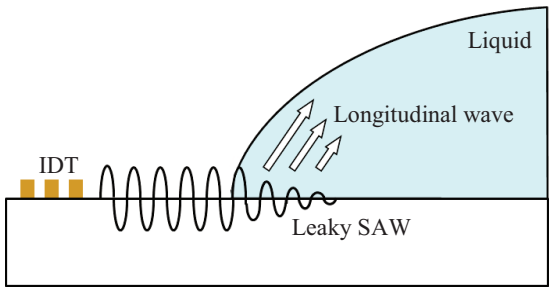


Figure 2. Schematic of SAW streaming.

Although there are several ways to achieve atomization, the size of SAW atomizer is rather smaller than those produced by other methods. Therefore it is employed to atomize the liquid-phase odor sample spouted from EO pumps into smell. In addition, this operation does not radiate heat to nearby environment unlike the heater.

The SAW device designed in this study consists of two ports of 10 finger pairs of an interdigital transducer (IDT) with width of 16μm, and 100 lines of a grating reflector as shown in Figure 3. All of the electrodes are made of gold and fabricated on a piezoelectric substrate made of a 128°rotated Y-cut X-propagation $LiNbO_3$. The device size is 11x19mm, and the center frequencies of this device are about 60.65MHz. Frequency characteristics of the SAW device measured by impedance analyzer (4291A, Hewlett Packard) are shown in Figure 4.

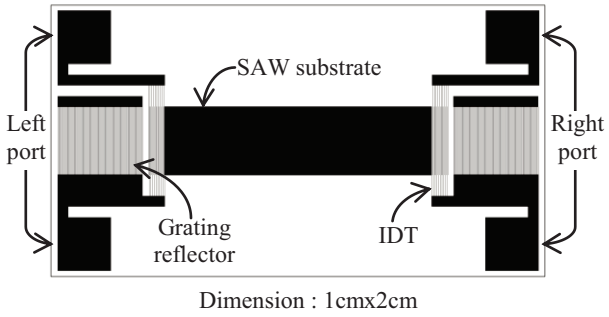


Figure 3. Pattern of the SAW device used in this study.

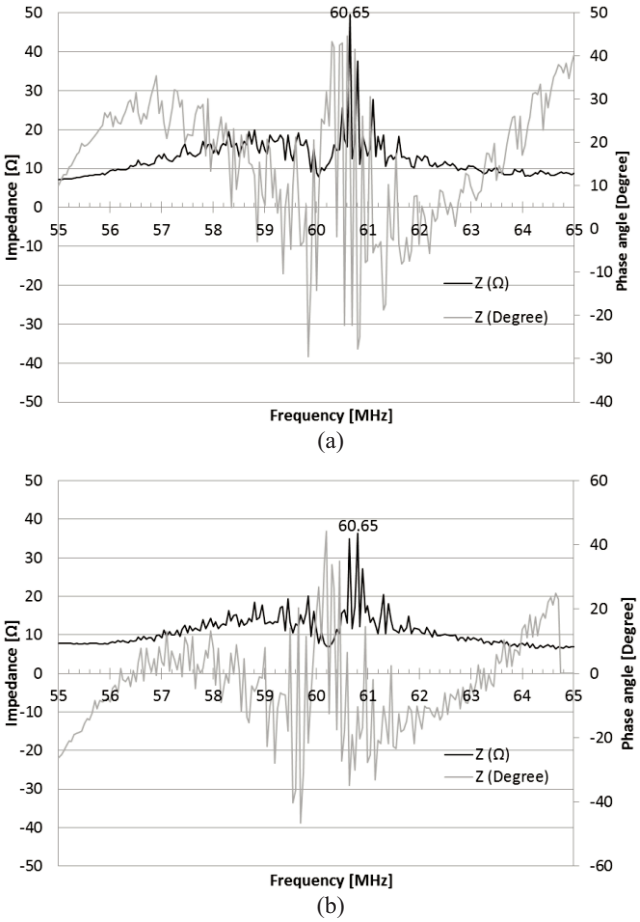


Figure 4. Frequency characteristic of the SAW device at the (a) left port, (b) right port.

2.3 Structure of the system

The structure of an olfactory display using an EO pump and a SAW device is shown in Figure 5. First, the target odor to be presented is dissolved into solvent that is capable to be driven by electroosmotic phenomenon using an EO pump (ex. ethanol, de-ionized water) with the reservoir on its top. By applying electrical pulse to the pump, a droplet of odor sample is supplied onto the surface of the SAW device located under the EO pump, where they are atomized forcibly by SAW streaming into smell and then are delivered to the user's nose by a small fan.

Based on this configuration, by adjusting the electrical voltage applied to an EO pump, we can adjust the volume of liquid droplet. As the odor intensity relies on the amount of liquid-phase odor sample to be atomized, the capability to control the intensity of a presenting odor can be achieved. Furthermore, when we use multiple EO pumps filled with different odor components the ability to blend odors can be realized by adjusting the volumes of liquid droplets in this manner.

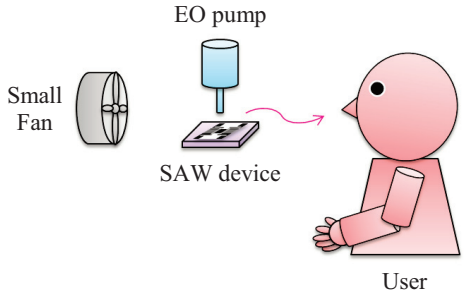


Figure 5. Schematic of an olfactory display using an EO pump and a SAW device.

3 EXPERIMENTAL SYSTEM

To evaluate the proposed olfactory display, our group has used an odor sensing system instead of using the sensory test since human perception is influenced by adaptation and varies from person to person while an odor sensing system can measure the temporal change of the odor intensity numerically with high reproducibility. The whole experimental setup is shown in Figure 6.

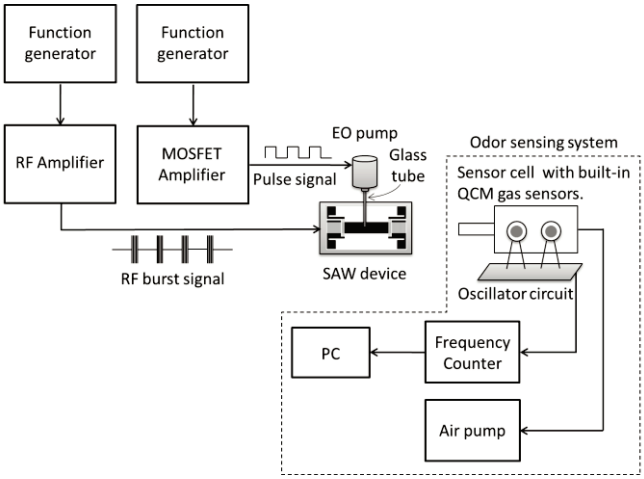


Figure 6. Schematic diagram of experimental set up.

The odor sensing system consists of gas sensors, an oscillator, and a frequency counter circuit. Gas sensors used here was the QCM (Quartz Crystal Microbalance) gas sensors. In the previous paper [16], a fan was used to blow the generated odor to the QCM gas sensor exposed directly to the ambient air. As that configuration create a lot of noise, in this study QCM gas sensors were built in the sensor cell which the generated odor was blew into by using an air pump connected to it. When the molecule of the odor is adsorbed onto the sensing film at the surface of QCM, the mass change causes the shift in resonance frequency of the quartz due to mass loading effect [19]. Therefore, the intensity of the presenting odor can be converted to the frequency shift. The

sensing films used in these experiments were Siponate DS-10 and PEG-1000. The electrodes on the QCM surfaces were made of gold, and the center frequencies of the sensors were approximately 20MHz. The frequency shift due to the odor molecule adsorbtion was measured by the frequency counter. Then the measured data were transferred to a computer via serial interface and were stored as a text file.

In the experiments, as an air pump was used to blow the generated odor into the sensor cell, a small fan as shown in Figure 5 was not actually used. The wind speed at the front end of the sensor cell was approximately 0.6m/s and the temperature control was not especially performed. The actual experimental environment is shown in Figure 7. Three EO pumps were prepared in this study, and the center one would be used if not specified in each section. In the previous paper [16], the SAW device was placed perpendicular to the EO pumps as is shown in Figure 5. However, the amount of liquid dropped on to the SAW device was significantly influenced by the distances between the tips of EO pumps and the surface of a SAW device. Thus the reproducibility deteriorated due to the difficulty in controlling this distance accurately when many EO pumps were used. Obviously, this problem needs to be overcome to realize the odor blending capability. Therefore, to solve the problem, in the present paper, the SAW device was placed in the parallel direction to the EO pumps as is shown in Figure 7. Furthermore, the glass tubes with 0.3mm inner diameters were connected to the EO pump tips and laid on to the SAW device's surface. Owing to this configuration, the difficulty in controlling the distance as described above can be avoided and thickness of the overall device can also be reduced. Besides, as the glass tube inner diameters were small, the droplets with tiny volume were obtained.

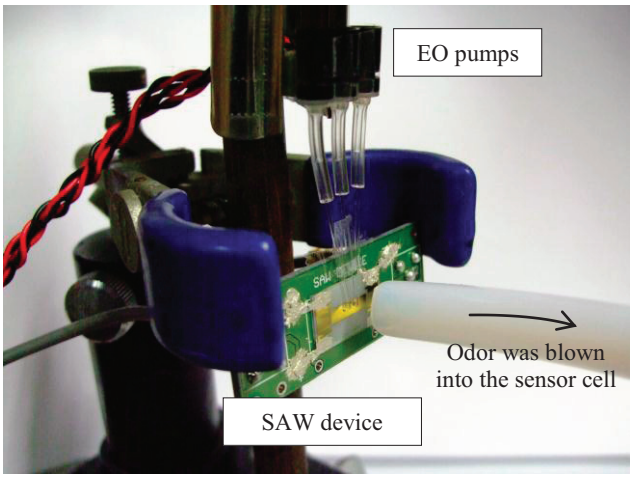


Figure 7. Actual experimental environment.

RF burst was used to drive the SAW device instead of the continuous signal since the latter might exceed the device's tolerance. In the previous paper [16], the duty cycle and repetition period of the RF burst were set to be 50% and 10ms respectively. However, the longer repetition period is, the significantly higher atomization performance is obtained. Thus, those parameters were set to be 1% and 100ms respectively if not specified in each section. And, as the SAW device requires sufficient driving voltage in RF frequency band to perform an atomization, a wide-band RF amplifier (ZHL-5W-1, Mini-Circuits) was used to amplify RF burst signal generated from a function generator (AFG 3251, Tektronix) into $60V_{pp}$ if not specified particularly.

MOSFET amplifier was used to amplify pulse signal from a function generator (33120A, Hewlett Packard) to 75V pulse to drive EO pumps. The amount of droplet dropped onto the SAW device's substrate was controlled by adjusting the duration of the applied pulse.

Samples used in the experiments were 10 % 2-hexanone (MW:100.2, b.p.:127℃), 10 % 1-butanol (MW:74.12, b.p.:117.2 ℃), and 10% beta-ionone (MW:192.3, b.p.:239℃) diluted with ethanol. Since 2-hexanone and 1-butanol are typical compounds with moderate volatility, they were used to evaluate the measurement system and the capability to present the typical scents of the olfactory display. On the other hand, beta-ionone has low-volatility which causes difficulty in smell presentation by conventional olfactory display [9][10]. It was used as a representative to evaluate the capability to present the scent of low-volatile substances of the olfactory display.

4 EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Odor presentation capability evaluation

To evaluate the odor presentation capability of the olfactory display, in the first experiment a 75 V electric pulse with duration for 3s was added to the EO pump. At that time 0.41μl of 2-hexanone solution was dropped on to the SAW substrate and then atomized abruptly. The sensor response to 2-hexanone is shown in Figure 8. The dashed line shows when the sample droplets were dropped onto SAW substrate.

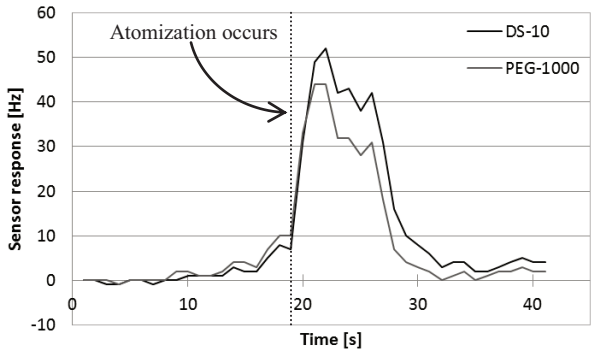


Figure 8. Sensor response to 2-hexanone generated by proposed olfactory display.

As a result, we can see that the abrupt peak of the sensor response was obtained. It means that the olfactory display can present the scent with moderate volatility suddenly and then disappeared in a short time without smell persistence around the sensor. Moreover, the noise of the sensor response was drastically reduced compared with the previous paper because of the sensor cell usage whereas the sensor, which is also sensitive to temperature and humidity variation, was directly exposed to the ambient air in the previous one.

Next, another experiment was conducted to confirm if the olfactory display can present low-volatile scents. In this experiment, the sensor response to scent generated from 0.27μl of beta-ionone solution atomization was measured. The result is shown in Figure 9. In comparison, the sensor response to the same sample presented by the conventional solenoid valve based olfactory display is shown in Figure 10. In that experiment, the QCM coated with Versamid900 was used.

As a result, we can see that the response time or the recovery time of the waveform in Figure 9 was much faster than that in Figure 10. It means that the olfactory display using atomization technique developed in this study can present even the scents of low-volatile substances abruptly and the scents then disappear in a short time while the conventional device requires very long time

to reach the steady state and also a very long time until the scents disappear due to the natural evaporation characteristic of the substance and the smell persistence inside the tube.

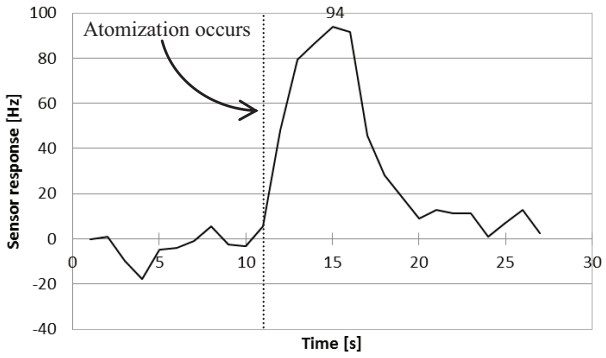


Figure 9. Sensor response to beta-ionone generated by proposed olfactory display.

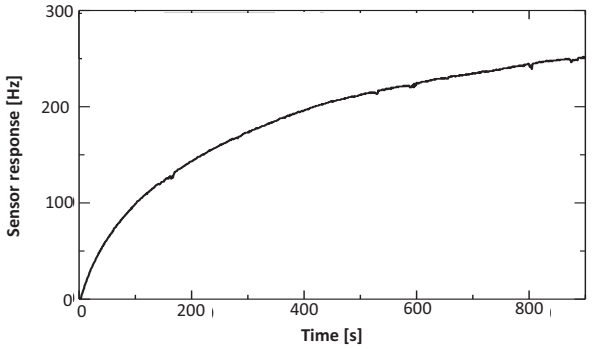


Figure 10. Sensor response to beta-ionone using a conventional solenoid valve type olfactory display (Flow rate:1350L/min).

Then, the next experiment was performed to verify the reproducibility of the odor presentation using developed olfactory display. In this experiment, the sensor response to scent generated from 0.41µl of beta-ionone atomization was measured 3 times. The result is shown in Figure 11. As a result, the certain repeatability of the sensor response can be observed. Thus, we consider that the olfactory display has sufficient reproducibility for practical usage.

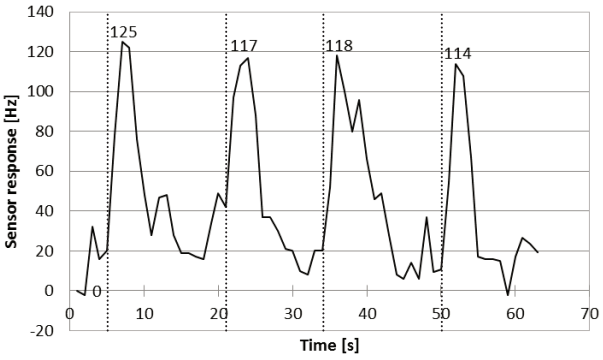


Figure 11. Repeatability of the odor intensity presented by the proposed olfactory display.

4.2 Optimal condition of duty cycle and RF voltage of SAW device

The atomization is caused by RF burst applied to the SAW device. The important parameters for the atomization are RF frequency, RF voltage, duty cycle and repetition period. It is known that the higher RF voltage and the more duty cycle are applied to the SAW device, the fiercer atomization occurs. Since the higher power consumption needs, the larger the size of the RF amplifier in the system requires. To make an olfactory display as miniature as possible, we need to find the way to drive the SAW device most effectively at the same power consumption. Here we made an experiment to investigate whether RF voltage or duty cycle comparatively governs the atomization performance under the condition of the same RF power. The result of this experiment would help us determine the valuable parameter to be adjusted to increase the atomization performance.

In this experiment, scents generated from 0.27µl of 2-hexanone solution atomization under three conditions that RF voltage and duty cycle of the burst wave are set to (1) $100V_{pp}$ and 2%, (2) $50V_{pp}$ and 4%, (3) $20V_{pp}$ and 10% were measured. The odor sensor used in this experiment was the QCM coated with PEG-1000, and the sensor response is shown in Figure 12.

As a result, we can see that RF voltage has more influence to the atomization performance than duty cycle under the same RF power condition. Therefore, according to the result we would set the duty cycle constant, and adjust the RF voltage if the fiercer atomization is required.

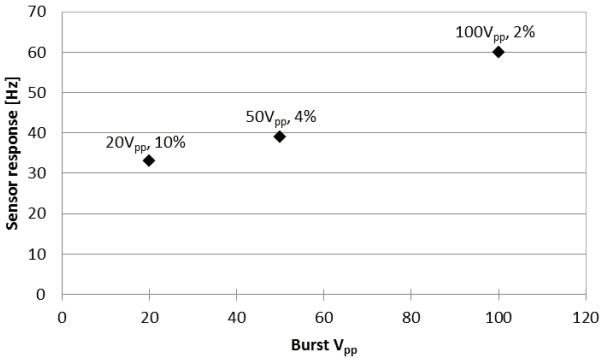


Figure 12. Relationship between RF voltage and sensor response under conditions of same RF power.

4.3 Odor intensity controllability Evaluation

In our previous work [16], we have already showed that the intensity of the presenting odor can be controlled by adjusting the number of liquid droplets injected onto the SAW device. However, the pulse width control requires shorter time than the pulse number control. Therefore, at this point, an experiment was performed to verify whether the intensity of the odor presented by the olfactory display in this study can be controlled by adjusting pulse width applied to the EO pump.

In this experiment, the four conditions that 75 V electric pulses with the duration for 1s, 2s, 3s, and 4s were applied to the EO pump to emit 0.14µL, 0.27µL, 0.41µL, and 0.55µL of 2-hexanone solutions onto the SAW substrate were performed to generate smells respectively. The experiment was done five times at every condition. The odor sensor used in this experiment was the QCM coated with Siponate-DS10, and the typical sensor responses are shown in Figure 13.

From the result, we can see that when the droplet volume increases by applying longer pulse width, the sensor response also increases proportionally. From this reason, we would be able to

control the intensity of the presenting odor by adjusting the pulse width applied to EO pump.

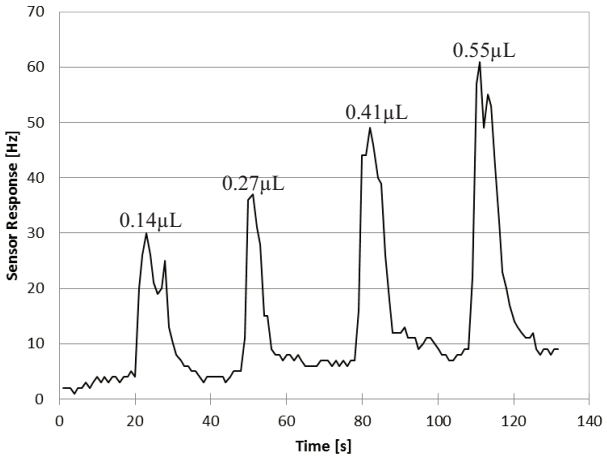


Figure 13. Relationship between pulse width applied to EO pump and sensor response.

4.4 Dependence of droplet position upon atomization

In section 4.3, we confirmed that the olfactory display developed in this study possesses the ability to control the odor intensity. Therefore, we expected the situation where multiple EO pumps were set up to spout different odor components on to any point of the SAW surface area and then atomized them together. In this situation, the mixture odor should be presented. To verify this assumption, in this experiment we divided the SAW area into left, center, and right region as shown in Figure 14. Then 2-hexanone solutions with the volume from 0.27μL to 0.95μL were dropped onto any region on the SAW surface to be atomized. Every condition was done twice. The RF burst used in this experiment was approximately $120V_{pp}$ and the odor sensor was QCM coated with Siponate-DS10. The average sensor responses are shown in Figure 15.

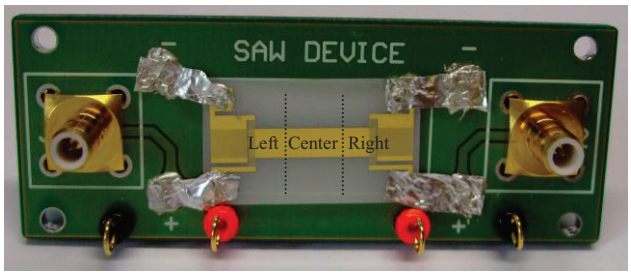


Figure 14. Photo of the SAW atomizer.

From the figure, it was found that there are only slight differences between the sensor responses from any region at any concentration level. As a result, we can conclude that the atomization performance does not change noticeably due to the atomization position. Therefore, it was proven that the multiple odors blending capability can be achieved if a lot of odor components are spouted on to the SAW substrate and then are atomized together.

4.5 Odor blending capability evaluation

Lastly, as the capability to blend smells make an olfactory display profitable for any applications, here the experiment to confirm the possibility to blend odors by using the proposed olfactory display was conducted. In this experiment, we used two EO pumps filled with 2-hexanone and beta-ionone, and presented its smell to the sensor at a certain level of 2-hexanone concentration. In the experiment, the QCM (20MHz, Ag electrode, AT-CUT) coated with Siponate-DS10 was used. The sensor response is shown in Figure 16. As a result, when the two different odors were presented at the same time, superposition characteristic was roughly observed. Thus, the fundamental blending function of the olfactory display developed in this study was confirmed.

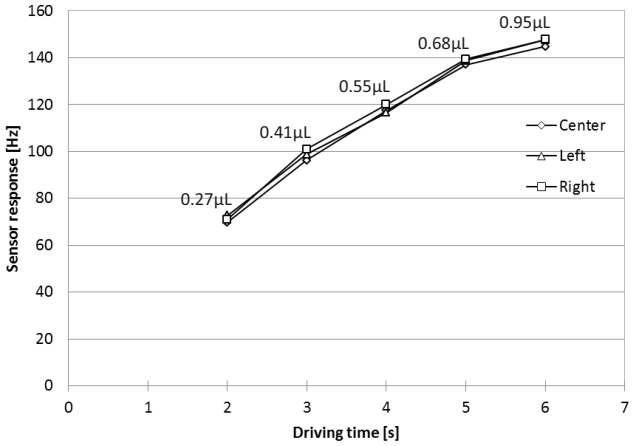


Figure 15. Comparison of the sensor responses when droplets were put onto the center and the right side of the SAW device.

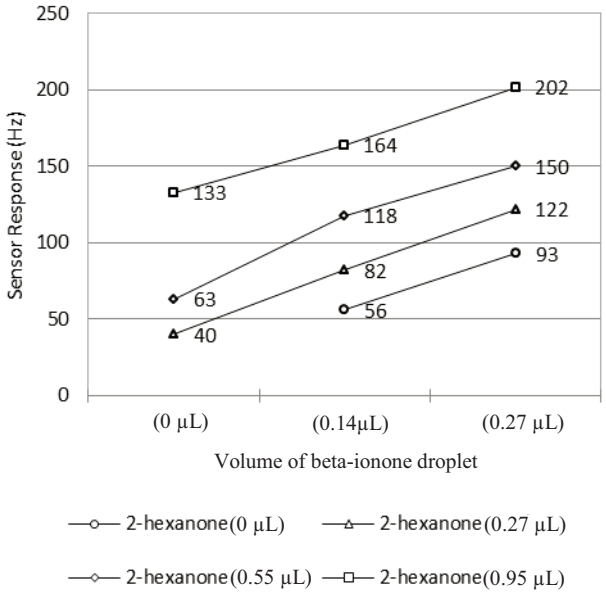


Figure 16. Superposition characteristic of binary-mixture presentation using proposed olfactory display.

5 CONCLUSION

In the present study, the olfactory display, utilizing SAW streaming to atomize liquid-phase odor sample forcibly to present smell, composed of a miniaturized EO pump and a SAW device was proposed. Several oriented validations were performed numerically to achieve its several key characteristics mentioned in

introduction using an odor sensing system. As a result, we showed that the scents of substances with wide variety of volatility can be generated abruptly and then disappear within a short time. Even the scents of low-volatile substances can be quickly presented and eliminated while its presentation is difficult using almost other techniques. The reproducibility of the amount of presented odors was also confirmed. Furthermore, the abilities to control the odor intensity and to blend smells were successfully realized by adjusting the driving parameters and by using a few EO pumps together. Moreover, as the device works soundlessly and does not radiate heat to the nearby environment, and its size is small and thin, it is expected to be integrated into other electronic apparatus in the near future to enhance the realistic feeling and enable more applications using olfaction information.

REFERENCES

[1] S.C. Shapiro, J. Anstey, D.E. Pape, Devdas T. Nayak, M. Kandefer and O. Telhan. A virtual reality drama using intelligent agents. AIIDE-05, Menlo Park, 2005.

[2] S. Jayaram, H. I. Connacher, and K. W. Lyons. Virtual assembly using virtual reality techniques. Computer-aided Design, 29(8):575–584, 1997.

[3] Psotka, J. Immersive training systems: Virtual reality and education and training. Instructional Science 23 (5–6), pages 405–423, 1996.

[4] Soferman, Z., Blythe, D., and John, N.W. Advanced graphics behind medical virtual reality: evolution of algorithms, hardware, and software interfaces. Proceedings of the IEEE, vol.86, no.3, pages 531–554, Mar 1998.

[5] T. Kai, Y. Kojima, Y. Hashimoto, and H. Kajimoto. Mechanism of pressure sensation generated by hot steam. ISVRI2011.

[6] Haruka Matsukura, Tomohiko Nihei, and Hiroshi Ishida. Multi-sensorial field display: presenting spatial distribution of airflow and odor. In Proceedings of VR'2011. Pages 119–122.

[7] M. Heilig. Beginnings: sensorama and the telesphere mask. in Digital Illusion (Ed.: C. Dodsworth, Jr.) ACM Press, pages 343–351, New York 1998

[8] F. Nakaizumi, H. Noma, K. Hosaka, and Y. Yanagida. SpotScents: a novel method of natural scent delivery using multiple scent projectors. Proc. IEEE Virtual Reality, Alexandria, Virginia, USA pages 207–214, 2006.

[9] T. Nakamoto, S.Utsumi, N.Yamashita, T.Moriizumi and Y.Sonoda, , Active gas sensing system using automatically controlled gas blender and numerical optimization technique, Sensors and Actuators B, 20 (1994) 131–137.

[10] Nakamoto, T. and Hai Pham Dinh Minh. Improvement of olfactory display using solenoid valves. Virtual Reality Conference, VR'07 IEEE, pages 179–186, 10-14 March 2007.

[11] T.Nakamoto, H.Takigawa, and T.Yamanaka. Fundamental study of odor recorder using inkjet devices for low-volatile scents. Trans. On IEICE, E87–C, pages 2081–2086, 2004.

[12] J. Sato, K. Ohtsu, Y. Bannai, and K. Okada. Effective presentation technique of scent using small ejection quantities of odor. Proc. IEEE Virtual Reality, Lafayette, Louisiana, USA, pages 151–158, 2009.

[13] Dong Wook Kim, Yeong Hee Cho, Nishimoto K., Kawakami Y., Kunifuji S., and Ando H. Development of aroma-card based soundless olfactory display. Electronics, Circuits, and Systems, ICECS 2009, 16th IEEE International Conference, pages 703–706, 13–16, Dec 2009.

[14] Yamada T., Yokoyama S., Tanikawa T., Hirota K., and Hirose M. Wearable olfactory display: using odor in outdoor environment. Virtual Reality Conference, pages 199–206, 25–29 March 2006.

[15] Yanagida Y., Kawato S., Noma H., Tomono A., and Tesutani N. Projection based olfactory display with nose tracking. Virtual Reality, Proceedings. IEEE, pages 43–50, 27–31, March 2004.

[16] Yossiri Ariyakul and Nakamoto Takamichi. Fundamental study of olfactory display using extremely small liquid pump and SAW atomizer, The virtual reality society of Japan, Vol. 15 (2010), No. 4 pages 589–594 [in Japanese].

[17] Takamura Y., Onoda H., Inokuchi H., Adachi S., Oki A. and Horiike Y. Low-voltage electroosmosis pump for stand-alone microfluidics devices. ELECTROPHORESIS, 24: 185–192, 2003.

[18] M. Kurosawa, T. Watanabe, T. Higuchi. Surface acoustic wave atomizer with pumping effect. Proceedings of the MEMS, pages 25–30, 1999.

[19] T. Nakamoto and T. Moriizumi. A theory of a Quartz Crystal Microbalance based upon a Mason Equivalent Circuit. Jpn. J. Appl. Phys. 29 (1990), pages 963-969

Subjective Image Quality Assessment of a Wide-view Head Mounted Projective Display with a Semi-transparent Retro-reflective Screen

Duc Nguyen Van[1] Tomohiro Mashita[1,2] Kiyoshi Kiyokawa[1,2] and Haruo Takemura[1,2]

[1] Graduate School of Information Science and Technology, Osaka University
[2] Cybermedia Center, Osaka University

ABSTRACT

This paper reports on a wearable Hyperboloidal Head Mounted Projective Display (HHMPD) and two user studies on the evaluation of visual quality of a wearable HHMPD. Using a hyperboloidal mirror, an HHMPD can provide a wide field-of-view (FOV), a large observational pupil, and optical see-through capability. We propose a simple head attached screen that is both retro-reflective and semi transparent thereby allowing the HHMPD to be used in a wearable situation. The two user studies have shown that our wearable HHMPD provides a virtual image with a visual acuity of around 20/200 at perceptually 2 to 3 meters away from the user.

KEYWORDS: wide view head mounted display, retro-reflective semi-transparent screens, wearable computing

1 INTRODUCTION

Our research goal is to realize a wearable computing system with a more intuitive and flexible information display by employing a wide FOV video display. An optical see-through head mounted display (HMD) is commonly used in a wearable computing system to enjoy a variety of IT services. With a see-through HMD, a computer can be used without interrupting the work at hand. Augmented reality (AR), that superimposes computational information onto the real objects, can also be realized with a see-through HMD. However, there is a major problem in most existing see-through HMDs; they can provide a very limited field of view (a horizontal viewing angle of 30-60 degrees) near the central visual field [1]. Sensics's piSight HMD provides 180 degrees of horizontal field of view, but is a closed HMD. To our knowledge, LinkSim. Train's optical see-through HMD, A-HMD, provides the largest horizontal field of view (110 degrees) in an optical see-through fashion.

Originally, human vision has a very wide field of view of 200 degrees horizontal and 125 degrees vertical. Peripheral vision plays an important role in determining situational awareness and action [2]. In a wearable environment, various advantages are obtained if a display device can present information to the peripheral visual field. For example, information can be superimposed by AR to the entire field of view. This can improve the efficiency and safety of the real-world tasks such as driving directions and monitoring. Moreover, considering the sensitivity of visual receptors, information can be presented more flexibly, e.g. to display critical information in the central view, and noncritical information in the periphery.

We have previously proposed a variation of a head mounted projective display (HMPD) that provides both a wide field-of-view and see-through capability, using a hyperboloidal half mirror (Hyperboloidal Head Mounted Projective Display, HHMPD) [3]. The original HHMPD requires a retro-reflective screen placed in the real environment and is unable to be used in a wearable environment. In order to make the HHMPD usable in a wearable environment, we have been building a prototype system with simple semi-transparent screens that combines semi-transparent and retro-reflective [4]. In this paper, we integrate an actually semi-transparent retro-reflective screen into the HHMPD, and report on subjective evaluation experiments conducted on visual acuity and perceptual distance of the projected image.

In the following, Section 2 briefly summarizes basic characteristics of a HHMPD design and its prototype [3]. Section 3 introduces a number of design considerations for a semi-transparent retro-reflective screen [4]. Sections 4 and 5 describe the conducted subjective experiments and Section 6 gives conclusions and future directions.

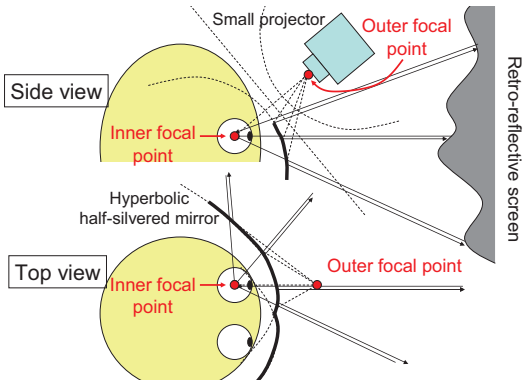


Figure 1. Schematic diagram of a HHMPD

2 CHARACTERISTICS OF HHMPD

The basic concept of the HHMPD is to employ a curved combiner rather than a planar combiner to diverge light rays to acquire a wider FOV. Every light ray reflected on the combiner should eventually travel back toward a single point, the user's eye. This constraint indicates that the combiner should be a hyperboloidal surface. Hyperboloidal mirrors have been widely used in computer vision [5]. However, our HMPD is thought to be the first display device to utilize a semi-transparent hyperboloidal mirror.

Figure 1 shows an overview of the design of the HHMPD. Projectors are placed at the outer focal points of the hyperboloidal semi-transparent mirrors, and the viewer observes stereo imagery from the mirrors' inner focal points. The axes of the hyperboloids are inclined to achieve a wide FOV without occlusion from the projectors. As described later in detail, an HHMPD can provide a

1-32 Machikaneyama, Toyonaka, Osaka 560-0043, Japan
Cybermedia Center, Osaka University
nguyen.van.duc@lab.ime.cmc.osaka-u.ac.jp,
{mashita, kiyo, takemura}@ime.cmc.osaka-u.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

very wide FOV with a normal projector that has a moderate projection angle.

A head mounted stereo prototype HHMPD was built using a pair of custom-made mirrors (see Figure 2) and two pocket projectors (3M MPro110, VGA, 17.7 by 14.4 degrees). It provides a 109.5-degree horizontal view angle and a 66.6-degree vertical view angle. As reported in [3], note that the HHMPD's optical design is theoretically capable of providing a horizontal field of view wider than 180 degrees, if appropriate mirror parameters and wider horizontal projection angles (~50 degrees) are given.

The primary advantages of the HHMPD include:

- Large observational pupil: As in the case of the conventional HMPD, a user observes a projected image on a retro-reflective screen a few meters away from the eyes. With an appropriately reflective screen, the observational pupil can be very large, making image visibility robust to eye rotation. This is important because eye rotation is likely to occur more frequently with a wide FOV image.
- Large binocular overlap: Owing to the curved shape, the HHMPD can provide a large binocular overlap, up to approximately 120 degrees, which is larger than that of a conventional HMPD.
- Small mirror size: Owing to the curved shape, the HHMPD can be much smaller for the same FOV with a more natural glasses-like appearance, compared to a conventional HMPD with a planar mirror.
- Wide range of applications: The HHMPD can be used, e.g., as an alternative to immersive projection technology (IPT) displays and for multi-user collaboration that requires wide FOV images. By adding a camera at the position of the projector using another optical combiner, taking wide FOV pictures from the user's viewpoint becomes possible [6], which is otherwise very difficult. The last example is useful for human activity analysis and attentive interfaces, for instance.

The main disadvantages of the HHMPD include:

- Low resolution: Since the entire FOV is covered by a single projector, the angular pixel resolution is decreased accordingly.
- Image distortion: Projected imagery has distortion caused by the curved mirror. However, this can easily be compensated by pre-distorting the rendering image.
- Defocus: The HHMPD, as well as a conventional HMPD, must project an image onto a retro-reflective screen without defocusing. Unlike a conventional HMPD, the basal plane of the projection frustum in the HHMPD is no longer planar, but is rather a curved surface. This means that keeping the entire projected image in focus is difficult. Dedicated projector optics, special screen geometry, or an anti-defocus projection is required to alleviate this problem.
- Last but not least, as in the case of a conventional HMPD, the HHMPD requires a retro-reflective screen, making it difficult to use in a wearable environment.

Figure 2. A stereo prototype of HHMPD

3 Semi-transparent Retro-reflective Screen

In principle, a retro-reflective screen has no transparency. In order to make a retro-reflective screen semi-transparent, many techniques have been proposed such as using an optical combiner [7], a rotational time division screen [8], or applying different optical principles [9]. We propose a simple pupil division screen and a vibrating screen. In the following, an overview and characteristics of each screen are described.

3.1 Pupil division screen

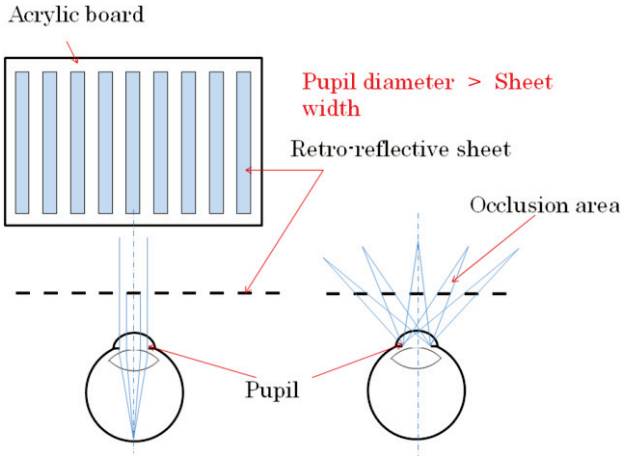


Figure 3. Pupil division screen

A pupil division screen is composed of many thin strips of retro-reflective material in a stripe whose width are smaller than a pupil diameter attached on a transparent substrate (such as an acrylic plate) to achieve both transparency and retro-reflection. This method does not need to move the screen so it is inexpensive and safe for the user. With a pupil division screen, the transparent part of the screen allows for viewing the real world, but there is a problem that the projected image on the transparent part is not retro-reflected and is missing. However, if the virtual image is distant from the screen, there will be no missing region in the observed image (Figure 3).

3.2 Vibrating screen

A vibrating screen is a type of moving screen. An example is shown in Figure 4. In this example, the screen is configured to move in a direction parallel to the transparent plate, and perpendicular to the stripe of the retro-reflective material. To

accommodate a wide FOV, a cylindrical screen can be moved along the arc as shown in Figure 4(c) or multiple screens can be used as shown in Figure 4(b).

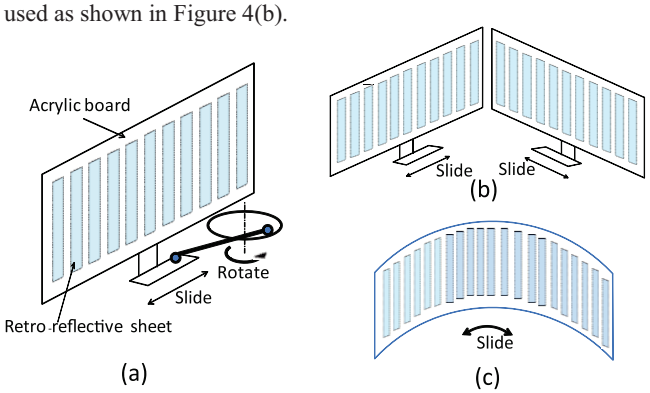


Figure 4. Vibrating screen

3.3 Prototype of semi-transparent retro-reflective screen

A vibrating screen and a pupil division screen were prototyped by attaching a retro-reflective sheet (3M Scotchlite High Gain Retro-reflective Sheeting 7610) cut into strips at a regular interval on an acrylic plate. The strip width and spacing between the strips are about 0.35mm for the pupil division screen and about 1.0mm for the vibrating screen. In addition, the base mechanism for vibration of the screen was produced by modifying a commercial CD drive unit. The vibration stroke of the screen is 36mm and the oscillation frequency is 5.5Hz. Thus, the average switching frequency between semi-transparency and retro-reflection is about 100Hz.

Photos were taken by a digital camera Sanyo Xacti HD1010 from an inner focal point of the HHMPD to verify that the prototype screens have characteristics of both retro-reflection and semi-transparency (see Figure 5). Figure 6 shows the captured pictures. Distance between the camera and the prototype screen is 15cm. The projected image is configured to focus on the screen. Distance between a reference real object (checkerboard) and the camera is 250cm. Focus F of the camera is set to 250cm and 15cm, and the shutter speed S is set to 1/8s considering the temporal characteristics of human motion perception [12]. The effective diameter of the lens is 3.5mm, close to the human pupil diameter under normal conditions. White fluorescent lights were used in the darkroom. In this experiment, a planar half mirror was used instead of a hyperboloidal half mirror.

As shown in Figure 6, it is clear that the virtual image (alphabet letters and numbers) and the reference real image (checkerboard) can be observed simultaneously. That is, it is confirmed that the prototype screens work as a semi-transparent retro-reflective screen that has both retro-reflective and transparent properties. With the pupil division screen, visibility of the real world is very poor when the focus F is near. And when F is far enough, the real images are able to be observed without missing regions. In addition, when F is near the screen gaps are apparent. When F is far the screen gaps are much less noticeable. In the latter case, the alphabet letters in the third row from the top can be recognized. The size of these letters is about 5mm by 3mm on the physical screens that is equivalent to visual acuity of around 0.07 or 14 minarc. Note that the visual field of each picture is about 30 x 20 degrees.

With the vibrating screen, when the focus F is far then the results are almost the same as those with the pupil division screen. On the other hand, when F is near, the visual quality significantly improves compared to that of the pupil division screen.

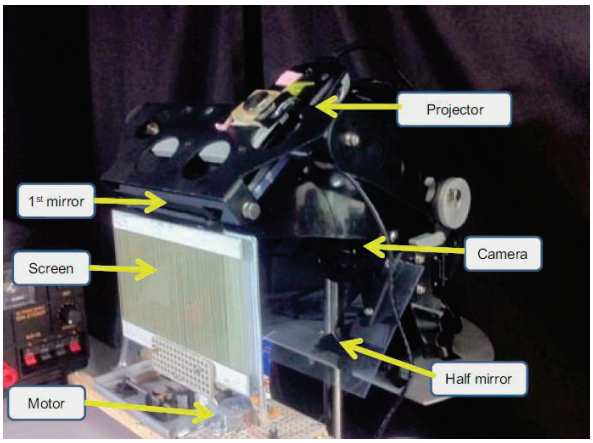


Figure 5. Vibrating screen with HHMPD

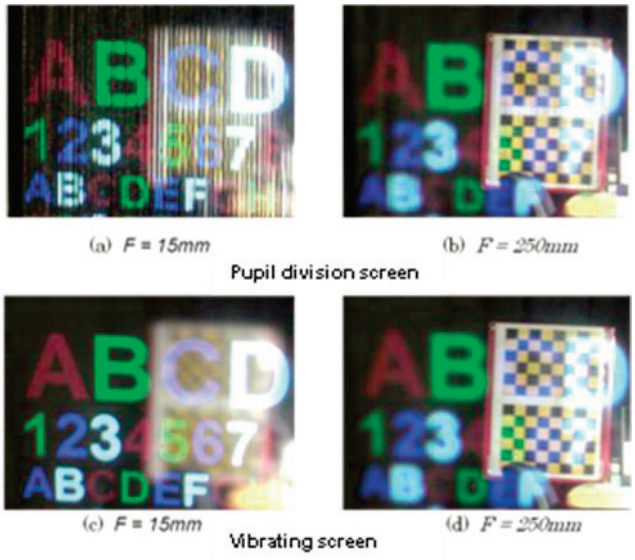


Figure 6. Photos taken through prototype screens

4 Experiment 1: Subjective evaluation on perceptual distance

4.1 Objective

The purpose of this experiment is to investigate the relationship between the physical observation distance in the real world and the perceived distance of the projected image using the HHMPD with a curved semi-transparent retro-reflective screen (hereinafter referred to as a wearable HHMPD, shown in Figure 7). In AR applications, it is very important to be able to simultaneously observe virtual information and the real environment that the virtual information is referring to. It is also often desirable that those two types of visual stimuli are observed perceptually at the same distance. However, it is unclear if this simultaneous observation is comfortably possible because the perceived distance of a projected image on the semi-transparent retro-

reflective screen, that is only 15cm in front of the user's eye, is expected to be very short. Note that this problem cannot be solved by a pin-hole projector or a laser-projector as the image is formed near the screen distance whereas the user needs to see further real objects. Stereoscopic viewing will help the user observe the projected image at an intended distance. However it is of our interest to investigate the fundamental properties in a monocular setup as a first step. Using a monocular setup, Zhang et al. [10] report that a perceived distance of the projected image is generally influenced by both the distance between the projector and the screen and the projector's focal length. In other words, a perceived distance of the projected image can be larger than the distance between the projector and the screen. However, such configurations require special retro-reflective materials such as high precision corner cubes that the current system does not have. Thus, the experiment is configured to use the projection distance consistent with the screen distance to investigate whether the projected image is perceived at a similar distance as the real environment that the projection is superimposed onto.

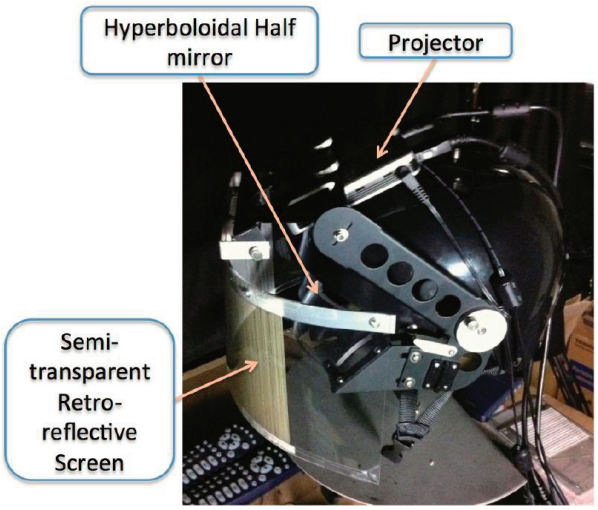


Figure 7. Wearable HHMPD

4.2 Procedure

Figure 8 shows the configuration of the experiment 1. Figure 9 shows visual stimuli presented to subjects (top) and the experiment environment (bottom). Each subject's head was fixed and equipped with the wearable HHMPD. Subjects observed the image with their dominant eye. The right half of the physical board presented a radial pattern, and the virtual image was superimposed in the left half. The degree of perceptual distance between the virtual and real images was subjectively evaluated in a five-step Likert scale (see Table 1). After exposing the subjects to the real pattern once at a distance of 1.0m and 4.0m, they were asked to indicate the level of agreement in perceptual distance between the virtual and real images by changing the position of the physical board. The board was placed at seven positions with 0.5m intervals, from 1.0m up to 4.0m. The experiment was then continued with the board being moved closer to the subject (in a reverse order, from 4.0m to 1.0m with 0.5m intervals). This procedure obtained ratings for 7 positions twice for each subject. Subjects could see the physical board being moved back and forth but had no information as to the actual distance it was placed at.

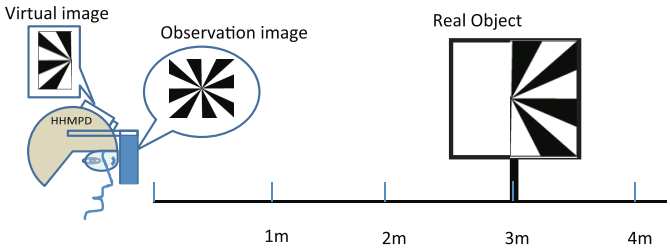


Figure 8. Configuration of experiment 1

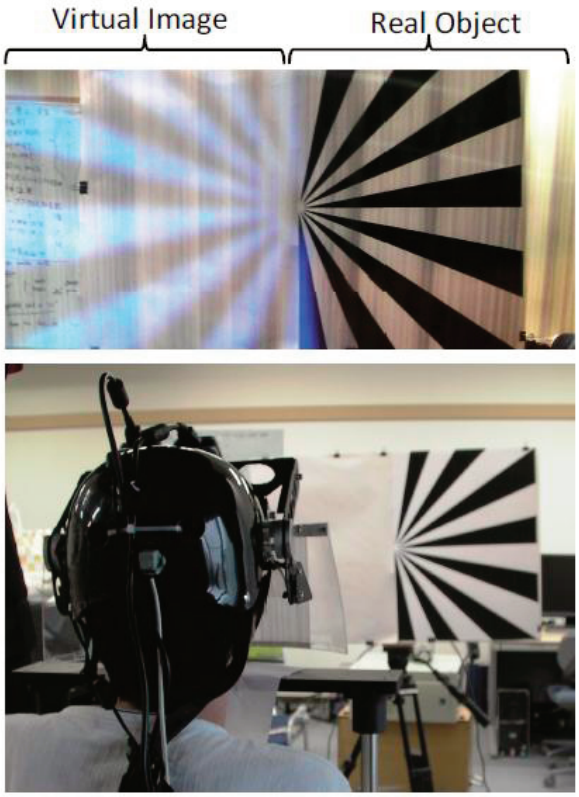


Figure 9. Experiment 1: example of visual stimuli (top) and experimental environment (bottom)

Table 1. Rating criteria in experiment 1

Does the real image appear to be at the same distance with the projected image?	Rating
Strongly agree	5
Agree	4
Neither agree nor disagree	3
Disagree	2
Strongly disagree	1

4.3 Result and discussion

We conducted this experiment with 10 test subjects (graduate and undergraduate students). Figure 10 shows the result including the averages and standard errors of the rating obtained from 20 samples for each distance. As shown in Figure 10, the level of agreement in perceptual distance between the virtual and real images is decreased rapidly with increasing observation distance.

This result shows that in this experimental configuration the virtual image is perceived at a similar distance as the real image only when the observation distance is within 2m. At the same time, this result also shows that subjects felt noticeable inconsistency between the virtual and real images only when the observation distance is beyond 3m. This result indicates that our simple prototype display is applicable to indoor and tabletop AR applications where the observation distance is relatively small.

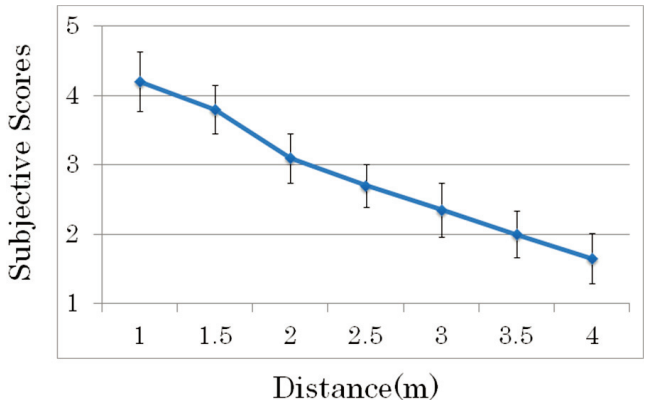


Figure 10. Results of experiment 1

5 Experiment 2: Subjective evaluation on visual acuity

5.1 Objective

The purpose of this experiment is to investigate the perceived visual resolution of the projected image when a user sees it while observing the real environment at the same time using the wearable HHMPD. It is expected to be able to observe the projected image in its highest resolution determined by angular resolution of the projected image and slit intervals of the screen, when focusing on the retro-reflective screen that is 15cm in front of a user. However, as described in the previous section, virtual and real images often need to be observed at the same time in many AR applications. The projected image will get blurred when focusing on the real environment and the real environment will get blurred when focusing on the projected image. The two types of visual stimuli will appear perceptually at different distances. Therefore it is practically of high importance to investigate how detail the projected image can be observed when focusing on the real environment. In this experiment, we use Landolt rings, commonly used for visual acuity test, as visual stimuli and investigate the minimum apparent size of the virtual and real Landolt rings that are presented in a short period of time and yet simultaneously observable.

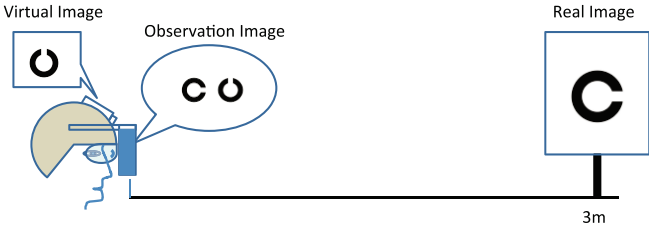


Figure 11. Configuration of experiment 2

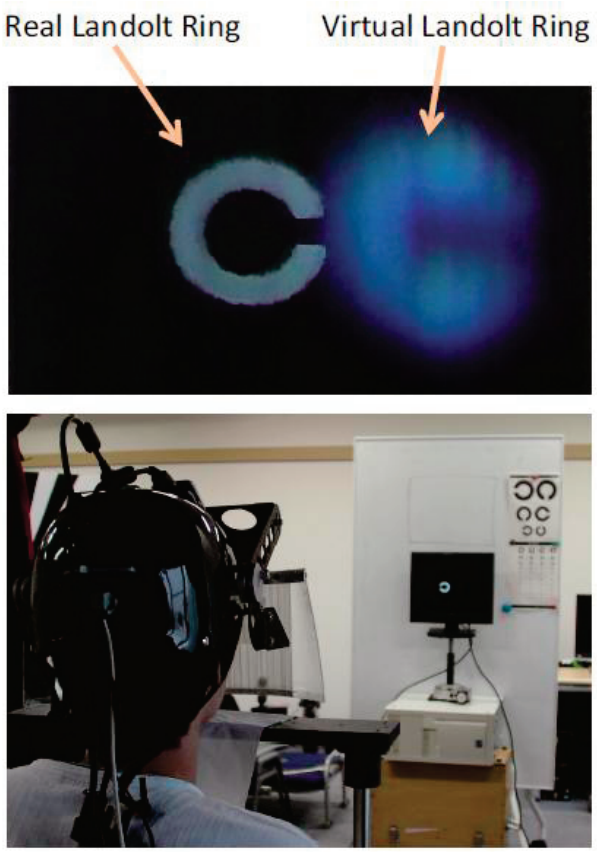


Figure 12. Experiment 2: example of visual stimuli (top) and experimental environment (bottom)

5.2 Procedure

Figure 11 shows the configuration of the experiment 2 while Figure 12 shows the entire environment of the experiment and an example of visual stimuli. Each subject's head was fixed and equipped with the wearable HHMPD. Subjects observed the stimuli with their dominant eye. The visual stimuli presented to the subjects at the same time are two Landolt rings. One is virtual, presented by the HHMPD and the other is real, displayed on an LCD monitor. The semi-transparent retro-reflective screen used in the experiment 2 is a pupil division screen same as the experiment 1, however this time is fixed on the vibrating mechanism. In this case the HHMPD is not wearable as the resulting vibrating screen is not fixed to it.

The real Landolt ring is presented on a 15-inch LCD monitor located at 3m from the HHMPD. The virtual Landolt ring is presented as a virtual image having the same apparent size as the real Landolt ring to its right. Each Landolt ring was presented in one of four orientations (up, down, left and right) randomly and the subjects had to answer as to whether or not the rings had the same orientation. The visual acuities corresponding to the apparent sizes of Landolt rings are 0.05, 0.075, 0.1 and 0.2 and were presented in this order. These parameters are chosen because of the fact that the visual acuity calculated from the angular resolution of the projection image is around 0.2, and that given by the slit intervals of the pupil division screen (0.35mm) and the screen distance from the eye (15cm) is around 0.25. We determine that a subject could observe Landolt rings in their size when three

or more answers are correct out of five trials. During the experiment subjects were not told if their answers were correct or not.

In this way, we determine the visual acuity of the projected image for each subject for each size. In each trial, the Landolt rings are presented to the subjects for approximately 400ms [10], to avoid observation of the two rings sequentially by changing their focus.

5.3 Result and discussion

We conducted this experiment with 8 test subjects (graduate and undergraduate students). Six of eight subjects joined the experiment 1. The results are shown in Table 2 and Figure 13. Table 2 shows the normal visual acuity of each subject measured just before the experiment. These results show the visual acuity of the projected image is in the range between 0.05 and 0.1 for all subjects and conditions. It also shows that the visual acuity is higher with the vibrating screen than with the static pupil division screen. In addition, there is a positive correlation between the visual acuity of the projected image with the vibrating screen and the subjects' natural visual acuity ($r = 0.76$) but no correlation was found when the static pupil division screen was used ($r = 0.23$). These results are comparable to the visual acuity of the projected image estimated from captured pictures (approximately 0.07) as written in Section 3.3. Through this experiment, it was confirmed that the wearable HHMPD is applicable to indoor AR applications if coarse resolution suffices.

Table 2. Subjects' visual acuity

	S1	S2	S3	S4	S5	S6	S7	S8
Normal	0.4	0.7	0.7	0.8	1.2	1.2	1.5	1.5
Pupil	0.05	0.05	0.075	0.05	0.075	0.1	0.075	0.1
Vibrating	0.1	0.1	0.1	0.075	0.1	0.1	0.1	0.075

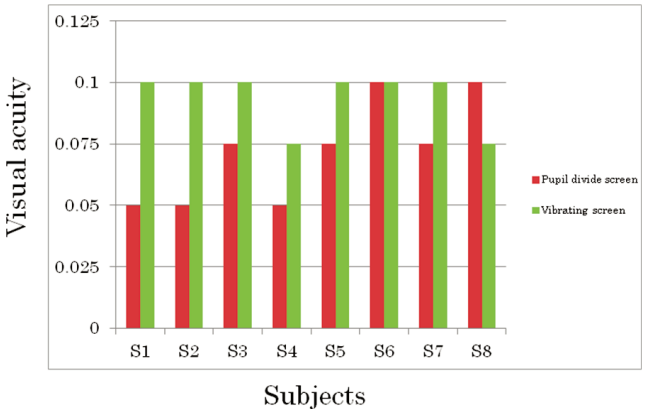


Figure 13. Result of experiment 2

6 Conclusion

In this paper, we reported a simple semi-transparent retro-reflective screen for a Hyperboloidal Head Mounted Projective Display (HHMPD) to be usable in a wearable scenario. A wearable HHMPD with the semi-transparent retro-reflective screen was built and the visual quality of the projected image was studied through subjective evaluation experiments. The experimental results show that subjects did not feel inconsistency in the perceived distance between the real environment and the superimposed projected image. The projected image was observable with the visual acuity of 0.05 to 0.1 when focusing on the real object at a distance of 3m. The visual acuity of the projected image estimated from captured pictures is around 0.07, and comparable results were acquired by the user studies. So these will be the upper bound of the visual acuity of the projected image observed by a human eye with the configuration of the prototype used. As future work we plan to improve the display in terms of visual quality, size and weight, and to investigate applicability of the wearable HHMPD in a wide mobile environment.

Acknowledgement

This research was funded in part by Grant-in-Aid for Scientific Research (B), #22300043 from Japan Society for the Promotion of Science (JSPS), Japan.

References

[1] Cakmakci, O. and Rolland, J., "Head-worn Displays: A Review," Journal of Display Technology, Vol. 2, No. 3, pp. 199-216, 2006.

[2] Arthur, K. W., "Effects of field of view on performance with head-mounted displays," ISBN:0-599-73372-1, University of North Carolina at Chapel Hill, Doctoral Thesis, 2000.

[3] Kiyoshi Kiyokawa, "A Wide Field-of-view Head Mounted Projective Display using Hyperbolic Half-silvered Mirrors," Proc. IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR), pp. 207-210, 2007.

[4] Nguyen van, D., Mashita, T., Kiyokawa, K. and Takemura, H., "Design Consideration for Semi-transparent Retro-reflective Screen for a Wide-view Head Mounted Projective Display using a Hyperboloidal Half-mirror," IEICE Technical Report, MVE2010-119 (non reviewed), 2011.

[5] Yamazawa, K., Yagi, Y. and Yachida, M., "Omnidirectional Imaging With Hyperboloidal Projection," Proc. the 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1029-1034, 1993.

[6] Sumiya, E., Mashita, T., Kiyokawa, K. and Takemura, H., "A Wide-view Parallax-free Eye-mark Recorder with a Hyperboloidal Half-silvered Mirror," Proc. ACM Symposium on Virtual Reality Software and Technology (VRST), pp. 19-22, 2009.

[7] Martins, R., Shaoulov, V., Ha, Y. and Rolland, J., "A Mobile Head-worn Projection Display," Optical Express, Vol. 15, No. 22, pp. 14530-14538, 2007.

[8] Nojima, T. and Kajimoto H., "A Study on a Flight Display using Retro-reflective Projection Technology and a Propeller," Journal of the Virtual Reality Society of Japan, Vol. 13, No. 2, pp. 217-226, 2008. (in Japanese)

[9] Kijima, R. and Watanabe, J., "False Image Projector for Head Mounted Display using Retrotransmissive Optical System," Proc. IEEE Virtual Reality, pp. 297-298, 2009.

[10] Zhang, R. and Hua, H., "Effects of a Retroreflective Screen on Depth Perception in a Head-mounted Projection Display," Proc. IEEE Symposium on Mixed and Augmented Reality (ISMAR), pp. 137-145, 2010.

[11] Wada, I., Murayama, N. and Neshige, R., "Visual Recognition and Event-Related Potentials(P300)," IEICE Technical Report. ME and Bio Cybernetics 95(501), pp. 31-36, 1996. (in Japanese)

[12] Fredericksen RE, Verstraten FA, van de Grind WA., "Spatio-temporal Characteristics of Human Motion Perception.," Vision Res. Vol. 33, No. 9, pp. 1193-1205, 1993.

An Evaluation of Augmented Reality X-Ray Vision for Outdoor Navigation

Arindam Dey*
Magic Vision Lab
University of South Australia

Graeme Jarvis†
Magic Vision Lab
University of South Australia

Christian Sandor‡
Magic Vision Lab
University of South Australia

Ariawan Kusumo Wibowo§
Magic Vision Lab
University of South Australia

Ville-Veikko Mattila¶
Nokia Research Center
Nokia

(a) (b) (c) (d) (e)

Figure 1: We have ported our recent AR X-ray prototype [16] to a mobile phone (a) and conducted an evaluation of its effectiveness for outdoor navigation, comparing it against standard mobile navigation applications. Participants had to complete a 900 meter route (b). Our core finding is that AR X-ray required significantly less context switches than other conditions. Heatmap visualizations indicate the number of context switches on the path, averaged over all participants: AR X-ray (c), North-up map (d), and View-up map (e).

Abstract

During the last decade, pedestrian navigation applications on mobile phones have become commonplace; most of them provide a birds-eye view of the environment. Recently, mobile Augmented Reality (AR) browsers have become popular, providing a complementary, egocentric view of where points of interest are located in the environment. As points of interest are often occluded by real-world objects, we previously developed a mobile AR X-ray system, which enables users to look through occluders.

We present an evaluation that compares it with two standard pedestrian navigation applications (North-up and View-up map). Participants had to walk a 900 meter route with three checkpoints along the path. Our main findings are based on the analysis of recorded videos. We could show that the number of context switches is significantly lowest in the AR X-ray condition.

We believe that this finding provides useful design constraints for any developer of mobile navigation applications.

Keywords: Augmented Reality, Evaluation, Visualization, Augmented Reality X-ray, Mobile Phone, Map, Navigation

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities H.1.2 [Models and Principles]: User/Machine Systems—Human factors

*e-mail: arindam.dey@unisa.edu.au
†e-mail: fanged@gmail.com
‡e-mail: chris.sandor@gmail.com
§e-mail: kusay003@mymail.unisa.edu.au
¶e-mail: ville-veikko.mattila@nokia.com

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

1 Introduction

During the last decade, pedestrian navigation applications on mobile phones have become commonplace; most of them provide a birds-eye view of the environment. There are many mobile applications providing navigation information along with landmarks to show points of interest such as Google maps and Nokia's Ovi maps. While these applications are widely used, they only provide an exocentric two dimensional view of the environment.

On the contrary, an environmental image being a combination of immediate sensation and past memory, is considered to be a strategic link in the process of way-finding and is used to interpret information and guide action [9]. This fact motivates us to provide a pictorial representation of the destination along with the immediate environmental image into a pedestrian navigation system which current map applications fail to provide.

Recently, mobile Augmented Reality (AR) browsers have become popular, providing a complementary, egocentric view of where points of interest are located in the environment. These applications commonly show points of interest on top of the real world, irrespective of their actual visibility. This causes several perceptual problems; most importantly, as occlusion is the most important depth cue [18], distances are hard to estimate. We have previously presented several AR see-through vision systems [2, 16, 17], which aim to improve the perception of occluded objects.

This paper first describes how we have ported our most recent AR X-ray system [16] to a mobile phone. Based on this platform, we have conducted an evaluation that compares it with two standard pedestrian navigation applications (North-up and View-up mobile map). Participants had to walk a 900 meter route with three check-

points along the path. We collected a large quantity of data during these trials: logged tracking data, completion time, and videos. Our main finding is based on the analysis of recorded videos. We were able to show that the number of context switches is significantly lowest in the AR X-ray condition. We believe that this finding provides useful design constraints for any developer of mobile navigation applications (see Section 5).

1.1 Related Work and Contribution

In this section, we first discuss related evaluations of mobile navigation applications. Second, we discuss related work on AR X-ray visualizations and their evaluation. Finally, we highlight our contribution based on the discussion of related work.

Since the first mobile pedestrian navigation application [1] was presented around 15 years ago, many evaluations have been conducted in this space. Typical topics of these studies were comparing 2D to 3D maps and also introducing novel navigation cues. A 3D map was found to be advantageous over a 2D map [14, 6]. While, 3D maps provide a more realistic and volumetric representation of the real environment, 2D maps enhance the use of previous knowledge effectively and reduce cognitive load [11]; for example, for an expert 2D map user, these advantages are minimal [8]. Several studies investigated the enhancement of common navigation aids through tactile feedback: paper map [12] and mobile maps [13]. In the same spirit as us, Rukzio and colleagues have evaluated common navigation aids against new paradigms for navigation: public displays and a rotating compass [15].

Various AR prototypes were built to provide location-based information; for example for tourist guide applications. A core challenges in these browsers is to show occluded points of interest. We have previously implemented several AR X-ray prototypes to address this challenge by experimenting with different visualization techniques: edge-overlay [2] and saliency [16]. We have also experimented with space-distorting visualizations to remove occluder objects in an intuitive way [17].

While our AR X-ray systems aim to create photorealistic renderings of occluded points of interest, most other research has focused on symbolic representations. Livingston et. al [7] have evaluated such a system through depth perception tasks. We have also previously evaluated two see-through visualizations using a handheld display [4]. Recently, a zooming interface for AR browsers was evaluated with an orientation task [10]. However, we are not aware of any evaluations that have evaluated the effectiveness of AR X-ray as a navigation aid.

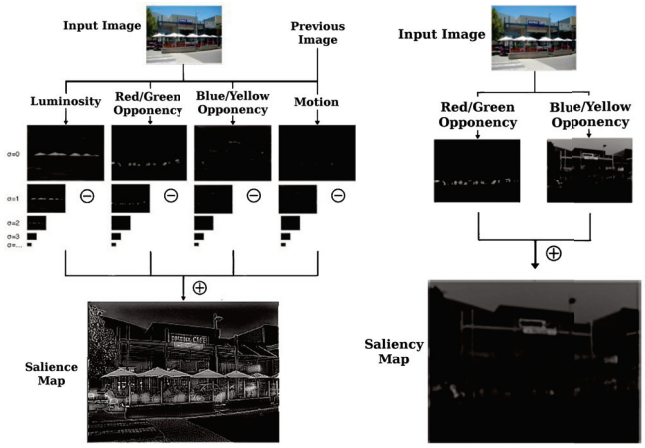
Contribution The core contribution of this paper is to present the first evaluation of an AR X-ray system in a navigation task. We could show that the number of context switches is significantly lower than with standard map applications on a mobile phone.

A side contribution of this paper is the porting of our previous AR X-ray system [16] to a mobile phone. This required us to perform several optimizations and adaption of our algorithms. Despite the limited computation power of mobile phones, we were able to achieve visually quite similar results to our previous prototype that ran on a laptop.

2 Experimental Platform

We have ported our previous AR X-ray system [16] to the Nokia N900 mobile phone. Our original system ran on a laptop and was developed using Python and OpenGL 2.1. The N900 port uses C++ and OpenGL ES 2. Built on Qt4.7, the application runs as a plug-in module for Nokia's proprietary Mixed Reality Framework (MRF). MRF exposes the various sensors available on the N900 in a manner that is much easier to use than the native API, streamlining the initial setup of an AR application. The device's pose is determined by an externally attached sensor box connected via bluetooth. The

Figure 2: Comparison of our previous prototype running on a laptop (left) and our N900 port (right). The visual appearance is quite similar.



(a) Laptop (b) Phone

Figure 3: Saliency computation

sensor box provides an 'orientation' software sensor, a fusion of data from hardware sensors; compass and accelerometer.

The porting of the AR X-ray system was successful, and produced results that are quite similar to our previous system running on a laptop (see Figure 2). In order to achieve an acceptable frame rate on the mobile phone, we had to perform three simplifications to our algorithm (see Figure 3); we removed three computations: mipmapping, motion saliency, and intensity saliency.

Mipmaps form an image pyramid, which provides multi-resolution images for feature detection and saliency calculation. This is a core part of the saliency calculation in our AR X-ray system, and must be run every frame. Benchmarking showed that mipmapping on our mobile phone accounted for approximately 500ms of rendering time per frame, which is significantly too slow. The reason for this (also confirmed by Nokia's driver developers) is that the mipmapping routine in the N900 is not very optimized. Typical mobile 3D applications, such as games, run the mipmapping step only once at startup, when loading the textures. We attempted several fixes to alleviate the mipmapping limitation without success including manual generation of mipmaps. Finally, as we could not overcome the performance problems, the mipmapping was removed completely, with the results being very comparable to using mipmapping previously.

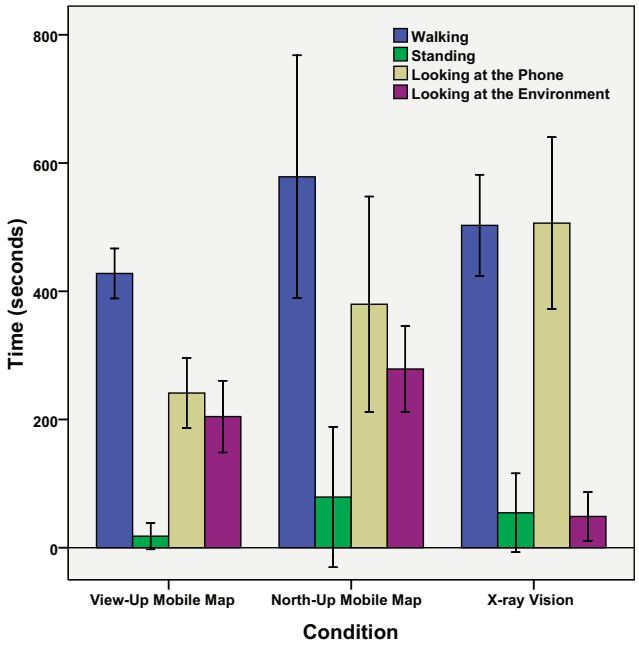
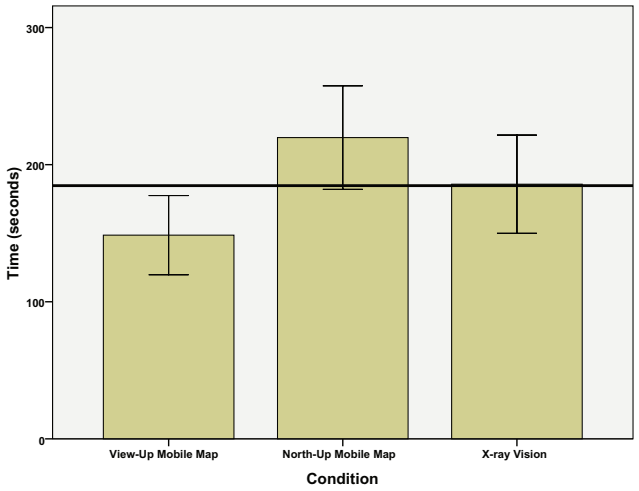
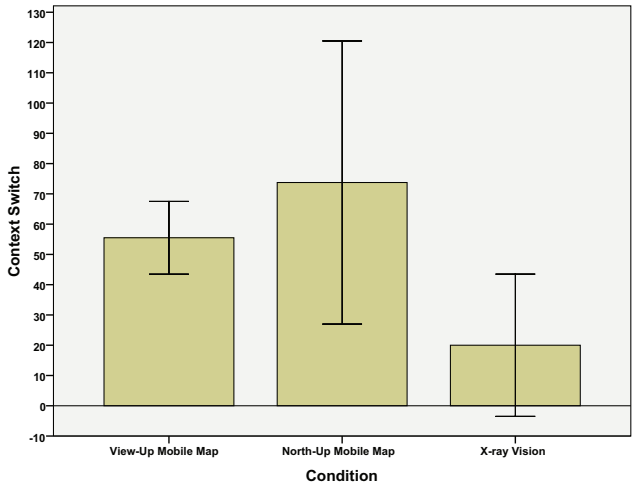


Figure 4: Video analysis: comparison of time spent in different activities. Compared to other conditions, participants looked significantly less at the environment and significantly more on the phone in the AR X-ray condition. Whiskers represent $\pm 95\%$ confidence intervals.



(a) Task completion time: Though there were no significant difference between AR X-ray vision and two other conditions; View-up map was significantly faster than North-up map.



(b) Video analysis: number of context switches. AR X-ray caused significantly less context switches compared to other conditions.

Figure 5: Further results. Whiskers represent $\pm 95\%$ confidence intervals and the thick Black lines represent overall mean.

3 Evaluation

12 voluntary participants (all male) with ages ranging from 22 to 40 years were recruited from the student population of the University of South Australia. In a between-subjects design, we divided the 12 participants into three groups of four participants; each group was exposed to one of the three conditions, as described in Section 3.1. We selected three different target locations on the campus of the University of South Australia. The locations were carefully chosen to be among the least accessed in the campus. The average length of path segments was 289 meters (SD=91.8). Each participant traveled to all of the target locations in the same order using the assigned condition. The entire experiment took about 30 minutes per participant.

We instructed participants to navigate to the target location as they would have done normally in their day to day life. We did not specify any predefined path, as we wanted to investigate the difference in choice of paths using the different conditions. We asked participants to speak out loud while navigating.

3.1 Conditions

AR X-ray vision as a navigation aid was the focus of our evaluation. In this condition, participants were provided with a mobile phone where only a photorealistic view of the target location was displayed through our AR X-ray vision. No other information such as route direction or distance to the target were provided. After participants reached a target location successfully, the next target location was loaded by the experimenter and presented to the participant. Overall, there were three target locations

As baseline conditions, we used two standard pedestrian navigation applications (North-up and View-up map). As North-up map, we ran Nokia's Ovi Maps on the N900. As View-up map, we used Apple's Maps application, which is preinstalled on iPhones, on an iPhone 3GS in View-up mode. We could not use Ovi Maps for this condition, as Ovi Maps does not support View-up maps. However, the appearance of both of the mobile maps were verified to have similar legibility. In the case of both of these mobile maps, only a target location was marked at one time on the map with a pin. Similar to the AR X-ray vision condition, once participants reached the location the next location was marked with a pin.

We collected three different types of data: task completion time, GPS tracks, and video recordings. Task completion time was measured using a stopwatch. Our main data source were video recordings of participants. We externally recorded participants throughout their travel to the target locations using a Canon 550D camera at 60 fps. Later, we analyzed the video by identifying different behaviors of participants.

4 Results

With regards to task completion time (see Figure 5(a)), the only significant difference was that View-up map was faster than North-up map (determined by a one-way ANOVA with $F(2,9) = 6.15; p = .017, \eta^2 = .58$). In the following, we focus on the results from the video analysis. We collected 243 minutes of video data for our 12

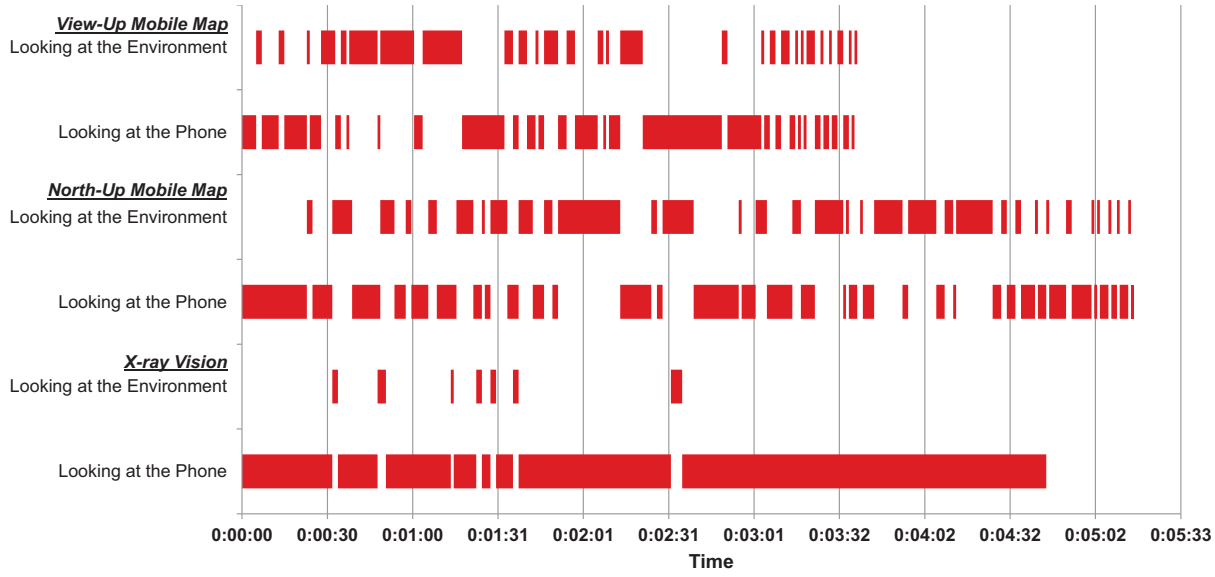


Figure 6: Video analysis: raw segmentation results for three typical trials. The number of context switches is clearly less for AR X-ray vision.

participants.

The video analysis yielded non-significant differences for disorientation and walking/standing across conditions. Walking/standing refers to the ratio of time that a participant spent in each mode. There was no significant interaction effect between walking/standing and the condition (see Figure 4). We define disorientation as participants standing at a fixed position for more than five seconds. All together, there were 82 occasions (View-Up map: 17, North-Up map: 40, AR X-ray: 25) when participants stopped while performing the navigation task. Out of these stops, the number of times when participants stopped for more than five seconds was: View-up map: 6, North-up map: 24, and AR X-ray: 13.

However, we could identify two significant effects in the video analysis: target of user's gaze (environment or mobile phone) and context switches. With regards to the target of the user's gaze, there was a significant main effect $F(2,9) = 6.21; p = .02; {\eta_p}^2 = .6$; in all conditions the environment was looked at less than the mobile phone. There was a significant interaction effect between condition and gaze target $F(2,9) = 25.56; p < .001; {\eta_p}^2 = .85$. In the AR X-ray condition, the gaze ratio of phone to environment was significantly higher than in any other condition (see Figure 4). A context switch was measured when participants switched their gaze from the mobile phone's screen to the environment. An one-way ANOVA showed a significant main effect of conditions on context switch $F(2,9) = 7.87; p = .011; \eta^2 = .64$ (see Figures 5(b), 6, and Figure 1(c-e)). AR X-ray had least context switch among all of the conditions. A Tukey's HSD post-hoc test revealed that the difference was significant (p=.009) with the View-Up map, but not with the North-Up map (p=.07). The average time after which a context switch occurred was: View-up map: 8.36 seconds, North-up map: 9.7 seconds, and AR X-ray: 31.6 seconds.

5 Discussion

In this paper, we have presented the first evaluation of an AR X-ray system in a navigation task. In order to perform this study, we have ported our previous AR X-ray system to a mobile phone.

The most important result of our evaluation is based on the analysis of recorded videos. In the AR X-ray condition, the number of context switches is significantly lowest; additionally, participants looked significantly more at the mobile phone than at their environment. Even with a lower number of participants in our experiment, the results showed a higher level of effect size. This result is not surprising, as the AR view on the mobile phone enables users to observe their environment and the navigation cues simultaneously as there is no need to look at the environment directly.

The number of context switches is closely linked to attention: more context switches consume more of the user's attention. The amount of free attention is positively correlated with perceptual, cognitive, and motor tasks. In our experiment, the AR X-ray condition required less eye-movements due to a better spatial relation between stimuli (AR X-ray depiction of the target) and responses (walking direction), also known as stimulus-response compatibility. A better stimulus-response compatibility is known to enable the user to perform more accurate actions [3].

So, we believe that our result is valuable, as it indicates the benefits of AR as a navigation aid, which consumes less attention of the user; therefore, resulting in a more efficient navigation. Any other task that requires the user's attention simultaneously to be on the environment and at the same time to be on some additional information about the environment can benefit from AR as well; for example, maintenance instructions while performing maintenance [5].

In the future, we want to further investigate the possibilities of using AR as a mainstream navigation aid particularly for pedestrians. Additionally, we plan to further improve the speed of our prototype. We also plan to compare AR X-ray against standard AR browsers and other mobile map applications. It will also be valuable to validate our study in a city center with more participants. It will be interesting to perform similar experiments in different social circumstances such as busy streets, unfamiliar location, and time critical situations and investigate the differences in results. We believe that our photorealistic depiction of points of interest will aid users significantly to build a richer mental model of their environment. However, we would like to develop an optimized visualization for the mobile phones despite its low resolution and limited processing power.

Acknowledgements

The authors wish to thank the voluntary participants of the evaluation. This research was funded by Nokia Research Center.

REFERENCES

[1] G. D. Abowd, C. G. Atkeson, J. Hong, S. Long, R. Kooper, M. Pinkerton, and U. Centre. Cyberguide: a mobile context-aware tour guide. *ACM Wireless Networks*, 3:421–433, 1997.

[2] B. Avery, C. Sandor, and B. Thomas. Improving spatial perception for augmented reality x-ray vision. In *Proceedings of the IEEE Virtual Reality Conference*, pages 79–82. IEEE, 2009.

[3] R. Chua, D. J. Weeks, and D. Goodman. The human-computer interaction handbook. chapter Perceptual-motor interaction: some implications for human-computer interaction, pages 23–34. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 2003.

[4] A. Dey, A. Cunningham, and C. Sandor. Evaluating depth perception of photorealistic mixed reality visualizations for occluded objects in outdoor environments. In *Proceedings of ACM Symposium on Virtual Reality Software and Technology*, pages 211–218, Hong Kong, China, November 2010.

[5] S. Henderson and S. Feiner. Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*, pages 135 –144, oct. 2009.

[6] K. Laakso, O. Gjesdal, and J. R. Sulebak. Tourist information and navigation support by using 3d maps displayed on mobile devices. In *Workshop on Mobile Guides, Mobile HCI 2003 Symposium*, 2003.

[7] M. A. Livingston, J. Swan, J. L. Gabbard, T. H. Höllerer, D. Hix, S. J. Julier, Y. Baillot, and D. Brown. Resolving multiple occluded layers in augmented reality. In *ISMAR '03: Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, page 56, Washington, DC, USA, 2003. IEEE Computer Society.

[8] R. Looije, G. M. te Brake, and M. A. Neerincx. Usability engineering for mobile maps. In *Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology*, Mobility '07, pages 532–539, New York, NY, USA, 2007. ACM.

[9] K. Lynch. *The Image Of the City*. The MIT Press, 1960.

[10] A. Mulloni, A. Dünser, and D. Schmalstieg. Zooming interfaces for augmented reality browsers. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, MobileHCI '10, pages 161–170, New York, NY, USA, 2010. ACM.

[11] A. Oulasvirta, S. Estlander, and A. Nurminen. Embodied interaction with a 3d versus 2d mobile map. *Personal Ubiquitous Comput.*, 13:303–320, May 2009.

[12] M. Pielot, N. Henze, and S. Boll. Supporting map-based wayfinding with tactile cues. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '09, pages 23:1–23:10, New York, NY, USA, 2009. ACM.

[13] M. Pielot, B. Poppinga, and S. Boll. Pocketnavigator: vibro-tactile waypoint navigation for everyday mobile devices. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, MobileHCI '10, pages 423–426, New York, NY, USA, 2010. ACM.

[14] I. Rakkolainen, J. Timmerheid, and T. Vainio. A 3d city info for mobile users. *Computers & Graphics*, 25:619–625, 2000.

[15] E. Rukzio, M. Müller, and R. Hardy. Design, implementation and evaluation of a novel public display for pedestrian navigation: the rotating compass. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 113–122, New York, NY, USA, April 2009. ACM.

[16] C. Sandor, A. Cunningham, A. Dey, and V. Mattila. An augmented reality x-ray system based on visual saliency. In *Proceedings of the IEEE International Symposium of Mixed and Augmented Reality*, pages 27–36, Seoul, Korea, 2010. IEEE.

[17] C. Sandor, A. Cunningham, U. Eck, D. Urquhart, G. Jarvis, A. Dey, S. Barbier, M. Marner, and S. Rhee. Egocentric space distortion visualizations for rapid environment exploration in mobile mixed reality. In *Proceedings of the IEEE Virtual Reality Conference 2010*, pages 47–50, Waltham, MA, USA, 2009. IEEE.

[18] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.

Multiple Camera Augmented Viewport: An Investigation of Camera Position, Visualizations, and the Effects of Sensor Errors and Head Movement

Thuong N. Hoang* and Bruce H. Thomas†

Wearable Computer Lab – University of South Australia

ABSTRACT

This paper presents an extension to our augmented viewport technique for action at a distance for outdoor AR systems by employing the use of different physical camera positions. The original technique augments the user's view with video images from a physical zoom lens camera to provide advantageous viewing windows into the augmented environment, through which the user can perform image plane manipulation of virtual objects. Our extended augmented viewport technique utilizes a range of camera positions, including remotely located cameras, head mounted zoom lens cameras, and tripod mounted zoom lens cameras, to offer several benefits: closer views of the scene of interest, novel and complementary viewing angles with multiple viewports, stability against sensor errors, and view-dependent interaction to enhance precision. We introduce new visualizations to assist in the discovery of the cameras. We conducted a user study to evaluate the effects of different camera viewpoints, sensor error, head movement, and the multiple viewports visualization on the usability of the augmented viewport.

KEYWORDS: Augmented viewport, interaction technique, image plane, outdoor augmented reality.

INDEX TERMS: H.5.2 [Information interfaces and Presentation]: Graphical User interfaces - Interaction styles; I.3.6 [Computer Graphics]: Methodology and Techniques - Interaction Techniques

1 INTRODUCTION

This paper presents our continued work on the augmented viewport technique [1] for action at a distance (AAAD) with outdoor augmented reality (AR) systems. AAAD is the problem of interacting with virtual objects that are located out of arm's reach. In our original augmented viewport technique, we demonstrated a set of techniques that augment the user's view with video images from a physical zoom lens camera to provide advantageous viewing windows into the augmented environment, through which the user can perform image plane manipulation of virtual objects located at a distance. The main benefit of augmented viewports is their support for precise interaction with virtual objects at a distance in an AR environment.

The results from our previous investigation posed a number of interesting questions:

1. The augmented viewport can utilize a range of different physical camera locations in the environment, so what effect do different camera viewpoints have on the usability of the technique?
2. Considering that the types of physical cameras include sensor tracked cameras and head mounted cameras, what effect do sensor error and user's head movement have on the usability of the technique?
3. When there are a number of physical cameras in the environment, how does the technique support the user in the discovery and utilization of physical cameras?

To answer these questions, we investigated three types of camera location and their effects on precise manipulation, in terms of head movement and sensor error in the zoom lens camera, and oblique viewing angle using the remote camera. We developed virtual visualizations to assist the user in discovering and selecting suitable cameras for the desired manipulation tasks, based on the location and the viewing area of each camera. We conducted a user study to evaluate various error effects and the multiple viewport visualization (see Figure 1) and present the results with post-study discussions.

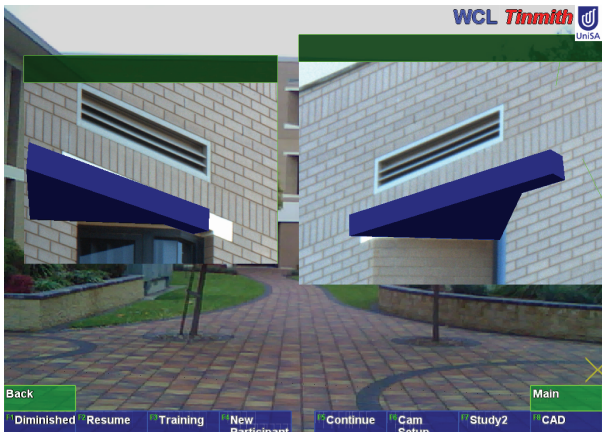


Figure 1. Multiple viewport visualization

1.1 Augmented Viewport Technique

The augmented viewport technique [1] enhances two common AAAD techniques for outdoor AR systems, the image plane [2] and AR working plane [3] techniques, which use the projection of the augmented environment as seen through a user's head mounted camera. Our technique leverages other cameras in the environment that can provide closer views of the distant location. There are two main types of cameras that can offer such an advantage: remotely located cameras and cameras with an optical zoom lens. In this paper, we investigate the use of remotely located cameras, and two variants of zoom lens cameras, namely head mounted and tripod mounted, for the augmented viewport technique. Remote cameras are mounted in a fixed remote location and orientation, while zoom lens cameras are located near the user and have adjustable orientation and position.

The augmented viewport shows a virtual window showing the video feed of a physical camera. The viewport window is overlaid with the view from a virtual counterpart of the physical camera, with the same intrinsic parameters, orientation, and location as the physical one. The combination of the video image and the overlay produces a windowed view into the AR environment through which the user can interact with virtual objects, using close body

*email: thuong.hoang@unisa.edu.au
†email: bruce.thomas@unisa.edu.au

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

interaction techniques. Figure 2 shows an augmented viewport with a blue virtual object overlaid on the physical background of a brick wall.

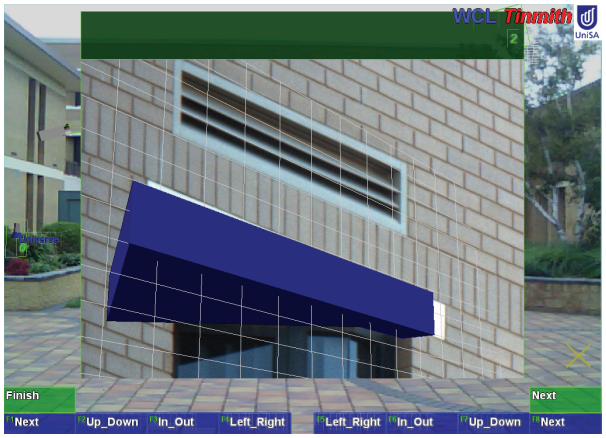


Figure 2. An augmented viewport

We previously investigated [1] three placements of the augmented viewport relative to the user's viewpoint based on a single tripod camera mounted next to the user. The placements of the viewports are defined as three different relative coordinate systems. *World relative* places the viewport at a fixed location in the world coordinate system (GPS), allowing the user to view the window from various angles by physically walking around the viewport. *Body relative* fixes the viewport in the coordinate system that takes the user's body as the origin, so that the viewport is always located at a fixed distance and orientation from the body; while *head relative* attaches the viewport to the user's head position and orientation, for a fixed and direct view of the viewport window. The three placements range in the flexibility of the viewing angle, with the *head relative* at the fixed viewing angle end, and the *world relative* at the flexible end of the scale. Our previous investigation found no difference in performance for each of the viewport placements.

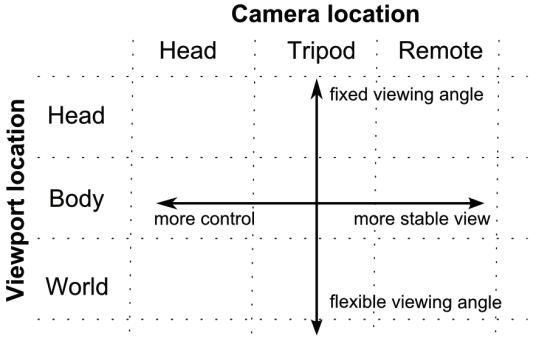


Figure 3. Variants of the augmented viewport

To better understand the relationship between the combinations of the three viewport placements and the three types of cameras locations, we place these concepts on a chart shown in Figure 3. The chart depicts nine combinations (3x3) between augmented viewport placement and physical camera position. The horizontal axis represents the camera location, and reflects the position of the camera relative to the user's view and the range of user control over the physical camera. A head mounted camera offers flexible control of the camera viewpoint using head movements. A sensor tracked tripod produces more stable viewpoints but requires a slower adjustment process. The remote camera offers the most stable view that is not affected by sensor errors, but its position is fixed and least flexible. As previously mentioned, our previous work [1] explored the effects of the viewing angle based on the viewport location (the vertical axis).

1.2 Contributions and Structure

This paper makes a number of contributions to AAAD manipulation techniques for outdoor AR:

1) A new set of the augmented viewport techniques and visualizations for the discovery and utilization of a range of physical cameras use for precise action at a distance manipulation tasks.
2) The results of a user study on the effects of different camera viewpoints, head movement and sensor errors on zoom lens cameras, and the multiple viewport visualization on the usability of the augmented viewport technique.

The paper starts with a description of the related research to AAAD and AR. Our extensions to the augmented viewpoint technique are then presented in detail. A description of the user study performed is given, followed by a discussion of the results. The paper finishes with a set of concluding remarks.

2 Background

The augmented viewport technique is based on image plane technique [2] for virtual immersive systems. The image plane technique [2] collapses the virtual world along the depth dimension onto a planar surface, and simplifies the interaction to two dimensions. This approach poses an inherent limitation of the inability to perform interaction along the depth/normal axis of the current viewing image plane. The AR working plane [3] extends the image plane approach to enable action at a distance interaction in AR environments. Instead of defaulting to the user's viewing plane, the AR working plane technique supports the creation of virtual planes for the input cursor and virtual objects to be projected onto. Both techniques are based on the image plane from the first person perspective using the user's head mounted camera. Our augmented viewport technique extends the image plane approach to use the viewpoints of other physical cameras in the environment. Augmented viewports are based on virtual environment viewports. SEAM [4] is a method of employing virtual viewports to intertwine multiple virtual environments in concert. The viewports are attached at various locations in the group of virtual worlds, acting as viewing platforms between two distant locations. Through-the-lens techniques [5] implement similar metaphor for navigation and object manipulation in VR. Interaction from two separate distant locations could be enabled by using multiple viewports [6].

Precise interactions are a major issue faced by many immersive modeling systems. HoloSketch [7] is a virtual environment sketching tool that relies on a highly accurate tracking and display system to function properly. 3DARModeler [8] and ARpm [9] are two hybrid immersive modeling systems that use desktop-based CAD systems for precision inputs. There are many factors affecting precision in direct manipulations in immersive environments. One such factor is the user's inability to perform constant and precise movements with their arms and hands. PRISM [10] is a manipulation technique that addresses this issue by applying a scaled mapping ratio between hand movements and virtual object displacements, based on the speed at which the hand travels. Our augmented viewport technique reduces such effects over the distance by offering a closer view of the remote scene.

Sensor error is another factor impacting on precise manipulation. Holloway's error model [11] identifies the sensor as the main source of most registration errors in AR systems. Sensor fusion is a common approach that combines various types of

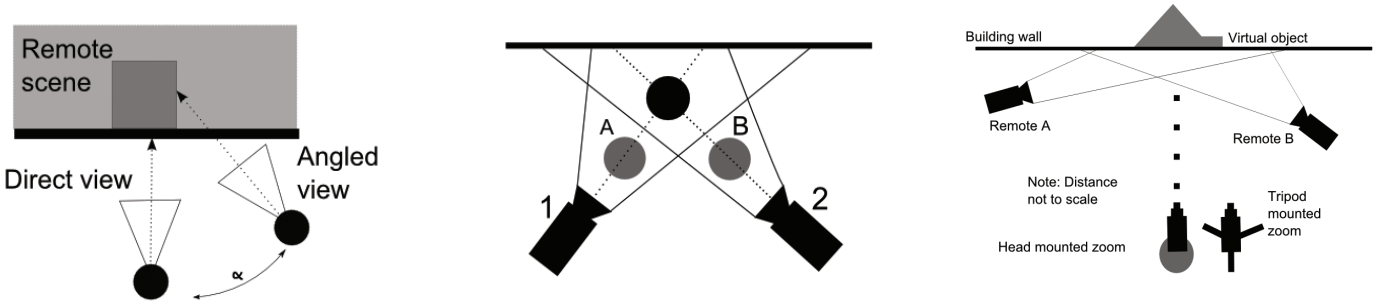


Figure 4. Left: View-dependent interaction Middle: Multiple viewport setup Right: User study setup

sensors to achieve better registration results. Effective combinations include inertial sensor and vision data [12, 13], with GPS tracker [14, 15], or fixed and head-mounted sensors [16]. The augmented viewport technique does not use sensor fusion for error correction; however, its usage with stationary remote cameras bypasses sensor errors for better precision.

3 Augmented Viewports with Different Camera Locations

In this section, we describe in detail the new extension to the augmented viewport technique to support multiple camera locations for the video image feed of the viewport, with regards to the benefits of the technique and the visualization used.

3.1 Benefits

The augmented viewport techniques with cameras in different physical locations have many potential benefits: novel viewpoints, the elimination of sensor errors, and view-dependent interaction.

3.1.1 Novel viewing angle

Depending on its physical location, remote cameras offer the advantage of a novel viewpoint at the scene of interest. Interaction techniques conventionally used for outdoor AR such as the image plane [2] and the AR working plane [3] are not effective along the normal axis of the head mounted display (HMD) camera. The augmented viewport enables remote image plane interaction using the imagery from the physical cameras, and has the same image plane limitation along the principal axis of the remote camera; however, this axis is not generally parallel to that of the user's head mounted display. Therefore, the use of the augmented viewport with remote cameras enables image plane interaction along the normal axis of the HMD by providing novel viewing angles of the scene of interest. The benefits of the novel viewpoints include: (1) a closer view to the scene, (2) a remote image plane interaction that is effective along the normal axis of the user's head mounted display to enable more precise manipulation along this depth axis, (3) a viewing angle that could not be obtained from the user's current location, due to physical constraints.

3.1.2 Stability against sensor noise

There are two main components to an augmented viewport: the physical and the virtual cameras. Physical remote cameras are often fixedly mounted and not tracked, thus an augmented viewport utilizing a remote camera is not affected by sensor errors and jittering.

The use of a remote stable camera for the viewport with *head or body relative* placement is also free from noise generated by the user's position and head orientation sensors (*head relative* only) in an outdoor AR system. A head relative augmented viewport using remote camera fixes a virtual window to the user's viewpoint, regardless of the current head orientation and location. *Body relative* placement offers similar stability against the location sensor noise, because the viewport window is attached to the user's body, regardless of their position. The alignment of the virtual and physical world inside the augmented viewport is also fixed, as mentioned earlier. Therefore, the user can perform manipulation tasks through the remote camera augmented viewport without being affected by the errors of the user's head orientation and location sensors.

The freeze-frame technique [17] is an example of eliminating sensor errors by capturing still snapshots of the environment, together with sensor data. The augmented viewport technique with a remote camera offers stability against sensor noise in a non-disruptive manner. The main video stream of the head mounted camera and all the sensors are kept running throughout the manipulation process. The augmented viewport is presented as a virtual window and does not cover the entire vision of the user. Stability of virtual objects inside the viewport is seamlessly achieved without the user's intervention. The freeze frame and similar techniques help reduce sensor jittering, but are still affected by sensor drifts. The accuracy of the virtual overlays at the time of freezing could be affected by accumulated drifts of the sensors running over time. The augmented viewport is not affected by such drifts, but only by the initial errors in the orientation and position measurements of the physical cameras.

3.1.3 View-dependent interaction

The implementation of the augmented viewport technique allows the user to achieve additional viewpoints of the virtual objects through the viewport, simply by adjusting the angle at which they interact with the viewport. The different placements of the viewport in head, body, and world relative coordinate systems enable adjustments of the viewport viewing angle.

The remote location is rendered in full 3D using OpenGL stencil buffers; thus, the user can gain extra viewing angles at the remote scene. When the viewport is placed in a head relative position, the user looks directly into the viewport (direct view in Figure 4 Left, where the shaded region represents the remote scene, and the dark box in the shaded region is a virtual object at the remote scene), gaining the view as if they were standing at the physical camera location. This viewport placement gives a constant direct view. In the body relative placement, the viewport window is fixed at a certain angle and distance from the body, enabling the user to look into the viewport constantly from an oblique α angle. The remote scene is then seen by the user as if the user was standing at a location that is rotated the same α angle about the remote scene from the physical camera location, as illustrated as the angled view in Figure 4 Left. This view allows the user to view different portions of the virtual object, such as the right side of the virtual box, as an example. The world relative placement supports both direct and angled views as it allows the user to walk around the augmented viewport.

This view-dependent viewport interaction allows the user to perform exploratory tasks to gain extra insights into the remote scene. For the best result, we suggest using image homography to generate the physical world view from camera images corresponding to the user's viewing angle of the viewport.

3.2 Visualizations

The augmented viewport can be used with many types of physical cameras available in the environment. Depending on the task requirements, one camera may offer more favorable viewing angle than the others. Therefore, we have introduced visualizations to support the discovery of physical cameras for the augmented viewport technique, as well as the utilization of multiple viewports setup.

3.2.1 Camera discovery

Upon the user arriving at an outdoor setting, information about the available cameras in the surroundings is downloaded to the wearable computer. There are many possible scenarios for camera positions: the camera itself can be (a) visible, or (b) not visible to the user; and the physical area the camera is looking at is either (a) visible, or (b) not visible to the user. Even if the cameras and its viewing areas are visible through the normal vision of the head mounted display, it is not clear to the user as to what the cameras are looking at.

Therefore, for each physical camera, we render a virtual overlay to highlight its position, orientation, identification, and viewing area. The overlay consists of a virtual model of the camera placed directly over the physical camera, a virtual frustum extending from the camera's position to the viewing area, and an identification number uniquely assigned for each camera in the surroundings. This visualization can be viewed in two different modes, namely immersive and orbital view.

In immersive mode, the virtual cameras are rendered in the first person perspective. This mode is mostly effective when the area of interest for the task is known, because the user can immediately identify if there are any cameras pointing at the required area and if their viewing angles are suitable for the manipulation tasks.

Orbital view [18], on the other hand, is a pure virtual viewpoint that gives an overview of the environment from a higher vantage point, allowing the user to explore the broader surroundings to discover more cameras. With the purely virtual nature of the view, the user can freely navigate around the environment. Wireframe models of physical buildings and landscapes, if available, could be rendered as reference for the relative positions of the cameras. If such models are not available, the orbital view could be taken from the viewpoint of a virtual camera that is fixed to the user's head orientation, but flew backwards and upwards to reach a higher perspective, so that the yaw orientations of the orbital and the immersive view are still aligned. Such alignment enables the user to switch between the two views without being disoriented about the locations of the physical cameras relative to the user.

3.2.2 Multiple viewports

The augmented viewport suffers from the same limitation as other image plane interaction techniques: ineffectiveness along the normal axis of the plane. We investigate and implement the usage of multiple augmented viewports to tackle this limitation.

Figure 4 Middle depicts an example scenario for the benefits of multiple augmented viewports. There are two cameras, 1 and 2, viewing the scene with a virtual sphere (dark circle) in its correct position, from different angles. The shaded circles, marked A and B, represent the possible erroneous positions of the virtual sphere that would potentially go undetected when using a single viewport only. At location A, the virtual object would appear as almost unchanged from the perspective of camera 1; in a similar manner that the object at location B would be mistaken as the correct position in camera 2. Both locations represent object displacement along the normal axes of the respective cameras. However, when both augmented viewports are visible, the user can detect such an anomaly and perform correction operations to put the virtual object in the correct position. The multiple viewports allow the users to build up a 3D model of the position of the virtual object, as single camera may not supply enough depth information for the user to understand the object's relative depth position.

4 User Study

Our previous study [1] evaluated the concept of the augmented viewport and showed an improvement in precision, time, and effort in manipulation tasks. In this paper, we extend the augmented viewport to support a wider range of physical cameras, which introduces several factors potentially affecting the usability of the technique. We were motivated to conduct a user study to evaluate the performance of different camera positions and determine the effects of different viewpoints, head movement and sensor noise, as well as the multiple viewport visualization.

4.1 Design

In order to separately examine the above-mentioned factors, we designed multiple task conditions in which the participants used the augmented viewport to perform common manipulation tasks. The aim of the base task is precise manipulation, by scaling or moving virtual objects to match with the size or position of a physical artifact, located at a distance, called the *distant scene*.

We implemented the following four different camera placement augmented viewports (head relative view), each characterized by a single or compounded evaluation factors (see Figure 4 Right):

1. **Remote** camera: This condition uses a single remote camera (remote A) looking at the distant scene from an oblique angle, which is different from the first person perspective viewing angle of the distant scene. The single factor of a different camera viewpoint is embedded in this condition.
2. **Head** mounted zoom lens camera: This condition uses a single zoom lens camera, controlled by the participant's head orientation. The participant's location and head orientation are tracked; therefore, this condition is compounded with the head movement and sensor error factors.
3. **Tripod** mounted zoom lens sensor tracked camera: This condition uses a single zoom lens camera, mounted on a tripod next to the user. This tripod is tracked with orientation and location (GPS) sensors, similar to that on the participant's HMD. The participant controls the orientation of the tripod. This condition is affected by a single factor of sensor errors, of both orientation and location sensors.
4. **Multiple** remote camera: This condition uses two remote cameras: remote A, and the second camera is another remote camera (remote B), mounted on the opposite side about the participant, pointing at the distant scene at a different oblique angle. This condition is compounded with the multiple viewport and different camera viewpoint factors.

We then add an additional augmented viewport as a baseline comparison condition:

5. **Fixed** tripod camera viewports: This condition uses a single zoom lens camera, mounted on a tripod whose position and orientation are fixed by calibration and not tracked, at the *same position* as tripod camera with sensors. This condition is not affected by any of the evaluation factors above.

In the experimental design, we ensure that there are separate conditions that use one of the three different camera positions, namely head, tripod, and remote mounted cameras, in order to compare overall effects of different positions of the cameras.

Table 1. Error in moving tasks (in meters) and scaling task (in unit) for five camera conditions

	Moving task (in m)						Scaling task (in unit)					
	Depth		*Side*		*Up*		*Depth*		*Side*		*Up*	
Condition	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Remote	0.41486	0.03694	0.58346	0.09957	0.21667	0.01284	0.0802	0.0020	0.1426	0.0035	0.0241	0.0001
Head	1.49689	0.71231	2.04042	1.01558	0.81769	0.19663	0.2480	0.0070	0.2459	0.0137	0.0372	0.0005
Tripod	1.20025	0.42868	1.16354	0.43478	0.39892	0.02749	0.2296	0.0246	0.2121	0.0110	0.0348	0.0002
Multiple	0.41545	0.09051	0.37281	0.03606	0.17453	0.00444	0.1082	0.0027	0.1436	0.0125	0.0232	0.0001
Fixed	0.30255	0.03173	0.03087	0.00056	0.06639	0.00356	0.0460	0.0012	0.1286	0.0044	0.0191	0.0006

Performance was measured with three different quantitative methods: *task error, completion time, and the number of mouse clicks required*. Task error was measured as the difference in the virtual object's final position and size from the actual position and size of the matching physical object, physically measured during calibration as the ground truth. The time to complete the task in microseconds was determined by the participant upon satisfaction of the task's result. Our hypotheses are as follows:

H1: There is a measurable reduction in the performance of the augmented viewport when affected by the different camera positions.

H2: There is a measurable reduction in the performance of the augmented viewport when affected by head movement.

H3: There is a measurable reduction in the performance of the augmented viewport when affected by sensor errors.

H4: There is a measurable reduction in the performance of the augmented viewport when affected by multiple viewport visualization.

4.2 Experiment

We had 16 participants (15 males and 1 female), aged 18 – 44 (mean: 25.37, SD: 7.71). Nine participants had never used an AR system or a wearable computer before. The participants were asked to wear the Tinmith wearable computer system to scale and move a virtual window lentil to match the size and position of a physical window lentil, using the augmented viewport technique. There were five different camera conditions, as described in the previous section and illustrated in Figure 4 Right (condition 3 and 5 were co-located at the Tripod mounted zoom location). In total there were ten tasks (2x5) to perform for each iteration. Each participant completed two iterations with randomized task orders, after one training session. There were breaks in between iterations.

The remote cameras were mounted with identical lenses with a focal length of 25 mm, while the head mounted camera and the tripod mounted camera used identical 75 mm fixed focal length lenses. Traditional variable focal length zoom lens was not used to reduce calibration errors. All cameras were set to capture at 640x480 resolution. Both remote A and B cameras were mounted on a fixed tripod, while the zoom lens camera located near the user was on an adjustable tripod. The user and the tripod were approximately 50 meters away from the building, while the remote cameras were mounted within 10 meters, so that each camera covers the same viewing area of the physical environment.

The user performed the tasks using a trackball mouse to control the onscreen cursor for direct manipulation, and a Bluetooth button box for command control. For each of the tasks, either scaling or moving, the participant could individually manipulate the object in the X, Y, or Z axes of the object's coordinate system, by clicking to select the object, moving the cursor to scale/move the object along the selected axis, and clicking again to release the object. At the start of each moving task, the virtual window sill object was misplaced at random positions, all equidistant from the correct position. For scaling tasks, the starting size of the window sill was randomly either smaller or larger than the correct size, all by an equivalent ratio. The randomization was done so that through the three iterations including training, the participant would not see the same starting position or size of the virtual object using the same camera, to reduce learning effects.

For each task, the time and number of mouse clicks required to complete the task was recorded. For the Tripod condition, this included the time the participant spent adjusting the tripod in order to complete the task. For the Head condition, the time to locate the physical window sill using the head mounted zoom lens camera was counted. The time to finish each task was decided by the participant when he/she was content with the correct position or size of the virtual object. The final position and size of the virtual object were recorded after each task. A questionnaire was completed at the end of three iterations for a qualitative evaluation of the participant's preferences among different camera positions.

5 Results and Discussion

Based on the GPS data recorded, we detected an outlier where the GPS position of one of the participants was displaced by a considerable amount from the actual position. Therefore, we discarded the data for this participant and performed analysis on the remaining 15 data sets. We performed ANOVA analysis on the measurement errors in position and size of the virtual object, as well as time to complete the tasks and the number of clicks required. ANOVA error analysis was done separately on error measurements in the X, Y, and Z axes, for both scaling and moving tasks. Based on the first person perspective, the X axis was the depth axes along the normal axis of the user's head mounted display image plane; the Y axis was the horizontal image plane axis, and Z was the vertical image plane axis. Therefore, scaling and moving errors along the X, Y, and Z axes will be referred to as depth, side, and up axes errors, respectively, in the results presented below.

5.1 Error analysis

For the error analysis, there was a significant effect ($p < 0.05$) over the five camera conditions, for all axes in both scaling and moving tasks, see Table 1. A post-hoc analysis on the error measurement was performed with a pairwise t-Test on six pairs of conditions, with a Bonferroni correction ($\alpha < 0.008$). We selected the following pairs to explore different error effects:

1. *Viewing angle:* Comparing the Remote camera and the Fixed tripod camera examines the effects of an oblique viewing angle, because the remote A camera looked at the scene from a 30 degree angle, while the fixed tripod camera shared a direct 90 degree angle view of the virtual object as the user's first person perspective.
2. *Multiple viewport*: Comparing the Remote camera with the Multiple remote cameras to examine the effects of multiple viewpoint against a single view. Both conditions shared the similar oblique viewing angles.

Table 2. Time (in s) and number of clicks (units) for moving and scaling tasks across five conditions

	Time (s)				Number of clicks (unit)			
	Moving task		*Scaling task*		*Moving task*		*Scaling task*	
Condition	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Remote	67.59	25.02	42.75	14.45	9.06	4.03	4.75	2.38
Head	98.38	41.78	82.29	35.34	6.88	2.95	4.03	1.84
Tripod	102.19	36.19	83.56	34.37	7.97	2.95	5.22	2.67
Multiple	110.00	54.00	63.53	23.85	13.53	10.45	5.34	2.83
Fixed	51.15	18.58	45.32	17.90	5.53	2.05	4.41	2.30

3. *Head movement*: Comparing the Head mounted zoom lens camera and Tripod adjustable camera examines the effect of head movement. In both conditions, the cameras were tracked by GPS and orientation sensors; the head mounted camera was additionally affected by head movement.
4. *Sensor noise:* Comparing the fixed tripod camera and the Tripod adjustable camera, in both conditions, the cameras had the first person perspective view, but the Fixed tripod was not affected by any sensor.
5. *Ranking Camera Position:* Remote camera and Head mounted cameras.
6. *Ranking Camera Position:* Remote camera and Tripod adjustable camera.

For the first pairwise t-Test to show the effects of oblique viewing angle, there was a significant effect ($p < 0.008$) to support the hypothesis H1 that the viewing angle adversely affected the precision in moving task in the side and up axes, and scaling task in the depth axis (Fixed tripod performed better than Remote).

For the second pairwise t-Test comparing the effect of multiple viewports, there was no effect ($p > 0.008$) that the multiple viewports adversely affected precision for the moving and scaling tasks in any axis. Hypothesis H4 was rejected.

For the third pairwise t-Test comparing the effects of head movement, there was a significant effect ($p < 0.008$) to support the hypothesis H2 that head movement reduced precision in the moving task *only* in the vertical image plane axis. However, it was noticed that there are no significant affects in any other axes for moving and scaling tasks.

For the pairwise t-Test to show the effects of sensor noise, there was a significant effect ($p < 0.008$) to support the hypothesis H3 that sensor noise degraded precision in the moving and scaling task across all three axes, except for scaling along the side axis.

For the ranking of camera positions, the remote camera showed a significant improvement in precision ($p < 0.008$) over the head mounted camera across all axes in both tasks, except for scaling in the up direction. The remote camera also showed a significant improvement in precision ($p < 0.008$) over the tripod camera across all axes in both tasks, except for scaling along the side axis.

5.2 Time analysis

For the time to complete the task, we performed an ANOVA analysis and found a significant effect ($p < 0.05$) among the five camera conditions for both scaling and moving tasks, see Table 2. A post-hoc analysis of the completion time was performed using a pairwise t-Test on seven pairs of conditions, with a Bonferroni correction ($\alpha < 0.0071$). We performed the analysis on the total time to complete the task, on the same six pairs as examined for error analysis, with an additional pair of Fixed camera and Head camera. The additional pair was added to evaluate the extra time taken to do the task due to the head movement as compared to the baseline Fixed camera condition.

There was a significant effect ($p < 0.0071$) to support the hypotheses H2, H3, and H4 for both scaling and moving tasks regarding the effects of head movement, error sensors using tripod adjustment, and the multiple viewport visualization. Hypothesis H1 was rejected for task time.

For the ranking pairs, there were significant effects that the remote camera improved in time to complete the task ($p < 0.0071$) over the head and tripod cameras for both scaling and moving tasks. There was no significant difference between the tripod and head mounted cameras.

5.3 Number of clicks

For the number of mouse clicks to complete the task, there was a significant effect ($p < 0.05$) over the five camera conditions, only in the moving tasks, see Table 2. A post-hoc analysis on the number of clicks in moving tasks was performed with a pairwise t-Test on seven pairs of condition, with a Bonferroni correction ($\alpha < 0.0071$). We selected the same seven pairs as in the time analysis tasks, to evaluate if the same factors caused the participants to perform more mouse clicks to complete the tasks.

There was a significant effect ($p < 0.0071$) that the participants were required to perform more mouse clicks in the moving task caused by sensor errors using tripod adjustment, and by the remote camera with oblique angle, as compared to a fixed person perspective view. In other words, hypotheses H1 and H3 were supported, while H2 and H4 were rejected. There was no significant difference between the three alternate pairwise tests among the remote, head, and tripod mounted cameras.

5.4 Questionnaire

The participants were asked to rank the three camera positions, head mounted, tripod mounted, and remote mounted (1 point for the most preferred, and 3 points for the lease preferred). Remote camera scored 23 points, tripod 25, and head 42 (the lower points the more preferred). This ranking mostly agrees with the error and task time analysis as presented above. The opinions fluctuated between the remote and the tripod condition. Most explanations for the higher rank of the remote camera were that the remote cameras were employed in a multiple viewport setting and assisted the user in completing the tasks. Among the participants who preferred the tripod camera, there were complaints about the confusion of the oblique angle presented by the remote camera.

5.5 Discussion

From the results of the study, we draw several conclusions regarding the types of errors, the different camera sources, and visualizations, specifically head movement, remote cameras, multiple viewports, and tripod cameras The following list summarizes the conclusions:

1. Head movement error is negligible in comparison to sensor noise.
2. Head movement error does not affect the estimation of size for manipulation tasks.

3. Head movement error on zoom lens cameras does not render the video image too blurry or too unstable for manipulation tasks.
4. Head movement error causes time delay, but does not complicate the manipulation tasks.
5. When the axes of manipulation of the virtual objects are not parallel to the horizontal and vertical axes of the image plane, precision in manipulation tasks is reduced.
6. Because of the discussion point listed above, remote cameras with oblique viewing angles require extra visualization cues to improve precision.
7. Simply adding another image plane from a different viewing angle does not increase precision (discussion point 5). Similarly, multiple viewport visualization on its own does not increase precision (discussion point 6).
8. The reduced mobility of tripod cameras may outweigh its benefits of stability, with the current sensor configurations used in the study.

5.5.1 Head movement

It may seem obvious that sensor noise and head movement would reduce the precision of the augmented viewport technique. However, with the error analysis of the study, we can conclude that sensor error causes reduction in precision to a greater extent than head movement. Sensor noise caused errors across more combinations of tasks and axes of operations than head movement did. The pairwise t-Tests on manipulation errors show that sensor noise had a significant effect in all but one axes in both moving and scaling tasks, while head movement only caused issues for the moving tasks along the vertical axes of the image plane. This can be explained by the fact that sensor noise included GPS that could report errors in the user's location causing precision errors on the depth axis, because the user's position fluctuated to be closer to or further from the remote location. Head movement does not have this issue in the depth axis.

Further investigation reveals the possibility that the significant effect of head movement in the up axis may have been caused by sensor calibration error instead. The zoom lens camera was mounted on the participant's head using an oval frame while the head's orientation sensor (Intersense InertiaCube) was separately mounted on the sunglass-style immersive display (Vuzix Wrap920AR). When the oval frame sat on top of the head, it was not possible to misalign the horizontal orientation (yaw) of the camera to the InertiaCube's, because the oval frame could not freely rotate left or right. However, it was highly likely that the oval frame could slip back and forth on the head and tilt the camera slightly upwards/downwards, due to the different shapes and sizes of the participants' heads. This caused an offset in the vertical orientation (pitch) between the camera and the InertiaCube. This offset eventually affected the error results in the moving task in the up axis, as noted in the post-hoc t-Test between the Head camera and Tripod camera in Section 0. Therefore, we are confident that head movement almost does not significantly affect precision.

The head movement only reduced the precision in the moving task in the up axes, but not affected scaling at all. For the scaling task, the position of the virtual object was fixed in the correct position, overlaying on top of the physical window sill. Head movement would cause the object to be displaced from the correct position; however, despite the misalignment error, the participants were able to complete the scaling task by estimating the size of the physical window in the background of the augmented viewport. Therefore, head movement does not affect the estimation of size for manipulation task.

During a prior pilot study, it was noticed that the use of a zoom lens camera for the head mounted display worsened the head movement at a distance, by the same ratio as the zoom lens bringing the closer view, causing precision error as well as blurry vision and rendering the background image too unstable to be useful. However, based on the results of the study, the participants completed the scaling task unaffected by head movement. Therefore, head movement together with zoom lens camera does not render the imagery blurry or unstable.

Comparing the analysis of completion times and the number of clicks reveals that head movement took a longer time to complete the task but did not require extra clicks. It can be deduced (and through observation during the study) that the participants spent most of the task time trying to stabilize the head mounted camera. Once a stable viewpoint is achieved, it took a similar number of mouse clicks as the fixed tripod condition. Therefore, it can be concluded that head movement does cause time delay, but does not complicate the manipulation tasks.

5.5.2 Remote cameras and multiple viewports

Using a remote camera caused the participants to use more mouse clicks but did not take longer time. From this discrepancy, we can explain that the participants performed a number of exploratory moving and scaling operations in short succession to get used to the oblique viewing angle. Thanks to the stability of this condition and extra mouse clicks, the participants could complete the scaling task without taking extra time and without sacrificing precision in scaling tasks. However, precision suffered for the moving tasks. As explained earlier in this paper, all augmented viewport suffers from the same image plane limitation of being ineffective along the normal axis of the viewport. The analysis indicates that with the seemingly rapid exploratory succession of virtual object movements, it was easy for the participants to move the object into an incorrect position along the normal axis of the viewport. Such an incorrect position could not be detected easily, which led the participant to believe that the task goal was completed. Therefore, it did not take longer time to perform this task, however, the precision suffered.

The opposite situation happened for the multiple viewports condition: taking longer in time but not extra mouse clicks. The extra time was spent on trying to understand the spatial relationship between the two camera viewpoints. There were not significantly more mouse clicks, possibly because the first few mouse clicks of moving or scaling the object introduced visual changes on both viewports. The confusion of the spatial relationship may have led the participants to conclude that extra mouse clicks may not be useful to comprehend the combination of two viewpoints. Therefore, they did not try any more exploratory extra clicks than the single remote viewport condition.

The multiple viewports, however, did not produce any more improvements in precision. It must be noted that the visualizations as described in Section 3.2.1 were not enabled in the study. We excluded the visualizations to reduce the confounding variables of the study. Therefore, the participants were left with only the two video streams from both cameras (see Figure 1) and the ability to perform exploratory manipulation on the virtual objects to figure out the spatial relationship of the viewports, which is what the visualizations described earlier, are designed to support. There were only a few participants that succeeded in the spatial relationship problem, after a few iterations of tasks. This reduction in performance is a well researched topic relating to situation awareness and mental workload, as explored by Veas et al. [19] in their work to present visualizations in assisting the understanding of multiple camera setups from the first person

perspective. Similar works in the area of video surveillance investigate different techniques to improve the spatial understanding of camera setups. Notable examples are the video flashlight technique [20], contextualized video [21], and the DOTS system [22]. Our study provides empirical results proving the needs for extra visualizations in a multiple camera setup. We are interested in applying these techniques to improve on the visualizations for the augmented viewport.

5.5.3 Tripod cameras

As can be seen from the results of the error analysis pairwise t-Tests, the sensor error caused the worst and most widespread effect on precise manipulation, agreeing with Holloway's error model [11]. Within the area of interaction research, it is more feasible to attempt to correct the head movement error instead, using vision-based image stabilization, for instance. In Figure 3 showing the variants of camera location, the head mounted camera is affected by head movement and sensor error, but providing the most flexible control of the camera. The tripod mounted camera introduces only sensor error, but takes a longer time to adjust. We also concluded from the study that head movement did not cause any more significant error than the sensors alone, and that the tripod camera required more mouse clicks to complete the same tasks, as well as a bulkier and less mobile setup (this condition required a physical tripod to be mounted next to the participant). Therefore, the advantage of the head mounted camera outweighs its drawbacks when compared to the tripod camera. *Based on this observation, it is suggested that we can focus the augmented viewport techniques on only using the head mounted cameras and existing remote cameras in the environment, thus making the technique more mobile and suitable for outdoor wearable computer systems.*

6 Conclusion

We have presented an extension of the augmented viewport framework of techniques and visualizations for the discovery and utilization of a range of physical cameras for precise action at a distance. The augmented viewport utilizes a range of cameras, including remotely located cameras, head mounted zoom lens cameras, and tripod mounted zoom lens cameras, to offer several potential benefits: closer views of the scene of interest, novel and complementary viewing angles with multiple viewports, stability against sensor noise, and view-dependent interaction to enhance precision. We also presented a user study to investigate the effects of different viewpoint, head movement and sensor noise on zoom lens cameras, as well as the multiple viewport visualization, on the usability of the augmented viewport action at a distance technique. The results of the study showed that head movement only causes minor reduction in precision, and extra visualizations are required to assist the user in understanding the spatial relationship among physical cameras.

Acknowledgement

We would like to thank Ross Smith for his help with building the Tinmith backpack for the study, Wayne Piekarski as the developer of Tinmith, and members of the WCL lab for proofreading this paper.

References

[1] T. N. Hoang and B. Thomas, "Augmented Viewport: An action at a distance technique for outdoor AR using distant and zoom lens cameras," in ISWC, 2010, pp. 117 - 120.

[2] J. S. Pierce, et al., "Image plane interaction techniques in 3D immersive environments," in I3D 1997.

[3] W. Piekarski and B. H. Thomas, "Augmented reality working planes: a foundation for action and construction at a distance," in ISMAR 2004, pp. 162-171, 2004.

[4] D. Schmalstieg and G. Schaufler, "Sewing Worlds Together With SEAMs: A Mechanism to Construct Complex Virtual Environments," Presence: Teleoperators & Virtual Environments,vol.8, pp. 449-461, 1999.

[5]S. L. Stoev, et al., "The through-the-lens metaphor: taxonomy and application," in IEEE VR 2002, pp. 285-286.

[6] K. Hirose, et al., "Interactive Reconfiguration Techniques of Reference Frame Hierarchy in the Multi-viewport Interface," in 3DUI 2006,pp 73-80.

[7] M. F. Deering, "HoloSketch: a virtual reality sketching/animation tool," ACM (TOCHI), vol. 2, pp. 220-238, 1995.

[8] T. V. Do and J. W. Lee, "3DARModeler: a 3D Modeling System in Augmented Reality Environment," International Journal of Computer Systems Science and Engineering, vol. 4, p. 2, 2008.

[9] P. Fiala and N. Adamo-Villani, "ARpm: an augmented reality interface for polygonal modeling," in ISMAR 2005, pp. 196-197, 2005.

[10] S. Frees and G. D. Kessler, "Precise and rapid interaction through scaled manipulation in immersive virtual environments," in IEEE VR, 2005, pp. 99-106.

[11] R. L. Holloway, "Registration error analysis for augmented reality," Presence: Teleoperators & Virtual Environments,vol.6, pp. 413-432, 1997.

[12] J. D. Hol, et al., "Sensor Fusion for Augmented Reality," in International Conference on Information Fusion, 2006, pp. 1-6.

[13] S. You, et al., "Orientation tracking for outdoor augmented reality registration," IEEE Computer Graphics and Applications, pp. 36-42, 1999.

[14] G. Schall, et al., "Global pose estimation using multi-sensor fusion for outdoor Augmented Reality," in ISMAR 2009, pp. 153-162.

[15] R. Behringer, "Registration for outdoor augmented reality applications using computer vision techniques and hybrid sensors," in IEEE VR 1999, pp. 244-251.

[16] W. A. Hoff, "Fusion of data from head-mounted and fixed sensors," in IWAR, 1998, pp. 167-182.

[17] S. Guven, et al., "Mobile augmented reality interaction techniques for authoring situated media on-site," in ISMAR 2006, p. 235.

[18] J. C. Chung, "A comparison of head-tracked and non-head-tracked steering modes in the targeting of radiotherapy treatment beams," in I3D 1992, pp. 193-196.

[19] E. Veas, et al., "Techniques for view transition in multi-camera outdoor environments," in Proceedings of Graphics Interface 2010,.

[20] H. S. Sawhney, et al., "Video flashlights: real time rendering of multiple videos for immersive model visualization," presented at the Proceedings of the 13th Eurographics workshop on Rendering, 2002.

[21] W. Yi, et al., "Contextualized Videos: Combining Videos with Environment Models to Support Situational Understanding," IEEE TVCG, vol. 13, pp. 1568-1575, 2007.

[22] A. Girgensohn, et al., "DOTS: support for effective video surveillance," presented at the Proceedings of the 15th international conference on Multimedia, Augsburg, Germany, 2007.

A User Study on Viewpoint Manipulation Methods for Diorama-Based Interface Utilizing Mobile Device Pose in Outdoor Environment

Masayuki Hayashi* University of Tsukuba
Itaru Kitahara † University of Tsukuba
Yoshinari Kameda ‡ University of Tsukuba
Yuichi Ohta § University of Tsukuba

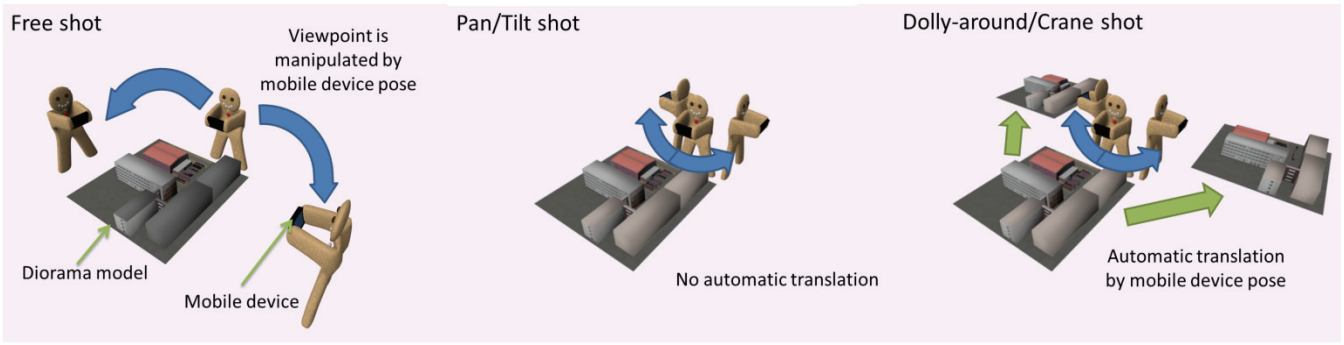


Figure 1. Three methods of viewpoint manipulation were compared in this paper. (Left) Free shot method works as if it was real diorama model placed on a table. (Middle) In pan/tilt shot method, orientation of the viewpoint corresponds to the orientation of mobile device. (Right) In dolly-around/crane (D/C) shot method, in spite of the rotation of the mobile device, the diorama model always appears in front of the mobile device.

ABSTRACT

Diorama-based interface that displays a point of interest (POI) on a miniature of real 3D world is a good approach to share the POI with people working in outdoor environment. The viewpoint to observe the diorama model should be manipulated by users to explore the diorama model and find the POI. However, poorly designed viewpoint manipulation method may cause difficulty to understand the corresponding point of the POI in the real world.

A viewpoint manipulation method should be able to manipulate the viewpoint freely and the viewpoint enables a user to understand the correspondence between the real world and the diorama model easily. In order to realize a viewpoint manipulation method with satisfying the above requirements, this paper compares three viewpoint manipulation methods (free shot, pan/tilt shot and dolly-around/crane shot) that utilize a mobile device pose. We have implemented an AR (Augmented Reality) test bench of the diorama-based interface with photorealistic diorama model to conduct the subjective evaluation experiment. As a result, we found that dolly-around/crane shot is superior to the others in aspect of performance to find POI and subjective impressions.

KEYWORDS: Outdoor mixed reality, User study, Viewpoint manipulation.

INDEX TERMS: H.4.3 [**Information Systems Applications**]: Communications applications – *Point of Interest sharing*; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *Artificial, augmented, and virtual realities*; I.3.6 [**Computer Graphics**]: Methodology and Techniques – *Interaction Techniques*

* e-mail: mhayashi@image.iit.tsukuba.ac.jp
† e-mail: kitahara@iit.tsukuba.ac.jp
‡ e-mail: kameda@iit.tsukuba.ac.jp
§ e-mail: ohta@acm.org

1 INTRODUCTION

When working in outdoor environments, people may need to find a three-dimensional point of interest (POI) indicated by other people. By putting a POI on a miniature diorama model (or a 3D map) of the environment, people could find the corresponding point to the POI in the real world. Hence, diorama-based interface such as *world in miniature* (WIM) [1] is a good approach to share a POI. In a design of the interface, the viewpoint observing the diorama model should be manipulated by users to explore the diorama model and find the POI. It is important to design the viewpoint manipulation method carefully.

With poorly designed viewpoint, users may not able to find out any cues that match their current view in the real world. As a result, they may feel difficulty to understand the corresponding point of the POI. Wingrave et al [2] proposed scaled and scrolling WIM (SSWIM) which is more suitable than originally WIM for large scale environment. However, it is difficult to apply SSWIM to outdoor augmented reality (AR) since its interaction is designed for immersive environment. Though some papers presented a diorama-based interfaces in outdoor AR context [3-6], there is still no user study on the viewpoint manipulation method for sharing a 3D POI at outdoor MR.

In this paper, we investigate the viewpoint manipulation methods of a diorama-based MR interface by a user study in outdoor environment. In particular, we compare three viewpoint manipulation methods: free shot, pan/tilt shot and dolly-around/crane shot. They are illustrated in the Figure 1. Since the

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

viewpoint manipulation methods utilize pose of mobile device, the way to handle the mobile device will make a difference.

Free shot is a commonly-used AR method which displays a diorama model on an ARToolKit marker [7] fixed to the real world. Pan/tilt shot is an egocentric viewpoint manipulation utilizing touch-screen instead of real-time position tracking of a mobile device. In dolly-round/crane shot, the viewpoint is moved on the surface of a virtual sphere surrounding a center of rotation by rotation of the mobile device.

These manipulation methods aim to manipulate the viewpoint freely and to make a user easily understand the correspondence between real world and the diorama model. Therefore, we align the viewpoint angles with the mobile device orientation in order to align the orientation of the diorama model with the real world.

We have conducted a user study to evaluate the three viewpoint manipulation methods in outdoor environment. Figure 2 illustrates the experimental situation. Our interface overlays a photorealistic diorama model of surrounding environment of the user on the video image captured by the camera of the mobile device. A POI is shown on the diorama model by an arrow-shaped icon. The premise of the experiment is that the diorama model of the environment, position of the users and position of the POI are given.

We exploited a photo-shooting task to evaluate performance to find out the POI. As a result of the experiment, we found that D/C shot is significantly superior to free shot in aspect of performance to find out POIs and subjective impressions. And D/C shot also tend to superior to pan/tilt shot.

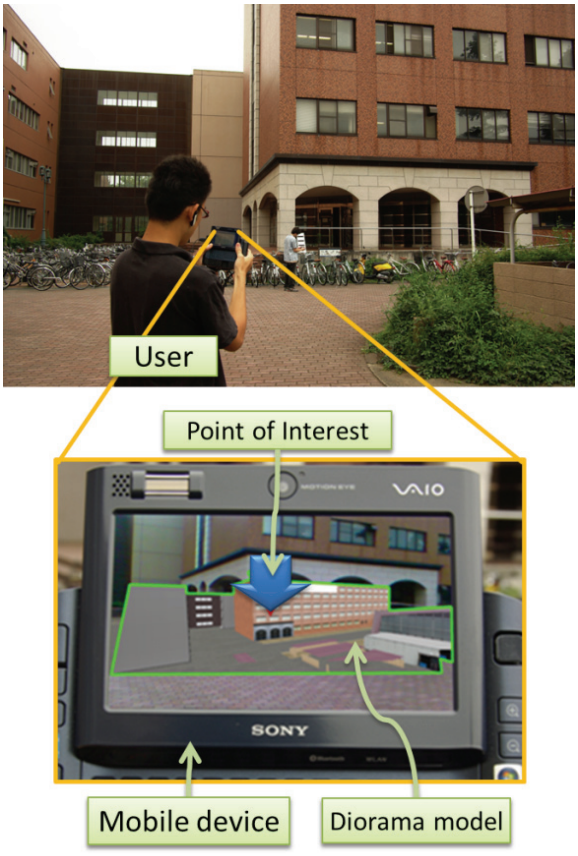


Figure 2. Our proposed diorama-based interface. A diorama model of surrounding environment is displayed on mobile device. An arrow-shaped icon represents the point of interest on the diorama model.

2 Related Work

POI in outdoor scene is sometimes used as geometric annotation. We review approaches of AR interface and user studies on diorama-based interface.

2.1 AR interface to point a place in outdoor environment

To indicate a POI in outdoor environment, directly putting annotation tag to the real world is common in AR [8], [9]. Although annotation tag makes it possible to intuitively understand the POI, there are two technical issues in practical use; "precise registration" and "good depth cue". Annotation tag could be misaligned according to camera registration error. However, precise and robust viewpoint registration in outdoor environment is still active research topics [10]. Even if registration has been done accurately, it is still difficult to perceive correct distance to annotation tag from user's view in the real world [8].

These problems can be avoided by using 3D map; the position of arrow-shaped icon indicating POI is represented in the diorama model coordinates. Thus, precise registration is not required. And user can perceive the correct distance from their position to the POI by exploring the 3D map. Designing interaction to the 3D map is important to realize effective navigation to POI [11]. Though there are user studies that display only 3D/2D maps [12], [13], we think that, diorama-based approach like WIM [1] is a good approach to share a POI related to the real world. We believe that a good interaction technique with mobile device pose can be more convenient than interaction with buttons, joystick or touch-screen in outdoor environment. Therefore, we utilize mobile device pose for our viewpoint manipulation methods of the diorama model.

2.2 User study on diorama-based mobile interface

A small-scaled CG model of surrounding real environment (i.e., diorama model) is useful to show the geometrical information. It visualizes a place where is not visible from a user's viewpoint. Thus there are many interfaces utilizes a diorama model. In this paper, we call them as diorama-based interface.

The concept of WIM, overlaying a diorama model of surrounding environment on user's view, was originally developed by Stoakley et al [1] to support navigation and interaction with virtual environment. Wingrave et al [2] conducted a detail user study on Scaled and Scrolling WIM. In AR context, Blaine et al [3] presented WIM with head mounted display. They used head-motion to manipulate the viewpoint. Höllerer et al [5] presented wire-frame rendered WIM aligned with the real world for pedestrian navigation using head mounted display. Okuma et al [6] conducted user study about viewpoint manipulation method for museum guide using 3D map.

However, there has not been any investigation about the suitable viewpoint manipulation method of WIM in outdoor environment. We compare three viewpoint manipulation methods utilizing mobile device pose. Our viewpoint manipulation methods can be considered as variations of Okuma's "bird's eye view + automatic rotation" with different interaction.

3 Viewpoint Manipulation Methods

In this section, we explain three viewpoint manipulation methods. For all methods, horizontal orientation of diorama model is aligned with the real world to reduce user's mental rotation [12].

3.1 Free shot

Free shot manipulation method fixes orientation and position of a diorama model to the real world. The viewpoint of the diorama model is fixed to the mobile device. Therefore, a user can

manipulate the viewpoint in 6 degree-of-freedom. It is also called as AR view. Figure 1 (left) illustrates a user's motion to look around the diorama model. The user can observe the diorama model displayed on the mobile device monitor as if the model is set on the table in front of the user. Therefore, we initially assumed that free shot is the most intuitive viewpoint manipulation method among the all methods. Pros and cons of this method are follows.

Pros: Intuitive method with highest degree-of-freedom.

Cons: A user needs to walk around the diorama model to translate the viewpoint.

3.2 Pan/tilt shot

Pan/tilt shot (as known as egocentric view) utilizes touch screen instead of the mobile device position to translate the viewpoint in horizontal directions. Hence a user does not need to walk around the diorama model. Only orientation of the viewpoint is manipulated by the mobile device pose. As shown in Figure 1 (middle), position of the diorama model with respect to the real world is fixed unless the user manipulates the horizontal position by dragging manipulation on the touch screen. A vertical dragging from top to down moves the viewpoint to forward. A horizontal dragging from left to right moves the viewpoint to rightward. To looking a place from opposite direction, the user has to turn the mobile device around, and then translate the viewpoint to where the place is seen. The manipulation of vertical position of the viewpoint is omitted in this method. Pros and cons of this method are follows.

Pros: A user does not need to walk around the diorama model to translate the viewpoint.

Cons: Less intuitive and fewer degree-of-freedom than free shot.

3.3 Dolly-around/Crane shot

Dolly-round/crane shot (D/C shot) can be considered as orbital viewing [14] without twist. The viewpoint of the diorama model is moved on the surface of a virtual sphere surrounding a center of rotation. Note that the gravity direction is always aligned with the real world. In D/C shot, the orientation of mobile device is mapped so as to move the viewpoint as shown in Figure 1 (right). It is easy to view the POI on a diorama model from different viewpoint if the center of rotation is placed on the same position of the POI. Quantitative experiment [15] shows that the method is preferable than other head-tracked and non-head-tracked methods.

D/C shot also utilizes dragging manipulation on the touch screen to translate the center of rotation in horizontal directions for exploring the diorama model. Pros and cons of this method are follows.

Pros: A user does not need to walk around the diorama model to translate the viewpoint. Moreover, the user does not need to drag touch screen to look around the center of rotation.

Cons: Less intuitive than pan/tilt shot and free shot.

4 User Study

We have conducted on a user study to compare the three different viewpoint manipulation methods presented above in outdoor environment. To compare the difference between these methods, we fixed the design of interface other than the viewpoint manipulation method by using a test bench interface. The test bench interface has been implemented on tablet PC (Sony VGN-UX92PS) with external inertial sensor (InterSense InertiaCube3). The pose tracking system for the test bench interface is simple. For free shot manipulation method, we used 15[cm] x 15[cm] sized marker set on a tripod stand of 75[cm] as high as common table top to track the mobile device. For pan/tilt shot and D/C shot, we used only inertial sensor since these methods require the orientation but not the position. Besides the viewpoint manipulation method, there are some design factors of our interface, such as the reality of the diorama model or visibility of the mobile device. These factors are not changed through the experiment using our test bench implementation in order to focus on the viewpoint manipulation method.

Figure 3 shows a screenshot of our test bench interface. The screen is divided into two parts; "diorama part" and "GUI part". In diorama part, a user can observe a human-shaped icon to show the user's current location and an arrow-shaped icon to indicate POI on a diorama model. When the icon is occluded from the viewpoint by the diorama model, the occluder object is rendered as a translucent object in order to keep the arrow visible. In GUI part, a slide-bar for controlling the scale of diorama model and mini camera image is displayed. The control for scaling is important to explore a large scale environment. The initial scale of the diorama model is set as 1/150 and could be varied 1/50 to 1/500.

Figure 4 shows the aerial view and a photorealistic textured diorama model of corresponding area used for the experiment. The experiment was conducted as paired comparisons. Participants compared three pairs of viewpoint manipulation methods ("Pan/tilt - Free", "D/C - Free", and "Pan/tilt - D/C"). We divided the participants into six groups in order to counterbalance the presentation order of the three pairs. We had at least two participants in each group so as to counterbalance the presentation order of the viewpoint manipulation methods in each pair. In each comparison, participants repeated user task ten times after few time practices. In the practice, we used same ar

After the repetition of the task, they answered following questions on a 5-point Likert scale.

Q1. Which was easier to understand the POI?

Q2. Which was easier to manipulate the viewpoint?

Q3. Which do you prefer to use?

Participants conducted 10 (trials per method) × 2 (methods in each pair) × 3 (combinations of viewpoint manipulation methods) = 60 trials through the experiment, and it took about 30 minutes.

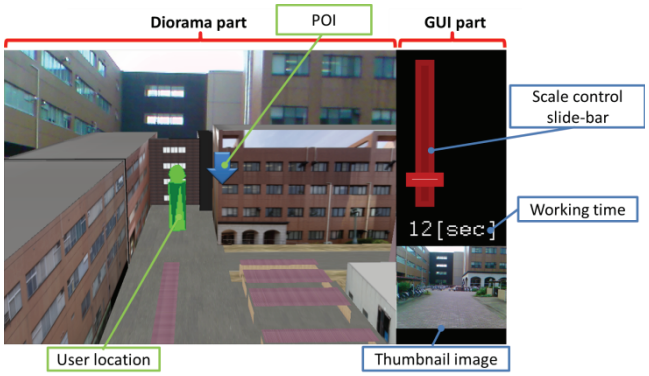


Figure 3. A screenshot of the mobile device display. The screen consists of two parts; one is diorama part where the diorama model, the real world and a user location and POI are displayed. The other is the GUI part with touch controllable slide-bar for controlling the scale, timer for measuring the working time for each task, and the thumbnail video image of the real world.

Figure 4. Experimental environment. (Left) An aerial photo. (Right) the photorealistic diorama model. Size of the modeled area is about 120[m] x 240[m].

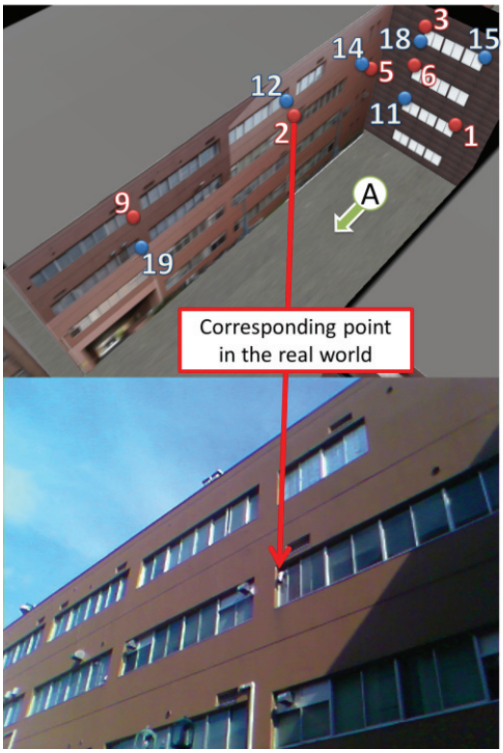


Figure 5. In our experiment, participants had to locate the corresponding point in the real world (top) to the POI in the diorama model (bottom), and took photo of the location to complete the task.

4.1 User task

The most important feature of the diorama-based interface is ability to indicate 3D POI. When a POI is indicated by an arrow on the diorama model, the user can easily understand the position of the indicated POI even if it is invisible from him/her. In order to do correct "photo-shooting task", the user has to understand the POI location accurately. Thus, we choose the task for evaluating the three types of viewpoint manipulation methods. In this task, a POI is indicated on the diorama model to participants. Then participants report the position by taking a photo of the POI. Participants were allowed to walk around the AR marker. They were requested to finish the task as soon as possible. Before starting the task, participants stood at initial position (point (A) in Figure 4) and turned the mobile device to its initial orientation (direction of the arrow). This initialization was done in each repetition, and hence the drift error of inertial sensor was very small during the task.

To complete the task, participants had to locate the corresponding point in the real world (top of Figure 5) to the POI in the diorama model (bottom of Figure 5), and took photo of the location. To evaluate performance, we measured the working time from appearing the diorama model and POI on the display to shooting the photo by a participant. We prepared two groups of ten POIs that are carefully selected in order not to be uneven distribution. In each viewpoint manipulation method, we measured the working time repeatedly ten times with POIs selected from one of the groups in random order. To evaluate subjective impressions, we used paired comparison between all the methods as explained in the section 3.

4.2 Results

Figure 6 shows box-plot of the working time of overall trials of photo-shooting task as a result of the performance evaluation with 15 participants (13 males and 2 females). We ran one-way ANOVA of the statistical software package SPSS for the results. We found significant difference between the means of working time for the viewpoint manipulation methods $F(2,925)=4.797$, $p=0.008$ with 5% significance level. A Tamhane post-hoc test revealed that D/C shot (12.2[sec]) was significantly faster than free shot (13.6[sec], $p=0.016$) and pan/tilt shot (13.7[sec], $p=0.030$), and there are no significant difference between pan/tilt shot and free shot ($p=1.000$).

Figure 7 shows the result of subjective questions for each pairs with 17 participants (15 males and 2 females). We analyzed the result of each question using Scheffe's method of paired comparison with Nakaya's variation. As a result, we found that D/C shot was significantly better than free shot in aspect of all questions at 1% significance level. We also found that participants preferred to use pan/tilt shot than free shot at 5% significance level about question Q3. Although we found no significant differences between pan/tilt shot and D/C shot, D/C shot was tend to be preferred than pan/tilt shot.

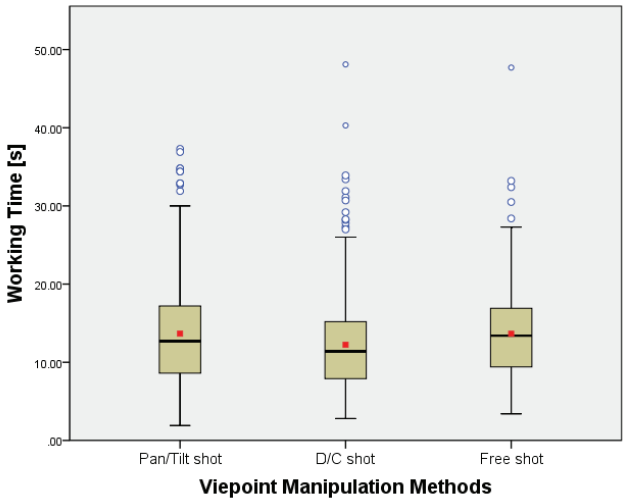


Figure 6. Result of "working time" which is measured as a performance evaluation. In the box plot, the red squares indicate the means and blue circles indicate outliers. Dolly-around/crane (D/C) shot was significant faster than free shot and pan/tilt shot (significance level is 5%).

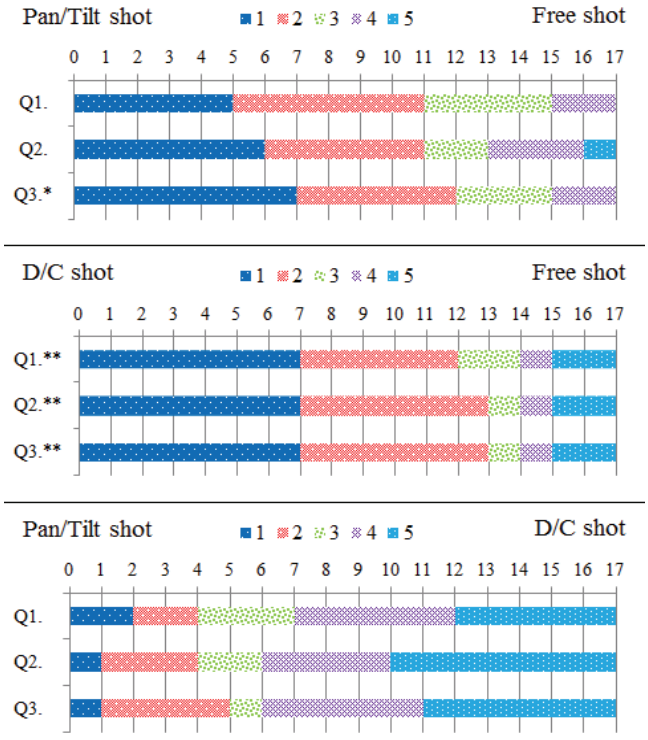


Figure 7. Result of subjective questions of paired comparison in the viewpoint manipulation methods. The graph shows the number of participants of each score. * and ** denotes significant at 5% and 1% level in correspond.

4.3 Discussion

In this experiment, the result shows that D/C shot was obviously superior to the other methods in aspect of performance and subjective impressions. By contrast, free shot could not obtain a good evaluation.

We initially expected that more participants prefer the free shot method in subjective impressions since this is the most intuitive method that has a real world metaphor of a diorama model on a table. One possible reason of the unexpected negative impression for this method is the limitation of implementation about the registration of a mobile device. Tracking of ARToolKit marker [7] was sometimes unstable in the user study, so that some participants reported that they feel discomfort to find the POI. Six participants reported that free shot was easy to use in early trials. However, the other two methods were better when they got used to the methods. Moreover, three participants were not preferred to walk around the AR marker. Therefore, there is also a possibility that the free shot was not suitable for finding a POI.

Pan/tilt shot have been preferred than free shot, though there is no difference in the working time. It seems that the time to translate the viewpoint by touch screen is as long as the time to walk around the AR marker (sometimes extended by tracking error). Therefore, using touch screen instead of walking to translate the viewpoint did not improve the performance evaluation in this experiment.

In D/C shot, over 70% (12) participants preferred this method than free shot, and almost 60% (11) participants preferred than pan/tilt shot. It seems that D/C shot makes it easy to compare POI and its corresponding point in the real world. The possible reason of the result is that diorama model was always visible in the center of the view. Hence it could be easy to compare POI and its corresponding point in the real world. This feature of D/C shot also has a bad effect. The diorama model often occluded a point of real world when they aimed to take a photo. Though it was possible to see thumbnail of camera image without diorama model, some participants demanded to control the visibility of the diorama model.

Though POI on the diorama model often was occluded by the diorama model during the experiment, it does not seem to be a big problem. Simple translucent rendering of occlude object seems effective in the experimental environment. We think we need further investigation to address the self-occlusion problem.

5 Conclusion

We have compared three viewpoint manipulation methods utilizing mobile device pose, free shot, pan/tilt shot and dolly-around/crane (D/C) shot for diorama-based interface. We have conducted a user study to evaluate the performance and participant's subjective impressions to find a 3D point of interest (POI). The experimental environment was a part of the campus of our university, and the photorealistic diorama model of the area. As a result, D/C shot was relatively superior to the other methods in aspect of performance to find a POI and subjective impressions. Our experiment has some limitations; the result of free shot was influenced by the instability of camera tracking. The effect of self-occlusion of the diorama model was limited since the experimental environment is not so complicated. We need further experiment with more robust tracking and more complex environment.

Acknowledgements

This work is partially supported by Grant-in-Aid for JSPS Fellows (23310).

References

[1] R. Stoakley, M. J. Conway, and R. Pausch, "Virtual reality on a WIM: interactive worlds in miniature," *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 265–272, 1995.

[2] C. A. Wingrave, Y. Haciahmetoglu, and D. A. Bowman, "Overcoming world in miniature limitations by a scaled and

scrolling WIM," *IEEE Symposium on 3D User Interfaces*, pp. 11-16, 2006.

[3] B. Bell, T. Höllerer, and S. Feiner, "An annotated situation-awareness aid for augmented reality," *Proceedings of the 15th annual ACM symposium on User interface software and technology (UIST)*, vol. 4, no. 2, pp. 213-216, 2002.

[4] T. Höllerer, S. Feiner, T. Terauchi, G. Rashid, and D. Hallaway, "Exploring MARS: developing indoor and outdoor user interfaces to a mobile augmented reality system," *Computers & Graphics*, vol. 23, no. 6, pp. 779-785, Dec. 1999.

[5] T. Höllerer et al., "User interface management techniques for collaborative mobile augmented reality," *Computers & Graphics*, vol. 25, no. 5, pp. 799-810, Oct. 2001.

[6] T. Okuma, M. Kourogi, K. Shichida, and T. Kurata, "User Study on a Position-and Direction-aware Museum Guide using 3-D Maps and Animated Instructions," *CD Proceedings The first Korea-Japan workshop on Mixed Reality (KJMR08)*, p. 8, 2008.

[7] H. Kato and M. Billinghurst, "Marker tracking and hmd calibration for a video-based augmented reality conferencing system," *Proceedings of the 2nd International Workshop on Augmented Reality (IWAR)*, pp. 85-94, 1999.

[8] J. Wither, S. DiVerdi, and T. Höllerer, "Annotation in outdoor augmented reality," *Computers & Graphics*, vol. 33, no. 6, pp. 679-689, Dec. 2009.

[9] C. Sandor et al., "Egocentric space-distorting visualizations for rapid environment exploration in mobile mixed reality," *Virtual Reality Conference (VR), 2010 IEEE*, pp. 47–50, Oct. 2010.

[10] J. Karlekar, S. Zhou, W. Lu, Z. C. Loh, Y. Nakayama, and D. Hii, "Positioning, tracking and mapping for outdoor augmentation," *Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 175–184, 2010.

[11] A. Nurminen and A. Oulasvirta, "Designing Interactions for Navigation in 3D Mobile Maps," *In L. Meng, A. Zipf, S. Winter (Eds.), Map-based Mobile Ser- vices: Design, Interaction and Usability, Springer, Lecture Notes in Geoinfor- mation and Cartography, London*, pp. 198-224, 2008.

[12] A. Oulasvirta, S. Estlander, and A. Nurminen, "Embodied interaction with a 3D versus 2D mobile map," *Personal and Ubiquitous Computing*, vol. 13, no. 4, pp. 303-320, Jul. 2008.

[13] K. Laakso, O. Gjesdal, and J. R. Sulebak, "Tourist information and navigation support by using 3D maps displayed on mobile devices," *Workshop on Mobile Guides, Mobile HCI 2003 Symposium, Udine*, no. September, pp. 3-8, 2003.

[14] D. Koller and M. Mine, "Head-tracked orbital viewing: an interaction technique for immersive virtual environments," *Proceedings of the 9th annual ACM Symposium on User Interface Software and Technology*, pp. 81-82, 1996.

[15] J. C. Chung, "A comparison of head-tracked and non-head-tracked steering modes in the targeting of radiotherapy treatment beams," *Proceedings of the 1992 symposium on Interactive 3D graphics (SI3D)*, no. 919, pp. 193-196, 1992.

Augmented fly-through using shared geographical data

Sandy Martedi*
KEIO University

Hideo Saito†
KEIO University

Abstract

This paper presents a development of augmented maps using shared geographical data. In general, augmented reality applications use predefined 3D models data. In our collaborative approach, we utilize shared geographical data from Google maps and city models from Google 3D Warehouse. Our system allows users to share and use maps and city models of any location in the world. The user then can print the maps and view the overlaid virtual city models through a web camera or head mounted display (HMD). We explore suitable tracking methods on three types of maps: normal (default), terrain and and satellite maps. Finally, we present an augmented fly-through application where the user can browse and view the 3D models on a paper map.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities;

1 Introduction

Some approaches and applications for visualizing the digital geographic data have been explored intensively. For example, geographical information system (GIS) accommodates the creation, management and visualization of geographic data such as digital maps and 3D city models. Generally, GIS data contain layers and localized data including numbers, vectors, images, and text. Existing GIS is used only for overlaying those 2D information on digital maps. State of the arts of GIS tried to enhance the visualization of maps by adding more informations including multimedia and 3D models in order to support geographical data analysis. For instance, Bing Map from Microsoft attempted to visualize GIS in form of surveillance and satellite maps in real time [3]. On the other hand, Google map, earth and 3D warehouse are developed to visualize geographical data seamlessly on desktop computers and mobile devices [5]. Especially, 3D warehouse supports the creation of city models and collaborative 3D modeling using geographic coordinate as the shared space.

To enhance 3D GIS visualization, augmented-reality-based systems have been developed. A system so called augmented maps combined digital layers and a paper map. Augmented maps overlays virtual data such as city names, region descriptions or 3D models of landmarks and buildings on top of a paper map.

In general, 3D models are prepared beforehand and dedicated only for a specific augmented maps application. It is feasible only for a particular region. However, to realize augmented maps for any location, preparing many maps and 3D models can be time-consuming.

To solve this issue, we are interested in the collaborative solution by sharing geographical data that available on the server such as Google Maps and 3D Warehouse. This solution allows users to download and use the shared data for their augmented maps application. Particularly, since Google maps and 3D warehouse lay on the same geographic coordinate, it is possible to merge or augment them straightforwardly.

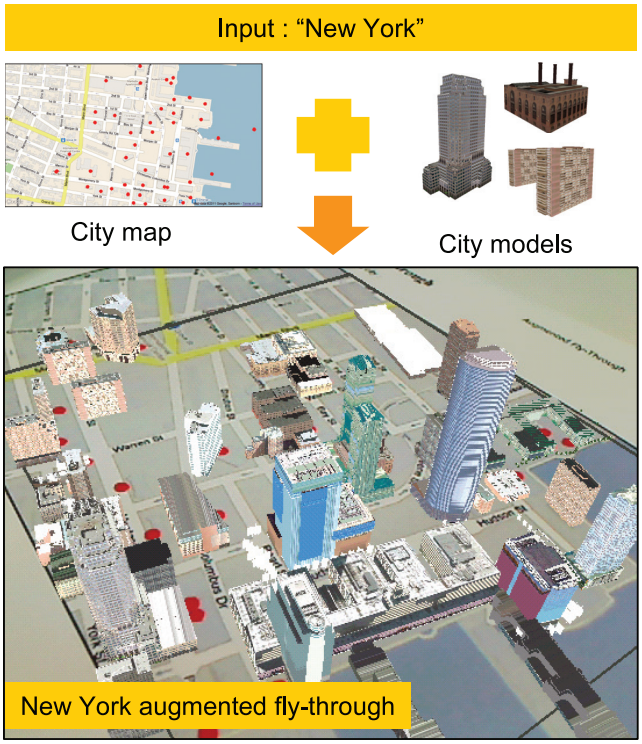


Figure 1: System overview. First, the user inputs a city name in our data extraction tool, for instance ”New York”. The user then downloads the 3D models from the 3D warehouse. The user prints the map and views the overlaid 3D models on the map.

We aware that the main difficulty of augmented maps is how to detect and track the paper map. Therefore, we need to study a suitable detection and tracking method for a paper map.

Image analysis for image tracking such as map indexing is feasible by separating the map into some layers such as roads, intersections or regions. It becomes easier to extract important features of map from those layers for detection and tracking. However, usually those layers are not accessible on a map server such as Google Map. Instead, only points data added by users and the map image are usually available. Therefore, additional image analysis and data preparations are necessary to create a track-able map for augmented maps.

We propose an approach on preparing a map for augmented maps by adding a tracking layer above the background map instead of separating layers from the map image. This approach gives us a freedom to choose any features for tracking regardless the type of maps we will use. Therefore, this approach is applicable for any type of map. We also develop a tool for retrieving the trackable maps and 3D city models from Google Maps and 3D Warehouse.

*e-mail: sandy@hvrl.ics.keio.ac.jp
†saito@hvrl.ics.keio.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

The user can download the desired maps and virtual contents by inputing a city name. The next step is printing the map and prepare a web camera or HMD. Finally, the user can view the 3D city models are overlaid on top of the paper map.

Our system improves the prior approaches on augmented maps. We also insist that our system is a novel system that combines the augmented reality with shared model data available on the Internet. To the best of our knowledge, there is no other research that has explored the augmentation of shared 3D city models by adding tracking layer above the background map. Finally, we demonstrate our approach by developing an augmented fly-through application that allows user to explore augmented maps using the web camera as a pointer.

The rest of the paper is organized as follows. The prior researches are described in the Sec. 2. We explain our proposed work in the Sec. 3 followed by the description of our augmented fly-through application in Sec. 4. We evaluate our scheme using some tracking methods in terms of the number of matches and computational cost in Sec. 5 and discuss it in Sec. 6. Finally, we conclude this paper in Sec. 7.

2 Related Works

AR toolkit has been introduced in the early stage of augmented reality [13]. AR toolkit is initially used for a collaborative conference system. There have been some researches on utilizing AR toolkit for developing augmented maps. Hedley et al. combined the augmented reality with geographical data visualization [11]. They also equipped the system with fingertip detection and interaction. Bobrich et al. also used the AR toolkit to track a planar map [7]. McGee at al. developed a collaboration system for augmented maps by placing four AR markers near the printed map [17]. The user can draw annotation on a paper map by using digital pen and share their modifications with the other users. However, AR toolkit is not robust against occlusions. The virtual contents are also limited to predefined data. On top of that, the marker obstructs the appearance of the map.

Many approaches tried to develop a marker that is visually related to the system. Some attempts used a map as the marker and extract its features for matching such as utilizing SIFT feature detector [15]. Recently, a fast keypoint detection using random ferns had also been developed for map tracking [21]. Another approach used mutual information between two map images for tracking [9].

Reitmayr et al. developed augmented maps used natural features tracking in table-top system equipped with a projector [23]. Their system could project the additional information on top of the map. Moreover, the user can select information using PDA device as a pointer. Similarly, Rohs et al. used patches in a paper map as the visual descriptor for detection and tracking [24]. Furthermore, they used mobile devices for displaying additional information.

In contrast to the texture-based map tracking above, keypoints based tracking has been also explored. Nakai et al. used random keypoints as features for camera pose estimation [20]. Their work fails on matching surface in extreme camera tilt because they do not use the previous successful tracking information. Their work had been improved by Uchiyama and Saito in so called the tracking by descriptor update [25]. The method updates the descriptor database and successfully tracks a paper in extreme camera tilt. By using the same method, they had developed augmented maps with intersections in a map as keypoints [27]. They colored the keypoints and extracted them using color detection. They introduced the random dots marker that successfully achieves robust and accurate tracking using thousand markers [26]. The random dots marker is also previously applied in the development of the foldable augmented maps [16].

In this research, we utilize the random dots for detecting and tracking the maps. Specifically, we can generate many dots from geographical data such as distribution of buildings and landmarks in a map. We assume that buildings and landmarks are distributed randomly over the map. As a result of using the random dots, we can track the map while preserve the appearance of the maps because the dots are relevant information in the map.

Besides tracking, researchers have drawn attentions to the collaboration on augmented reality. Generally, augmented reality application contains predefined 3D models which are not reusable for another application. Instead of creating 3D models from scratch, some AR applications retrieve them from the Internet. For instance, AR sights system allows users to download available markers from its website and 3D models from Google 3D Warehouse [1].

Live videos augmentation on aerial map was explored by Kim et al. [14]. They also applied the real lighting condition estimated directly from the maps image. Similarly, Bing from Microsoft integrates maps, panorama pictures, and live videos submitted by users in one spatial augmentation [3]. Both works use collaboration among users. However, they only augment on the digital map instead of paper maps.

Similar to our approach, Morrison et al. used the natural features to track a printed Google map and visualized additional information on top of it [19]. However, they did not augment the 3D city models onto the printed map. Similarly, Paelke et al. integrated a paper map and additional information on mobile devices [22]. Gruber et al. provided a dataset for tracking using city models from Google 3D warehouse [10]. However, their work only covered virtual contents preparation and omitted map detection and tracking.

All approaches above used conventional way where the user has to prepare the map and the virtual data separately and connect them manually. To solve these issues, we propose an integrated data preparation procedure for retrieving shared 3D models from Google 3D warehouse together with the map image from Google Maps. Since they are in the same geographical coordinate, both are automatically connected. Therefore, we can augment the 3D models directly onto the paper map. Furthermore, we can use the map for tracking as replacement of fiducial marker. Since we can query any location in Google Maps, as a result, we can build the world augmented maps using our approach.

3 Proposed System

Our proposed system focuses on the data access to maps and 3D models database for realizing the augmented maps. We also explore the suitable tracking method for the maps. In our system, we use Google Maps and 3D Warehouse as our resources. The flow of our proposed system is illustrated in Fig. 2.

3.1 Map Data

We retrieve maps from the Google map. Three types of map are available: default, terrain, and satellite maps as illustrated in Fig. 3.

3.1.1 Map production

We create a map by combining 3D model locations as a tracking layer on top of a background map as illustrated in Fig. 4. We define the tracking layer as colored dots. This layer is then extracted back in the initialization step using color detection.

We create a tool for retrieving the produced maps (tracking layer + background map) from Google Maps and 3D models from 3D Warehouse as illustrated in Fig. 5. Our tool receives a city name as the input and displays locations of 3D models on top of the background map the city as colored dots. The right panel beside the map enlists the name of the model appears on the map. The user can download the local database of the city map with the associated 3D city model. The output of our tool are a trackable map, a set of 3D building models inside the map area and a text file contains the position of 3D building in geographic coordinate. The text file is used for detecting the paper map in the initialization step 3.2.1.

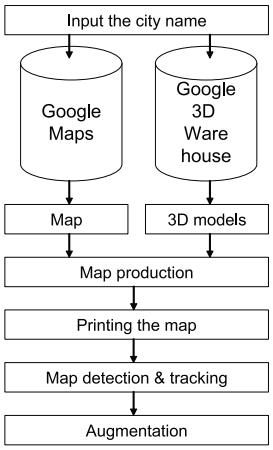


Figure 2: Flow of the proposed system. First, the user inputs a city name. The system requests to servers for a map and 3D models. The trackable map is then made in the map production. The user then prints the map. At online phase, the paper map is detected and tracked. The 3D models are then augmented.

3.1.2 3D Data

Google 3D Warehouse allows users to create and share their 3D models. Since all 3D models in Google 3D Warehouse are made in geographic coordinate, they can be augmented directly onto map that also uses geographic coordinate. For our system, we only use 3D models of buildings and landmarks in a city. Currently, the user downloads the maps and 3D models using our tools beforehand. However, in the future, the 3D models can be downloaded simultaneously when the application runs.

3.2 Map Tracking

For augmenting 3d models on top of a map, a camera pose estimation is necessary because the view should follow the orientation and the movement of the camera. One way to estimate camera pose is placing a planar marker on the map and estimate it using homography. Thus, whenever the markers is detected, the camera pose is estimated and the 3D models can be rendered correctly.

However, this fiducial marker approach is not robust against occlusion. It also obstructs the appearance of the map. Therefore, we utilize the random dots marker approach because it is robust against occlusion and the random dots marker can be generated from the geographical data. First, we prepare keypoints database that includes the location of the 3D models. By using the matching method in the random dots marker method, the geographic coordinate of the map is acquired. Then, we start to track the map using the random dots marker or another method such as SIFT, SURF or random ferns.

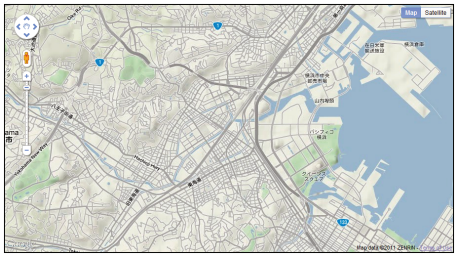
3.2.1 Initialization

The random dots are generated from the 3D models location from the Google 3D warehouse. We use these dots as the keypoints for matching. These keypoints are then stored to a file that can be loaded when the system starts as illustrated on Fig. 6. The initialization is the key of the coordinate transformation from the geographic coordinate to image coordinate. Since the dot markers can be retrieved in both coordinates using Google API, then the transformation from image to geographic coordinate can be estimated. By knowing the geographic coordinate, we can load the 3D models that exists near the position.

Keypoints are matched in the initialization step as illustrated in Fig.7. In offline phase, we create a descriptor database from the text

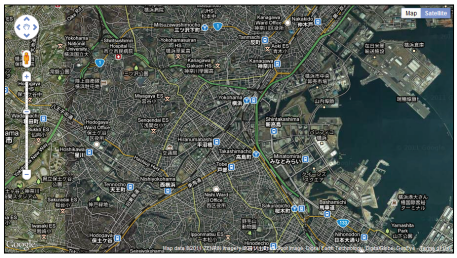
©2011 Google - Map Data ©ZENRIN

(a)



©2011 Google - Map Data ©ZENRIN

(b)



©2011 Google - Imagery ©2011 Cnes/Spot Image, Digital Earth Technology, DigitalGlobe, GeoEye, Map data ©2011 ZENRIN

(c)

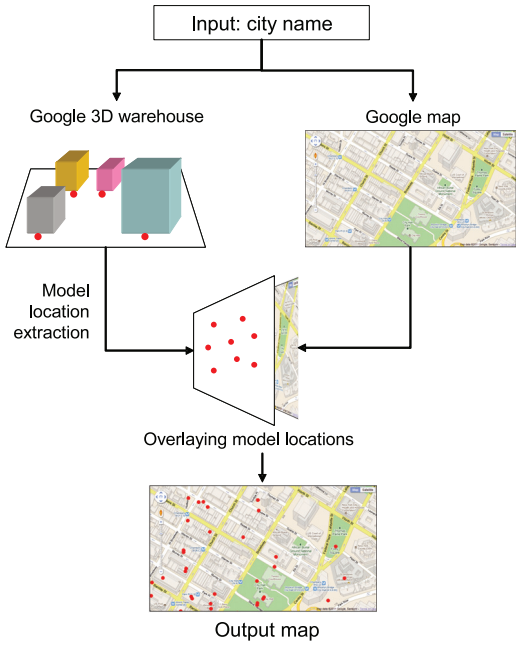
Figure 3: Three types of Yokohama map. (a) Default map. It consists some labels. (b) Terrain map. It consists dense edges and lines. (c) Satellite map. The real captured image from the satellite.

file that contains the location of buildings (dots) in the geographic coordinate. The descriptor of a point are computed by estimating its relationship to the neighboring points.

In online phase where a camera captures the printed map, we extract dots in tracking layer and output a binary image. We then calculate the descriptor for each dots followed by matching the calculated descriptors with the descriptors in the database. At this stage, many matches are established and the map is detected.

3.2.2 Tracking

After the matches are established in the initialization, we start to track the map on succeeding frames. For tracking using random dots markers, we only updates the descriptor database using descriptors calculated in current frame. For tracking, we also apply another tracking methods such as SIFT, SURF and random ferns. In contrast to random dots marker that uses keypoint features, SIFT, SURF and random ferns use the texture of map image for matching. Theoretically, all four methods are applicable for augmented maps. However, each method requires different initialization. For instance, creating the index of the keypoints database is necessary for the random dots marker. Whereas random ferns method requires a learning process. The preparation time varies to each method. We compare the results of each method in terms of the number of suc-



©2011 Google - Map Data ©2011 Google, Sanborn

Figure 4: Map production flow. First, a city name is queried. The tool extracts the map and 3D models of the city. The local database for initialization are then built. The 3D model positions are overlaid as the colored dots.

cessful tracking and the computational cost in Sec. 5.

4 Interaction for the Augmented Fly-Through Application

We develop an augmented fly-through that allows user to browse and view the 3D city models through camera or HMD on a paper map. The user also can select the information on the map, by moving the camera as a pointer (see Fig. 8). We prepare the building name as the virtual information that will appear when the center of the camera approaches the models.

5 Results and Evaluation

The outputs of our system are the augmentation of any locations in the world as illustrated in the Fig. 9. The virtual contents in our system fully depends on the models availability in Google 3D warehouse.

For tracking, we apply four detection and tracking methods: random dots markers, SIFT, SURF, and random ferns. We count the successful tracking based on three types of map: default, terrain and satellite map. We also estimate the computational cost for convincing that our system runs in real time. For our experiments, we use a web camera with resolution 640×480 and we print the map in A4 size paper.

We implemented our system using OpenCV library [6]. The city models are loaded using Open Asset Import Library (Assimp)[2]. We calibrate the camera using the Calibration tool [4] that is based on the implementation of Zhang calibration method [28]. For experiments, we use a laptop computer with specifications: Intel I7 Quad Core 2.80GHz and 4GB memory.

5.1 Tracking Robustness

In this section we show the percentage of successful tracking on image sequences contains the image map. For this experiments, we

©2011 Google - Map Data ©2011 ZENRIN

Figure 5: A tool for extracting city map (Yokohama) from Google Maps and 3D warehouse. The red dots on the map represent the location of 3D building models. The list on the right panel shows the available 3D building models.

use Kyoto map that consists 84 mesh models. Accordingly, we have 84 dots on the map. We prepare three image maps: default, terrain and satellite maps. We capture those three maps using web camera and record them as image sequences. We then applied each tracking method on the image sequences. Successful tracking occurs if the system can detect the map on the image sequences. We then reproject the border of the map using the homography to the map image. We count the frame of which the projected border is near to the actual border of the map in the paper divided by the number of frames as illustrated in the Fig. 10.

According to the results, texture-based tracking using SIFT, SURF and random ferns are robust on default and terrain maps. We note that default and terrain map have strong edges and distinctive colors that help the successful tracking. On the other hand, the robustness of tracking drops on the satellite maps. In the satellite maps, the texture is relatively uniform that makes the matching becomes difficult.

As we expected, the random dots marker method could track the paper maps regardless the type thanks to the tracking layer. Surprisingly, the tracking robustness even increased on the satellite maps . We realize that the dots become distinctive enough on the satellite map to make them easy to extract.

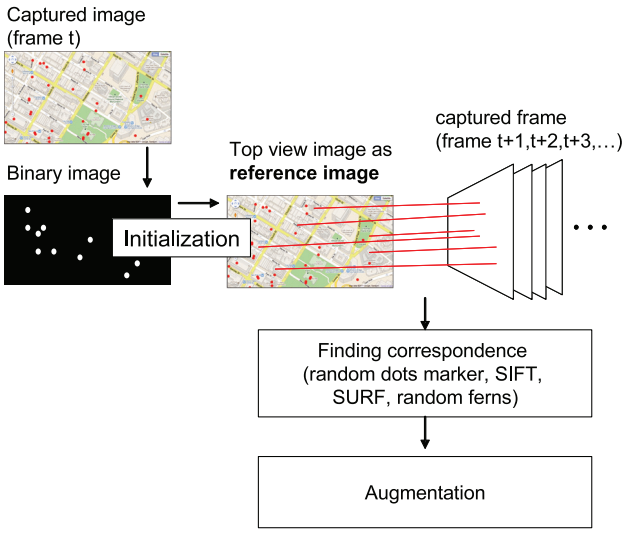
5.2 Computational cost

We evaluate the computational cost of our application on the matching and rendering process. First, we provide a map with some models for augmentation. We compute the average time for matching the map with the reference map and rendering the models as listed in Table. 1.

Table 1: Computational cost based on the type of the map and tracking method.

Method	default map (ms)	terrain map (ms)	satellite map (ms)
Random dots marker	45.41	52.9	51.49
SIFT	853.39	844.91	926.02
SURF	462.61	440.06	827.46
Random ferns	174.26	168.44	202.14

We can see from the results that random dots method works faster than the other method because it depends only on the color



©2011 Google - Map Data ©2011 Google, Sanborn

Figure 6: From initialization to tracking. First, a frame is captured. Each frame is binarized and the keypoints are extracted using color detection. The descriptors are then computed and matched with the keypoints database. When the frame is matched, the homography is computed. The frame is warped using inverse homography to get the top view of the map image as reference for tracking.

detection. In addition, it uses a hashing technique for fast descriptor lookup. On the other hand, the matching method that utilizes the texture information such as SIFT, SURF and random ferns requires longer time. SURF has better performance than SIFT thanks to the integral image approach. Random ferns works best among the three methods. However, random ferns requires around 10 minutes learning or building database beforehand. This 10-minutes-long learning is not suitable for our purpose because preparing learning data every time the user download the map will make our application unpractical.

6 Discussion

In our experiments, tracking using SIFT, SURF and random ferns can work robustly for default and terrain maps. We can choose those method if we want to use default map and terrain maps. On the other hand, if we want to use the satellite maps, random dots marker can be the alternative for tracking.

Moreover, random dots marker contributes less time than the other method thanks to the simple extraction method and hashing. The computational cost is significant for deciding the suitable tracking method for augmented maps. Comparison of another feature descriptor for tracking paper map such as BRIEF [8], or GLOH [18] and its variant are the next step of this research. Furthermore, it is interesting to explore on combining the random dots marker with the other texture based method for realizing the best tracking method for augmented maps.

Technically, we add a tracking layer on a background map for initialization. This solution is practical to implement. Currently, we create the tracking layer by overlaying colored dots over the background map. In order to create more realistic map, instead of colored dots, we can print colored icons.

On the other hand, instead of adding tracking layer on the background map, we are interested on different approach by extracting specific features on the original map. Since there are terrain and satellite map, feature extraction will be different for each type and

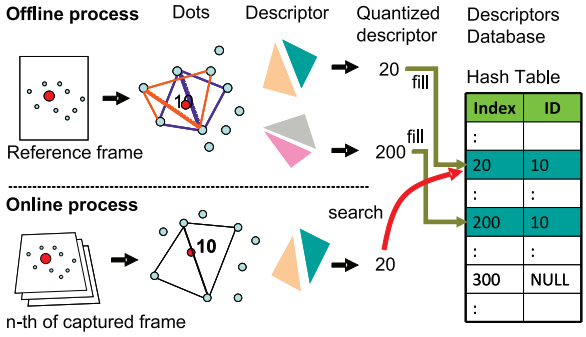
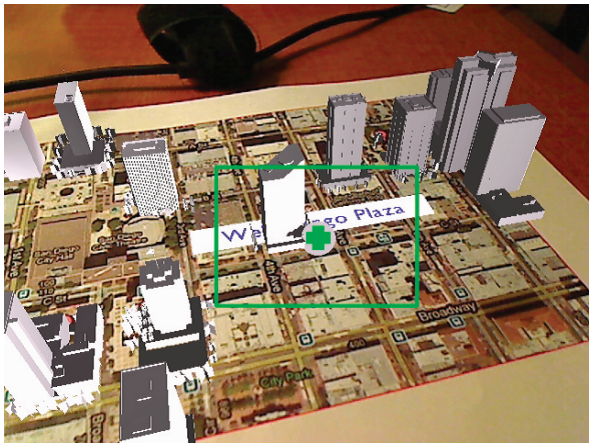


Figure 7: Keypoints matching. There are two main process on the matching. Offline process estimates descriptors of all dots/keypoints and store them to a descriptor database as the index of dots id. The online process extract dots from captured image and compute the descriptor of the dot followed by searching the dots id based on the descriptor.



©2011 Google - Imagery ©2011 Google, Map data ©Google

Figure 8: Data selection using center of the camera image. The user moves the camera to select and display the information of the San Diego map. The name appears when the center of the camera image approaches a building model.

we lose the generality for each type of map. However, defining features for particular type of map is interesting issue to explore.

We are also interested on bringing this research more on the collaboration aspects. Through our application, we have shown that the geographic coordinate can be utilized as the shared space for augmented reality. However, our current implementation only covers the viewing workspace of the collaboration. To add the individuality features for collaborative AR environment [12], it is necessary to let user to create and modify the shared 3D models on the augmented maps. Thus, each user can view coherent 3D models on their site.

7 Conclusion

We have presented our augmented fly-through using shared geographical data. The users can use and share virtual contents from the maps and 3D model database server. Therefore, they can view any location in the world through augmented reality. We have also presented our study on the characteristics of map and its usage for developing augmented maps. Finally, we proved that random dots

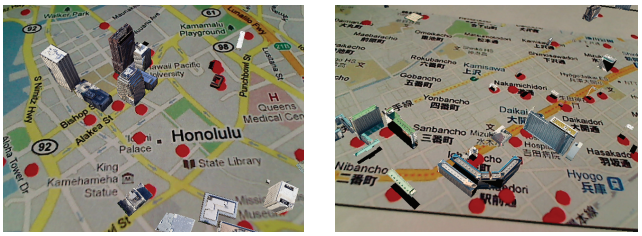


Figure 9: Augmentation results. 3D building models are augmented on top of printed maps. (a) Honolulu (b) Kobe (c) Park Avenue (d) San Diego.

marker is suitable for our system in terms of the minimum computational cost.

Providing an instant way for retrieving the maps and virtual data for augmented maps remains as our main challenges. In the future, on-line connection to the database server and the cloud architecture are promising outlooks to access the maps and virtual data efficiently. As a result, we can achieve the robust and rich augmented maps application. We will improve our user interaction by implementing robust finger detection. We will also work on occlusion handling and realistic rendering for augmented maps.

Acknowledgement

This work is supported in part by a Grant-in-Aid for the Global Center of Excellence for high-Level Global Cooperation for Leading-Edge Platform on Access Spaces from the Ministry of Education, Culture, Sport, Science, and Technology in Japan and Grant-in-Aid for JSPS Fellows. We also thank Hideaki Uchiyama for providing the random dots marker source code.

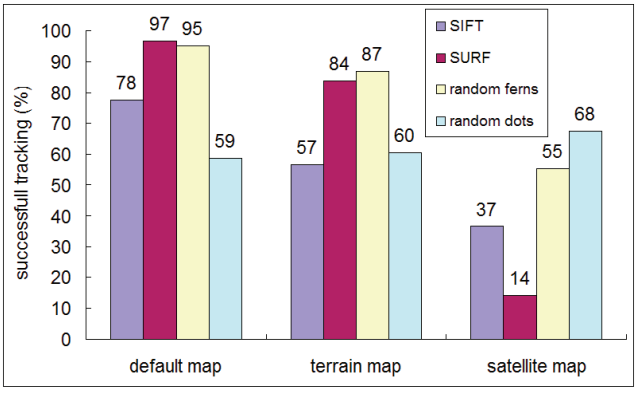


Figure 10: Successful tracking rate. Random dots marker works on any type of map.

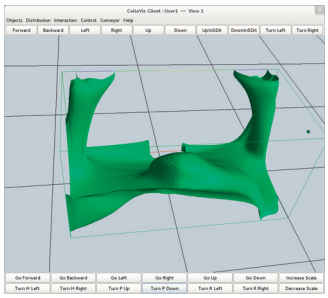
References

[1] AR Sights. http://www.arsights.com/.
[2] Assimp. http://assimp.sourceforge.net/.
[3] Bing Maps. http://www.bing.com/maps/.
[4] Camera Calibration Tools. http://www.doc.ic.ac.uk/ dvs/calib/main.html.
[5] Google Map and 3D Warehouse. http://www.google.com/.
[6] OpenCV. http://sourceforge.net/projects/opencvlibrary/.
[7] J. Bobrich and S. Otto. Augmented maps. In *ISPRS*, 2002.
[8] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. BRIEF: Binary Robust Independent Elementary Features. In *Proc. ECCV*, September 2010.
[9] A. Dame and E. Marchand. Accurate real-time tracking using mutual information. In *Proc. ISMAR*, pages 47–56, 2010.
[10] L. Gruber, S. Gauglitz, J. Ventura, S. Zollmann, M. Huber, M. Schlegel, G. Klinker, D. Schmalstieg, and T. Höllerer. The City of Sights: Design, construction, and measurement of an Augmented Reality stage set. In *Proc. ISMAR*, pages 157–163, 2010.
[11] N. R. Hedley, M. Billinghurst, L. Postner, R. May, and H. Kato. Explorations in the use of augmented reality for geographic visualization. *Presence*, 11:119–133, 2002.
[12] A. Ismail and M. Sunar. Survey on collaborative ar for multi-user in urban studies and planning. In *Learning by Playing. Game-based Education System Design and Development*, volume 5670 of *Lecture Notes in Computer Science*, pages 444–455. 2009.
[13] H. Kato and M. Billinghurst. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *Proc. IWAR*, pages 85–94, 1999.
[14] K. Kim, S. Oh, J. Lee, and I. Essa. Augmenting aerial earth maps with dynamic information. In *Proc. ISMAR*, 2009.
[15] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60:91–110, 2004.
[16] S. Martedi, H. Uchiyama, G. Enriquez, H. Saito, T. Miyashita, and T. Hara. Foldable Augmented Maps. In *Proc. ISMAR*, pages 65–72, 2010.
[17] D. McGee, X. Huang, P. Barthelmess, and P. Cohen. Poster: The neteyes collaborative, augmented reality, digital paper system. In *Proc. 3DUI*, pages 145 –146, March 2008.
[18] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1615–1630, 2005.
[19] A. Morrison, A. Oulasvirta, P. Peltonen, S. Lemmela, G. Jacucci, G. Reitmayr, J. Näsänen, and A. Juustila. Like bees around the hive: a comparative study of a mobile augmented reality map. In *Proc. CHI*, pages 1889–1898, 2009.
[20] T. Nakai, K. Kise, and M. Iwamura. Camera based document image retrieval with more time and memory efficient LLAH. In *Proc. CBDAR*, pages 21–28, 2007.
[21] M. Ozuysal, M. Calonder, V. Lepetit, and P. Fua. Fast keypoint recognition using random ferns. *TPAMI*, 32:448–461, 2010.
[22] V. Paelke and M. Sester. Augmented paper maps: Exploring the design space of a mixed reality system. *ISPRS*, 65:256–265, 2010.
[23] G. Reitmayr, E. Eade, and T. Drummond. Localisation and interaction for augmented maps. In *Proc. ISMAR*, pages 120–129, 2005.
[24] M. Rohs, J. Schoning, A. Kruger, and B. Hecht. Towards real-time markerless tracking of magic lenses on paper maps. In *adjunct Proc. Pervasive*, pages 69–72, 2007.
[25] H. Uchiyama and H. Saito. Augmenting text document by on-line learning of local arrangement of keypoints. In *Proc. ISMAR*, pages 95–98, 2009.
[26] H. Uchiyama and H. Saito. Random dot markers. In *Proc. IEEE VR*, pages 35 –38, march 2011.
[27] H. Uchiyama, H. Saito, M. Servieres, and G. Moreau. AR GIS on a physical map based on map image tetrieval using LLAH tracking. In *Proc. MVA*, pages 382–385, 2009.
[28] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *Proc. ICCV*, pages 666–673, 1999.

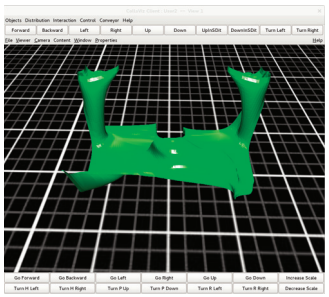
PAC-C3D: A New Software Architectural Model for Designing 3D Collaborative Virtual Environments

Thierry Duval *
Université de Rennes 1, IRISA UMR CNRS 6074, Rennes, France

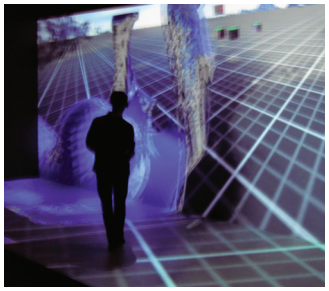
Cédric Fleury†
INSA de Rennes, IRISA UMR CNRS 6074, Rennes, France



The Java3D Visualizer



The jReality Visualizer



The Immersive jReality Visualizer

Figure 1: Three different visualizers sharing the same virtual environment

Abstract

We propose PAC-C3D as a new software model for 3D Collaborative Virtual Environments (CVE). This model merges the results from two research fields: distribution models for CVE and HCI design for computer-supported cooperative work. PAC-C3D proposes to describe each part of a shared virtual object through explicit interfaces to ensure a strong separation between the core functions of a virtual environment, its (visual) representations, and its collaborative features such as synchronization and consistency maintenance between remote users. PAC-C3D makes it possible to design a CVE with low dependency between the core functions, the distribution mode and the 3D graphics API used. It explicitly deals with the main distribution modes encountered in CVE. It makes it easy to use different 3D graphics API for different nodes involved in the same collaborative session, providing interoperability between these 3D graphics API. It also makes it possible to integrate other kinds of 3D representations such as physics engines into the CVE.

Index Terms: H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces—Theory and methods; H.5.3 [Information Interfaces and Presentation (e.g., HCI)]: Group and Organization Interfaces—Computer-supported cooperative work (CSCW); I.3.7 [Computer Graphics]: 3-Dimensional Graphics and Realism—Virtual reality; D.2.11 [Software Engineering]: Software Architectures—Patterns I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques, Device independence

1 Introduction

The design of 3D Collaborative Virtual Environments (CVE) merges the design of interactive 3D applications and the design of distributed collaborative applications. This task is complex because it must address 3D interaction and immersion issues as well as collaborative issues dealing with distribution, synchronization, and consistency maintenance of the shared virtual environment.

The configuration (adaptation to the hardware deployment systems) of CVE is complex because it must address various network characteristics (from high bandwidth on professional or experimental networks to low bandwidth on personal networks) as well as various displays and 3D interaction devices (from 6-face *Caves* [8] to simple workstations or even to simple interactive tablets). All these configurations can even be used at the same time in a single deployment in order to make asymmetric collaboration possible between remote users using different input and output devices.

To meet all these requirements, these CVE must be designed according to a software architectural model that makes it possible to adapt the distribution mode of a CVE to solve the network interoperability issues. Such a model should also make it possible to design software components that encapsulate the hardware 3D graphics requirements, in order to be able to choose at run-time the best components for each hardware configuration. Existing solutions focus either on how to manage distribution and consistency maintenance for 3D CVE, or on how to manage independence between core functions and graphics API for 2D CSCW. For now, design models for CVE deal neither with independence to 3D graphics API nor with efficient management of distribution modes.

So we propose to merge these two main research fields in order to provide a new solution, the PAC-C3D model, which offers a better way to design and implement CVE. PAC-C3D meets two requirements: ensure synchronization and consistency maintenance of CVE, and ensure independence of CVE from 3D graphics engines to make interoperability possible between such 3D engines.

In this paper, section 2 presents the context of our work: the need to design CVE with various distribution models. Section 3 presents the software architectural models used in the field of HCI and CSCW. Section 4 presents our new software architectural model and how its instances communicate together. Consistency maintenance is explained in section 5 for each of the main distribution modes. Then section 6 describes how this model can be used to address the problem of interoperability between 3D API, for 3D graphics or physics engines. Finally, section 7 gives some implementation examples illustrating how our model faces adaptation to different distribution situations and to different 3D engines.

*e-mail: thierry.duval@irisa.fr
†e-mail: cedric.fleury@irisa.fr

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

2 The Context: distributed architectures for CVE

The location of the virtual environment data (i.e. geometric data, textures, etc.) is a critical decision when designing a CVE system [23]. It determines which nodes (usually the users' computers) store this data, which nodes execute the processing related to each virtual object, and how the synchronization of the distributed objects is achieved. We distinguish three data distribution modes: homogeneously replicated, centralized, and partially replicated [19], which is similar to the approaches presented in [12]. A more complete overview of synchronization and data distribution within CVE can be found in [10, 11, 20].

2.1 Duplication: homogeneous replicated world

In replicated CVE systems all nodes are initialized with the same database that contains all the informations about the virtual environment (geometric models, textures, object behaviors, etc.). During a session, the database evolves independently on each node and all object behaviors are executed locally. Object modifications are performed locally before being sent to the other nodes by using update messages.

So latency is very low during user interactions. However, inconsistencies between users' states of the virtual environment can appear because of delays or losses during messages transmissions. Additional mechanisms must also be used to manage concurrent access to the objects to avoid conflicts.

2.2 Centralization: shared centralized world

In centralized systems all the CVE data is stored on a central server (client/server network architecture). Similarly, virtual object behaviors are executed on this server. When a user wants to modify an object, he sends a request to the server. The server processes the modification request, then transmits the up-to-date state of the object to all the nodes, including the one that has asked for modification.

This method implicitly ensures consistency between all the nodes and avoids data replication. However, this architecture can introduce latency during user interactions because each modification request has to pass through the server, and a performance "bottleneck" can appear on the server when there are many users.

2.3 Hybrid solution: partially replicated world

Many CVE systems choose hybrid solutions between totally centralized and totally replicated data distributions, mixing features to meet particular requirements of consistency and responsiveness. These solutions distribute the data and their processing among the nodes. Most of the time, a referent/proxy paradigm is used for each object. Proxies maintain a local copy of the virtual environment, only receiving update messages from the referent.

This distribution mode makes a trade-off between the advantages and drawbacks of the two other data distributions.

2.4 Dynamic solution: mixing all the modes

Some distribution mechanisms, such as the distribution model of the Collaviz system [19] even propose a mix of these three distribution modes, allowing each object to change dynamically at run-time its own distribution mode.

2.5 Synthesis about distribution modes

As each distribution mode has its own advantages and drawbacks, a good software architectural model for CVE should be able to manage these three main distribution modes and should be flexible enough to provide solutions for evolution toward new distribution or synchronization modes.

3 Related Work: models for HCI and CSCW

A lot of research work about architectural models for human-computer interaction (HCI) deals with separating clearly the graphics part of interactive software from its core part. Some of these models have been adapted to the context of computer-supported collaborative work.

3.1 Software architectural models for HCI

The most commonly used software architectural models for HCI are based either on the Model-View-Controller (MVC) model [26, 21] or on the Presentation-Abstraction-Control (PAC) model [7]. Both of them have inspired many models dedicated to particular situations: for example for Struts web-based applications (MVC-2 [9]), for C++ or Java applications (Model-View-Presenter (MVP) [25]) or for multi-modal applications (PAC-Amodeus [24]).

MVC divides interactive components into three parts: the *Model*, the *View* and the *Controller* (see figure 2(a)).

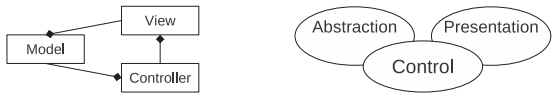


Figure 2: (a) The MVC model — (b) The PAC model

"The *Model* represents data and the rules that govern access to and updates of this data. ... The *View* renders the contents of a *Model*. It specifies exactly how the *Model* data should be presented. If the *Model* data changes, the *View* must update its presentation as needed. This can be achieved by using a push model, in which the *View* registers itself with the *Model* for change notifications, or a pull model, in which the *View* is responsible for calling the *Model*. ... The *Controller* translates the user's interactions with the *View* into actions that the *Model* will perform. ..." [16]

So the *Model* and the *View* of MVC can be closely coupled, contrary to the formal separation achieved by PAC between these two kinds of components.

PAC divides interactive components into three parts: the *Presentation*, the *Abstraction* and the *Control* (see figure 2(b)). At first look, one could consider that PAC is just another name for MVC where *Presentation* could stand for *View*, *Abstraction* could stand for *Model* and *Control* could stand for *Controller*. In practice, the PAC components have a quite different behavior than the MVC components.

"The *Presentation* defines the concrete syntax of the application, i.e. the input and output behavior of the application as perceived by the user. The *Abstraction* corresponds to the semantics of the application, it implements the functions that the application is able to perform. ... The *Control* maintains the mapping and the consistency between the abstract entities involved in the interaction and implemented in the *Abstraction*, and their *Presentation* to the user. It embodies the boundary between semantics and syntax. It is intended to hold the context of the overall interaction between the user and the application." [7]

So the *Abstraction* and the *Presentation* are not allowed to communicate directly: the *Control* acts as a mediator and filters all the communications between its *Abstraction* and *Presentation*, and with the other *Controls*.

In fact, most of the MVC-like models propose also this separation between the *Model* and the *View*, as detailed in the Oracle/Sun interpretation of MVC [16], which is very similar to the PAC model.

To ensure a better independence between these three kinds of components, the Arch model [27] proposes to add adaptor components between them (see figure 3). This model is also considered as a meta-model for other software models, which should follow this generic separation between facets of interactive components.

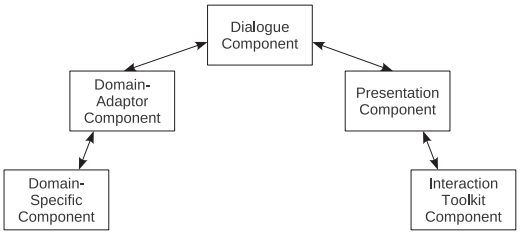


Figure 3: The Arch model

These models must now be extended to manage the collaborative aspects of CVE.

3.2 Models for collaborative HCI

Several adaptations of the PAC and Arch models have been proposed to cope with these collaborative features. These approaches rely on Ellis's conceptual model of groupware [17] or on the clover conceptual model [22]. Ellis's model proposes three complementary components or models: the ontological model, the coordination model, and the user-interface model. The clover conceptual model proposes to divide the services of collaborative software into three main parts: production, communication and coordination (see figure 4(a)). The ontological model and the production space refer to the shared virtual objects of a CVE. The coordination model and the coordination space cover the consistency maintenance in the CVE. The user-interface model refers to the representation of human-computer interaction while the communication space refers only to the communication between the users of a CVE.

PAC* [6] (see figure 4(b)) dispatches these three kinds of functions across the three PAC facets. To our opinion, this is a problem for designing *Abstractions* independently from the collaborative aspects.

Figure 4: (a) The clover concepts — (b) The PAC* model

Clover [22] (see figure 5(b)) is an extension of PAC* that relies on Dewan's "generic multi-user architecture"[12] (see figure 5(a)), which is a collaborative extension of the Arch model. Here again, each unit of the model can contain three sub-components about production, communication and coordination, especially the higher-level units that correspond to the core of the CVE.

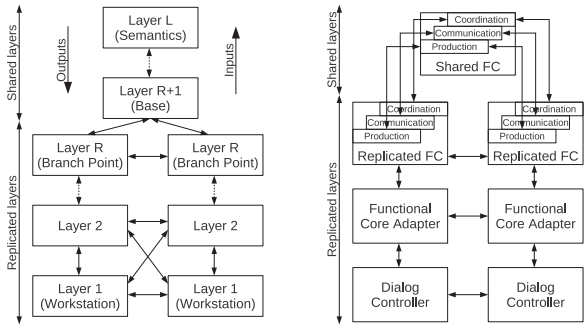


Figure 5: (a) Dewan's model — (b) The Clover model

3.3 Synthesis about HCI models and collaboration

Software architectural models for HCI propose to divide interactive components in three kinds of components that should be as independent as possible from each other. Some of these models have been extended to address the design of CVE, according to the clover conceptual model, but they do not address how to cover the three main distribution modes of the CVE. Furthermore, they spread the collaborative aspects over all the components of the models, which is a problem for designing *Abstractions* that should not be aware of the collaborative aspects.

This is why we need a new model for designing 3D CVE, which would ensure the best possible separation between core functions, visualization (3D graphics API and libraries) and collaboration aspects, and which would provide explicit solutions to achieve these different synchronization modes.

4 PAC FOR COLLABORATIVE 3D APPLICATIONS

4.1 Interfaces for independence between components

In order to make the PAC facets independent from each other, we choose a special interpretation of the PAC model that proposes interfaces to specify the features of each facet of the model. An important feature of this model is that the *Control* is a *Proxy* (GoF207)[15] of its associated *Abstraction*.

We present figure 6 a new interpretation of this model in order to allow the presence of several *Presentations* associated to the same *Control*. Each virtual shared object will be described through 3 interfaces:

- *Interface for the Abstraction* (**IA**): it declares the methods in charge of the object behavior and the methods allowing to set and get its attributes.
- *Interface for the Presentation* (**IP**): it declares the methods allowing to set and get the attributes of the representation of the object (for example the position of its 3D visualization).
- *Interface for the Control* (**IC**): it declares all the methods of the *Interface for the Abstraction*, as the *Control* will be used to manage the access to the *Abstraction* (the *Control* will be the proxy of the *Abstraction*) to maintain consistency between the *Abstraction* and all the *Presentations*, and some methods dedicated to the communication with its *Presentations* and the other *Controls*.

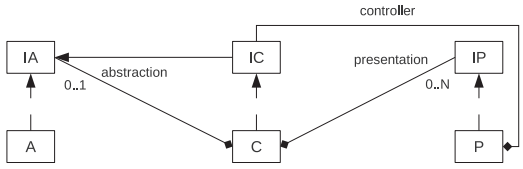


Figure 6: The PAC model with interfaces between facets

As these interfaces will be implemented by the real facets of the PAC components, at run time these facets will be instances of:

- *Abstraction* (**A**): it implements the object model and behavior, and the setters and getters.
- *Presentation* (**P**): it implements the object representation: for example it can use a 3D graphic API to visualize the object and its properties.
- *Control* (**C**): it implements the consistency maintenance between the *Abstraction* and the *Presentations*, and it regulates the access to the *Abstraction*.

Very often, a *Presentation* is closely coupled to a 3D graphics API, but thanks to the *Interface for the Presentation*, the *Control* will be totally independent of this 3D API.

In the same way, thanks to the *Interface for the Control*, the *Presentations* and *Abstraction* will be totally independent from the implementation of the *Control*.

4.2 Adapting PAC to collaboration

As for the PAC* model [6] and the Clover model [22], here again the PAC model will be the basis of our proposition, but unlike these two models, the collaborative parts will not be spread out into all the components of the model, but only into the *Control* of the PAC components.

Indeed, we consider that the objects of the production space should remain in the core parts of a 3D CVE, and that their coordination should be achieved by the *Control* of the PAC components. Last, we consider that communications between users should be either totally integrated within a 3D CVE through shared virtual objects, or totally independent of the 3D CVE, so these communication aspects are not central to a model dedicated to the design of 3D CVE. In our opinion, the distributed aspects should not impact the *Presentations* and *Abstraction* of a PAC component, in the same way that the 3D graphics details should be limited inside the *Presentations* and that the core concepts should remain in the *Abstraction*. This independence is possible thanks to the three interfaces of our model.

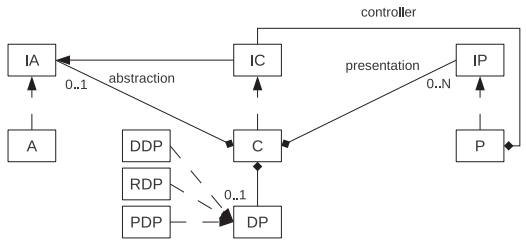


Figure 7: Adaptation of the PAC model for 3D CVE

The *Control* is associated to one distribution policy, dedicated to synchronization and consistency maintenance. There are three distribution policies:

- *Referent distribution policy* (**RDP**): implementation of what is needed to manage the set messages and to distribute updates to other *Controls*: each time the value of a parameter of the object is set, this policy makes the referent *Control* send (for example using multicast) an update message to the distributed proxy *Controls* so that they can also update their *Presentations*.

- *Proxy distribution policy* (**PDP**): implementation of what is needed to transmit the set messages toward the referent *Control*, and to manage the update messages returned by the referent *Control*: each time the value of a parameter of the object is set, this policy makes the proxy *Control* send a set message to its referent *Control*, and this value will be effectively set only when the proxy *Control* will receive an update message from its referent *Control*.

- *Duplicated distribution policy* (**DDP**): same management of the set messages as the RDP, and same management of the update messages as the PDP.

These components will be distributed across the network according to the number of nodes in a collaborative session and to the architecture chosen for the distribution.

5 Dealing with distribution modes

In this section we will detail the behavior of PAC-C3D *Controls* according to their associated distribution policy, in order to show that this behavior is able to deal with the three main distribution modes for CVE.

We will consider a modification of the value of a parameter of a virtual object occurring from a presentation component where a user will have made an action upon a shared virtual object, and we will trace the subsequent communications between the PAC-C3D components and facets.

5.1 PAC-C3D and duplicated architecture

In a CVE with a typical duplicated architecture, for each shared virtual object there will be as many instances of *Abstractions*, *Presentations* and *Controls* with a duplicated distribution policy (**DDP**) as there are visualization nodes embedding one or several representations of the shared virtual universe.

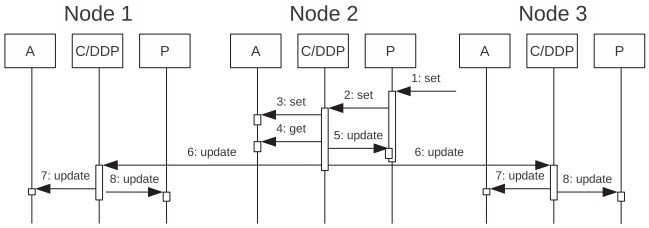


Figure 8: PAC-C3D and duplicated architecture

Figure 8 presents a typical duplicated architecture with three nodes. The exchanges between the facets of the PAC-C3D components will be as follow:

1 An action occurs upon the *Presentation* of a virtual object, on one node, to set the value of one attribute of this virtual object. This *Presentation* does not set the attribute, but instead asks its *Control* for a modification.

2-5 This *Control* receives the set request and transmits it to its *Abstraction*. This *Abstraction* processes the set requests in its own way: the final value of the attribute of the *Abstraction* can be different from the value proposed by the *Control*, for example because a proposed value could put the *Abstraction* into an incorrect status. This is why the *Control* then asks its *Abstraction* for the effective value of the attribute and updates its *Presentation* with this new value. Then the *Control* transmits this value to the other duplicated *Controls* by sending them an update message (this sending can be synchronous but it is more efficient to make it asynchronous to allow all the duplicated *Controls* to process the update at the same time).

6-8 Finally on each other node a duplicated *Control* receives the update message and asks its *Abstraction* and then its *Presentation* for an update.

As we have seen in section 2.1, the main advantage of this architecture is that when a user interacts with an object, he obtains an immediate feedback. Then all the other users may perceive the result of the interaction at the same time, with a delay corresponding to the network latency. The main drawback of this architecture is that it must ensure a strong synchronization between the nodes because of the potential autonomous behavior of some shared virtual objects. It is also quite impossible to allow several users to interact directly at the same time on a same shared virtual object.

5.2 PAC-C3D and centralized architecture

In a CVE with a typical centralized architecture, for each shared virtual object:

- there is only one instance of *Abstraction*, on the server,
- there are as many instances of *Presentations* as there are visualization nodes embedding one or several representations of the shared virtual universe,
- there is only one instance of *Control* with a referent distribution policy (**RDP**), without *Presentation*,
- there are as many instances of *Control* with a proxy distribution policy (**PDP**) as there are *Presentation* instances.

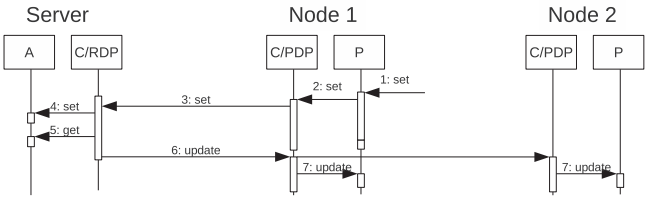


Figure 9: PAC-C3D and centralized architecture

Figure 9 presents a typical centralized architecture with one server and two clients. The exchanges between the facets of the PAC-C3D components will be as follow:

1-2 On one client an action occurs upon the *Presentation* of a virtual object, and this *Presentation* asks its proxy *Control* for a modification. This proxy *Control* transmits the set request to its referent *Control* (this transmission can be synchronous or asynchronous).

3-5 The referent *Control* transmits the value to its *Abstraction*. Here again this *Abstraction* processes the set requests in its own way and then the *Control* asks its *Abstraction* for the effective value of the attribute. Then it transmits to its proxy *Control* by sending them an update message (here again this sending should be asynchronous).

6-7 Finally on each client a proxy *Control* receives the update message and asks its *Presentation* for an update.

As we have seen in section 2.2, the main advantage of this architecture is that all the users may perceive the result of the interaction at the same time, but the drawback is that the delay for the semantic feedback is about twice the network latency. Another interesting property of this distribution policy is that it is not absolutely necessary to have a strong synchronization between all the nodes as all the behaviors are executed on the server node. Last, it is easy to allow several users to interact at the same time with the same object as the referent *Control* will be in charge of centralizing all the interactions coming from all the nodes: it can integrate all the concurrent propositions to compute a single result.

5.3 PAC-C3D and hybrid architecture

In order to answer more quickly to a user's interaction with a virtual object, it can be interesting to locate the *Abstraction* of this object on this user's node, which means to allow the referent to be on a client node rather than to stay on a centralized server. So we can consider that the hybrid architecture is a simple evolution of the centralized architecture where all the referents are not necessarily on the same node and where a centralized server is not absolutely necessary any longer.

In such a case, there will be two different situations while interacting with a virtual object, as described figure 10: either the *Abstraction* of the object is on the same node than the user, either it is on another node.

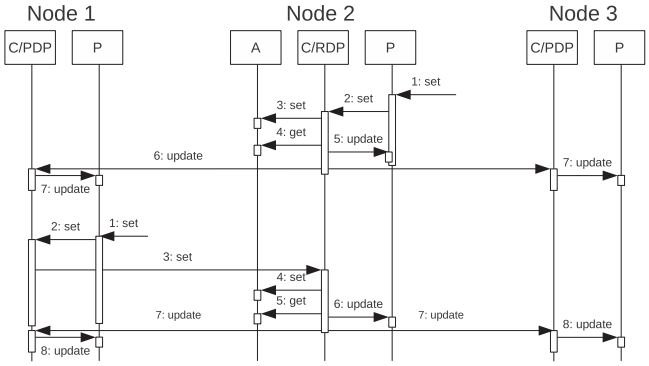


Figure 10: PAC-C3D and hybrid architecture

If the *Abstraction* of the object is on the node of the interacting user, this user will obtain an immediate interaction feedback, while the other users will perceive the interaction with a small lag mainly due to the network latency. This is very similar to the behavior of the duplicated architecture.

If the *Abstraction* of the object is not on the node of the interacting user, the user on the node of the *Abstraction* will be the first to perceive the result of the interaction, and all the other users (even the one who is interacting) will see the result of the interaction at the same time, with a delayed feedback about twice the network latency. This is quite similar to the behavior of the centralized architecture.

In both cases this hybrid solution offers the same possibility as the centralized architecture to enable several users to interact at the same time with a same object. And as we have seen in section 2.3, it also has the same main drawback as the duplicated architecture about the necessity to synchronize all the nodes because each node can be in charge of the behavior of some virtual objects.

5.4 Adapting distribution policies

To change the distribution mode of a system designed according to our model, for example to transform a centralized architecture to a hybrid architecture or to a duplicated architecture, we only have to change the distribution policy of the *Controls*: it impacts neither the *Abstractions* nor the *Presentations*. It is even possible to change dynamically the distribution policy of a *Control*, by replacing its current distribution policy by a new one, which allows to meet the requirements of the Collaviz system [19].

In the same way, with a hybrid system it is possible to enable some *Abstractions* to migrate from one node to another and to change the distribution policies of the associated *Controls* to offer a better interaction to a user by placing the *Abstraction* of a virtual object on the user's node. And in the case of concurrent interaction of two users with the same object, it is better for equity to make the *Abstraction* of the co-manipulated object migrate to a third node.

Last, thanks to a precise description of the basic network services, all the network communication details are also limited to basic network policy components. These components implement the communications between the PAC-C3D *Controls* using network facilities such as RPC, RMI, TCP communications (Unicast or Multicast) or HTTP communications. So, to change the basic network layer used by the *Controls*, we only have to provide a new set of distribution policies that rely on the new network layer.

5.5 Creation of the shared virtual objects

To ensure an easy evolution of a CVE, we must use the *Abstract Factory* design pattern (GoF87) [15] for object creation. This design pattern makes it possible to let the *Abstractions* create new objects without any knowledge of collaboration by asking an abstract factory to create these objects. The real instance of this abstract factory, called PAC-C3D factory, will deliver *Controls* (which are *Proxies* of their *Abstraction*) instead of *Abstractions*.

In the same way, the *Controls* must use several factories in order to create their associated *Abstractions* and *Presentations*. The PAC-C3D factory will give each *Control* one factory for *Abstraction* creation, and a list (which can be empty) of factories for *Presentations* creation, corresponding to each kind of existing *Presentation* on its node. If allowed by its distribution policy, then the *Control* will ask the *Abstraction* factory to create a real *Abstraction*. Next, the *Control* will ask each *Presentation* factory to create a *Presentation*. This is illustrated figure 11 for the creation of a PAC-C3D object on one node that is hosting two kinds of *Presentations*.

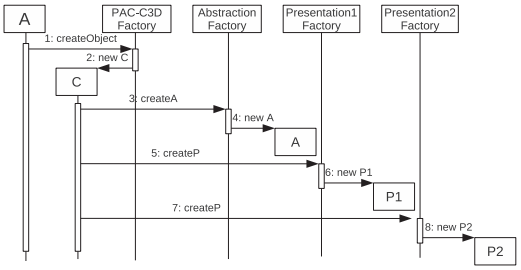


Figure 11: Creation of the PAC-C3D facets

Then, according to its distribution policy, this *Control* can send a message to the other nodes of the CVE to create also local *Controls*: each local PAC-C3D factory will allow the creation of a local *Control* with the appropriate set of local *Presentations*.

6 Adaptation to different representations

As our model offers great independence between the facets of each PAC-C3D object, it makes it easy to provide several kinds of *Presentations* for a same virtual object.

First, for each distribution mode, the *Controls* can be associated to different kinds of *Presentations*, dedicated to a particular visualization of the shared virtual environment. For example, with *Controls* written in Java, on one node the *Presentation* could rely on Java3D [2] while on another node it could rely on JMonkey [4] or jReality [5]. As the implementation details of the 3D graphics API are encapsulated within the *Presentations*, for example it is also possible to use any C++ 3D graphics API without perturbing *Abstractions* and *Controls*.

To avoid code duplication, all the code relative to high-level interaction with virtual objects should be removed from the *Presentations* and written once in some *Abstractions* of PAC 3D interaction tools. But for an optimal efficiency, it is also possible to use built-in interaction and navigation metaphors that come with a 3D viewer.

Several kinds of *Presentations* can also be associated with the same *Control* in order to provide several representations of a shared virtual environment to a user. Some of these representations can also be a 2D visualization of the CVE. This can be extended to any kind of presentation, which could be non-visual, as a sound or a physical representation.

To benefit fully from "active" *Presentations* such as physics engines (for example they can react to an update because of collision detection when trying to move a 3D object), the behavior of the PAC-C3D *Controls* should be slightly adapted, otherwise the "naive" use of such engines could introduce more latency and some small inconsistencies on other *Presentations* in the worst situation about distribution (when the physical *Presentation* is not on the same node than its associated *Abstraction*)(see figure 12)). This adaptation could consist in updating first the "active" *Presentations* and taking their results into account before updating the "passive" *Presentations*.

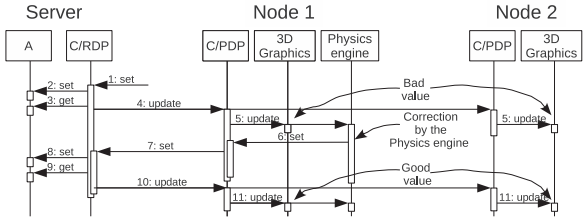


Figure 12: "Naive" use of a physics engine

7 PAC-C3D implementation examples

In this section we will make a short point about the current implementations of PAC-C3D, then we will illustrate this model through 3 examples, in the context of the Collaviz framework [13, 1]. The first one explains in details how PAC-C3D can help the designer of a new 3D collaborative visualizer to reuse existing interaction tools. The second one describes more generally how PAC-C3D has been used to design the IIVC concept. The third one describes how PAC-C3D allowed us to integrate a physics engine within the Collaviz framework.

7.1 Current implementations of PAC-C3D

The first implementation of our model is dedicated to student VR projects and has already been used for two years. This simple implementation has been made with Java3D [2] and JMonkey [4] as 3D rendering engines, and implements only the referent and proxy distribution policies, with migration capabilities. The proxy distribution policy uses Java RMI for communication with its referent, and the referent distribution policy uses Multicast facilities to communicate with its proxies *Controls*.

The second implementation is dedicated to industrial collaborative scientific visualization, it has been implemented in the context of a collaborative project called Collaviz [13, 1]. It relies on both Java3D (for desktop visualization) and jReality (for desktop and immersive visualization) as illustrated figure 1. The *Controls* can use the three distribution policies, and these distribution policies use either TCP or HTTP for communication [19]. All these distribution policies can be changed at run-time. This implementation has also been coupled to the JBullet [3] Physics Engine which appears as another *Presentation* associated to some *Controls* of the system.

7.2 The 2DPointer/3DRay

Here is a full example of the benefits to use our model that shows the complementarity of the PAC-C3D separation between abstraction and presentation and of the PAC-C3D collaboration through the controls. This exemple is the implementation of the 2DPointer/3DRay metaphor [14]: a 3D ray for 3D selection and interaction which orientation is computed so that the user always sees this 3D ray as a 2D pointer on the screen, to be used as easily as a classical 2D pointer, but to be seen as a 3D ray by the other users of the shared virtual environment.

We first implemented this 2DPointer/3DRay metaphor in our Java3D visualizer, this cursor was driven with mouse events provided by Java3D. We took care to clearly separate the behavior of the 2DPointer/3DRay (the computation of its orientation according

to its position, that has been isolated in an abstraction component) from the Java3D presentation code in charge of the Java3D mouse events and of the 3D picking for object selection. As a first result, this 2DPointer/3DRay can be driven by any other input device able to provide a position, for example a wiimote or an ART tracking device (see figure 13)).

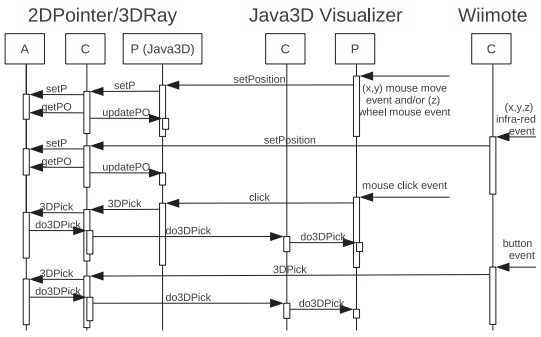


Figure 13: Driving the same abstraction with different input devices

Then, we worked with jReality to provide another 3D viewer, mainly dedicated to immersive visualization devices such as workbenchs or CAVE. As this viewer was not first dedicated to desktops, we did not want to waste time to write a small presentation component able to deal with mouse events, which would be useless for immersive situations as we would use an ART tracking system for interaction. But for testing this new jReality visualization component, some work had to be done in desktop mode, and some interaction could be useful. We decided to use the 2DPointer/3DRay metaphor driven by a wiimote (see figure 14).

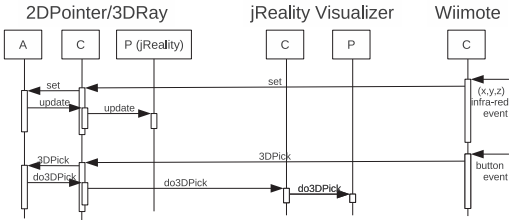


Figure 14: Visualizing an abstraction with another 3D API

Once the 2DPointer/3DRay metaphor was driven by a wiimote within our jReality visualizer, the only remaining problem was to allow this metaphor to select and manipulate 3D objects. The more natural solution was to enable a 3D picking service in jReality.

If such a picking service could not have been realized with jReality, we could also have enabled the 3D picking thanks to our Java3D visualizer. As PAC-C3D has been used to design the Collaviz framework architecture, the Collaviz system allows to share a virtual environment between several visualizers, so it is possible to instantiate a Java3D visualizer and a jReality visualizer to share a common virtual environment. The abstraction of the 2DPointer/3DRay interaction metaphor of the jReality Visualizer can be instantiated on the process of the Java3D visualizer, that owns also a control component and a Java3D presentation component for this metaphor, while the process of the jReality visualizer owns only a proxy control and a jReality presentation component for this metaphor. As illustrated figure 15, this distribution mode would allow the proxy control on the jReality side to send the picking request to the referent control on the Java3D side, which would be able to achieve the picking thanks to the 3D picking service of the Java3D visualizer.

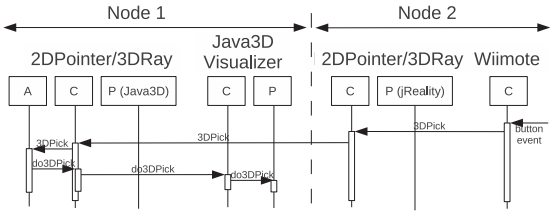


Figure 15: Delegating behavior to abstraction and other presentation

7.3 Other interaction and navigation tools

The whole architecture of our 3D visualizers is designed thanks to the PAC-C3D model, this enables us to provide a common architecture for navigating and interacting within 3D virtual environments that we call the Immersive Interactive Virtual Cabin (IIVC) [18]. All the operators that have been proposed for this IIVC (such as dedicated navigation modes or interaction tools) are implemented within abstraction components linked to dedicated presentation in charge of the visualization of their actions upon the virtual environment through control components.

It makes it possible to use the same input devices (for example a 2D GUI or a joystick) for navigation whatever the 3D graphics API is used for a 3D visualizer: the navigation orders are sent to the abstraction of what we call a "conveyor", which supports several virtual objects including a virtual viewpoint, which position and orientation are changed whenever the conveyor moves in the world. These changes occur in the abstraction of the virtual viewpoint, then its control component is in charge of propagating these changes to its associated local presentation component and to its distributed controls if any.

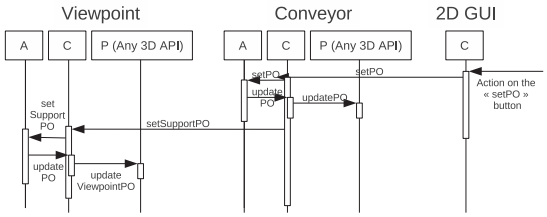


Figure 16: Delegating navigation to the abstraction of a viewpoint

Figure 16 illustrates these local exchanges with the camera of a 3D visualizer: the position and orientation changes of the conveyor are sent to the abstraction of the virtual viewpoint, which can compute its own new position and orientation before sending them to its presentation through its control component. The presentation of the virtual viewpoint can be linked to the camera of a 3D visualizer, which impacts the rendering of the 3D visualizer.

So, any particular way of navigation (for example a "travelling" along some interesting object, or an "examine" navigation mode allowing to turn around a virtual object, or a "walk" or a "fly" navigation mode allowing different kinds of exploration of a virtual environment) has only to be coded once, in the abstraction of the conveyor, to be available for any 3D visualizer, whatever the 3D graphics API used.

In the same way, a conveyor can support any 3D interaction tool (such as our 2D pointer / 3D ray, or classical 3D virtual rays, virtual hands or virtual 3D cursors), and here again these interaction tools offer the same behavior whatever the 3D graphics API used.

7.4 Coupling a physics engine to a virtual environment

To make collision detection possible within our 3D visualizers, we chose to integrate the JBullet Physics Engine [3] into the Collaviz framework. The most straightforward way to achieve this is to place the JBullet engine component on the central collaboration server process, in order to be able to provide the physics services (for example collision detection or mechanical constraints) in the same way to any Visualizer client. Otherwise, this JBullet component could be placed on any node of the shared virtual environment. So, for any virtual object for which we want to offer physics, we declare it as a physical object, with an additional JBullet presentation component, linked to the JBullet engine. Each move of the virtual object in the virtual environment will make its physical JBullet presentation move in the JBullet world (see figure 17), which will allow to take into account the results (collision or constraints) provided by the JBullet engine, as already presented figure 12.

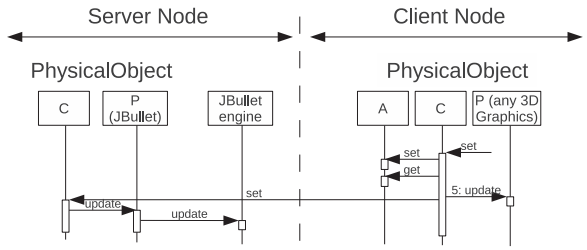


Figure 17: Maintaining consistency between Virtual and Physical worlds

Once again, this allows to offer physics for the objects of the virtual universe whatever the 3D graphics API used for the 3D Visualizer: We have implemented generic physical 3D cursors that behave exactly in the same way in our two main visualizers, based on Java3D and jReality.

8 Conclusion and Future Work

The PAC-C3D architectural model is an explicit evolution of the PAC model dedicated to 3D collaborative virtual environments. Each shared virtual object of a CVE must be decomposed into three main kinds of components described by three interfaces. The *Abstraction* is in charge of the core data and behavior of the object, the *Presentations* are in charge of the presentation of the object to the user, and the *Control* is in charge of the consistency maintenance between *Abstraction* and *Presentations*, and between all the distributed *Controls* of the shared object.

PAC-C3D can deal with the main distribution modes encountered in CVE and it has been validated with several distribution policies, on local area networks and on wide area networks over the internet.

PAC-C3D also makes it possible to design a CVE with very small dependency on a 3D graphics API, and it makes it easy to use different 3D graphics API for different nodes involved in the same collaborative session, providing easy interoperability between 3D graphics API. It is also possible to rapidly couple other 3D engines (for example physics engines) by adding another *Presentation* (related to the engine) to PAC-C3D objects.

The next steps are to take explicitly into account "active" *Presentations* and to couple PAC-C3D objects with other kinds of engines, for example Artificial Intelligence behavior libraries that could be used in the same way as a physics engine to drive virtual objects.

Acknowledgements

This work was partly funded by the French Research National Agency project named Collaviz (ANR-08-COSI-003-01).

References

[1] The Collaviz website. http://www.collaviz.org/.
[2] The Java3D website. http://java3d.java.net/.
[3] The JBullet website. http://jbullet.advel.cz/.
[4] The JMonkey website. http://jmonkeyengine.org/.
[5] The jReality website. http://www3.math.tu-berlin.de/jreality/.
[6] G. Calvary, J. Coutaz, and L. Nigay. From Single-User Architectural Design to PAC*: a Generic Software Architecture Model for CSCW. In *Proceedings of CHI 97, ACM publ*, pages 242–249, 1997.
[7] J. Coutaz. PAC: An Object Oriented Model for Implementing User Interfaces. *SIGCHI Bull.*, 19(2):37–41, 1987.
[8] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti. Surround-screen projection-based virtual reality: the design and implementation of the cave. In *Proceedings of SIGGRAPH'93*, pages 135–142, New York, NY, USA, 1993. ACM.
[9] M. Davis. Struts, an open-source MVC implementation. http://www.ibm.com/developerworks/library/j-struts/, february 2001.
[10] D. Delaney, T. Ward, and S. McLoone. On consistency and network latency in distributed interactive applications: A survey – part I. *Presence: Teleoperators and Virtual Environments*, 15(2):218–234, 2006.
[11] D. Delaney, T. Ward, and S. McLoone. On Consistency and Network Latency in Distributed Interactive Applications: A Survey – Part II. *Presence: Teleoperators and Virtual Env.*, 15(4):465–482, 2006.
[12] P. Dewan. Architectures for Collaborative Applications. *Trends in Software, special issue on Collaborative Systems*, pages 169–193, 1999.
[13] F. Dupont, T. Duval, C. Fleury, J. Forest, V. Gouranton, P. Lando, T. Laurent, G. Lavoué, and A. Schmutz. Collaborative Scientific Visualization: The COLLAVIZ Framework. In *JVRC Demos*, 2010.
[14] T. Duval and C. Fleury. "An asymmetric 2D Pointer/3D Ray for 3D interaction within collaborative virtual environments". In *Proceedings of the Web3D'09 conference*, pages 33–41, 2009.
[15] E. Gamma, R. Helm, R. Johnson, J. Vlissides. *Design patterns: Elements of reusable Object-Oriented Software*. Addison-Wesley, 1995.
[16] R. Eckstein. Java SE Application Design With MVC. http://www.oracle.com/technetwork/articles/javase/mvc-136693.html, march 2007.
[17] C. Ellis and J. Wainer. A conceptual model of groupware. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, CSCW '94, pages 79–88, New York, NY, USA, 1994. ACM.
[18] C. Fleury, A. Chauffaut, T. Duval, V. Gouranton, and B. Arnaldi. A Generic Model for Embedding Users' Physical Workspaces into Multi-Scale Collaborative Virtual Environments. In *Proc. of ICAT*, pages 1–8, 2011.
[19] C. Fleury, T. Duval, V. Gouranton, and B. Arnaldi. A New Adaptive Data Distribution Model for Consistency Maintenance in Collaborative Virtual Environments. In *Proc. of JVRC*, pages 29–36, 2010.
[20] C. Fleury, T. Duval, V. Gouranton, and B. Arnaldi. Architectures and Mechanisms to efficiently Maintain Consistency in Collaborative Virtual Environments. In *Proc. of the 3rd IEEE VR 2010 Workshop on Software Engineering and Architectures for Realtime Interactive Systems (SEARIS 2010)*, pages 87–94, 2010.
[21] A. Goldberg. Information models, views, and controllers. *Dr. Dobb's J.*, 15:54–61, May 1990.
[22] Y. Laurillau and L. Nigay. Clover architecture for groupware. In *Proceedings of the Conference on Computer-Supported Cooperative Work*, pages 236–245. ACM, 2002.
[23] M. R. Macedonia and M. J. Zyda. A taxonomy for networked virtual environments. *IEEE Multimedia*, 4(1):48–56, Jan-Mar 1997.
[24] L. Nigay and J. Coutaz. Building User Interfaces: Organizing Software Agents. In *Proceedings of Esprit'91*, pages 709–717, 1991.
[25] M. Potel. MVP: Model-View-Presenter — The Taligent Programming Model for C++ and Java. http://www.wildcrest.com/Potel/Portfolio/mvp.pdf, 1996.
[26] T. Reenskaug. The original MVC reports. http://heim.ifi.uio.no/˜trygver/2007/MVC_Originals.pdf, 1979.
[27] . UIMS 1992. A metamodel for the runtime architecture of an interactive system: the uims tool developers workshop. *SIGCHI Bull.*, 24(1):32–37, 1992.

TouchMe: An Augmented Reality Based Remote Robot Manipulation

Sunao Hashimoto[1,*] Akihiko Ishida[1,2,†] Masahiko Inami[1,3,‡] Takeo Igarashi[1,4,§]

[1]JST ERATO Igarashi Design Interface Project
[2]Tokyo University of Science
[3]Keio University
[4]The University of Tokyo

ABSTRACT

A general remote controlled robot is manipulated by a joystick and a gamepad. However, these methods are difficult for inexperienced users because the mapping between the user input and resulting robot motion is not always intuitive (e.g. tilt a joystick to the right to rotate the robot to the left). To solve this problem, we propose a touch-based interface for remotely controlling a robot from a third-person view, which is called "TouchMe". This system allows the user to manipulate each part of the robot by directly touching it on a view of the world as seen by a camera looking at the robot from a third-person view. Our system provides intuitive operation, and the user can use our system with minimal user training. In this paper we describe the TouchMe interaction and its prototype implementation. We also introduce three scheduling methods for controlling the robot in response to user interaction and report on the results of empirical comparisons of these methods.

KEYWORDS: Remote robot control, third-person view, augmented reality, touch screen, direct manipulation.

INDEX TERMS: H.5.2 [Information Interfaces and Presentation]: User Interfaces — Interaction styles; I.2.9 [Artificial Intelligence]: Robotics — Operator interfaces

1 INTRODUCTION

There are many environments where it is hard for humans to work, such as in water, high places, high/low temperature environments and environments contaminated with poison or radioactivity. Various robots have been developed to perform tasks in these dangerous environments. Fully autonomous robot operation is desired, but it is difficult because of recognition problems. One way to alleviate recognition problem is to put tags onto objects, and use a pre-structured environment model. However, these methods are not applicable to unstructured environments and human supervisor controls are necessary.

A robot that can grab and deliver physical objects generally has a multi-DOF (degree of freedom), but it is not easy to control them for an inexperienced operator. For example, a robotic arm has generally 4 or 6 DOF. When the robotic arm is mounted on a mobile vehicle, the total DOF is increased. The most popular method for the control of multi-DOF robots is a joystick and a gamepad; however, in these control devices, the number of controllable DOF is limited by the number of buttons and axes of the devices. The controllable DOF can be increased by using some combinations of 2 or 3 keys, but it will make operability more difficult and demand a longer training time from the user. To facilitate controlling a multiple-link robot, inverse kinematics is widely employed. Generally a robotic arm using inverse kinematics maps the control of the end-effector to a joystick, and the angles of each joint are calculated appropriately. However, in a general joystick based controller, the moving velocity of the robot is given in proportion to the timing or degree of key pressing. This also needs user training.

We propose a tele-operating system that allows the user to manipulate a multi-DOF robot intuitively with touch interaction from a third-person view, which we call "TouchMe". The system is shown in Figure 1. This is a touch screen based interface, and it displays an image acquired from a camera observing the robot from a third-person view. The user can directly specify the desired pose and position of the robot by touching and dragging the part that he/she wants to control. The camera image showing the robot is overlaid with a computer graphics (CG) model synchronized with the user's manipulation to help the user predict how the robot will move. This is an augmented reality application applying the direct manipulation of a posing tool of a virtual human such as Poser to the multi-DOF robot in the real-world. Typical remotely controlled robots use a robot-mounted camera providing a first-person view, but we use a third-person view camera because it allows the user to understand the situation of the entire working space (the controlled robot, target objects, and obstacles). We discuss advantages and disadvantages of various third-person view camera settings for the proposed method such as a fixed surveillance camera, a flying camera, and other robot's eyes.

In this paper, we describe the TouchMe interaction and our prototype implementation. We also introduce three scheduling methods for the robot control in response to user interaction. One is to move the robot after the touch, one is to move the robot during the touch, and one to move the robot during and after the touch. We compared these three methods in an empirical evaluation and report on the results.

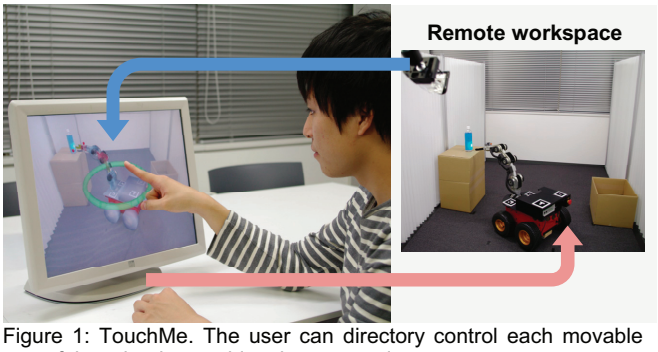


Figure 1: TouchMe. The user can directory control each movable part of the robot by touching the camera image.

*e-mail: hashimoto@designinterface.jp
†e-mail: ishida@designinterface.jp
‡e-mail: inami@inami.info
§e-mail: takeo@acm.org

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

2 RELATED WORKS

There are various video-based interfaces for controlling a robot and appliances.

Tani et al. presented interactive video techniques that allow interaction with objects in live video on the screen, by having models of the objects monitored by cameras [1]. They explored two strategies for modeling objects imaged by cameras in 2D and 3D. They implemented a system called HyperPlant for monitoring and controlling an electric power plant, by using 2D modeling. Seifried et al. developed a video-based interface for controlling home appliances in the project CRISTAL [2]. In their work, the camera is mounted for a top-down view of the living room, and the image is displayed on a multi-touch tabletop surface. This system allows multiple users to operate multiple appliances collaboratively. We also control a device through a camera image, but our controlled object is a multi-DOF robot, and we aim to achieve more complicated tasks using it.

Top-down view has been used in several video-based robot control interfaces. Sakamoto et al. proposed a video-based Tablet PC interface to control vacuum cleaning robots [3]. In this system, ceiling mounted cameras provide the user a top-down view that allows the user to control robots and design their behaviors by sketching using a stylus pen. Kato et al. developed a multi-touch tabletop interface for controlling multiple robots [4]. They proposed a method to control multiple mobile robots simultaneously by manipulating a vector field on a top-down view from a ceiling camera. Guo et al. presented two interfaces for remotely interacting with multiple robots using toys on a large tabletop display showing a top-down view of the workspace [5]. This research shows the fact that a top-down view is easy for controlling the locomotion of mobile robots on a 2D surface, however it is hard to control multi-DOF robots.

There are several interfaces for controlling a robot through a first-person view (robot's-eye view). Sekimoto et al. proposed a simple driving interface for a mobile robot using a touch panel and first-person view images from the robot [6]. Once the operator gives a point of the temporary goal position by touching on the monitor displaying the front view of the robot, the system generates a path to the goal position and the vehicle is controlled to follow the path to reach the goal position autonomously. Fong et al. also developed a similar system on a handheld device (PDA) [7]. Correa et al. proposed a handheld tablet interface for operating an autonomous forklift, where users provide high-level directives to the forklift through a combination of spoken utterances and sketched gestures on the robot's-eye view displayed on the interface [8].

Third-person view is also used in video-based remote robot control systems. Hosoi et al. proposed a robot control technique using a camera-mounted mobile device such as a PDA and a mobile phone, which is called Shepherd [9]. In this system, the operator holds a camera-mounted mobile device in his/her hand, and he/she instructs the robot how to move by moving the device. Sugimoto et al. proposed a visual presentation system for controlling a robotic vehicle remotely, which is called Time Follower's Vision [10]. This allows the operator to control a remote rescue robot by observing a virtual third-person view which is created from a first-person view camera mounted on the robot. They show the effectiveness of a third-person view to allow even inexperienced operators to easily control the robot.

There are several robot interfaces using augmented reality and mixed reality techniques. Nawab et al. proposed a method that overlays a color-coded coordinate system on the end-effector of the robot using augmented reality to help the user to understand the key mapping of a joystick [11]. Kobayashi et al. developed a mixed reality environment which can overlay internal statuses of a humanoid robot such as recognition results and planning results [12]. Their method enables the operator to understand the robot internal statuses intuitively, which is helpful for debugging and actual operation. Chen et al. also developed a mixed reality environment for performing robot simulations involving physical and virtual objects [13]. Drascic et al developed an augmented reality through graphic overlaying on a stereo video [14]. In their application, the user wearing a data glove controls a robotic arm by manipulating a virtual cursor overlaid on the video image. Xiong et al also developed a tele-robotic system based on augmented reality to control a six DOF robotic arm [15]. In this system, a virtual robot works as an interface between the operator and the real robot, mitigating the problem of time-delay between user operation and real robot action. This idea is also used in our research, but we use a touch screen for the interface and we empirically compare three touch interaction methods, while they used a head-mounted display, a data glove, and voice commands for their interface.

3 USER INTERACTION

TouchMe is an augmented reality interaction technique for remote robot control. The system overview is shown in Figure 2. The camera captures the image of the workspace in real-time, and the image is shown on the touch screen with a CG model of the real robot. The CG model is overlaid on the robot, and it is shown semi-transparently (like a ghost). The user controls the robot by touching the overlaid CG model. The user touches the part of the CG model where he/she wants to move, and he/she then drags it to the desired position and direction. For example, the user slides the body to move the robot to a specific position, and then drags the top of the arm to reach an object. This is similar to manipulations performed in 3D modeling and posing software, in which a 3D object is manipulated using user operations on a 2D image plane. As the user moves the CG model, it eventually moves away from the physical robot on the screen, and the system drives the robot so that it matches with the CG model.

3.1 Third-person view

We use a third-person view camera because it allows the user to understand the situation of the entire work space composed of the controlled robot, target objects and obstacles. A typical approach is to use a first-person view image obtained from the robot-mounted camera, but we did not use this because it is difficult to avoid collisions with obstacles on the side or behind the robot when the robot is rotating or moving backwards. We will now discuss various third-person view camera. We only implemented and tested the first method. The implementation of the remaining two methods is our future work.

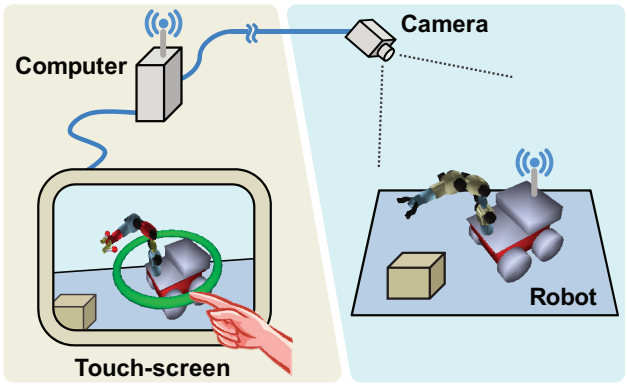


Figure 2: Overview of the system.

Fixed surveillance camera: Fixed surveillance cameras are already installed in various places such as roads, parks, stations, museums, factories, stores and homes, for security and recording. A surveillance camera is mounted in a high position that is higher than human height and provides a bird's-eye view. The advantage of a surveillance camera is that it gives the user a good stable view for understanding the entire surrounding environment. However, the movement of the fixed camera is limited to panning, tilting and zooming, making it difficult to resolve possible occlusion.

Flying camera: Various unmanned aerial vehicles (UAV) such as a remote-controlled helicopter and an airship with a camera are used for scouting. A UAV's camera also provides a bird's-eye view, and it can move freely unlike a fixed camera. A flying camera can use viewpoint operation (such as the CG modeling software) in the real-world. Moreover, the camera can track the target robot automatically to keep the robot in the field of view. This gives the operator a view like a 3D action game where a third-person view camera follows after the game character such as in Nintendo's Super Mario 64. The disadvantage of this camera is that it demands very stable and highly precise control for having such free viewpoint movement.

Another robot's camera: When two or more robots are employed in a workspace and one of them has an eye (a first-person view camera), we can operate the other robots in a third-person view by borrowing the view of the robot. If all robots have cameras, the user can operate the robots while switching first-person views and third-person views freely. For example, a first-person view is used when the user operates the hands of the target robot, and a third-person view is used when the user wants to move the target robot to another position avoiding obstacles.

The viewpoint operation can be performed by touching on the region outside of the robot or touching a special icon for manipulating the view point. We guess that automatic camera control would be useful. For example, when the user manipulates the end-effector of a robotic arm, the camera moves to the position where it can give good the operator a good view, and is zoomed in automatically.

3.2 Virtual handles

Virtual handles are user interfaces to make the CG model easy to manipulate. It is useful for understanding the controllable direction of the mounted part. Figure 3 shows two types of virtual handles. The ring type is used for manipulating a rotating part (e.g. rotation of the body, rotation of a link of the arm). The lever type is used for manipulating a small part such as an end-effector. These ideas are used widely in CG modeling software [16]. We apply them to the real robot by using an augmented reality technique.

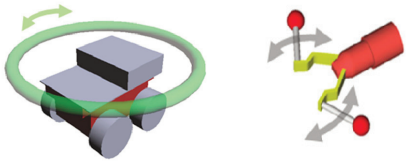


Figure 3: Virtual handles. Ring type for rotation of the vehicle, and lever type for manipulating end-effector of a robotic arm.

3.3 Inverse kinematics

We use inverse kinematics (IK) to facilitate controlling a multi-joint robot. When one of the links is manipulated, related links are moved automatically. For example, when the user pulls the wrist of the arm, the elbow and the shoulder (or the body) are controlled by the system automatically. We manipulate 3D multi-joint structures in 3D space on a 2D display surface, and we use an IK method that is used in general posing tools for virtual human models such as Poser.

3.4 Scheduling of robot motion

The robot only moves with a limited speed, so the CG model and the robot on the screen do not always match during the user interaction. The system resolves this mismatch by moving the robot towards the CG model, but there are multiple ways to achieve this. Here we introduce three possible scheduling methods.

Move-after-touch: The robot does not move while the finger is touching the screen and is manipulating the CG model. When the user releases their finger, the CG model is fixed and the robot begins to move toward the CG model. The robot stops when the pose matches with the CG model (Figure 4).

Move-during-touch: The robot begins to move toward the CG model immediately after the finger begins to manipulate the CG model by touching the screen. While the finger touches the screen, the pose and position of the CG model is continuously updated, and the robot continuously tracks the CG model. When the finger is released, the robot stops immediately and the CG model pose is set to the robot pose at the time of release (Figure 5).

Move-during-and-after-touch: This is a combination of the above two methods. The robot begins to move toward the CG model immediately after the finger begins to manipulate the CG model by touching the screen and continues moving during user manipulation. When the user releases their finger, the CG model is fixed to the pose at the time of release and the robot continues moving toward the fixed CG model. The robot stops when the pose and position of the robot matches with those of the CG model (Figure 6).

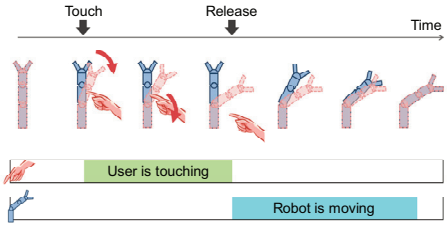


Figure 4: Move-after-touch.

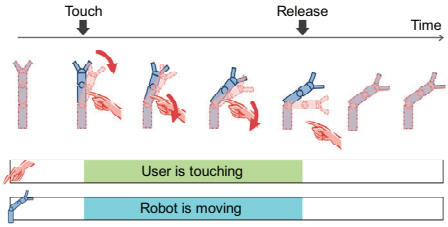


Figure 5: Move-during-touch.

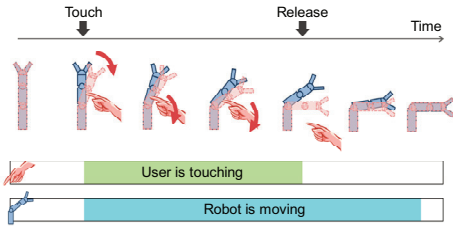


Figure 6: Move-during-and-after-touch.

4 PROTOTYPE SYSTEM

We developed a prototype system in which the user controls a robot vehicle using our proposed interface.

4.1 Robot

We used a robotic vehicle (MobileRobots PIONEER3-DX) equipped with a robotic arm (Neuronics Katana). Figure 7 shows our robot and its DOF. The vehicle has a mechanism for locomotion using four wheels. It allows the user to rotate and to move forward or backward (2DOF). The mounted robotic arm has 6DOF but we limited controllable parts to the hand (1DOF) and the three joints (3DOF) to simplify the operation. Therefore the whole robot has 6 DOF in total.

We made the CG model of this robot, and gave two kinds of virtual handles to facilitate control of the robot; ring type for the rotation of the vehicle and lever type for manipulating the hand of the robotic arm. When the user manipulates the arm, all or part of the three joint angles (joint 3, 4 and 5, shown in Figure 7) are updated according to the result of IK computation.

The vehicle and the mounted arm are controlled remotely by a host computer. The host computer controls the joint angles of the arm individually, and obtains each angle's value. In our prototype, the host computer communicates with the robot via USB wired connection. To help the user grab an object with the arm, we mounted a green flashlight on the wrist of the robotic arm to light the target when it is in front of the hand. This is important because depth information is missing in the single camera view.

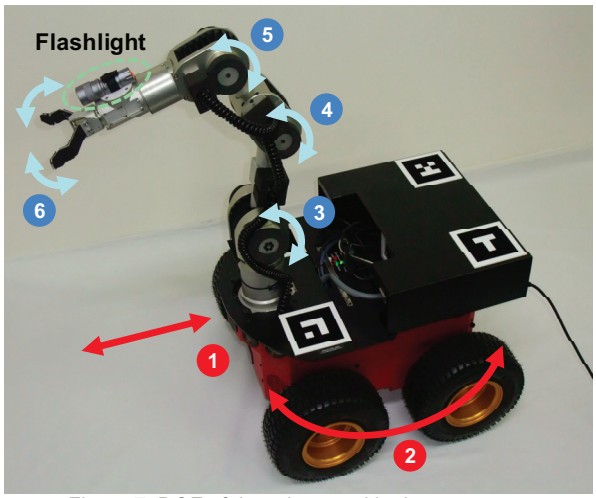


Figure 7: DOF of the robot used in the prototype.

4.2 Camera view and registration

We use a commercial webcam (Logicool QCAM-200V) as the third-person view camera fixed in the workspace. The camera image is displayed on a 19 inch LCD desktop touch screen. The resolution of the image is 800 × 600 pixels, and the frame rate is 15fps. The camera does not support any physical movements such as panning and zooming.

We use fiducial markers (ARToolKit [17]) for registration between the real robot and virtual robot (CG). We put four markers (10 × 10 cm^2) on the top of the vehicle. At the initial state and when the robot stops, the system gives the CG model the actual joint angles obtained from the robotic arm, and physical position and direction obtained from the fiducial markers. The markers are also used for visual feedback when the robot moves to the specified goal.

5 EMPIRICAL COMPARISON OF THE SCHEDULING METHODS

We ran a user study using our prototype system to test general usability of the system and to compare the three scheduling methods. Figure 8 shows the experimental workspace displayed on the touch screen with robot and overlaid CG model. A 190 × 250 cm^2 workspace is divided by partition walls. In this space, a blue labeled plastic bottle (with diameter of 7 cm, 25 cm high) is placed on a rack with a height of 58 cm, and a trash-box (42 × 33 × 31 cm^3) is placed on the opposite side. The camera is fixed at 123 cm high from the floor.

We recruited 12 people aged 20-25 years old, 8 males and 4 females, to participate in our study. All of them are students from a university, and they use a computer in their daily lives. Most of them had no experience with robot control, and they were not familiar with our robot. The sessions lasted about an hour.

We gave each participant the task of controlling the robot to pick up a blue bottle and drop it in the trash-box using our touch screen interface. The hand angle is limited to a pre-defined angle to prevent breakdown of the hand when it grabs the bottle.

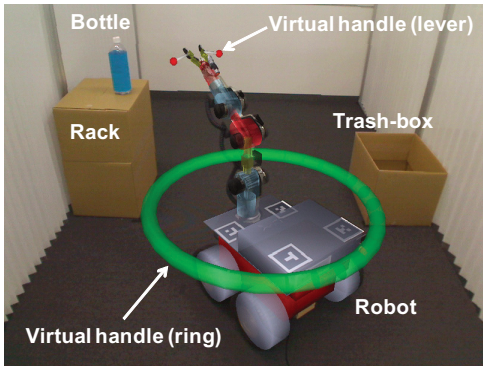


Figure 8: The superimposed image displayed on the touch screen.

5.1 Conditions

We conducted our user test for three conditions, move-after-touch (A), move-during-touch (D), and move-during-and-after-touch (DA).

We did not allow the participants to enter or see the workspace directly, therefore the workspace was a completely unknown environment for them. They tried to operate the robot by only observing the camera image. When the user test began, we explained to the participant how to control the robot, the DOF of the robot, and the fact that the flashlight was mounted on the wrist of the arm for aiming. The comparison was performed as within-subjects, where each participant tested all conditions. Each participant performed a task on three conditions in balanced order. For each condition, we gave the participant a training time of up to five minutes before the trial. All objects and the robot were placed in their initial positions for each trial. If the robot dropped the bottle on the floor due to an operator's mistake, we recorded the trial as a failure. If it was caused by a system error, we gave the participants a chance to retry. For each trial, we recorded the task completion time and asked the participant to answer a questionnaire. After three trials we interviewed them.

5.2 Results

All 12 participants except one person succeeded in the task. The one who failed dropped the bottle in all trials. Table 1 lists the time to complete the task for three conditions (only successful cases). The tasks were finished in approximately two minutes. The results indicate no significant differences between conditions (by ANOVA, p=0.77).

The results from the questionnaires (seven-point Likert scale with high scores positive) are shown in Figure 9, and we show the detailed questionnaires in Table 2. The only negative question is Q4. After ANOVA, Ryan's method was performed for the results. Statistically significant results ($p<0.05$) are shown in Q2, Q5 and Q6. Condition A got the highest positive value of all the questions, and there are no significant differences between D and DA through all the results. The result of Q2 shows that the participants controlled the robot with stronger confidence in A than D ($p=0.01$). The result of Q5 shows that most participants expected that many people can control the robot easily by using A, stronger than D ($p=0.01$). The result of Q6 shows that the participants could control the robot as they expected by using A, stronger than D ($p=0.001$).

Table 1: Average time to complete tasks.

Condition	Time (m:ss)	St.Dev. (m:ss)
A	1:52	0:30
D	2:00	0:22
DA	1:53	0:22

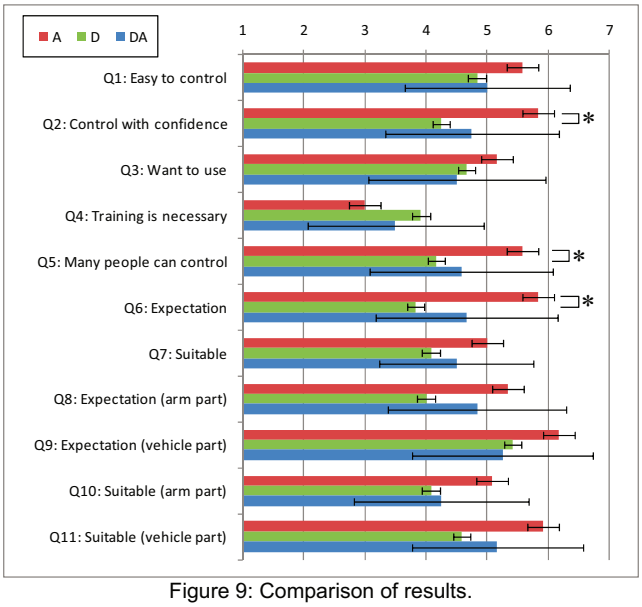


Figure 9: Comparison of results.

Table 2: Questionnaires for the conditions.

Q1	This method is easy to control the robot.
Q2	I controlled the robot with confidence.
Q3	I want to control the robot using this method.
Q4	A lot of training is necessary for using this method.
Q5	Many people can control the robot easily using this method.
Q6	I could control the robot as I expected.
Q7	This method is suitable for controlling the robot.
Q8	I could control the arm part as I expected.
Q9	I could control the part of vehicle as I expected.
Q10	This method is suitable for controlling the arm part.
Q11	This method is suitable for controlling the vehicle part.

5.3 Discussion

The comparison results show that A tends to be supported most among the three conditions, in particular, in Q2, Q5 and Q6. We found no significant differences about the easiness (Q1). However, in the interview, 9 of 12 people answered that A is the easiest, and the main reason was that they could operate calmly without being rushed. On the other hand, 9 of 12 people answered that A might take the longest time to complete a task among the three conditions in the interview, despite the fact that the task completion time showed no significant difference between conditions in the trials. They mentioned "D and DA are faster than A because these track manipulated CG model immediately". The condition A seems that it gives the impression that it is not efficient mentally because the robot moves after the finger is released from the touch screen.

In Q4 (Training is necessary), there are no significant differences; however, all conditions were less than four. In addition, 11 of 12 participants completed the tasks with training taking less than five minutes. These results show that the proposed system needs less time for training than for using.

We found no significant differences in Q3 (want to use) and Q7 (suitable). In the interview, we asked them about these criteria in detail. Several people answered that they wanted to use A in an actual situation and that A is the most suitable method. The main reason was that A was the easiest to use for them. People who preferred D or DA found great value in the physical feedback that the real robot moves immediately according to the user's operation. Particularly, people who preferred D liked a property that allowed the user to stop the robot when he/she wants to stop it by releasing his/her finger from the screen. We also got several negative comments for D as follows: "D makes me tired (or it is troublesome) because I should keep touching until the robot arrives to the destination", "Sometimes I have released my finger by mistake".

5.4 General observations

All participants understood how to use the virtual handles, and they used them efficiently when they rotated the robot and when they manipulated the hand of the arm. Participants commented about manipulation of the CG model as follows: "I want to give an inertia to CG model because it might make the operation more light", "It was hard to recognize that the hand is grabbing the target object or not, therefore some visualization is necessary, e.g. hand's color is changed when target object is grabbed".

For the IK based arm control, several people said that it was not clear how the arm's pose transforms when they manipulated the end of the linked parts. This fact shows that it is necessary to visualize the moveable range of the links of the arm. They also wanted to use both IK and FK (forward kinematics) for controlling the robotic arm.

The third-person view camera was well received by all participants through the experiments. In particular it was fully appreciated by the people who have experience playing 3D action games using a third-person view camera. The biggest negative comment about the third-person view is that it is hard to understand the depth of the space from the image obtained from the single fixed camera. The moveable part would be uncontrollable when the view direction was orthogonal to the rotation axis of the moveable part, or when the part was occluded by another part. We believe that these problems can be solved by using a moveable camera or multiple cameras. The participants requested 1 or 2 additional viewpoints such as a top-down view, a side view or a view on the wrist for our future development. Zooming was also requested to observe the working area more

closely. One of the participants requested a moving camera that follows the robot from behind such as those seen in a third person shooter game.

Several people requested a stylus pen. The main reasons are that it might allow the user to have more precise manipulation and that the display area hidden by a pen is smaller than that of a finger. A multi-touch screen was also requested for the pointing device, with the requested gestures being a pinching gesture for zooming in and out of the view, a pinching gesture for open-close manipulation of the robotic hand, and a two fingers gesture where one finger rotates a link of the arm while another finger holds the anchor point of the joint. Two-finger interaction might be a good method for switching between IK and FK for controlling a robotic arm.

6 Limitations

We will now discuss the current limitations of this work. The proposed method needs a third-person view camera, and the moveable area of the controlled robot is limited to the field of view of this camera. The camera needs to keep a certain amount of distance from the controlled robot to give the image of the robot including the controlled part. The possible position where the camera is put is limited physically in real-world environments, by obstacles and limited small spaces. As a result, it might cause bad views in which it is hard to operate the robot. The resolution of operation depends on the display resolution, and the amount of operation given by a pixel is different if the controlled part is near or far from the camera. To control a CG model by touching, the controlled part needs a certain amount of surface area, and the display also needs a certain amount of physical area to accommodate the CG model.

7 Conclusion and Future Work

In this paper, we presented the design, implementation and an initial evaluation of an augmented reality interface for controlling a multi-DOF robot. TouchMe allows the user to manipulate each part of the robot by directly touching it on a view of the world as seen by a third-person view camera. We compared three scheduling methods on our first prototype system. Most participants found that the easiest method was when the robot began to move after the participant's finger was released from the touch screen.

The results of the user study provided further design recommendations for future iterations of TouchMe and for similar robot control systems. The virtual handles were well received by all participants. The users requested richer visualization for understanding the state of the robot. We found that both IK and FK are desirable for controlling a robotic arm. The third-person view was well received by all participants in the user study, though they also claimed that the third-person view caused the problems such as occlusions and rotation axis aligning with the camera view. Several people requested a stylus pen for more precise manipulation, and also requested a multi-touch screen for advanced manipulation.

Our immediate work in the future is to solve the viewpoint problem that causes occlusions and uncontrollable situations in the third-person view. We expect that this problem can be solved by employing a moveable camera or multiple cameras. Introducing a multi-touch screen is also our future work. We expect that it will allow multiple users to manipulate multiple robots on a screen collaboratively. Moreover, in the future, a model-based tracking would be introduced to relate the virtual robot to the real robot, instead of fiducial markers used in the current prototype.

Our proposed method has a flexible scalability. We can apply this method for various kinds of robots such as humanoids, tabletop robots, bulldozers, power shovels and cars. We expect that extremely small robots can be controlled by touching on a microscope image, and a very large robot could also be controlled by viewing from a distance. We plan to extend our implementation and explore the applicability for various platforms.

References

[1] M. Tani, K. Yamaashi, K. Tanikoshi, M. Futakawa and S. Tanifuji, Object-oriented video: interaction with real-world objects through live video, In *Proceedings of the CHI'92*, pp.593-598, 1992.

[2] T. Seifried, M. Haller, S. D. Scott, F. Perteneder, C. Rendl, D. Sakamoto and M. Inami, CRISTAL: A Collaborative Home Media and Device Controller Based on a Multi-touch Display, In *Proceedings of the Tabletop'09*, pp.33-40, 2009.

[3] D. Sakamoto, K. Honda, M. Inami and T. Igarashi, Sketch and Run: A Stroke-based Interface for Home Robots, In *Proceedings of the CHI'09*, pp.197-200, 2009.

[4] J. Kato, D. Sakamoto, M. Inami and T. Igarashi, Multi-touch Interface for Controlling Multiple Mobile Robots, In *Proceedings of the CHI'09*, pp.3443-3448, 2009.

[5] C. Guo, J. E. Young and E. Sharlin, Touch and toys: new techniques for interaction with a remote group of robots, In *Proceedings of the CHI'09*, pp.491-500, 2009.

[6] T. Sekimoto, T. Tsubouchi and S. Yuta, A Simple Driving Device for a Vehicle Implementation and Evaluation, In *proceedings of the IROS'97*, pp.147-154, 1997.

[7] T. Fong, C. Thorpe and B. Glass, PdaDriver: A Handheld system for Remote Driving, In *Proceedings of the ICAR'03*, 2003.

[8] A. Correa, M. R. Walter, L. Fletcher, J. Glass, S. Teller and R. Davis, Multimodal Interaction with an Autonomous Forklift, In *Proceedings of the HRI2010*, 2010.

[9] K. Hosoi and M. Sugimoto, Shepherd: A Mobile Interface for Robot Control from a User's Viewpoint, In *Proceedings of the ROBIO'06*, pp.908-913, 2006.

[10] M. Sugimoto, G. Kagotani, H. Nii, N. Shiroma, M. Inami and F. Matsuno, Time follower's vision, In *Proceedings of the SIGGRAPH'04*, p.29, 2004.

[11] A. Nawab, K. Chintamani, D. Ellis, G. Auner and A. Pandya, Joystick mapped Augmented Reality Cues for End-Effector controlled Tele-operated Robots, In *Proceedings of the IEEE Virtual Reality'07*, pp.263-266, 2007.

[12] K. Kobayashi, K. Nishiwaki, S. Uchiyama, H. Yamamoto, S. Kagami and T. Kanade, Overlay what Humanoid Robot Perceives and Thinks to the Real-world by Mixed Reality System, In *Proceedings of the ISMAR'07*, pp.1-2, 2007.

[13] I. Y. H. Chen, B. MacDonald and B. Wünsche, Mixed reality simulation for mobile robots, In *Proceedings of the ICRA'09*, pp.922-927, 2009.

[14] D. Drascic, J. J. Grodski, P. Milgram, K. Ruffo, P. Wong and S. Zhai, ARGOS: A Display System for Augmenting Reality, In Proceedings of the INTERACT'93, p.521, 1993.

[15] Y. Xiong, S. Li and M. Xie, Predictive display and interaction of telerobots based on augmented reality, *Robotica(2006)*, 24, Cambridge University Press, pp.447-453, 2006.

[16] B. D. Conner, S. S. Snibbe, K. P. Herndon, D. C. Robbins, R. C. Zeleznik and A. Van Dam, Three-dimensional widgets, In *Proceedings of the SI3D'92*, pp.183-188, 1992.

[17] H. Kato and M. Billinghurst, Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System, In *Proceedings of the IWAR'99*, pp.85-94, 1999.

A First Look at a Telepresence System with Room-Sized Real-Time 3D Capture and Life-Sized Tracked Display Wall

Andrew Maimone* Henry Fuchs†

Department of Computer Science
University of North Carolina at Chapel Hill

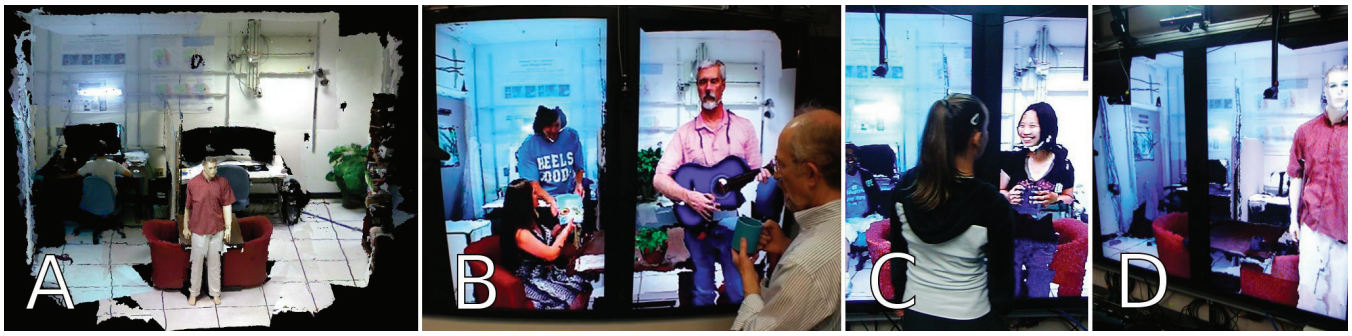


Figure 1: Left to Right: A) System Kinect Coverage. B-C) Users collaborating with remote participants. D) View from far right side.

Abstract

This paper provides a first look at a telepresence system offering room-sized, fully dynamic real-time 3D scene capture and continuous-viewpoint head-tracked display on a life-sized tiled display wall. The system is an expansion of a previous system, based an array of commodity depth sensors. We describe adjustments and improvements made to camera calibration, sensor data processing, data merger, rendering, and display, as required to scale the earlier system to room-sized.

Keywords: teleconferencing, virtual reality, sensor fusion, camera calibration, color calibration, filtering, tracking

Index Terms: H.4.3 [Information Systems Applications]: Communications Applications—Computer conferencing, teleconferencing, and videoconferencing

1 Introduction

The unification of two remote workspaces through a shared virtual window, allowing remote participants to see each other's environment as a continuation of their own, has been a long-standing goal of telepresence [3, 7].

A recent system by the authors [4] progressed toward this goal by providing fully dynamic 3D scene capture and continuous-viewpoint head tracked 3D display, but the impression that the remote environment was an extension of the viewer's own was limited by the relatively small capture volume (a small office cubicle) and display area (0.43 m^2).

Previous capture systems have demonstrated real-time acquisition of larger volumes (the size of a small room) with various compromises. A 2002 UNC/UPenn system [10] presented an office sized volume, but only the remote collaborator was dynamic; the rest of the scene was a scanned static 3D model. A more recent system by Petit et al. [6] also captures only the remote collaborator, but utilizes a multi-camera setup to offer a larger capture volume. Systems based on interpolation between densely placed 2D cameras, such as the 2004 MERL 3DTV system [5] and the 2010 Holografika system [1] also offer larger capture volumes but do not support continuous viewpoints or vertical parallax.

These limitations of 2D camera systems can be eliminated by providing depth estimates, as in the proposed 3DPresence [8] and Extended Window Metaphor [11] systems and in the demonstrated free-viewpoint television systems of Nagoya University [9], but to our knowledge these systems have not yet demonstrated real-time capture at room scale.

Our new display system, a pair of large (65") conventional 2D display panels with user tracking, is a temporary compromise. It has high resolution (4 MP) and supports continuous viewpoints through encumbrance-free tracking but does not provide a stereo image nor support for multiple tracked users. In comparison, a state-of-the art 3D display, the Holovizio [1], alleviates these issues, but introduces others – lower resolution, a limited field of view and minimum user distance, and lack of vertical parallax. The Holovizio also comes at a much higher cost and complexity.

In this paper, we present an updated telepresence system that supports fully dynamic capture of a small room ($\sim 15\ m^2$) that can be rendered from any of the continuous viewpoints of a tracked user. We believe our system to be the first to incorporate these characteristics at room scale. Furthermore, we have fitted our system with a large tracked display wall that allows the user to become more immersed in the remote scene.

2 Background and Contributions

The system described in this paper is an extension of earlier work [4] based on an array of Microsoft Kinect™ sensors, widely available, inexpensive ($150) devices that provide matched color images and depth maps. Multiple Kinect sensors were strategically placed and calibrated to provide a unified mesh of the 3D scene which is rendered from the perspective of a tracked user.

*e-mail: maimone@cs.unc.edu
†e-mail: fuchs@cs.unc.edu

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

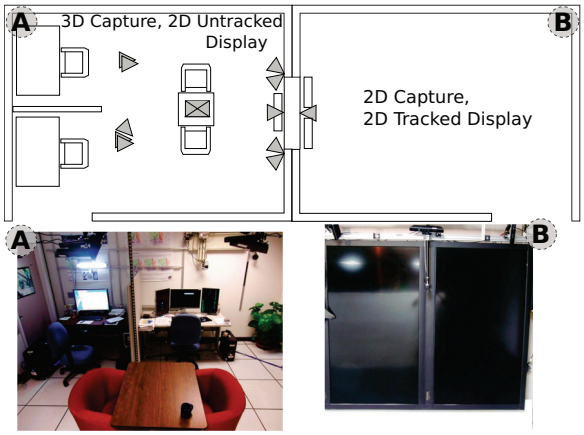


Figure 2: Layout of demonstrated system. Top: Layout of capture and display rooms showing virtual shared wall area. Bottom: Photos of actual capture and display rooms.

To support a much larger capture volume than our previous system [4], the following improvements were made to the system:

1. Calibration procedures were improved to reduce Kinect misalignment, which was more evident in our new larger configuration.

2. Depth data processing was enhanced to reduce the effects of noise caused by increased distances between surfaces and Kinects and by the interference that occurs when multiple Kinects have overlapping views.

3. The software system was enhanced to allow Kinects with a view of only static surfaces (upper walls, ceilings, etc) to be turned off or physically removed, increasing performance and providing more coverage than the total number of physical Kinects allow.

3 System Overview

3.1 Physical Layout

Figure 2 shows the layout of our system. Room A, which offers 3D capture and 2D untracked display of room B, measures 4.3 m $\times$ 4.7 m $\times$ 2.4 m with approximately 75% of the total floor area ($\sim$ 15 m^2) in the capture zone. Room B features 2D capture and a head-tracked perspective 2D display of room A. The two rooms are physically separated, but a view of room A can be seen "through" the display of room B as if the spaces were aligned with a shared hole in the wall (see Figure 2, top). This configuration allows us to demonstrate 3D capture and tracked 2D display while requiring only one set of Kinects and tracked displays.

Figure 4 shows our "ideal" configuration – 3D capture and multi-user autostereoscopic displays are supported in both rooms (C,D).

3.2 Hardware Configuration

Both rooms in our proof-of-concept system share a single PC with a quad-core CPU and a Nvidia GeForce GTX 295 graphics board. Eleven Microsoft Kinect sensors are connected to the PC. The 2D display wall consists of two 1080p 65" LCD panels. We avoided networking and audio in this version of our system since both rooms are served by a single PC and are in close proximity. We plan to address these omissions in a future system.

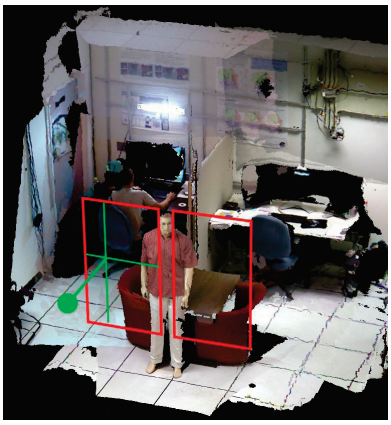


Figure 3: Virtual position of displays (red rectangles) and typical user eye position (green spot) in capture room A of Figure 2.

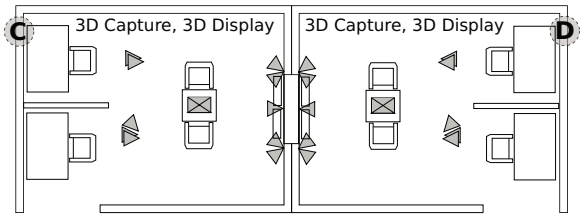


Figure 4: Virtual layout of ideal system

3.3 Software Overview

Rendering We used the same basic rendering pipeline as in our original system [4], but without stereo rendering:

1. When new data is available, read color and depth images from Kinect units and upload to GPU

2. Smooth and fill holes in depth image.

3. For each Kinect's data, form triangle mesh using depth data.

4. For each Kinect's data, apply color texture to triangle mesh and estimate quality at each rendered pixel; render from the tracked user's current position, saving color, quality, and depth values.

5. Merge data for all Kinect units using saved color, quality and depth information.

Tracking We used the same eye tracking method as in our original system [4] – 2D eye detection, depth data, and motion tracking are combined to create a markerless 3D eye position tracker. For the images in this paper and in the supplemental video, the filming camera was tracked in 3D space using a color-coded marker (Figure 5) and the Kinect's depth information.

4 System Enhancements

4.1 Camera Calibration

As previously described [4], we used Zhang's method [13] (as implemented in the OpenCV library) to obtain an initial calibration of the color cameras in the Kinect units using a checkerboard target. However, our new system required several additional considerations:

1. There was no single Kinect unit that shared views with all other units.

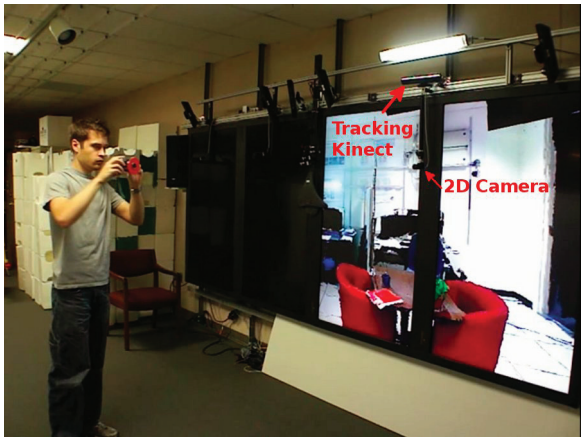


Figure 5: Video camera with colored ring that is detected for tracking, used to provide the tracked images in this paper and in the supplemental video. (Unmarked cameras and deactivated displays visible in image are not used in our system.)

2. Since Kinects are more sparsely placed, the checkerboard target is often located farther from the Kinects, reducing resolution and increasing checkerboard corner detection error.

3. Pairs of Kinects often have overlap at the edges of each other's fields of view, where radial distortion is greatest.

4. Since the user has a greater range of viewing positions, there is more opportunity to look at the scene farther from any one Kinect's line of sight, making depth error more apparent. Since only the pose and distortion parameters of the Kinect color cameras are calibrated, there is no opportunity to correct possible distortions in Kinect depth imagery.

5. Since there are more Kinects than in our previous system, there is a greater opportunity for calibration error to propagate between units.

To address item 1, a camera calibration hierarchy was established that minimized the number of transforms between each Kinect and the reference Kinect. In our demonstrated configuration, at most two transforms were required to transform a Kinect into the reference view.

To address item 2, the Kinect's low framerate/high resolution camera mode was used during calibration and the computed calibration parameters were converted for the low resolution/high framerate mode used during scene capture. To further reduce error resulting from motion blur and from the lack of multi-Kinect synchronization, the checkerboard target was placed at rest before capturing each image.

To address item 3, radial distortion was corrected during scene capture using the distortion coefficients computed during intrinsic calibration of the color camera. Since the the API we are using to communicate with the Kinect, OpenNI[1], automatically registers the depth image to the color image, the same distortion coefficients were applied to both images.

To address items 4 and 5, a supplemental calibration procedure was established. The procedure operates on 3D points as measured by the depth sensor, rather than on 2D projections of points as seen by the color camera, to allow correction of biases in the Kinect's depth readings. The procedure aims to minimize the distance between 3D points measured by all Kinects, reducing the effect of error propagation between Kinects.

[1] http://www.openni.org/

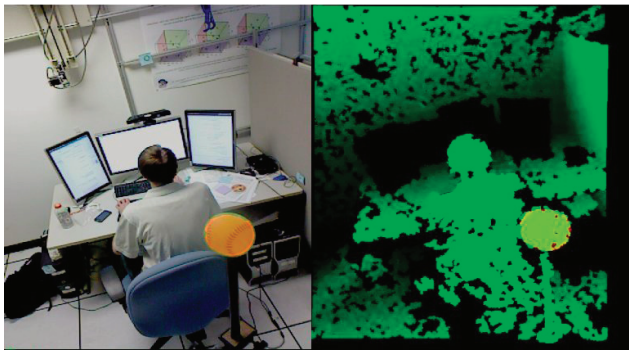


Figure 6: Sphere used for 3D calibration. Left: Green ball detected by its color and highlighted orange in software. Right: Corresponding depth values of ball highlighted orange in depth map.

The procedure is performed follows:

1. An initial calibration is performed using Zhang's method as described above.

2. A spherical object is placed in multiple positions in the capture area. At each position, the sphere is segmented by its color for all Kinects in view and the associated depth values are recorded. This step is illustrated in Figure 6.

3. The depth values from the previous step are fitted to sphere models using RANSAC [2] to eliminate outliers. If the computed radius is too far from true value or if the distances of the sphere center between Kinects are too far apart, the data is rejected.

4. The data for each Kinect is fitted to a affine transform that minimizes the distance between its detected sphere locations and the center of the sphere locations as seen by all other Kinects. RANSAC is used to eliminate outliers. An affine transform fitting was selected over a rigid transform in order to allow linear biases in the Kinect depth imagery to be corrected by scaling.

5. The previous step is repeated until convergence. In practice, between 10 and 100 iterations were performed.

4.2 Depth Data Processing

As previously described [4], interference between Kinect sensors with overlapping views causes holes and additional noise, which can be filled and smoothed in software. These effects are more prominent in our room-sized system as there is more overlap between Kinects – the ones in the rear of the room interfere with others in the rear as well as those in the front. In addition to interference, the Kinects are typically located farther from surfaces in our new system, causing additional depth noise. Another undesirable artifact of our old system was raggedness on edges that represent depth discontinuities. Since the lengths of the ragged edges are related to depth noise, this issue was also more pronounced in our new system.

To reduce the effects of extra noise, the previously described hole filling and smoothing algorithm [4] was enhanced to allow multiple passes of fine scale smoothing (by median filter), which is controlled by parameter N in the revised Algorithm 1. To reduce the appearance of ragged edges, we remove any data that does not meet the previously described hole filling criteria and smoothing criteria [4]; such data occurs at object boundaries along depth discontinuities. The trimming operation is applied to the first t_{trim} passes

of the N passes of the algorithm, allowing control of the degree to which edges are trimmed.

Algorithm 1 Modified N-Pass Median Filter for Hole Filling

```
for pass = 1 to N do
  for i = 1 to numPixels do
    depth_out[i] ← depth_in[i]
    if depth_in[i] = 0 or pass > 1 then
      count ← 0, enclosed ← 0
      v ← {}, n ← neighbors(depth_in[i], radius_pass)
      min ← min(n), max ← max(n)
      for j = 1 to n.length do
        if n[j] ≠ 0 then
          count ← count + 1
          v[count] ← n[j]
          if on_edge(j) then
            enclosed ← enclosed + 1
          end if
        end if
      end for
      if max − min ≤ t_r and count ≥ t_c and enclosed ≥ t_e then
        sort(v)
        depth_out[i] ← v[v.length/2]
      else if pass > 1 and pass <= t_trim then
        depth_out[i] ← 0
      end if
    end if
  end for
  depth_in ← depth_out
end for
```

4.3 Static Kinects

Although we believe that fully dynamic scene capture allows users to better communicate by utilizing surrounding objects, there are often parts of the scene that very rarely change (such as the upper walls and ceiling of a room) but still contribute to the sense of immersion. Providing real-time updates of these static surfaces decreases performance and contributes to interference that occurs between multiple Kinects. Our updated software improves upon this scenario in one of two ways:

1. A Kinect can be temporarily disabled, leaving the last captured frame in the scene. Frame rates increase as the data must no longer be processed and uploaded to the GPU at each frame.

2. A Kinect's last frame can be saved to disk, and the data can be incorporated into future capture sessions even if the Kinect is physically removed. This allows a limited number Kinects to be utilized more effectively. Note that the camera must be returned to its original position and new static data must be captured if the system is recalibrated.

In the latter case, the Kinect data is saved with all other Kinects turned off, eliminating any multi-Kinect interference and improving image quality. In either case, the saved or paused data is rendered just as live data and is color-corrected to match subsequent lighting changes.

We expect that this system could be further improved by automatically and dynamically switching on and off based on movement detected in a scene, trading system performance for latency in the activation of paused Kinects.

Figure 7: Color-coded coverage of our ten Kinect capture units.

Table 1: 3D Positional Error.

Case	RMS error (cm)
After initial 2D calibration	3.26
Points refitted using 3D affine transform	1.14

4.4 Display

In the new system introduced here, we replaced our autostereoscopic 3D display with a much larger and higher-resolution pair of tiled 2D displays, allowing the remote participant to appear life-sized at a natural interaction distance of 1 m. The tiled display area is approximately 2.5 m^2 (1.76 m x 1.43 m), 8% of which is covered by a 14.6 cm wide bezel. Combined display resolution is 2160×1920 pixels. As before, our display supports only a single head tracked user. In the future, we plan to return to an autostereo 3D display and seek one that can support multiple users and scale to the size of our current 2D display; one such possibility is the Random Hole Display [12].

5 Results

5.1 Kinect Coverage and Calibration Results

Kinect Coverage Figure 1A and Figure 7 show the coverage obtained with the ten Kinect capture units in our updated system. As shown, coverage is provided for most of the surfaces that can be seen by a viewer who is standing near the pictured standing mannequin, facing into the scene.

3D Positional Error 3D Positional error was measured by placing spheres throughout the capture area and measuring the distance between their detected centers between all Kinects with the sphere in view. Figure 8 shows the detected initial locations of each sphere and the locations after refitting using the method described in Section 4.1. These values are quantified in Table 1 – the error values listed are the RMS differences between the detected location of the sphere as seen by each Kinect and the center of the cluster of all Kinects with the sphere in view. In this data set, 98 sphere locations were recorded, seen by an average of 2.23 Kinects each. Each Kinect had between 14 and 40 of the 98 total spheres in view. Figure 9 shows an example of improved calibration using our new methods.

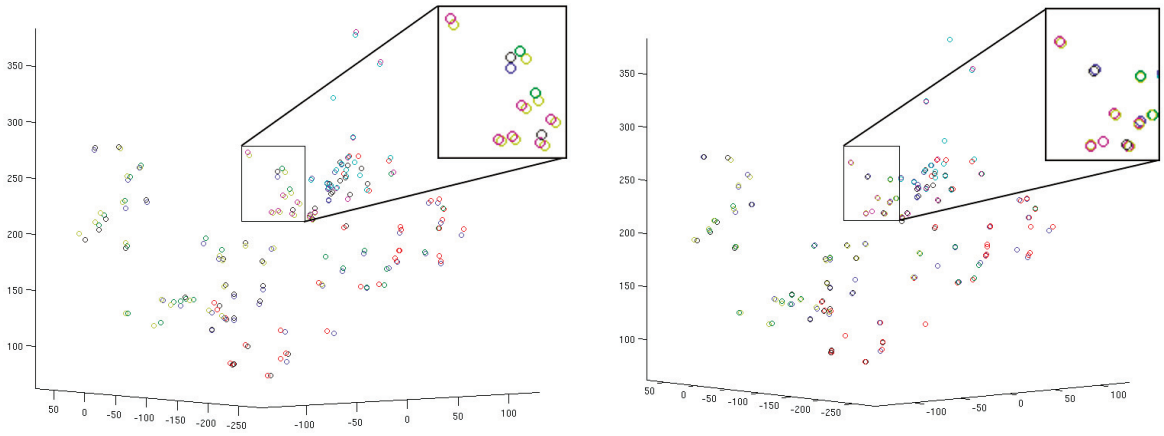


Figure 8: 3D Calibration using detected spheres. Left: Detected locations before adjustment (RMS Error: 3.26 cm). Right: Locations after adjustment (RMS Error 1.14 cm).

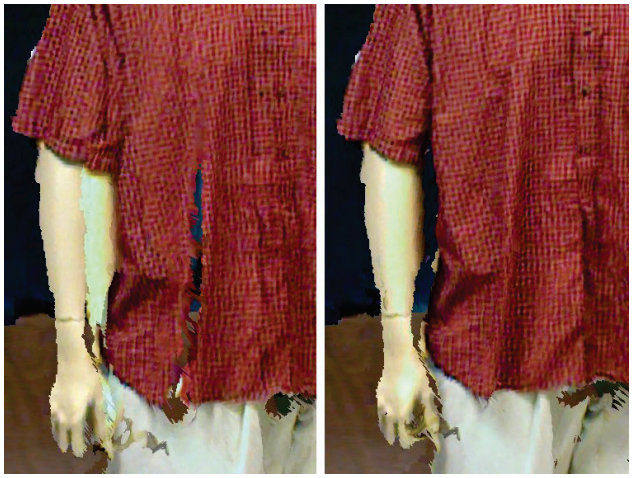


Figure 9: Improvement with additional 3D calibration. Left: Initial calibration resulted in misalignment near arm. Right: Arm misalignment improved with additional 3D calibration.

Figure 10: Depth filtering improvement. Left: Depth filtering from [4]. Right: Enhanced depth filtering.

Table 2: Display rates (frames per second)

Rendering Mode	FPS
No enhancements (raw colored point cloud)	62
Meshing, Hole Filling, Data Merger, Tracking	14

5.2 Depth Data Processing Results

Figure 10 shows a comparison of the old and new depth filtering algorithms. In the figure, the revised algorithm removed some of the noisy edges on the mannequin's head and shoulder when set to perform 3 hole filling and smoothing (parameter N) passes and 2 trimming (parameter t_{trim}) passes.

5.3 Display and Tracking

Figures 1B-1D show the system from the perspective of a tracked video camera. Remote users appear life-sized and tracking shows that the view appears correct from several angles, creating a window-like appearance.

5.4 System Performance

Table 2 lists the performance achieved with our test system in two rendering configurations. The system was configured to use data from 7 live capture Kinects, 3 static data Kinects (as described in Section 4.3), and 1 tracking Kinect to render a 2160×1920 view for the display wall. With all enhancements on, display rates remained interactive.

6 Conclusions and Future Work

We have presented solutions to some of the problems related to expanding an earlier telepresence system to room sized: improved calibration techniques, improved data filtering methods, and selectively using live and static data to improve performance.

Using the described methods, we have demonstrated a telepresence system that is able to capture a dynamic, room-sized 3D scene while allowing a remote user to look around the scene from any viewpoint on a life-sized display wall. Using a single PC our system was able to maintain interactive rendering rates.

There are several areas that we would like to improve in our room-size system. Image quality should be further enhanced – images tend to have a noisy look that could be improved with more advanced depth data processing techniques. Our system would also feel more immersive if rendering rates were raised to 30+ Hz.

We also intend to expand our test setup into the "ideal" system shown in Figure 4 by supporting 3D capture and autostereo 3D display for multiple users in both rooms.

ACKNOWLEDGEMENTS

The authors would like to thank Herman Towles, Andrei State, and Jonathan Bidwell for technical discussions and advice, and John Thomas for helping construct some of the camera apparatus. This work was supported in part by the National Science Foundation (award CNS-0751187) and by the BeingThere Centre, a collaboration of UNC Chapel Hill, ETH Zurich, NTU Singapore, and the Media Development Authority of Singapore.

REFERENCES

[1] T. Balogh and P. T. Kovács. Real-time 3d light field transmission. volume 7724, page 772406. SPIE, 2010.

[2] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24:381–395, June 1981.

[3] S. J. Gibbs, C. Arapis, and C. J. Breiteneder. Teleport towards immersive copresence. *Multimedia Systems*, 7:214–221, 1999. 10.1007/s005300050123.

[4] A. Maimone and H. Fuchs. Encumbrance-free telepresence system with real-time 3d capture and display using commodity depth cameras. In *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, oct. 2011.

[5] W. Matusik and H. Pfister. 3d tv: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. *ACM Trans. Graph.*, 23:814–824, August 2004.

[6] B. Petit, T. Dupeux, B. Bossavit, J. Legaux, B. Raffin, E. Melin, J.-S. Franco, I. Assenmacher, and E. Boyer. A 3d data intensive tele-immersive grid. In *Proceedings of the international conference on Multimedia*, MM '10, pages 1315–1318, New York, NY, USA, 2010. ACM.

[7] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs. The office of the future: a unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '98, pages 179–188, New York, NY, USA, 1998. ACM.

[8] O. Schreer, I. Feldmann, N. Atzpadin, P. Eisert, P. Kauff, and H. Belt. 3dpresence -a system concept for multi-user and multi-party immersive 3d videoconferencing. In *Visual Media Production (CVMP 2008), 5th European Conference on*, pages 1 –8, nov. 2008.

[9] M. Tanimoto. Overview of free viewpoint television. *Signal Processing: Image Communication*, 21(6):454 – 461, 2006. Special issue on multi-view image processing and its application in image-based rendering.

[10] H. Towles, W.-C. Chen, R. Yang, S.-U. Kum, H. F. N. Kelshikar, J. Mulligan, K. Daniilidis, H. Fuchs, C. C. Hill, N. K. J. Mulligan, L. Holden, B. Zeleznik, A. Sadagic, and J. Lanier. 3d tele-collaboration over internet2. In *International Workshop on Immersive Telepresence, Juan Les Pins*, 2002.

[11] M. Willert, S. Ohl, A. Lehmann, and O. Staadt. The Extended Window Metaphor for Large High-Resolution Displays. pages 69–76.

[12] G. Ye, A. State, and H. Fuchs. A practical multi-viewer tabletop autostereoscopic display. In *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*, pages 147 –156, oct. 2010.

[13] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 1, pages 666 –673 vol.1, 1999.

Augmented Reality Enhanced Image-Guided Surgery System Using CT and Ultrasound Registration for Brain-Shift Estimation

Wei-Chic Huang[1], Chung-Hung Hsieh[1], Chung-Hsien Huang[1], Shin-Tseng Lee[2] , Chieh-Tsai Wu[2]
Yung-Nien Sun[3], Yu-Te Wu[4] and Jiann-Der Lee[1]*

[1]Department of Electrical Engineering, Chang Gung University, Tao-Yuan, Taiwan
[2]Deptartment of Neurosurgery, Chang Gung Memorial Hospital, Tao-Yuan, Taiwan
[3]Department of Computer Science and Information Engineering, National Cheng Kung University,Tainan, Taiwan
[4]Department of Biomedical Imaging and Radiological Sciences, National Yang-Ming University, Taipei, Taiwan

ABSTRACT

In neurosurgery, brain shift, usually caused by the gravity or the changes of intra-cranial pressure, is the main factor to affect the accuracy of tumor removal. To deal with this problem, an image-guided surgery system which corrects the brain shift from the pre-operative CT images by using intra-operative ultrasound images is presented. First, with reconstructing 2-D free-hand ultrasound images to 3-D volume data, the system applies a Mutual-Information based registration algorithm to estimate the deformation between pre-operative and intra-operative ultrasound images. The estimated deformation transform describes the shifts of soft tissues and is then applied to the pre-operative CT images. The brain-shift correction procedure was validated with a brain phantom. When the shift of an artificial tumor is from 5mm ~ 12mm, the overlapping rates can be improved from 32% ~ 45% to 87% ~ 95%. In addition, the system displays the fusion of the corrected CT images or the real-time 2-D ultrasound images with the patient in the physical space through a head mounted display device, providing an immersive augmented-reality environment.

KEYWORDS: Augmented Reality; Image-Guided Surgery; Brain-Shift Estimation; Medical Image Registration

1. INTRODUCTION

Nowadays, image-guided surgery (IGS) has become an important part for neurosurgery. The anatomical information observed from pre-operative medical images such as Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) helps neurosurgeons to diagnosis the status of disease, locate tumor and plan a surgery. With the assistance of IGS, each voxel on the pre-operative images can be linked to a 3-D position in the physical space and be reached by surgical tools, improving the accuracy of target localization and reducing surgical time.

However, pre-operative medical images only provide the information of the patient before surgery but not the up-to-date one. Some surgical targets such as soft tissues may be shifted during the surgery. The shift usually causes inaccurate target localization of IGS system. A simple and ordinary method to solve this problem is using intra-operative CT or MRI; however, this procedure would interrupt the process of operation and may not be practical due to the environmental limitation in operating room. Recently, a significant body of work appears on using intra-operative ultrasound scanner to obtain real-time images, especially on brain surgery [1], cardiac surgery [2], lung surgery [3], or liver surgery [4]. Since ultrasonic imaging is less damaging than CT and MRI and low cost, intra-operative ultrasound can conveniently be applied during operation. For example, the ultrasound images acquired before and during surgery can be compared with each other in order to estimate the brain shift. The transformation estimated can be further applied to the pre-operative CT or MRI, which provides better and detailed anatomical information. This process helps neurosurgeon to map pre-operative information onto intra-operative situation efficiently.

In the last two decades, Augmented Reality (AR) has been drawn much attention and been adopted in various fields such as education, entertainment, and medical applications. AR could provide surgeons a mapping visualization instead looking away from a patient to consult a manual operation. The study in [5] may be referred as the pioneer of applying AR in the operating room. In [6], a research group at UNC Chapel Hill presented an AR system for ultrasound-guided needle biopsy of breast. In this study, an AR enhanced IGS system using ultrasound and CT registration for brain shift estimation is presented. In the setup of the IGS system, a spatial digitizing device is attached to the probe to obtain the spatial location and orientation of the ultrasound probe. After performing the calibration between the digitizing device and the ultrasound probe, 3-D ultrasound volume data can be reconstructed by using a pixel-based interpolation algorithm. When a patient lies down on the operation table, the coordinate of the patient in the physical space, the preoperative CT of the patient, and of the ultrasound images can thus be integrated into a unified coordinate system. When the brain shift occurs during surgery, the patient is scanned again by the ultrasound scanner. The intra-operative ultrasound images are then utilized to register with the per-operative ultrasound images for the estimation of deformation, which is subsequently applied to update the pre-operative CT, producing better anatomical information and closer to the intra-operative situation.

To validate the performance of the procedure of the brain-shift correction, a brain phantom made by silicone was utilized for testing and evaluation. We simulated the spatial shift with 5mm, 8mm and 12mm by squeezing the brain phantom. The estimated transformation was applied on the pre-operative CT images and the transformed images were compared with ground-truth intra-operative CT images.

Moreover, in general IGS system, the anatomical information resolved from images is usually displayed on the screen. In the proposed system, the medical images are augmented with the patient on the real scene by using an AR head-mounted display (HMD)[7]. The AR display provides more immediate and direct visual experience to neurosurgeons. In addition, the visual display modes can be changed under requirements.

The paper is organized as follows. Section 2 describes the proposed system in detail. Section 3 reveals our experimental

Chang-Gung University, 259 Wen-Hwa 1st Road, Kwei-Shan Tao-Yuan, Taiwan, 333, R.O.C., *corresponding author:jdlee@mail.cgu.edu.tw

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

results and AR visualization. Conclusions are drawn in Section 4.

2. METHODS

System and Flowchart

Figure 1 shows the presented AR-enhanced IGS system includes image modalities, hardware devices, and their spatial relationships. The following components are involved: preoperative CT images, a portable ultrasound scanner, a digitizing device, a movable camera embedded with an HMD, and a designed black-and-white pattern for AR visualization.

For the IGS installation, firstly, the digitizing device is attached to the ultrasound probe. A commercial digitizing system, NDI Polaris Vicra System [8], abbreviated as NDI hereafter, is adopted as the digitizing device. The coordinates of the NDI (C_{NDI}) and the ultrasound probe (C_{US}) are calibrated with a calibration box, as shown in Fig. 2 (a). Therefore, real-time spatial tracking of the ultrasound probe can be achieved. In addition, the coordinate of the AR pattern (C_{AR}) is also calibrated with the NDI system. When the AR pattern appears in the field of view of the moving camera CAM, the extrinsic parameters of the camera can be estimated through the observed shape of the pattern. As a result, the AR-enhanced visual display can be provided.

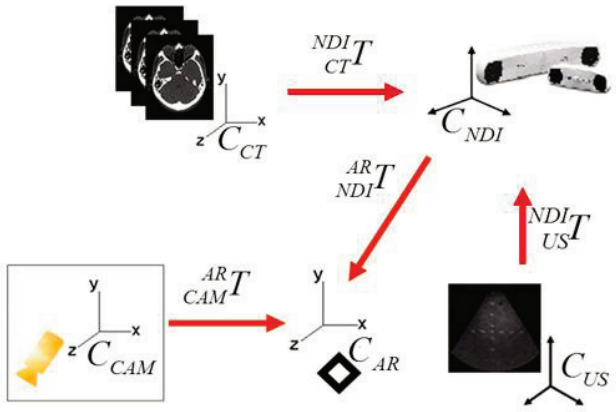
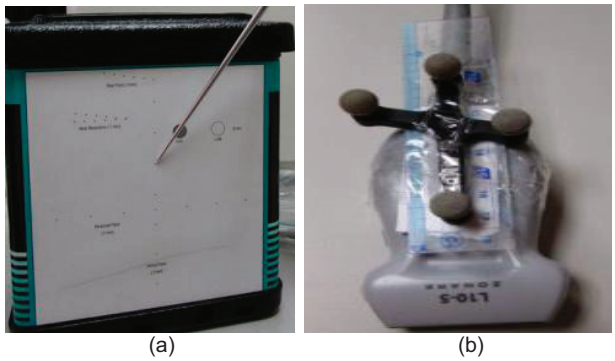


Figure 1. Components involved and their spatial relationships



(a) (b)

Figure 2. (a) Calibration box for ultrasound/DRF/NDI calibration; (b) Ultrasound probe attached with DRF

Figure 3 shows the flowchart of the medical image registration and AR visualization of the proposed system. Two types of medical imaging modalities are involved: one is ultrasound and the other is per-operative CT (Pre-CT). The ultrasound provides real-time but noisy and low-resolution pre-operative (Pre-US) and intra-operative (I-US) images, while the Pre-CT provides high-resolution but pre-operative anatomical information. Notably, the soft tissues such as brain observed from the Pre-CT may have some spatial shifts due to the surgical operation or gravity during surgery. In this study, to compensate the shifts we update the Pre-CT via the transformation which was estimated by registering the Pre-CT/Pre-US with Pre-US/I-US images.

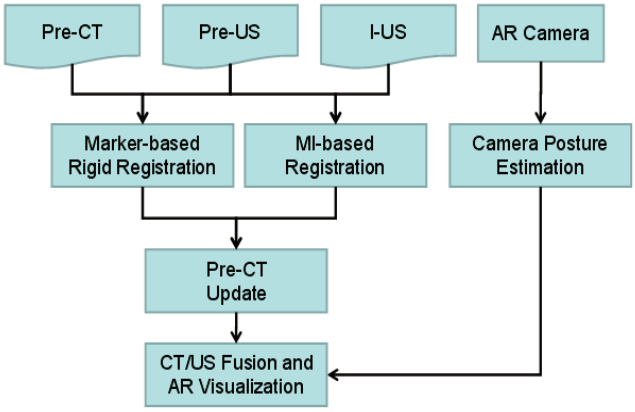


Figure 3. Flowchart of the proposed system for data fusion AR visualization

2.1 Ultrasound/NDI Calibration

In order to map the pixels on the ultrasound image to the physical space, as reported in [9], the ultrasound probe is attached to a trackable device named Dynamic Reference Frame (DRF) of the NDI digitizing system, as shown in Fig. 2 (b). With ultrasound/DRF/NDI calibration, each pixel (x_{US},y_{US}) at the C_{US} can be mapped to a real 3-D location (x_{NDI},y_{NDI},z_{NDI}) of C_{NDI} by applying a 4×4 transform ${}^{NDI}_{US}T$, which is defined in Eq. (1). The ${}^{NDI}_{US}T$ is a composition of two transforms ${}^{NDI}_{DRF}T$ and ${}^{DRF}_{US}T$, *i.e.*, ${}^{NDI}_{US}T = {}^{NDI}_{DRF}T \cdot {}^{DRF}_{US}T$, where ${}^{NDI}_{DRF}T$ is the DRF posture, which can be read directly from the NDI system. The ${}^{DRF}_{US}T$ is estimated by using least squares method (LSM) [10] with sixteen co-selected points from a calibration box, as shown in Fig. 3(b), by using the NDI digitizer and selecting the corresponding pixels on the ultrasound image. Readers are referred to [2] for the details of the calibration process.

$$\begin{bmatrix} x_{NDI} \\ y_{NDI} \\ z_{NDI} \\ 1 \end{bmatrix} = {}^{NDI}_{US}T \begin{bmatrix} x_{US} \\ y_{US} \\ 0 \\ 1 \end{bmatrix} \quad (1)$$

2.2 3-D Reconstruction of Ultrasound Images

Since the ultrasound scanner provides only 2-D images, a 3-D reconstruction algorithm is applied on a sequence of free-hand 2-D scans to obtain 3-D volume data. In this study, the freehand 3-D ultrasound reconstruction algorithm using Pixel-Based Methods (PBM) proposed by Solberg et al. [11] is employed. The PBM applies a 3-D Gaussian kernel around a voxel, and drives the impact of the voxel to its neighbor voxels as a weighting function. Therefore, the voxel needed to be interpolated can be reconstructed according to the weighted contribution from its neighbors.

2.3 Pre-CT/Pre-US Registration

The Pre-CT/Pre-US registration is accomplished by selecting N landmarks, denoted as P_{NDI}, on the patient by NDI and their corresponding pixels, denoted as P_{CT}, on the CT images. The landmarks could be any artificial skin markers glued externally to the patient or natural feature points of the patient. Therefore, the transform ${}^{NDI}_{CT}T$ can be calculated by Eq. (2) with LSM if we have at least four landmarks.

$$P_{NDI} = {}^{NDI}_{CT}T \cdot P_{CT} \quad (2)$$

2.4 Pre-US/I-US Registration and Pre-CT Update

To estimate brain shifts, inspired by the work proposed by Letteboer [1], we estimate a free-form deformation (FFD) transformation by using a nonlinear Pre-US/I-US registration and then update the Pre-CT by the estimated transformation. The basic idea of FFD involves manipulating an underlying mesh of a set of control points to obtain the prostate deformation model. In this study, B-Spline Mutual Information (MI) based algorithm [12] is applied for the task of the nonlinear Pre-US/I-US registration.

B-Spline is a free-form deformation and its basic idea is locating an object within a mesh so that the object deforms according to the deformation of the mesh. B-Spline registration manipulates the mesh via a set of control points and naturally lends itself for multi-resolution registration.

MI is defined to maximize the common information shared by the two images to be registered and to reduce the information in the combined image. The more correlated the two images, the lower joint entropy. The implementation of the B-Spline MI-based algorithm is accomplished by using the Insight Segmentation and Registration Toolkit library (ITK) [13] and can be simply expressed by the following equation.

$$\hat{T} = \arg\max_{T_{B\text{-Spline}}} \mathrm{MI}(\text{Pre-US}, \text{I-US}, T_{B\text{-Spline}}) \quad (3)$$

Through the step of the Pre-US/I-US registration, the non-rigid brain shifts could thus be compensated, *i.e.*, using the non-linear transformation estimated by the B-Spline MI-based algorithm. The estimated transformation can then be applied to the Pre-CT, obtaining the intra-operative CT (I-CT'), *i.e.*, $\text{I-CT}' = \hat{T}(\text{Pre-CT})$.

2.5 AR Visualization

After performing the mentioned image-to-patient (section 2.4) and image-to-image (section 2.5) registrations, the Pre-CT and I-US are integrated to the physical space of the patient, *i.e.*, the NDI coordinate system. With the use of ARTOOLKIT [14], an HMD device attached with a CCD camera is applied to provide an immersive AR environment for uses. The pose of the camera can be estimated through observing a designed black-and-white AR pattern. Note that the AR pattern should be calibrated with the NDI coordinate system before surgery. The calibration procedure is accomplished by selecting the four corners around the AR pattern by NDI, and the estimating the transform ${}^{AR}_{NDI}T$ by LSM as well. Figure 4 (a) shows the HMD and the camera, and Fig. 4 (b) is the AR pattern. The adopted HMD is Iwear-VR920 made by VUZIX Inc [15].

3. Results

A phantom made by silicone was utilized to validate the procedure of the brain-shift correction. The shape of the phantom was reconstructed according to a brain 3-D model segmented from a set of MRI images. In addition, a balloon filled with glycerin was attached beneath the brain surface as a simulation of brain tumor. Figure 5 (a) and (b) show the phantom and its 3-D reconstruction model, respectively. The phantom was placed within a plastic cubic and performed CT scan four times. The first scan was under normal situation, while in the other scans the surface of phantom was squeezed with a plastic rod to simulate the effect causing by brain shifts, as shown in Fig 5(c), (d), and (e), where the red circles indicate the location of the glycerin balloon. The first scan was regarded as the Pre-CT, and the others were regarded as intra-operative CT (I-CT) and denoted as Scan-A, Scan-B and Scan-C. We estimated the overall shift of the glycerin balloon by measuring the distance between its gravity centers calculated in Pre-CT and I-CTs. The overall shifts of Scan-A, Scan-B and Scan-C are 5mm, 8mm, and 12mm, respectively. The volume size of CT image is 512 × 512 × 324 and the voxel size is 0.4 × 0.4 × 0.8 mm^3, while the volume size of US image is 230 × 350 and the pixel size is 0.6 × 0.6 mm^2. The ultrasound device is made by ZONARE Medical Systems Inc.

(a) (b)

Figure 4. Components for AR visualization (a) HMD and the camera attached; (b) AR pattern

3.1 Evaluation on Image Registration

Figure 6 illustrates an instance of the CT/US registration results with Scan-A. In Fig.6 (a), the Pre-US and Pre-CT are shown by gray scale and green, respectively. For better visualization, the contour of the glycerin balloon was extracted manually from the Pre-CT and laid on the Pre-US, as shown in Fig. 6 (d). Similarly, Fig. 6(b) and (e) show the fusion results of the Pre-CT and I-US. The shift of the glycerin balloon caused by squeezing brain can be observed easily from the fusion images. The B-Spline MI-based registration was applied on Pre-US and I-US, resulting in a nonlinear transform $\hat{T}$, as defined in Eq. (3). The transform was utilized to update the Pre-CT so that it can be deformed based on the estimated brain shifts and denoted as I-CT'. The results of I-US/I-CT' fusion are shown in Fig. 6 (c) and (f) where the I-CT' is colored in blue.

Since the CT scans of the phantom with different deformations were performed, they could be adopted as the ground truth to evaluate the performance of compensating the brain shift. Figure 7 (a) and (b) show the chessboard-like image fusion and the contours of the glycerin balloon, respectively. The green line indicates the contour extracted from the Pre-CT image, while the red line is from the I-CT image, *i.e.*, Scan-A. After updating the Pre-CT by the transform $\hat{T}$, the shift effect is thus corrected and

illustrated in Fig. 7 (c) and (d). The blue line in Fig. 7(d) indicates the contour of the glycerin balloon extracted from the updated Pre-CT, *i.e.*, I-CT'. It can be seen that the I-CT's provides more accurate anatomical information than Pre-CT.

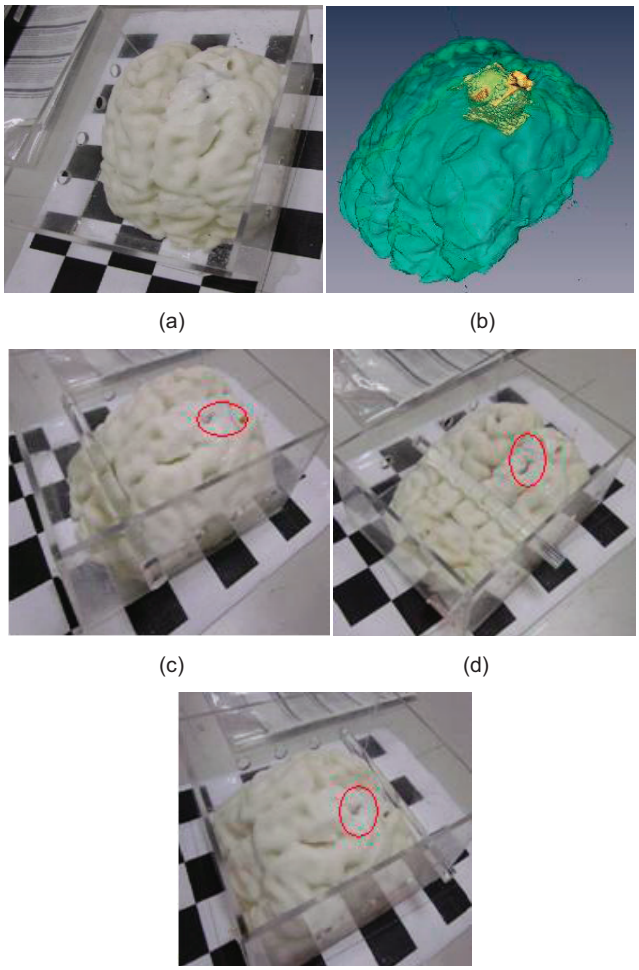


Figure 5. Brain phantom for evaluation (a) the phantom; (b) 3-D model of the phantom; (c), (d) and (e) the phantom squeezed under different pressure.

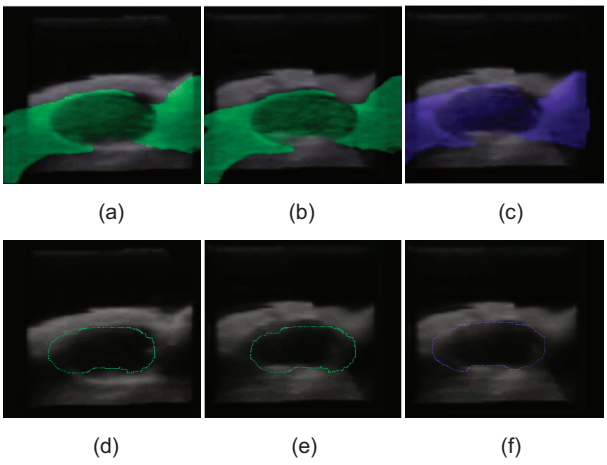


Figure 6. An instance showing the image fusion results of CT/US registration. (a) Pre-CT/Pre-US fusion; (b) Pre-CT/I-US fusion; (c) image fusion of I-US and updated Pre-CT; (d) overlapping the Pre-CT contour on Pre-US; (e) overlapping the Pre-CT contour on I-US; (f) overlapping the updated Pre-CT contour on I-US.

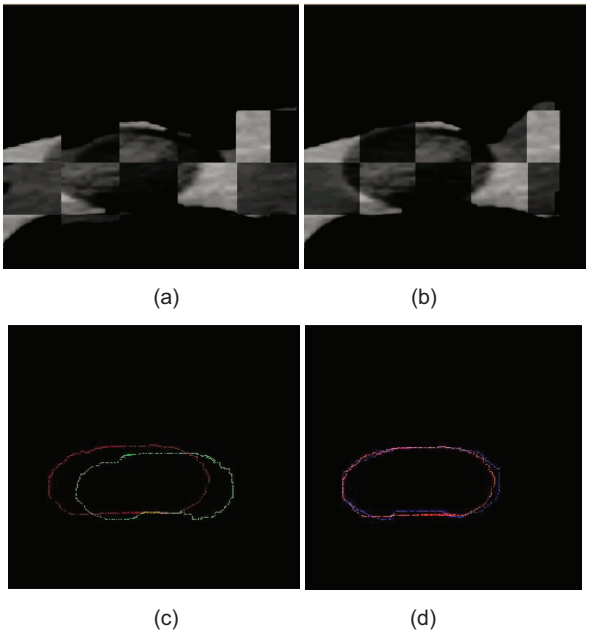


Figure 7. Comparison among Pre-CT, I-CT and I-CT' (a) chessboard-like fusion of Pre-CT and I-CT; (b) chessboard-like fusion of I-CT' and I-CT; (c) contours of the glycerin balloon extracted from Pre-CT (green) and I-CT (red); (d) contours of the glycerin balloon extracted from I-CT' (blue) and I-CT (red)

More precisely, we measure the overlapping rate of the glycerin balloon between the I-CT' and I-CT to validate the performance of brain-shift estimation. The overlapping rate is defined as $(A \cap B)/(A \cup B)$, where A is the volume of the glycerin balloon in the I-CT and B is its volume in the I-CT'. Figure 8 shows the overlapping rate before and after performing brain-shift estimation of the three CT scans. The overlapping rates of Scan-A, Scan-B and Scan-C were 97%, 95%, and 87%, respectively. It is anticipated that the larger is the brain shift, the lower is the compensation accuracy.

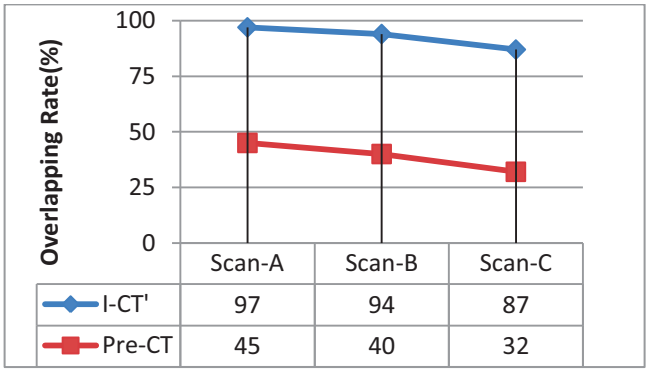


Figure 8. Comparison of brain-shift estimation among Scan-A, Scan-B and Scan-C.

3.2 Augmented Reality Display

Finally, we demonstrate the immersive AR display provided by the proposed system. The AR display mode is switchable according to the user's need. The available modes include displaying the reconstructed CT 3-D model, the real-time ultrasound image, or a mixture of them. Figure 9 (a) shows the fusion of the phantom and the I-CT', Fig.9 (b) is with the I-US, and Fig. 9 (c) shows the glycerin balloon extracted from I-CT' with I-US.

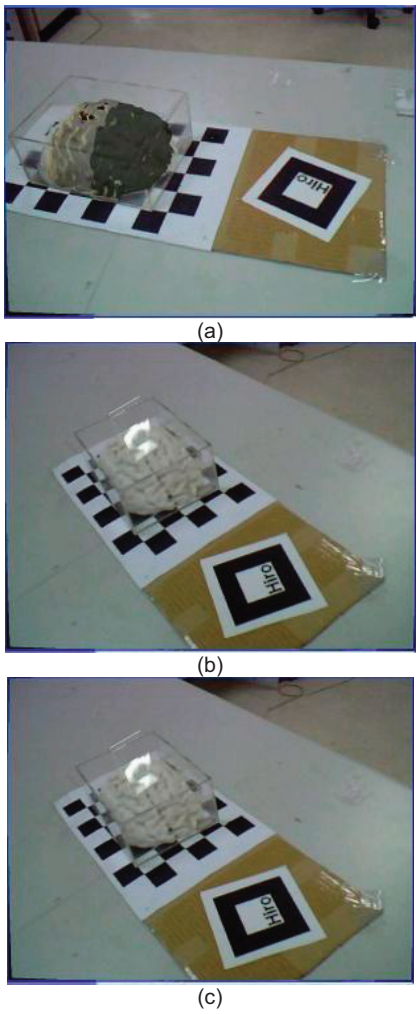


Figure 9. Results of AR display (a) augmented with I-CT'; (b) augmented with I-US; (c) augmented with a mixture of I-CT' and I-US.

4. Conclusion

In this study, we have presented an AR-enhanced IGS system using preoperative CT and intra-operative ultrasound. The preoperative CT provides anatomical information of patient, the intra-operative ultrasound produces real-time images during the surgery, and the AR display provides user a direct and integrated visualization experience.

Furthermore, the application of proposed system to the brain surgery application for brain-shift estimation was also investigated. In order to estimate the brain shift, a non-rigid image registration method based on Mutual Information and free-form B-spline deformation was adopted. The registration method was applied to estimate the deformation between pre-US and I-US, and then update pre-CT by the estimated transformation. Thus the updated pre-CT showed better representation on the current situation of the patient than the pre-CT.

So far, the experiments for validating the proposed system are only performed on phantoms, future works will test the system on animal trials. In addition, the brain-shift estimation model can be further connected with a brain biomechanical model for a more realistic estimation.

Acknowledgement

This study is supported by the technology development program of Ministry of Economic Affairs, R.O.C (serial number: 99-EC-17-A-19-S1-035) .

References

[1] M. M. Letteboer, P. W. Willems, and W. J. Niessen, "Brain Shift Estimation in Image-Guided Neurosurgery Using 3-D Ultrasound," IEEE Transactions on Biomedical Engineering, vol. 52, no. 2, pp. 268-286, 2005

[2] X. Huang, J. Ren, G. Guiraudon, D. Boughner, and T. M. Peters, "Rapid Dynamic Image Registration of the Beating Heart for Diagnosis and Surgical Navigation," IEEE Transactions on Medical Imaging, vol. 28, pp. 1802-1814, 2009

[3] N. A. Sadeghi, R. V. Patel, and A. Samani, "CT-Enhanced Ultrasound Image of a Totally Deflated Lung for Image-Guided Minimally Invasive Tumor Ablative Procedures," IEEE Transactions on Medical Imaging, vol. 57, no. 10, pp. 2627-2630, 2010

[4] W. Wein, O. Kutter, A. Aichert, D. Zikic, A. Kamen, and N. Navab, "Automatic Non-Linear Mapping of Pre-procedure CT Volumes to 3D Ultrasound," In Proceedings of the 2010 IEEE international conference on Biomedical imaging: from nano to Macro, pp. 1225-1228, 2010

[5] W. Lorensen, H. Cline, C. Nafis, R. Kikinis, D. Altobelli and L. Gleason, "Enhancing Reality in the Operating Room," in Proceedings of Visualization'93, pp. 410-415, 1993.

[6] A. State, M. A. Livingston, W. F. Garrett, G. Hirota, M. C. Whitton, E. D. Pisano and H. Fuchs, "Technologies for Augmented Reality Systems: Realizing Ultrasound Guided Needle Biopsies," in Proceedings of SIGGRAPH'96, vol. 30, pp. 439-446, 1996.

[7] T. Blum, S. M. Heining, O. Kutter, and N. Navab, "Advanced Training Methods Using Augmented Reality Ultrasound Simulator," In Proceedings of IEEE International Symposium on Mixed and Augmented Reality, pp. 19-22, 2009

[8] Northern Digital Inc., http://www.ndidigital.com/

[9] H. Zhang, F. Banovac, K. Cleary, and A. White, "Freehand 3D Ultrasound Calibration Using an Electromagnetically Tracked Needle," In Proceedings of the SPIE, vol. 6141, pp. 775-783, 2007

[10] W. M. Donald, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," Journal of the Society for Industrial and Applied Mathematics, pp. 431-441, 1963

[11] O. V. Solberg, F. Lindseth, H. Torp, R. E. Blake, and T. A. Hernes, "Freehand 3D Ultrasound Reconstruction Algorithms - A Review," Ultrasound in Med & Biol, vol. 33, no. 7, pp. 991-1009, 2007

[12] Y. Jin and G. Ma, "Investigation and Evaluation of Optimal Registration for Medical CT Images," International Congress on Image and Signal Processing, pp. 2794-2797, 2010

[13] Insight Segmentation and Registration Toolkit., http://www.itk.org/

[14] Artoolkit 2002, http://www.washington.edu/ARTOOLKIT/

[15] Vuzix Inc., http://www.vuzix.com/home/

3-Dimensional Visual Navigation for Repetitive Transcranial Magnetic Stimulation Treatment

‡Yoshihiro Yasumuro*
Faculty of Environmental and Urban Engineering, Kansai University

Tatsuya Ogino
Faculty of Environmental and Urban Engineering, Kansai University

Masahiko Fuyuki
Faculty of Environmental and Urban Engineering, Kansai University

‡Atsushi Nishikawa
Faculty of Textile Science and Technology, Shinshu University

‡Masaki Sekino
Dept. of Electrical Engineering and Information Systems, Grad School of Engineering, The University of Tokyo

‡Taiga Matsuzaki
Home Healthcare Research & Development Department, Teijin Pharma Limited

Kouichi Hosomi
Dept. of Neuromodulation and Nuerosurgery, Osaka University‡

Youichi Saitoh
Dept. of Neuromodulation and Nuerosurgery, Osaka University‡

ABSTRACT

Repetitive transcranial magnetic stimulation (rTMS) is a non- invasive method for treating various neurological and psychiatric disorders. This paper focuses on a treatment for neuropathic pain that can be caused by a lesion or disease of the central or peripheral nervous system, including stroke, trauma or surgical operation. With the growing demands of neuropathic pain patients and their increasing numbers, rTMS treatment tools are becoming more necessary. rTMS uses electromagnetic induction to induce weak electric currents by rapidly changing the magnetic field. Targeting a specific part of the brain to locate the magnetic field works as a treatment for pain relief. However, the current style of rTMS treatment is still developing and is so technically specialized that only a limited number of hospitals and only a handful of specialists can provide this therapy. The existing systems of rTMS are based on an optical maker-based 3-dimensional (3D) sensing technique for positioning the stimulation coil to target the small spot in the region of interest in the brain, and for referring pre-scanned MRI data to check the target position. Furthermore, this system requires the patient to be fixed on a bed in which optical markers for 3D sensing are placed during the treatment to maintain positioning precision. We propose a constraints-free style of the rTMS system, which employs a markerless method for positioning the patient's head with imaging sensors. Utilizing the patient's face shape instead of the positioning markers, this paper shows the potential for achieving a relaxed curative environment and an easy-to-handle system framework.

Index Terms: H.5.2 [INFORMATION INTERFACES AND PRESENTATION]: User Interfaces—Graphical user interfaces (GUI); I.4.8 [IMAGE PROCESSING AND COMPUTER VISION]: Scene Analysis—Tracking; J.3 [LIFE AND MEDICAL SCIENCES]: Medical information systems—

1 INTRODUCTION

Repetitive transcranial magnetic stimulation (rTMS) has been gathering attention as a non-invasive method for treating various neurological and psychiatric disorders including strokes, Parkinson's disease, and depression. This paper focuses on a new treatment to alleviate neuropathic pain that can be caused by a lesion or disease of the central or peripheral nervous system, including stroke, trauma or surgical operation. rTMS uses electromagnetic induction to induce weak electric currents by rapid change of a pulsed magnetic field; thus, rTMS treatment is capable of stimulating specific parts of the brain, the primary motor area for the neuropathic pains for instance, with minimal discomfort in a non-invasive manner [5, 3].

The existing treatment method of rTMS is achieved by an optical 3- dimensional (3D) sensing technique for positioning the stimulation coil to target the spot in the brain that needs to be cured [12, 10, 8]. As shown in the Figure 1, before the treatment, the doctors are supposed to measure the 3D position of the feature points on the patients head; the nasal point, nose top and anterior auricular points, for instance. These positions allow the system to register the MRI data of the preoperative examination in the sensor coordinate by using the corresponding feature points in the MRI data. Then the positioning sensor tracks the coil, using a set of optical markers installed on it. Showing the relative position of the coil and the registered MRI data helps the doctor to target the spot in the brain. To ensure the positioning accuracy during the treatment, the patient must be immobilized on the bed throughout the treatment, because the registered patient's feature points are based on the optical markers fixed on the bed.

On the other hand, since the effects of rTMS last only several hours, rTMS treatments need to be available whenever the patient needs these treatments. In fact, rTMS treatments are only available at specialized clinics such as university hospitals, because only experienced physicians in a limited number of hospitals can use the expensive and complicated rTMS system in its present circumstances. Fukushima et al. proposed a magnetic navigation system designed for home use of rTMS [2], using inexpensive and small magnetic sensors. By collecting the spatial data samples of the magnetic field to record the proper position and orientation of the stimulation coil at the initial treatment by an expert doctor, the system can help the user navigate and reproduce the coil position to target the treatment spot for subsequent treatments at home. This system theoretically allows any users, even those without medical knowledge and techniques to easily locate the coil for subsequent treatment. The magnetic sensors are fixed on a glasses-style mount device which the user can easily wear when applying the treatment

Figure 1: Current rTMS Procedure Example

*e-mail: yasumuro@kansai-u.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

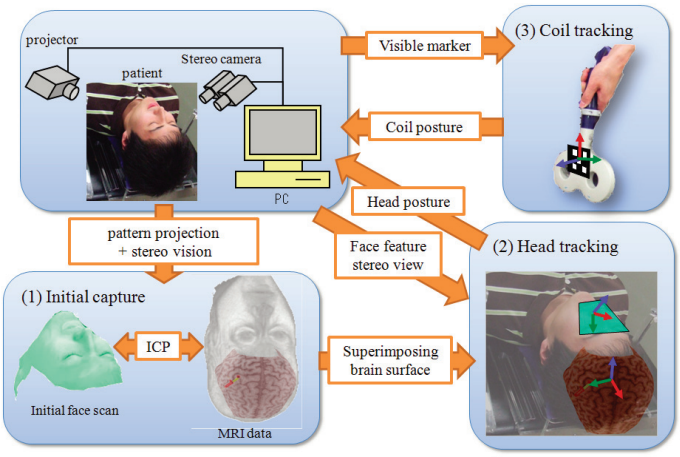


Figure 2: Proposed System Process

with no need to be constrained on the bed. Since the positioning accuracy depends on the reproducibility of wearing the glasses-style mount device, that device must be elaborately designed for adjustment to the individual patient in practical steps.

This paper proposes a new system framework that achieves both a constraints-free manner for patient use and stable positioning accuracy independent of the sensor mounting conditions.

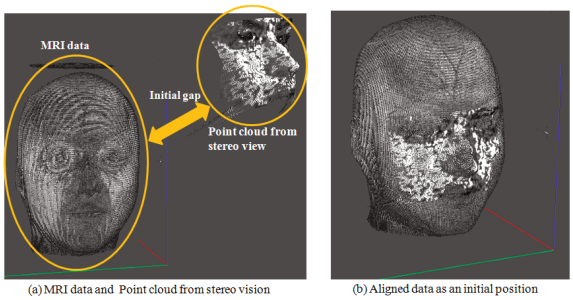
2 Approach

Our system utilizes image sensing for non-contact 3D positioning, employing a stereo-video camera. The basic idea is to use a set of visible features of the patient's face as a tracking target so the positioning system is independent from both the marker setup and the sensor mount on the patient. Figure 2 shows an overview of the proposed system. The principal components are initial alignment, head-tracking and coil-tracking functionalities.

Initial alignment procedure fits the MRI volume data to the user's face surface shape so that the brain geometry in the MRI volume is properly located to the live position of the user's head. The human face has many smooth areas without clear image features such as the cheeks and forehead. Passive measurement of the stereo camera may produce sparse shape data from such area. To acquire dense shape data for a stable fitting, we project a random pattern of dots from a projector onto the users face while the stereo camera captures the initial face shape. This shape-fitting procedure is done by an iterative computation to find a transform that gives the best match of the two-shape data sets, and the successive head motion is a relative transform of this initial head alignment.

The head tracking requires a realtime rate process for monitoring the live head motion of the patient without using physical constraints, assuming that the patient may sit at the doctor for face-to-face communication for instance, during the treatment. To achieve realtime head tracking, we apply the image feature tracking to give the corresponding points between the consecutive stereo camera frames over time. Each image feature, including eye corners, nose top, and corners of the mouth are tracked on both cameras, and their 3D coordinates are computed. Assuming that the set of face features has no deformation, a 3D transform to give the best match between the two face feature sets is found. Because the camera frame interval is short enough to capture a sitting patient, the displacements between the frames are not so large. Therefore, the iteration of this computation can reach convergence in a limited time.

For the coil tracking, we use a marker-based method, since installing and permanently maintaining the markers on the coil is allowed. Visible features of the markers can be registered and tracked



(a) MRI data and Point cloud from stereo vision (b) Aligned data as an initial position

Figure 3: Initial Alignment

for their positioning by the identical stereo camera, and thus the MRI data, the patient head and the coil geometries are integrated in the same camera coordinate. Relative positions and the orientations can be monitored and used for navigating the coil to the target spot on the brain.

3 Initial Alignment

For the initial alignment, the surface of the MRI is prepared by extracting the boundary surface dots between the areas of the air and the human skin in the MRI volume preliminarily. Using a random pattern projection to provide visible features on the patient's face surface, the stereo camera is capable of capturing the face shape as a set of dense points. Both the surface shape of the MRI and the patient's face are unorganized 3D point sets. To fit these point sets, we use the ICP (iterative closest point)[1].

The ICP is an iterative algorithm often employed to find a rigid-body motion by minimizing the difference between two shape data sets of point clouds a and b as shown in equation(1). Since the human face shape has a certain variation around eyes, nose, and mouth, the ICP is expected to be applicable for searching the proper corresponding points to fit together between the point sets.

$$E = \sum_{i=1}^{N} (Ra_i + t - b_i)^T (Ra_i + t - b_i) \quad (1)$$

The inputs are the two point sets from the MRI and the stereo camera. The initial estimation of the transformation to fit the MRI data to the stereo camera data is the output, which is a refined result of transformation with rotation R and translation t to fit them together as in the following steps:

1. Associate points by the nearest neighbor criteria.
2. Estimate transformation parameters using a mean square cost function (1).
3. Transform the points using the estimated parameters.
4. Re-associate the points and iterate.

4 Head Tracking

Initial alignment gives the 3D geometrical relation between the live face and the brain in the MRI. Transformation between the initial and current faces virtually superimposes the MRI brain onto the current physical brain in the live patient's head. Our head tracking to find the 3D transform of the face data is based on Matsumoto's method [9]. We selected the corners of the eyes, nose, head, and corners of the mouth as clear image features in the patient's face instead of putting in artificial markers. The criterion of the difference Equation (2)) to be minimized is similar to that of ICP, but computation is much simpler, since the corresponding pairs of the points are known and the number of points is small.

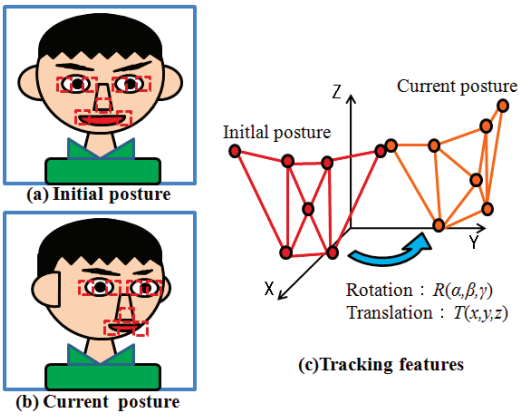


Figure 4: Head Tracking based on Face Features

$$J = \sum_{i=1}^{M} w_i (Rx_i + t - y_i)^T (Rx_i + t - y_i), \tag{2}$$

where, N is the number of the tracking points ($N = 7$ in this paper), x_i is the initial 3D coordinates for the feature i, y_i is the current 3D coordinate for the feature and w_i is the weight factor for each corresponding pair. The searching process for finding corresponding points of the 7-face features is based on template matching, in which a small image region is defined as a template for each feature point in the initially captured frame. We use normalized correlation coefficient (NCC) for the matching criterion and the weight factor of w_i as in Matsumoto's method. NCC works as a reliability of the corresponding points and thus the higher w_i gives higher priority to the more reliable corresponding features.

$$w_i = \frac{\sum_{v=1}^{M}\sum_{u=1}^{N} \phi(u,v)\varphi(u,v)}{\sqrt{\sum_{v=1}^{M}\sum_{u=1}^{N} \phi(u,v)^2 \sum_{v=1}^{M}\sum_{u=1}^{N} \varphi(u,v)^2}}, \tag{3}$$

$$\phi(u,v) = I(u,v) - \frac{\sum_{v=1}^{M}\sum_{u=1}^{N} I(u,v))}{NM} \tag{4}$$

$$\varphi(u,v) = T_i(u,v) - \frac{\sum_{v=1}^{M}\sum_{u=1}^{N} T_i(u,v))}{NM} \tag{5}$$

where, the current frame image is I, the template image is T_i (size:$M \times N$).

For effectively narrowing the feature searching area, we apply face- tracking on the frame before the template matching. We use a combination of a Haar-Like feature [7] and an Adaboost training algorithm [14] implemented in OpenCV library [15].

5 Coil Tracking

Since the stimulation coil is specialized for medical use, the coil can be maintained with a specific marker installed for the curing process (Figure 5). Detecting the marker from both images of the captured frame pair by the stereo camera, corner points of the marker can be easily specified and triangulated for acquiring a 3D coordinate to monitor the coil position and orientation. Image processing for the marker detection is as follows (see also Figure 6):

1. Make binary images from the captured frames.
2. Labeling the dark closed areas.
3. Select rectangles by counting the corners for each closed area.

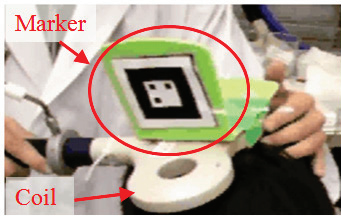


Figure 5: Coil with a Marker

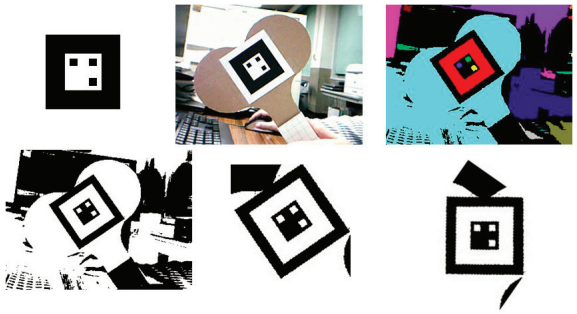


Figure 6: Marker recognition process example

4. Compare the registered pattern and the regularized rectangles to recognize the marker.

The implementation of AR toolkit has a similar process for detecting and recognizing registered binary markers [4]. We use some of the library codes in the AR toolkit for low-level image processing whose outputs correspond to the points pairs which are used for further triangulation by the stereo camera in our system.

6 Brain Appearance Model

To help navigate the doctor in handling the coil to target the specific spot, we must describe the live spatial relation of the coil and the patient's brain. We prepared a 3D model of the brain on which the current coil position is mapped on for navigation display. We used software MRIcro [13] to convert the MRI data in DICOM format to a set of intersectional images of a bitmap format. The easiest way to show the 3D brain is volume rendering with a colored point cloud, directly using the pixels of the bitmap intersections. However, the interest region for rTMS is the surface with a folded or wrinkled pattern, which shows brain mapping to locate the stimulation spot, depending on the pain region, for example. Volume rendering is not suitable for our visual context, because internal information not only is unnecessary, but also interferes with surface pattern visibility. Therefore, we create a polygon model of the surface shape with proper texture of the wrinkled pattern of the brain.

6.1 Texture Extraction

First, we manually extract the brain region in each intersection image, excluding the region of skin, bones, spinal fluid and etc. Secondly, Setting a polar coordinate with its origin at the center of the brain, the Cartesian coordinate (x,y,z) of every point from the intersectional image pixels can be mapped onto 2-dimensional angular space $\phi - \theta$, according to the following equations;

$$r = \sqrt{x^2 + y^2 + z^2} \tag{6}$$

$$\phi = cos^{-1} \frac{y}{\sqrt{x^2 + y^2 + z^2}} \tag{7}$$

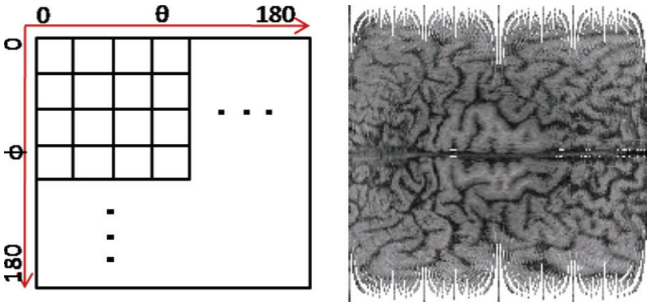


Figure 7: Brain Texture and its Coordinate

$$\theta = cos^{-1}\frac{x}{\sqrt{x^2+z^2}}, \quad (8)$$

where, the parameter r is the distance from the origin. The point with the largest r is the most distant from the center, namely the most nearest from the brain surface among the points in the same orientation bin of (ϕ, θ). Finally, a 2-dimensional 180×180 array is prepared to contain the color of the point which is the nearest from surface at the (ϕ, θ). This array can be used for the surface color pattern texture as shown in Figure 7(right).

6.2 Surface Mesh

To generate a surface of the brain, we utilize the contours of the original intersectional images in the MRI data. A rough approximation of the each contour with a polygon is used, and the vertices can be mapped onto the $\phi - \theta$ plane and re-sampled to be contained in another 2-dimensional 180×180 array as shown in Figure 8(left). Using 2-dimensional Delaunay triangulation, a triangle mesh network can be generated 8(middle). In this network drawing the triangle mesh with the original 3D coordinate creates the surface shape of the brain. The $\phi - \theta$ map can be used as the texture coordinate and a surface shape is rendered with the corresponding texture as shown in Figure 8(right).

A navigation display can be prepared as shown in Figure 9, integrating the brain model, the coil position and orientation.

7 Experiments

7.1 Implementation

We implemented the proposed method for constructing a prototype system, using the devices and software listed in Table 1. The stereo camera was assembled with two cameras that have an IEEE1394 bus connection. For the stereo matching process, we applied hardware programming with CUDA, GPU (graphical processing unit) architecture for parallel processing using software utilities by NVIDIA[11]. In this implementation, the highest computation cost in searching for correspondence between a stereo image pair is assigned to the CUDA processing with C-based coding.

We used 64 threads to compute the sum of squared differences (SSD) for every 11×11 block-size window in the left camera image to search the corresponding area in right-camera image. The single thread computes the SSD for one column and the SSD of the both sides of the column are summed up to get a total block SSD and search the minimum SSD position effectively.

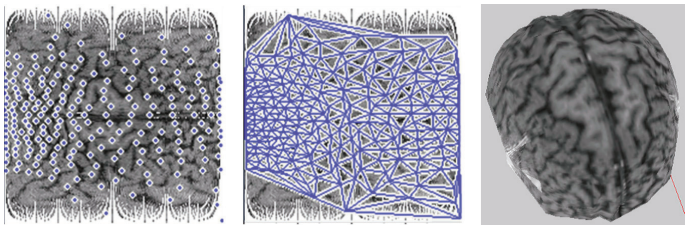


Figure 8: Brain Surface Model

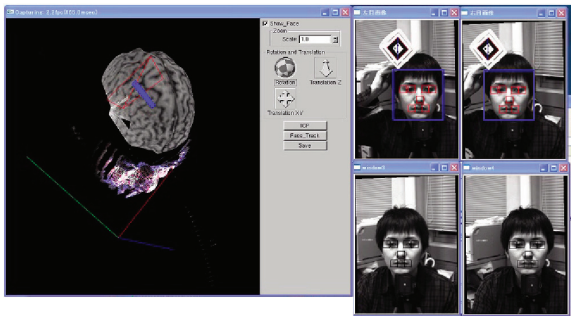


Figure 9: Navigation Display: The blue line shows the current coil position and orientation(left) and the marker detection and the face tracking are processed from the same image pair in parallel(right)

Table 1: Devices and Softwares for Prototyping

DLP Projector	TAXAN U6-232, Kaga Component Inc. 2500 lm 1024 x 768 pix
Camera	Dragonfly Express, Point Grey Research Inc. 640 x 480 pix
PC	CPU: Core2 Quad 2.66 GHz 3.0 GB RAM
Video Card	Quadro FX1700, Nvidia Inc.
Library	MS Visual studio on windows XP CUDA, OpenCV, OpenGL, ARToolkit, VTK[6]

To capture the dense point cloud of the face shape for the initial alignment, we prepared a projector just behind the stereo camera to project a random dot pattern, which is created by dividing the VGA screen into 16×12 of 64 pixel $\times$ 64 pixel blocks. We randomly put a 2×2 pixel size dot within each block.

7.2 Positioning Experiments

To clarify the result of the initial alignment, intersections are set up as shown in Figure 11.The red lines show the point clouds recorded by the stereo camera. The maximum residual displacement between the MRI surface and the point cloud from the stereo camera was 6.0 mm (Figure 11).

Using a human-like head model and a protractor (Figure 12), we investigated the tracking error. As shown in Figure 10, by assuming that the display monitor faces right in front of the patient and the stereo camera looks down the patient's face, the camera and the projector are set up. While the head model is rotated for every 2.5 degrees within $\pm$ 12.5 degrees range, the estimated tracked position by the proposed system and the directly measured positions were compared (Figure 13). For the tracking coil, the average error was 1.1 degrees rotation around each axis and 5.0 mm translation along each axis. Considering the target spot size is about 10 mm in diameter[3], achieved performance is feasible, but may require interactive trial-and-error to reach the proper spot.

8 Conclusion

This paper proposed a navigation system for visualizing the rTMS target by virtually superimposing MRI volume data onto a live patient's head position, using image measurement for 3D tracking in realtime. The combination of markerless tracking for the patient's

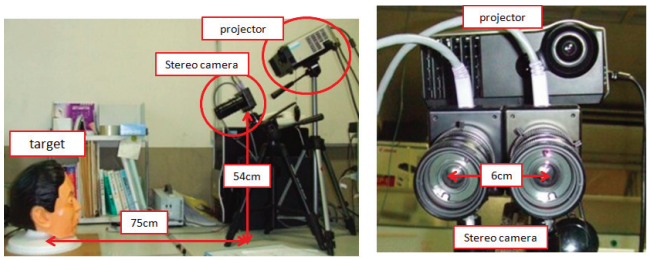


Figure 10: Stereo Camera Setup

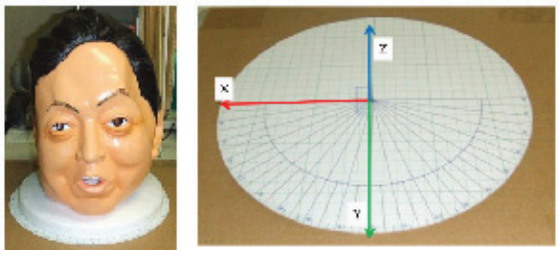


Figure 12: Experimental target

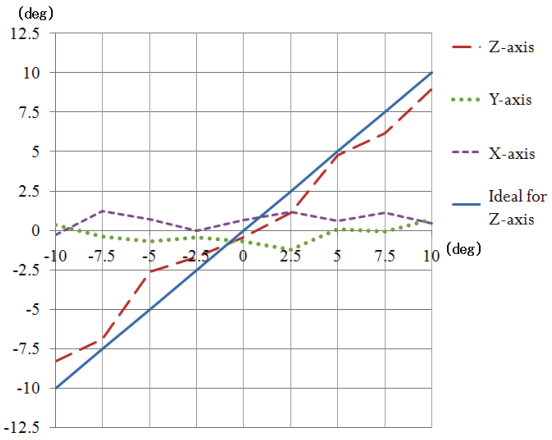


Figure 13: Tracking Error for Rotational Movement

head and marker tracking for the coil by an identical stereo camera showed good potential and feasibility for constraints-free rTMS treatment.

Our next step includes improving the stereo camera precision by an intensive sub-pixel process for disparity computation with a limited base-line length. Developing a user-friendly graphical user interface for not only specialized doctors but also generic doctors and patients themselves is also within our focus.

Acknowledgements

This work was supported in part by a grant from Teijin Pharma Limited.

References

[1] P. Besl and N. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:239–256, 1992.

[2] T. Fukushima, A. Nishikawa, F. Miyazaki, K. Uchida, M. Sekino, and Y. Saitoh. Magnetic guided transcranial magnetic stimulation: Newly convenient navigation system. In *the 24th International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS2010)*, volume 5(supplement 1), pages S36–S37, June 2010.

[3] A. Hirayama, Y. Saitoh, K. H, T. Shimokawa, S. Oshino, M. Hirata, A. Kato, and T. Yoshimine. Reduction of intractable deafferentation pain by navigation-guided repetitive transcranial magnetic stimulation of the primary motor cortex. In *Pain*, volume 122, pages 22–27, 1-2 2006.

[4] H. Kato and M. Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality*, pages 85–95, Washington, DC, USA, 1999. IEEE Computer Society.

[5] G. Kindlmann. Semi-automatic generation of transfer functions for direct volume rendering. Master's thesis, Cornell University, 1999.

[6] Kitware, Inc. *The Visualization Toolkit User's Guide*, January 2003.

[7] R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. In *IEEE ICIP 2002*, pages 900–903, 2002.

[8] The Magstim, Inc. *Magstim http://www.magstim.com/.*

[9] Y. Matsumoto, J. Heinzmann, and A. Zelinsky. The essential components of human-friendly robot systems. In *in Int. Conference on Field and Service Robotics*, pages 43–51, 1999.

[10] NDI, Inc. *POLARIS, http://www.ndigital.com/medical/polarisfamily.php/.*

[11] Nvidia, Inc. *CUDA http://developer.nvidia.com/category/zone/cuda-zone/x.*

[12] T. Paus. Imaging the brain before, during, and after transcranial magnetic simulation. *Neuropsychologia*, 37,2:219–224, 2001.

[13] Principal Investigator; Neuropsychology Lab, Atlanta GA, USA. *MRIcro http://www.cabiatl.com/mricro/.*

[14] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:511, 2001.

[15] Willowgarage. *OpenCV; http://opencv.willowgarage.com/.*

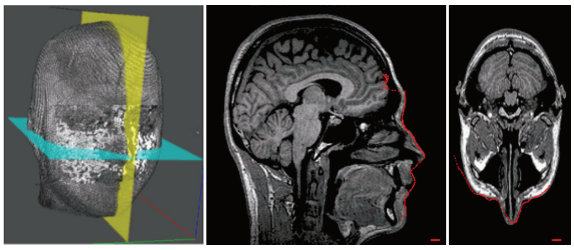


Figure 11: Initial Alignment Result: Red lines show the point cloud by the stereo camera.

Interactive Cutting Method for Physics-based Electrosurgery Simulation

Yoshihiro Kuroda* Osaka University
Shota Tanaka† Osaka University
Masataka Imura‡ Osaka University
Osamu Oshiro§ Osaka University

Abstract

Virtual reality-based electrosurgical simulators are demanded for training of the major skills in laparoscopic surgery to reduce complications. However, foregoing studies have never proposed the physics-based real-time electrosurgery simulation, because of the complexity of the phenomena and computational costs. The aim of this study is to construct an interactive surgical simulator with a unified physics-based modeling of whole processes of electrosurgery. In this paper, we proposed an interactive simulation method of physics-based electrosurgical cutting. Especially, pre-processing independent of contact information enabled simulation with user's interactive manipulation by reducing real-time calculation. The results of simulation showed that the combination of the proposed method and parallelization reduced calculation time for electric potential in the real-time process by 59.7%. The proposed method enabled interactive electrosurgical cutting with a unified physics-based modeling of whole processes of electrosurgery.

Keywords: virtual reality, medical information systems, finite element methods.

Index Terms: I.6.3 [Simulation and Modeling]: Applications; C.3 [Special-purpose and Application-based Systems]: Real-time and embedded systems—

1 Introduction

The recent progress of computer technologies has enabled an interactive physics simulator that responds to a user's manipulation and displays visual and haptic information. In clinical medicine, electrosurgery is a fundamental operation, and over 90 percent of Minimally Invasive Surgeries (MIS) utilize electrosurgery [6]. Electrosurgery consists of a series of physical phases: electrical, thermal, and structural phases, to cut soft tissue, as shown in Fig. 1. However, the foregoing electrosurgery simulators ignored underlying physical phenomena, due to their complexity and required update rate (graphics:>30Hz, haptics:>300Hz). The simulator removed the material when the temperature reached 100 °C [8, 3, 4]. The interactive electrosurgical cutting simulation has not been reported on yet.

The aim of this study is to construct an interactive surgical simulator with a unified physics-based modeling of the whole processes of electrosurgery. We have already reported on the simulation models, and found the similarity of the temperature change in electrosurgical cutting between the simulation results and real porcine livers [7]. In this paper, we propose an interactive simulation method of physics-based electrosurgical cutting. In particular, pre-processing, independent of contact information, reduces real-time calculations and enables an interactive simulation.

*e-mail: ykuroda@bpe.es.osaka-u.ac.jp
†e-mail: s-tanaka@bpe.es.osaka-u.ac.jp
‡e-mail: imura@bpe.es.osaka-u.ac.jp
§e-mail: oshiro@bpe.es.osaka-u.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

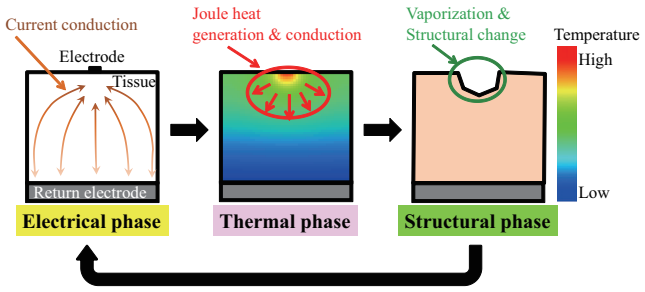


Figure 1: A series of physical phases in electrosurgery

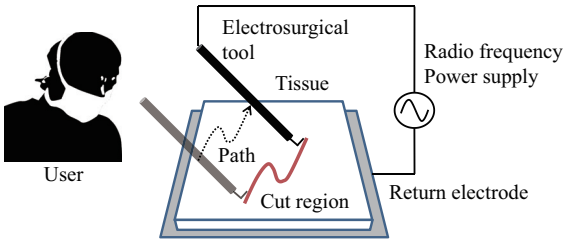


Figure 2: Electrosurgical unit

2 Physics-based Modeling of Electrosurgery

2.1 Electrosurgical unit

An electrosurgical unit is a device for cutting soft tissue by Joule heat caused by the current of high frequency, as shown in Fig. 2. The electrosurgical tool collides with tissue at a small contact area, whereas the return electrode is in contact with a large area of tissue. The electrical density around the contact area of the tool becomes high. Vaporization caused by resistive heating is thought to destruct biological tissue, although biophysical mechanisms in electrosurgery remain under investigation [2].

2.2 Electrical and thermal phases

The electrical phase determines the current density distribution caused by the contact between the electrosurgical tool and the object. The electric potential of the return electrode and the electrosurgical tool are given as boundary conditions. Although the return electrode is fixed on the tissue, the electrosurgical tool is moved, and thus, the contact area changes interactively. The thermal phase determines the temperature distribution caused by Joule heating and temperature conduction. The governing equations of electric conduction and heat transfer are given by Laplace and heat equations.

$$\nabla^2 V = 0 \tag{1}$$

$$c\rho \frac{\partial T}{\partial t} = \lambda \nabla^2 T + \mathbf{J} \cdot \mathbf{E} \tag{2}$$

where V is voltage, T is absolute temperature, $\mathbf{J}$ is current density, $\mathbf{E}$ is electric field, and c, ρ, λ are specific heat, density, and thermal conductivity, respectively. Eq. 1 gives the electric potential distribution. Eq. 2 calculates the temperature distribution.

2.3 Structural phase

It is reported that material destruction in electrosurgery is caused from mechanical rupture [1]. The structural phase determines the structural change caused by vaporization and stress concentration. The stress is derived from the expansion of the volume caused by water vaporization. The volume expansion at the position $\mathbf{r}$ at the time t is given by

$$\triangle v(\mathbf{r},t) = \triangle v_{liq}(\mathbf{r},t) + \triangle v_{gas}(\mathbf{r},t) \tag{3}$$

where $\triangle v_{liq}(\mathbf{r},t), \triangle v_{gas}(\mathbf{r},t)$ are the expansion of water from the initial temperature and the expansion caused by the transition of water from liquid to vapor, respectively. This expansion leads deformation and the stress concentration. The equilibrium, strain-displacement relation, stress-strain relation, and constitutive equations are solved to calculate stress distribution. Finite Element Method(FEM) is used for the numerical solution. In the processes, Young modulus, Poisson's ratio, and breaking stress are considered. If the Mises stress is greater than the breaking stress, the elements in the area are removed to represent material destruction. The contact area between the tool and the tissue is updated from the moving direction of the tool, and the simulation is carried out repetitively by reconstructing the mechanical structure.

3 Interactive electrosurgical cutting simulation

3.1 Approach

This section describes a pre-processing method of electric potential. The approach is to reduce the computational cost in real-time processing by pre-processing independent of prior contact information. Fig. 3 shows the difference of the conventional and proposed methods. It can be assumed that the number of contact nodes is small, compared with the number of other nodes, i.e. non-contact nodes. Therefore, high-cost matrix operations of other nodes, such as the matrix inversion and factorization, require a large calculation time. The conventional method requires the contact information in the high-cost matrix operation, while the proposed method does not require the contact information in the high-cost matrix operation. Thus, the proposed method excludes the high-cost matrix operation from real-time processing to pre-processing.

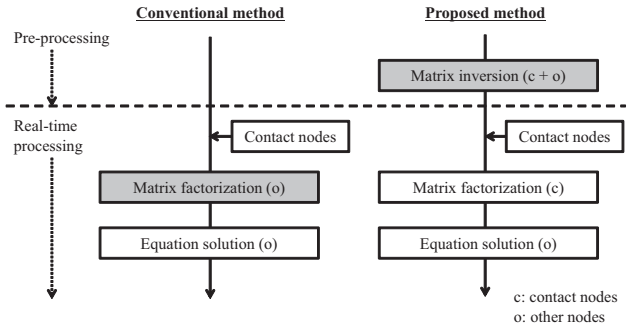


Figure 3: The difference of the conventional and proposed methods to solve a simultaneous equation for electric potential. The processes colored gray have high computational costs.

3.2 Conventional method: matrix factorization requiring prior contact information

In general, Laplace equation, as shown in Eq. 1, is solved by LU factorization, or similar matrix factorization methods. However, the contact nodes have to be known in advance. In other words, the change of contact nodes requires matrix factorization once again. Since the computational cost of matrix factorization is high, interactive manipulation that allows for the change of contact nodes becomes difficult.

First, the conventional method is described. After the contact nodes are given, the nodes are categorized into contact and other nodes, in order to solve Laplace equation,

$$\mathbf{K}\mathbf{u} = \mathbf{f} \tag{4}$$

$$\begin{pmatrix} \mathbf{K}_{oo} & \mathbf{K}_{oc} \\ \mathbf{K}_{co} & \mathbf{K}_{cc} \end{pmatrix} \begin{pmatrix} \mathbf{u}_o \\ \mathbf{u}_c \end{pmatrix} = \begin{pmatrix} \mathbf{f}_o \\ \mathbf{f}_c \end{pmatrix} \tag{5}$$

where $\mathbf{K}, \mathbf{u}, \mathbf{f}$ are the coefficient matrix, and the vectors of electric potential and electric density, respectively. The suffixes c, o represent the components of contact and other nodes, respectively. Here, $\mathbf{f}_o = \mathbf{0}$ gives:

$$\mathbf{K}_{oo}\mathbf{u}_o = -\mathbf{K}_{oc}\mathbf{u}_c \tag{6}$$

$$\mathbf{u}_o = -\mathbf{K}_{oo}^{-1}\mathbf{K}_{oc}\mathbf{u}_c \tag{7}$$

The order of Eq. 6 is the number of other nodes. As a result, the high computational cost of matrix factorization makes it difficult to achieve real-time simulation that allow for the interactive change of contact nodes.

3.3 Proposed method: matrix inversion independent of prior contact information

The proposed method inverts the whole coefficient matrix, before the contact nodes are given. This approach has the advantage of not requiring prior information regarding the contact nodes and thus reduces real-time computation through pre-processing time-consuming calculations. The process is as follows:

1. (Pre-processing): The inverse of coefficient matrix $\mathbf{L} = \mathbf{K}^{-1}$ is calculated, where $\mathbf{K}, \mathbf{u}, \mathbf{f}$ are the coefficient matrix and the vectors of electric potential and electric density, respectively.

$$\mathbf{K}\mathbf{u} = \mathbf{f} \tag{8}$$

$$\mathbf{u} = \mathbf{L}\mathbf{f} \tag{9}$$

2. (Real-time processing): Given the contact nodes between a tool and an object by interactive manipulation, $\mathbf{L}$ are arranged with the contact and other nodes. $\mathbf{f}_c$ is represented by a part of matrix L, $\mathbf{L}_{cc}$, where the suffixes c, o represent the components of contact and other nodes, respectively. Eq. 1 induces $\mathbf{f}_o = \mathbf{0}$.

$$\begin{pmatrix} \mathbf{u}_o \\ \mathbf{u}_c \end{pmatrix} = \begin{pmatrix} \mathbf{L}_{oo} & \mathbf{L}_{oc} \\ \mathbf{L}_{co} & \mathbf{L}_{cc} \end{pmatrix} \begin{pmatrix} \mathbf{f}_o \\ \mathbf{f}_c \end{pmatrix} \tag{10}$$

$$\mathbf{f}_c = \mathbf{L}_{cc}^{-1}\mathbf{u}_c \tag{11}$$

3. (Real-time processing): electric potential at non-contact nodes $\mathbf{u}_o$ is calculated.

$$\mathbf{u}_o = \mathbf{L}_{oc}\mathbf{f}_c \tag{12}$$

$$= \mathbf{L}_{oc}\mathbf{L}_{cc}^{-1}\mathbf{u}_c \tag{13}$$

Although the matrix inversion is computationally high, the order of Eq. 12 is only the number of contact nodes. Eq. 13 requires a low computation time because the matrix-vector multiplication is low cost. In summary, the first process in pre-processing is computationally high, and the second and third processes in real-time processing require low computational costs.

3.4 Process flow

The process flow of the proposed method is shown in Fig. 4. In real-time processing, the calculation of electrical-potential, temperature, and stress distribution are carried out after the collision between a virtual tool and a tissue model is detected. If a rupture is found in any element of the tissue model, the reconstruction process is launched. While the reconstruction is being processed in the back ground, the real-time processing is kept up. After the reconstruction, the time progress from the beginning of reconstruction to the end is re-calculated.

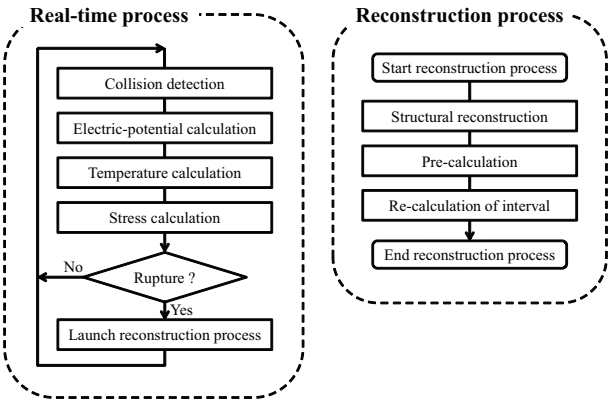


Figure 4: The process flow of electrosurgical cutting.

4 Experiments

4.1 Experimental setups

The simulation system was equipped with Intel CPU (Core i7 3.07GHz), 12GB main memory, and an nVidia GeForce GTX 580 graphics board. Intel Math Kernel Library was used for numerical solution including matrix operations. The object used in the simulation had 1013 nodes (tetrahedral mesh). The object size was 100 mm $\times$ 100 mm $\times$ 10 mm. Fig. 5 shows the shape of the object. Table 1 shows the applied physical parameters, which are derived from the studies using porcine liver [5, 9, 10]. The shape of the return electrode was a circle with a radius of 40 mm. In the simulation, the water evaporates at 100 °C, assuming that the effect of the stress on the vaporization is small enough to be ignored. The boundary conditions in the simulation are given in Table 2, where $\mathbf{n}$ is the normal vector, σ is the electric conductivity of the nodes, and λ is the heat conductivity.

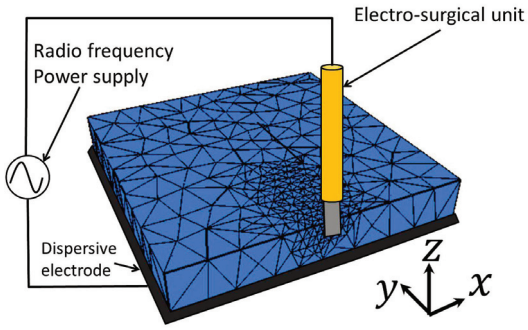


Figure 5: 3D object used in the simulation

Table 1: Main physical parameters in the simulation

Specific heat $\times$ density	$3.82 \times 10^6 \mathbf{J}/(^\circ C \cdot m^3)$
Heat conductivity	0.502 W/(m·° C)
Electric conductivity	0.144 S/m
Young modulus	4.75×10^4 Pa
Poisson's ratio	0.400
Critical stress	2.45×10^4 Pa

Table 2: Boundary conditions in the simulation

(a)Electrical phase

	Electric potential
Nodes contact with tool	V = 72 V
Nodes contact with return electrode	V = 0 V
Other boundary nodes	$\mathbf{n} \cdot (\sigma \nabla V) = 0$

(b)Thermal phase

	Temperature
Nodes contact with tool	$\mathbf{n} \cdot (\lambda \nabla T) = 0$
Nodes contact with return electrode	$\mathbf{n} \cdot (\lambda \nabla T) = 0$
Other boundary nodes	$\mathbf{n} \cdot (\lambda \nabla T) = 0$

4.2 Simulation results

Fig. 6 shows the simulation results of the calculation phases: (a) the distribution of the electric potential, (b) the distribution of the temperature, and (c) distribution of the stress. The electric potential around the contact nodes with the electrosurgical tool was high. The temperature of the nodes was increased by the Joule's heat, while the temperature was diffused in the object as time progressed. The stress increased after the vaporization of water in the elements. The elements whose stress was over the criteria were removed. The simulation was successfully continued during the reconstruction of the model. Fig. 7 shows the repetitive structure change by surgical cutting from the left side to the right side of the object.

Fig. 8 shows the interactive cutting simulation of electrosurgery. The objects used in the simulation had 643 and 1013 nodes, respectively. The contact nodes between the object and surgical tool were determined from the distance, and updated interactively. The surface was colored red, when the temperature was increased. The increased temperature is decreased gradually, because the heat diffusion is simulated based on Eq. 2.

4.3 Calculation time

Fig. 9 shows the calculation time of a series of the physics simulation. Fig. 9(a) is the calculation time of electric potential in the case of the proposed and conventional methods. The effect of parallelization of matrix operations was also examined. The result showed that the parallelization with three threads reduced 26.6% of the calculation time in the case of the conventional method. The result also showed that the proposed method reduced 45.1% of the calculation time, compared with the conventional method. The combination of the proposed method and parallelization reduced 59.7% of the calculation time in total. Fig. 9(b) shows the calculation time of each process and the total time. The result showed that the calculation time in the real-time processing was 30ms in the case of 506-noded object, so that the real-time simulation is achieved without a larger than 506-noded object. The pre-processing time was 114ms in the same simulation.

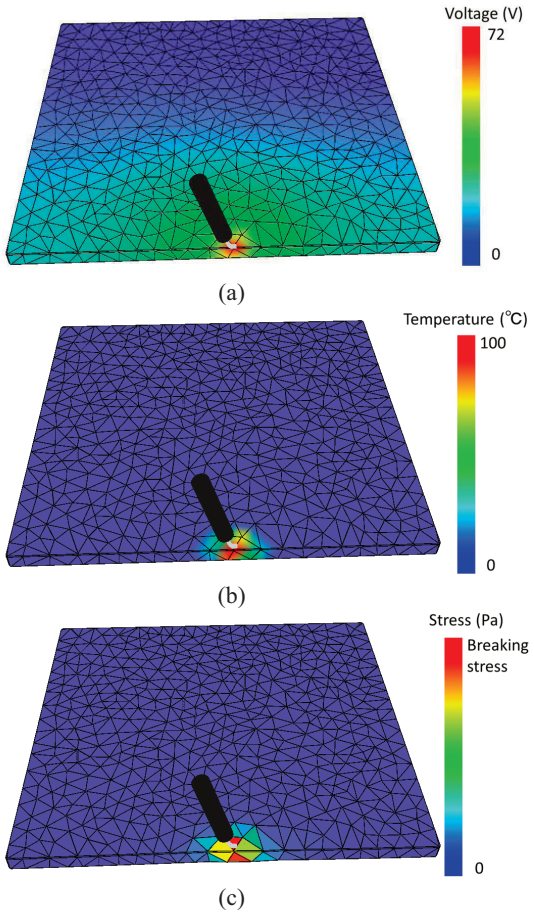


Figure 6: Simulation results: (a) the distribution of electric potential, (b) the distribution of temperature, and (c) the stress distribution

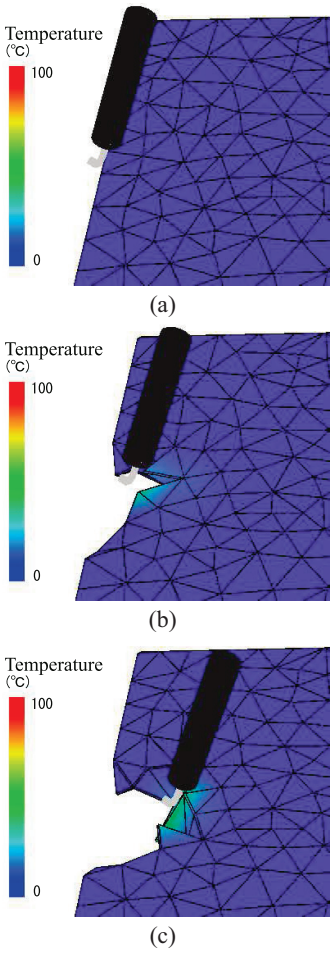


Figure 7: Simulation results of repetitive structure change in electrosurgical cutting

5 Conclusion

In this paper, we proposed an interactive simulation method of physics-based electrosurgical cutting. In particular, pre-processing, independent of contact information, reduced real-time calculations. The simulation results showed that the proposed method enabled an interactive simulation with consideration of a series of physical phases: electrical, thermal, and structural phases. In the clinical situation, it is occasional that the forces are applied on the object from the side. The simulation with the applied forces is a future work.

Acknowledgements

This study was partly supported by the Global COE Program "*in silico* medicine" at Osaka University.

References

[1] C. L. A. Ward and G. Collins. Material removal mechanisms in monopolar elecrosurgery. In *IEEE EMBC*, pages 1180–1183, 2007.

[2] E. J. Berjano. Theoretical modeling for radiofrequency ablation: state-of-the-art and challenges for the future. *Biomedical engineering online*, 5, 2006.

[3] C. Chen, M. Miga, and R. Galloway Jr. Optimizing electrode placement using finite element models in radiofrequency ablation treatment planning. *IEEE Trans Biomed Eng*, 2008.

[4] S. Cotin, H. Delingette, and N. Ayache. A hybrid elastic model allowing real-time cutting, deformations, and force feedback for surgery training and simulation. *The Visual Computer*, 16(7):437–452, 2000.

[5] C. Gabriel, S. Gabriel, and E. Corthout. The dielectric properties of biological tissues: I. literature survey. *Physics in Medicine and Biology*, 41(11):2231, 1996.

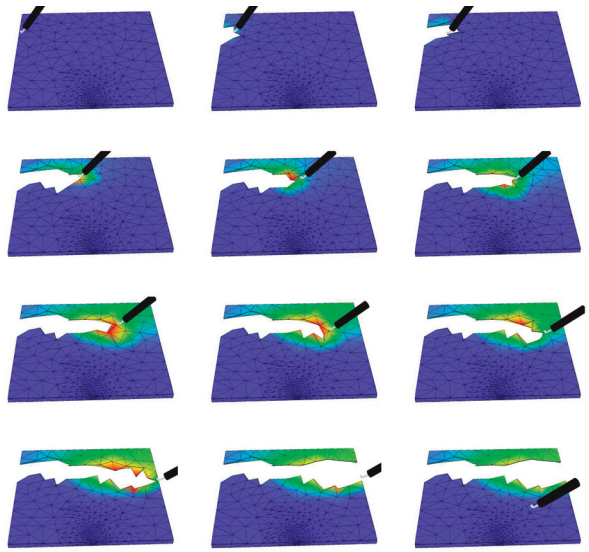
[6] G. Grinstein, D. Keim, and M. Ward. Laparoscopic electrosurgical complications. www.encision.com, October 2002.

[7] Y. Kuroda, S. Tanaka, M. Imura, and O. Oshiro. Electrical-thermal-structural coupling simulation for electrosurgery simulators. In *IEEE EMBC 2011*, 2011.

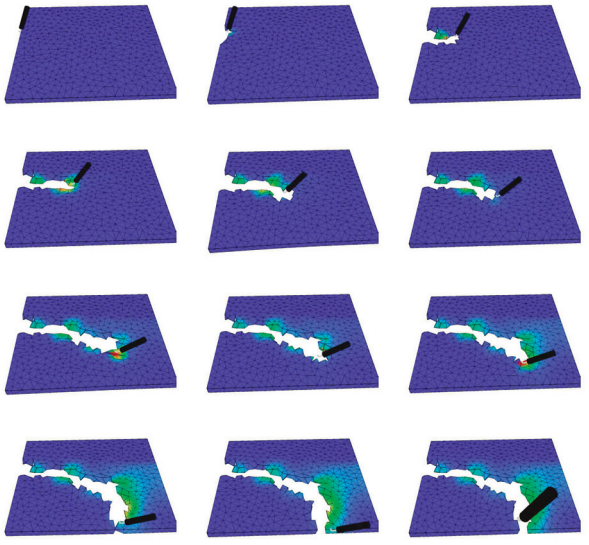
[8] A. Maciel and S. De. Physics-based real time laparoscopic electrosurgerysimulation. In *Medicine Meets Virtual Reality 16*, pages 272–274, 2008.

[9] S. H. Oh, B. I. Lee, E. J. Woo, S. Y. Lee, T.-S. Kim, O. Kwon, and J. K. Seo. Electrical conductivity images of biological tissue phantoms in mreit. *Physiological Measurement*, 26(2):S279, 2005.

[10] J. W. Valvano, J. R. Cochran, and K. R. Diller. Thermal conductivity and diffusivity of biomaterials measured with self-heated thermistors. *International Journal of Thermophysics*, 6:301–311, 1985. 10.1007/BF00522151.

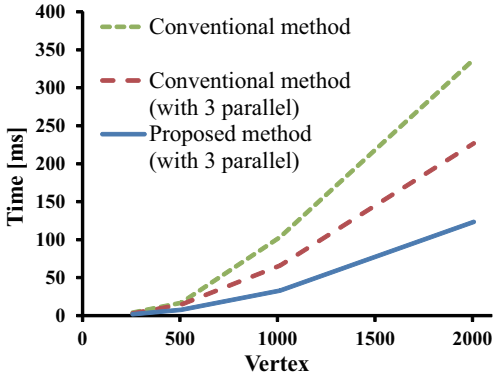


(a)

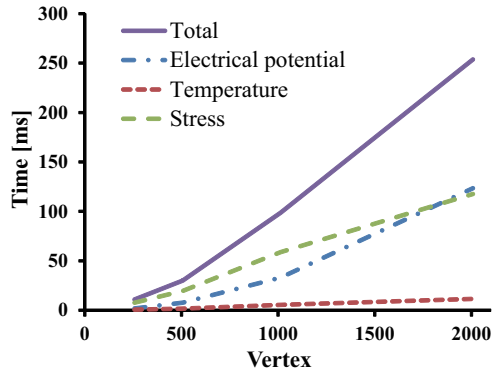


(b)

Figure 8: Interactive cutting simulation of electrosurgery: (a) 643 nodes (b) 1013 nodes



(a)



(b)

Figure 9: Calculation time of simulation: (a)electrical potential, (b)each procedure and whole procedures

Controller Reduction for Pseudo-Reference in High-Degree of Freedom Control System

Masaaki Watanabe*
Dept. of Mechanical Sciences and Engineering, Tokyo Institute of Technology, Tokyo, JAPAN

Masafumi Okada†
Dept. of Mechanical Sciences and Engineering, Tokyo Institute of Technology, Tokyo, JAPAN

Dong Dung Ngyuen‡
Dept. of Mechanical and Control Engineering, Tokyo Institute of Technology, Tokyo, JAPAN

ABSTRACT

Dance teaching, sports teaching and rehabilitation require instruction of motion from an expert to a beginner. So far, we have proposed "Pseudo-Reference" based on attractor design method. The pseudo-reference is a virtual reference which is derived from a controller of an autonomous control system and shows a target posture to execute the motion. However, for high-degree of freedom systems, because the controller has to be a high order function, it is not easy to be obtained. In this paper, we propose a controller reduction method for attractor design. Based on the correlation of motion data, the principal component analysis gives us appropriate low dimensional space of the motion. The effectiveness is evaluated by using the tap dancing robot, and pseudo-reference is applied to human motion.

Index Terms: I.2.9 [ARTIFICIAL INTELLIGENCE]: Robotics—Kinematics and dynamics; I.2.8 [ARTIFICIAL INTELLIGENCE]: Problem Solving, Control Methods, and Search—Control theory

1 INTRODUCTION

For human-human motion instruction, we sometimes use dance notation or time sequence posture variation. The dance notation was developed to hand down a traditional dance to posterity. However the dance notation is difficult to understand for beginners, because it is for experts who acquaint themselves with terpsichorean art. Moreover, for example, the time sequence posture of the long jump is illustrated in the textbook of gymnastics as shown in Figure 1. Because the figure represents only the kinematic information of the motion, some annotations are used to explain the dynamical characteristic of the motion. The instructor will say 'Jump like running up stairs' or 'Put your head forward'. We may imagine the corresponding postures, but it is difficult for beginners because the annotations include personal inspirations. The annotations frequently represent a knack of motion which is an important factor to make efficient motion. This concept is similar to human knowledge which consists of explicit and implicit knowledge[10]. For smooth communication, the appropriate representation of implicit knowledge plays an important role[4]. The kinematic postures and the athlete's annotations analogize with explicit and implicit knowledge respectively. From these considerations, the embodiment of the athlete's instinct with a posture data will contribute to the effective motion instruction.

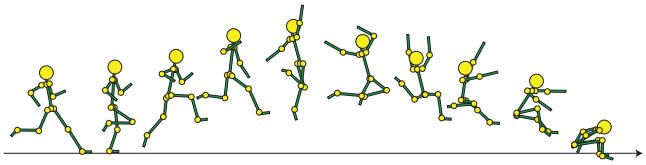


Figure 1: The time sequence posture of long jump

Some results with a same concept have been reported for robot control. Hasegawa et al.[13] and Cortesao et al.[11] divided human skill of peg-in-hole into three motions and they had the robot realize this task with appropriate selection of motion. Dordevic et al.[3] defined human skill from motion elements of experts. These methods focused on the representative motions to execute the given task effectively. Kuniyoshi et al.[15] showed a knack of motion to be obtained from a lot of measured data. Kawamura et al.[5] focused on the turning points of rotation, angular velocity and acceleration in the motion data, and defined another type of knack of motion. These methods give us an important motion key frame from dynamical point of view. In terms of annotation embodiment, we proposed a "Pseudo-reference"[6] based on modeling of human motion with autonomous controlled system. It gives us the timing and amplitude of the input torque which represents a knack of motion. In this method, the human motion is modeled by an autonomous controlled system based on orbit attractor[8], and the implicit reference is embodied as a pseudo-reference from a dynamical point of view.

However, the autonomous system is difficult to be obtained for a high-degree of freedom system because of the higher order controller. Some controller reduction methods have been reported for robot control. Moore[7] focused on model reduction of minimal realization for linear system by using principal component analysis. Villemagne et al.[2] focused on controller reduction by using canonical interaction analysis for linear system. Anderson et al.[1] discussed the many approaches for controller reduction with linear state-space equation. Though these methods are applied to controllers in linear system, it is difficult to apply to a nonlinear state feedback controller.

In this paper, we propose a reduction method of nonlinear controller in attractor design for high-degree of freedom system. Based on the high correlation between joint angle data of humans[12], a state-space projection is obtained based on principal component analysis. The effectiveness of the proposed method is evaluated by using a tap dancing robot, and pseudo-reference is applied to human walking motion in sagittal plane.

2 ATTRACTOR DESIGN AND PSEUDO-REFERENCE

2.1 Controller design based on orbit attractor

To obtain pseudo-reference, a given motion (for example, motion capture data) is modeled by an autonomous system[8] based on attractor design method[9]. In this section, we summarize the attractor design method simply. Consider the robot dynamics represented by the following nonlinear difference equation in discrete time do-

*e-mail: watanabe@micro.mep.titech.ac.jp
†e-mail: okada@mep.titech.ac.jp
‡e-mail: nguyen.d.aa@m.titech.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

main;

$$\boldsymbol{x}[k+1] = f(\boldsymbol{x}[k]) + g(\boldsymbol{x}[k], \boldsymbol{u}[k]) \tag{1}$$

where $\boldsymbol{x} \in R^n$ is a state variable, $\boldsymbol{u} \in R^m$ is an input at time stamp k. The controller is designed by a nonlinear function of $\boldsymbol{x}$ as follows;

$$\boldsymbol{u}[k] = h(\boldsymbol{x}[k]) \tag{2}$$

so that $\boldsymbol{x}$ is entrained to a specified closed curved line Ξ as;

$$\Xi = \left[\begin{array}{cccc} \xi_1 & \xi_2 & \cdots & \xi_N \end{array} \right] \quad (\xi_{N+1} = \xi_1) \tag{3}$$

Ξ is assumed to be realizable which means there exists an input sequence that realizes motion Ξ for the dynamics. The block diagram of the autonomous system using equation (1) and (2) is shown in Figure 2. The robot realizes the motion Ξ without external input.

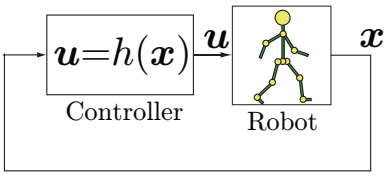


Figure 2: Autonomous control system

For a nonlinear function in equation (2), a ℓ-th order polynomial of $\boldsymbol{x}$ as;

$$\begin{aligned} \boldsymbol{u} &= a_0 + a_1\boldsymbol{x} + a_2\boldsymbol{x}^2 + \cdots + a_\ell \boldsymbol{x}^\ell & (4)\\ &= \Theta\phi(\boldsymbol{x}) & (5)\\ \Theta &= \left[\begin{array}{cccc} a_0 & a_1 & \cdots & a_\ell \end{array} \right] & (6)\\ \phi(\boldsymbol{x}) &= \left[\begin{array}{ccccc} 1 & \boldsymbol{x}^T & \boldsymbol{x}^{2T} & \cdots & \boldsymbol{x}^{\ell T} \end{array} \right]^T & (7) \end{aligned}$$

is utilized. a_i $(i = 1, \cdots, \ell)$ and Θ are coefficient matrices of the polynomial function, and $\phi(\boldsymbol{x})$ expands $\boldsymbol{x}$ to the power vector. For example, $\boldsymbol{x} \in R^3$ and $j = 2$ defines $\boldsymbol{x}^j$ as;

$$\boldsymbol{x}^2 = \left[\begin{array}{cccccc} x_1^2 & x_1x_2 & x_1x_3 & x_2^2 & x_2x_3 & x_3^2 \end{array} \right]^T \tag{8}$$

By setting realizable pairs of $(\boldsymbol{u}_i, \boldsymbol{x}_i)$, Θ is obtained by the least mean square approximation as;

$$\begin{aligned} \Theta &= \boldsymbol{U}\Phi^{\#} & (9)\\ \boldsymbol{U} &= \left[\begin{array}{cccc} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_N \end{array} \right] & (10)\\ \Phi &= \left[\begin{array}{cccc} \phi(\boldsymbol{x}_1) & \phi(\boldsymbol{x}_2) & \cdots & \phi(\boldsymbol{x}_N) \end{array} \right] & (11) \end{aligned}$$

where $[\,\cdot\,]^{\#}$ means a pseudo inverse matrix defined by;

$$\Phi^{\#} = \Phi^T \left(\Phi\Phi^T\right)^{-1} \tag{12}$$

2.2 Pseudo-reference design

In this section, we summarize the pseudo-reference design method[6]. The closed loop system in Figure 2 does not have any external input. By using the decomposition of the controller, the system of Figure 2 is changed into that of Figure 3 which has a virtual reference $\boldsymbol{x}^{ref}$ inside the controller. By considering the comparison between autonomous system and linear controlled systems, $h(\boldsymbol{x})$ is decomposed. The state variable $\boldsymbol{x}$ converges to Ξ by the attractor design as $k \to \infty$. On the other hand, there are two methods which realize $\boldsymbol{x} = \xi$. One is a two degree of freedom control system (model matching) [14] as shown in Figure 4. P is a plant, P_m^{-1} is a inverse dynamical model of P, K is a feedback controller that

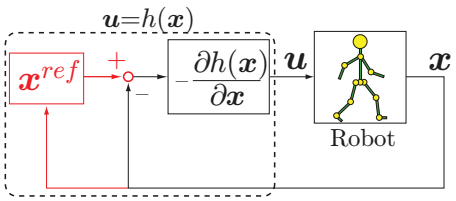


Figure 3: Robot control system using the pseudo-reference

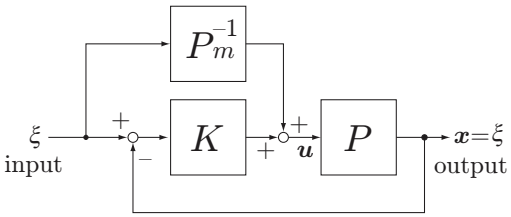


Figure 4: Two DOF model matching control system

stabilizes P and $\boldsymbol{u}$ is an input to P. Because the transfer function from the input to the output of the closed loop system is equal to 1 with the assumption $P_m = P$, $\boldsymbol{x}$ converges to ξ as $k \to \infty$. In this system, $\boldsymbol{u}$ is obtained by;

$$\boldsymbol{u} = P_m^{-1}\xi + K(\xi - \boldsymbol{x}) \tag{13}$$

The other method, which realizes $\boldsymbol{x} = \xi$, is appropriate selection of reference in the one degree of freedom control system as shown in Figure 5. K is the same feedback controller in Figure 4, and $\boldsymbol{x}^{ref}$ is the reference motion pattern. By setting $\boldsymbol{x}^{ref}$ as;

$$\boldsymbol{x}^{ref} = \frac{1 + PK}{PK}\xi \tag{14}$$

$\boldsymbol{x}$ converges to ξ as $k \to \infty$. In this system, $\boldsymbol{u}$ is represented as;

$$\boldsymbol{u} = K(\boldsymbol{x}^{ref} - \boldsymbol{x}) \tag{15}$$

The first order Taylor expansion of equation (2) around $\boldsymbol{x}$ using $\xi = \boldsymbol{x} + \delta$ $(\delta \ll 1)$ gives the following equation.

$$\boldsymbol{u} = h(\xi) - \frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}}(\xi - \boldsymbol{x}) \tag{16}$$

$\partial h/\partial \boldsymbol{x}$ is differential of $h(\boldsymbol{x})$ with respect to the each element of $\boldsymbol{x}$. For example, $\boldsymbol{x} \in R^3$ defines $\partial h/\partial \boldsymbol{x}$ as;

$$\frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}} = \left[\begin{array}{ccc} \frac{\partial h(\boldsymbol{x})}{\partial x_1} & \frac{\partial h(\boldsymbol{x})}{\partial x_2} & \frac{\partial h(\boldsymbol{x})}{\partial x_3} \end{array} \right]^T \tag{17}$$

By comparing equation (13) and (16), because the first term is concerned to ξ and the second term is concerned to $\xi - \boldsymbol{x}$, we can regard

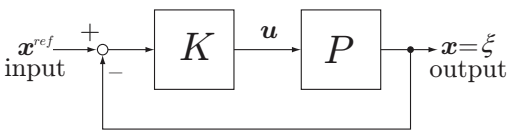


Figure 5: One DOF control system

K as;

$$K = -\frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}} \tag{18}$$

This equation means that to realize the autonomous system the controller K is a nonlinear function of $\boldsymbol{x}$. By substituting equation (18) into (15), the input $\boldsymbol{u}$ of the 1DOF feedback system is represented as;

$$\boldsymbol{u} = -\frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}}(\boldsymbol{x}^{ref} - \boldsymbol{x}) \tag{19}$$

By considering that the inputs $\boldsymbol{u}$ in equation (2) and (19) are the same, we obtain;

$$\boldsymbol{x}^{ref} = -\left(\frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}}\right)^{\#} h(\boldsymbol{x}) + \boldsymbol{x} + \left(\frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}}\right)^{\perp} \alpha \tag{20}$$

where $[\,\cdot\,]^{\perp}$ means basis of null space and $[\,\cdot\,]^{\perp}\alpha$ means an arbitrary vector that belongs to the null space. We call $\boldsymbol{x}^{ref}$ in equation (20) the pseudo-reference of the autonomous system. It is a virtual reference inside the controller, which means that an implicit reference is obtained based on the current state variable. Here we remark that $\boldsymbol{x}^{ref}$ does not always coincide with ξ because $\boldsymbol{x}^{ref}$ is obtained as a virtual reference from dynamical point of view.

3 CONTROLLER REDUCTION

3.1 Controller reduction using principal component analysis

A larger ℓ in equation (5) causes a larger number of terms in polynomial function. The number of terms L is calculated as follows;

$$L = 1 + \sum_{i=1}^{\ell} {}_nH_i = \sum_{i=0}^{\ell} \frac{(n+i-1)!}{(n-1)!i!} \tag{21}$$

where ${}_nH_i$ is a repeated combination, n and ℓ are the dimension of $\boldsymbol{x}$ and the power respectively. Table 1 shows the value of L with respect to n and ℓ. L increases rapidly according to the increase of

Table 1: The value of L with respect to n and ℓ

n \ ℓ	3	4	5	···	10
3	20	35	56		286
4	35	70	126		1001
5	56	126	252		3003
⋮					
18	1330	7315	33649		$10^7>$

n and ℓ. For example, in the case of the $n = 18$ which represents a planar human legged model, more than several thousand terms are required. High order dimension makes it difficult to design an autonomous system. Actually, calculation amount in polynomial function is as follows;

$$c(n,\ell) = m\left[\sum_{i=0}^{\ell}\left\{\frac{(n+i-1)!}{(n-1)!i!}(i+1)\right\} - 1\right] \tag{22}$$

where m is the dimension of $\boldsymbol{u}$. Then by using the Stirling's approximation, the order of $c(n,\ell)$ is required more and more calculation cost as follows;

$$O^{(c)} \approx O\left({}^m\sqrt{\frac{n+\ell-1}{(n-1)\ell}}\frac{(n+\ell-1)^{(n+\ell-1)}}{(n-1)^{n-1}\ell^{\ell}}(\ell+1)\right) \tag{23}$$

To overcome this problem, we define a new controller formulation as;

$$\begin{aligned} \boldsymbol{u} &= a_0 + a_1\boldsymbol{x} + \widehat{a}_2\widehat{\boldsymbol{x}}^2 + \cdots + \widehat{a}_\ell\widehat{\boldsymbol{x}}^\ell && (24)\\ &= \widehat{\Theta}\widehat{\phi}(\boldsymbol{x},Q) = \widehat{h}(\boldsymbol{x},Q) && (25)\\ \widehat{\phi}(\boldsymbol{x},Q) &= \left[\begin{array}{cccccc} 1 & \boldsymbol{x}^T & \widehat{\boldsymbol{x}}^{2T} & \cdots & \widehat{\boldsymbol{x}}^{\ell T} \end{array}\right]^T && (26)\end{aligned}$$

$\widehat{\boldsymbol{x}}$ is a linear projection of $\boldsymbol{x}$ as;

$$\widehat{\boldsymbol{x}} = Q\boldsymbol{x} \quad \in R^r \tag{27}$$

where $r < n$ and $Q \in R^{r\times n}$ is a constant matrix. Q is obtained by principal component analysis of Ξ in (3) as follows;

$$\begin{aligned} \Xi &= USV^T && (28)\\ &= \left[\begin{array}{cc} U_1 & U_2 \end{array}\right]\left[\begin{array}{cc} S_1 & 0 \\ 0 & S_2 \end{array}\right]\left[\begin{array}{c} V_1^T \\ V_2^T \end{array}\right] && (29)\\ S_1 &= \mathrm{diag}\{\ s_1 \quad s_2 \quad \cdots \quad s_r\ \} && (30)\\ S_2 &= \mathrm{diag}\{\ s_{r+1} \quad s_{r+2} \quad \cdots \quad s_n\ \} && (31)\end{aligned}$$

By assuming $s_r \gg s_{r+1}$, Q is obtained by;

$$Q = U_1^T \tag{32}$$

In equation (26), $\boldsymbol{x}$ is hold and more than second power of $\boldsymbol{x}$ are replaced to $\widehat{\boldsymbol{x}}$ because of the conservation of observation. Projection

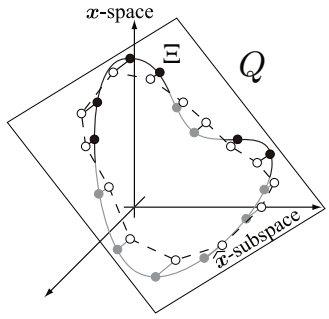


Figure 6: Projection of Ξ in $\boldsymbol{x}$-space on $\widehat{\boldsymbol{x}}$-subspace

with Q means that the most probable subspace is calculated to approximate the motion Ξ as shown in Figure 6. In the reduced controller, the value of terms $\widehat{L}$ is calculated as;

$$\widehat{L} = 1 + n + \sum_{i=2}^{\ell} {}_rH_i = 1 + n + \sum_{i=2}^{\ell} \frac{(r+i-1)!}{(r-1)!i!} \tag{33}$$

For example, in the case of $n = 18$ and $\ell = 4$, $L = 7315$ (cf. Table 1), the amount of increase of $\widehat{L}$ is smaller than that of L as shown in Table 2. Measured computational times also gradually decrease when r become smaller by using INTEL Core2Duo 2.4GHz processor for calculation. Calculation amount of reduced controller is as follows;

$$\hat{c}(n,\ell,r) = m\left[\sum_{i=2}^{\ell}\left\{\frac{(r+i-1)!}{(r-1)!i!}(i+1)\right\} + 2n - 2\right] \tag{34}$$

By using the Stirling's approximation, the order of $c(n,\ell,r)$ is calculated as follows;

$$O^{(\hat{c})} \approx O\left({}^m\sqrt{\frac{r+\ell-1}{(r-1)\ell}}\frac{(r+\ell-1)^{(r+\ell-1)}}{(r-1)^{r-1}\ell^{\ell}}(\ell+1)\right) \tag{35}$$

Table 2: The value of $\widehat{L}$ and computational time with respect to r

	n=18, ℓ=4						n
r	4	5	6	...	10	...	17	18
$\widehat{L}$	84	139	222	...	1009	...	5986	7315
time[ms]	2.2	4.5	8.1	...	53.2	...	469.9	596.3

When r is smaller than n, the order of $\hat{c}$ is sufficiently smaller than that of c. This result shows the effectiveness of the proposed reduction method.

It has been already shown that Θ is calculated by equation (9), and it requires inverse of $\Phi\Phi^T \in R^{L\times L}$. By the same way, $\widehat{\Theta}$ is obtained by;

$$\widehat{\Theta} = \boldsymbol{U}\widehat{\Phi}^{\#} \quad (36)$$

$$\widehat{\Phi} = \left[\begin{array}{cccc} \widehat{\phi}(\boldsymbol{x}_1, Q) & \widehat{\phi}(\boldsymbol{x}_2, Q) & \cdots & \widehat{\phi}(\boldsymbol{x}_N, Q) \end{array}\right] \quad (37)$$

and it requires the calculation of inverse of $\widehat{\Phi}\widehat{\Phi}^T \in R^{\widehat{L}\times\widehat{L}}$, which is much smaller than $\Phi\Phi^T$. Moreover, the pseudo-reference with the controller reduction requires calculation of pseudo inverse of $\partial\widehat{h}/\partial\boldsymbol{x} \in R^{\widehat{L}\times n}$ which is much smaller than that of $\partial h/\partial\boldsymbol{x} \in R^{L\times n}$ in equation (20).

3.2 Validation of controller reduction

3.2.1 Tap dancing robot

The proposed controller reduction is applied to the tap dancing robot which is a simple system for evaluation. The tap dancing robot is shown in Figure 7-(a) and its dynamical model is shown in Figure 7-(b). The detail of the dynamic equation of tap dancing robot is shown in [8]. It steps continuously by changing the stance

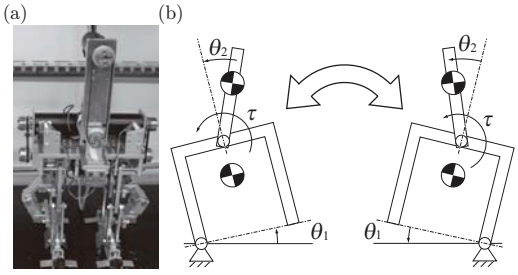


Figure 7: Tap dancing robot[8] and its dynamical model

foot and is stabilized by shaking the head. The input is the torque τ and the state variable $\boldsymbol{x}$ consists of lower body rotational angle θ_1, head rotational angle θ_2 and their angular velocities $\dot{\theta}_1$, $\dot{\theta}_2$ as follows;

$$\boldsymbol{x} = \left[\begin{array}{cc} \boldsymbol{x}_p^T & \boldsymbol{x}_v^T \end{array}\right]^T \in R^4 \quad (38)$$

$$\boldsymbol{x}_p = \left[\begin{array}{cc} \theta_1 & \theta_2 \end{array}\right]^T, \boldsymbol{x}_v = \left[\begin{array}{cc} \dot{\theta}_1 & \dot{\theta}_2 \end{array}\right]^T$$

Here we assume that the impact of foot on the ground is completely inelastic collision. The referenced motion Ξ in equation (3) is generated by giving step reference to PD controlled θ_2. The robot moves dynamically but unstably. By clipping one cycle motion, Ξ is obtained and this motion is modeled by the attractor design method.

Because $n = 4$ in this robot, L and $\widehat{L}$ are calculated for different orders of polynomial ℓ as shown in Table 3. $r = 3$ is utilized for controller reduction. $\ell = 2$ dose not stabilize the robot by both original and reduced controller (which is indicated by "failed" in Table 3). This result show that controller reduction method requires only 21 terms which is smaller value than 35 terms to stabilize the robot by using the conventional method. Therefore, $\ell = 4$ is used

Table 3: L and $\widehat{L}$ of polynomial function

The order of polynomial ℓ	2	3	4
The value of L (Original controller $n = 4$)	15 failed	35	70
The value of $\widehat{L}$ (Reduced controller $n = 4$, $r = 3$)	11 failed	21	36

to design the pseudo-reference in the following.

The pseudo-reference is designed by using the conventional controller and the reduced controller. Because $\boldsymbol{x}^{ref}$ is not decided uniquely as shown in equation (20), the objective function J_t is set as;

$$J_t = \left\|W_1(\boldsymbol{x}_p^{ref} - \boldsymbol{x}_p)\right\|^2 + \left\|W_2\boldsymbol{x}_v^{ref}\right\|^2 \quad (39)$$

where W_1 and W_2 are weighting matrices. The first term aims at making the distance small between $\boldsymbol{x}_p^{ref}$ and $\boldsymbol{x}_p$ to avoid a radical variation of $\boldsymbol{x}^{ref}$. The second term aims at making $\boldsymbol{x}_v^{ref}$ small so that the pseudo-reference shows posture information with small velocity. Figure 8 and 9 show the motion of the tap dancing motion with the conventional controller and the reduced controller respectively. (a) shows the locus of $\boldsymbol{x}$ with a solid line and the pseudo-reference with dots respectively in θ_1, $\dot{\theta}_1$ and θ_2 space. (b) shows the motion of the tap dancing robot. Indicated numbers in (b) correspond to those in (a). From these results, we can see that the

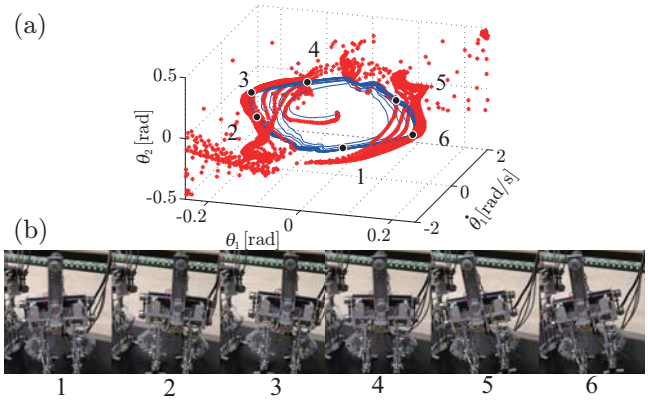


Figure 8: Tap dancing motion and pseudo-reference with original controller

stable motion is realized with the reduced controller, and pseudo-references are designed similarly with the conventional method.

4 PSEUDO-REFERENCE FOR HIGH-DEGREE OF FREEDOM SYSTEM

4.1 Human legged model

In this section, pseudo-reference is applied to human legged motion using the controller reduction. Human legged model is shown in Figure 10. The model is a high-degree of freedom model in

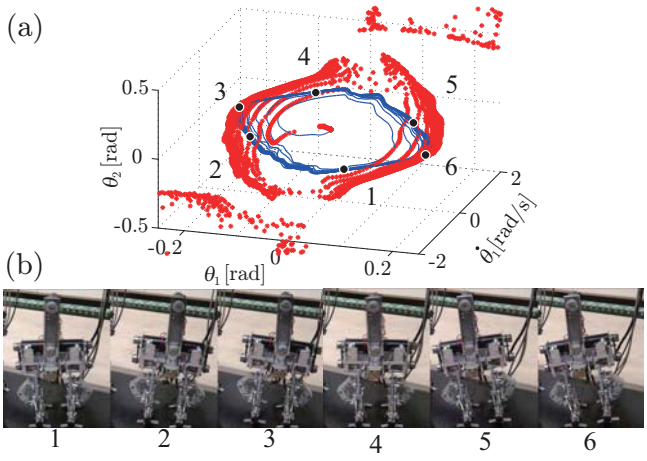


Figure 9: Tap dancing motion and pseudo-reference with controller reduction

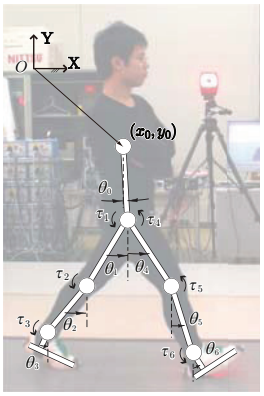


Figure 10: High-degree of freedom model for human legged motion

sagittal plane. XY coordinates are absolute coordinates, (x_0, y_0) represents the center of mass of the upper body and θ_i, τ_i $(i = 0, 1, 2, 3, 4, 5, 6)$ represent joint angles and joint torques respectively. The state variable $\boldsymbol{x}$ consists of x_0, y_0, θ_i and their velocities $\dot{x}_0$, $\dot{y}_0$, $\dot{\theta}_i$ as follows;

$$\boldsymbol{x} = \left[\begin{array}{cc} \boldsymbol{x}_p^T & \boldsymbol{x}_v^T \end{array} \right]^T \in R^{18} \quad (40)$$

$$\boldsymbol{x}_p = \left[\begin{array}{ccccc} x_0 & y_0 & \theta_0 & \cdots & \theta_6 \end{array} \right]^T \quad (41)$$

$$\boldsymbol{x}_v = \left[\begin{array}{ccccc} \dot{x}_0 & \dot{y}_0 & \dot{\theta}_0 & \cdots & \dot{\theta}_6 \end{array} \right]^T \quad (42)$$

The input $\boldsymbol{u}$ consists of each joint torque as follows;

$$\boldsymbol{u} = \left[\begin{array}{cccc} \tau_1 & \tau_2 & \cdots & \tau_6 \end{array} \right]^T \in R^6 \quad (43)$$

The controller is designed by using $\bar{\boldsymbol{x}}$ as follows;

$$\boldsymbol{u} = \widehat{\Theta}\widehat{\phi}(\boldsymbol{x}, Q) \quad (44)$$

Human walking motion is measured by an optical motion capture system and it is projected on sagittal plane to obtain Ξ in equation (3). The linear projection Q is obtained using $r = 5$. Table 4 shows relationship between the number of terms in polynomial. The original controller (without reduction) contains 7315 terms with $n = 18$, on the other hand, the reduced controller has only 139 terms by employing $r = 5$.

Table 4: Number of terms of polynomial function

Exponential number ℓ	4
Number of terms (Original controller $n = 18$)	7315
Number of terms (Reduced controller $n = 18$, $r = 5$)	139

4.2 Pseudo-reference of walking motion

Based on the reduced controller, pseudo-references of walking motion is obtained. To determine $\boldsymbol{x}^{ref}$ in equation (20) uniquely, the following objective function J_ℓ is employed and minimized.

$$\begin{aligned} J_\ell &= \left\| W_1(\boldsymbol{x}_p^{ref}[k] - \boldsymbol{x}_p[k]) \right\|^2 \\ &+ \left\| W_2(\boldsymbol{x}_p^{ref}[k+1] - (\boldsymbol{x}_p^{ref}[k] + T\boldsymbol{x}_v^{ref}[k])) \right\|^2 \end{aligned} \quad (45)$$

where W_1 and W_2 are weighting matrices and T is sampling time. The first term is same as previous section. The second term aims at

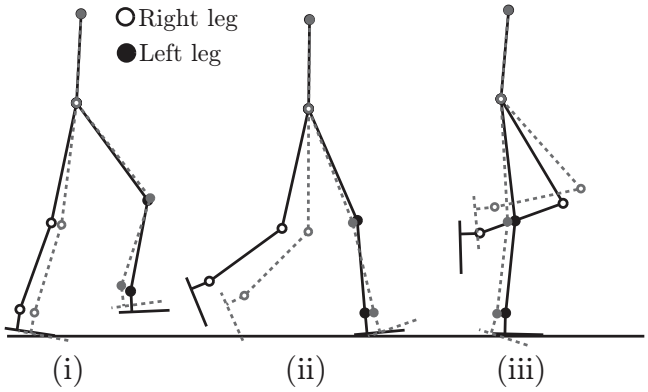


Figure 11: Human walking motion and its pseudo-reference

satisfaction of dynamic consistency between $\boldsymbol{x}_p^{ref}$ and $\boldsymbol{x}_v^{ref}$. Because of the symmetry of walking motion, the half walking motion in the case of the right supporting leg is represented in Figure 11. The measured walking motion and pseudo-reference are represented by a solid line and a dashed line respectively. The right leg joints are marked by white circles. The posture of pseudo-reference is illustrated so that its hip joint position and orientation coincides to the measured data. From this result, pseudo-reference gives us the following information for human walking.

- In (i), the kicking leg (right leg) moves forward to slow down the body speed.
- In (ii), the swing leg is moved forward strongly for the next gait. The landing leg is stretched to support the weight of the body.
- In (iii), the swing leg is continuously lifted up to prepare for the next foot landing. The ankle of support leg is bended to obtain the propulsion force.

Here we note that

(a) Unfortunately, the obtained reduced controller could not stabilize the legged system. Because Ξ obtained from a motion

capture system projecting in the sagittal plane, Ξ may not satisfy the dynamic constrain of the planner walking. Transformation of Ξ considering dynamical consistency will be required.

(b) Pseudo-reference obtained from the reduced controller should be compared to original pseudo-reference obtained from the original controller. However, the original controller is not calculated because of the shortage of computational memory because of large number of L.

5 Conclusions

In this paper, we proposed the controller reduction method to design an autonomous controlled system, and pseudo-reference is applied to human motion based on the proposed reduction method. The results of this paper are summarized as follows;

1. Controller reduction of an autonomous control system based on an orbit attractor is proposed by using correlation of motion data.

2. The proposed method is evaluated by the tap dancing robot. The same tap dancing motion and pseudo-reference are realized by using either the proposed controller or the conventional one, which means the effectiveness of the proposed reduction method.

3. The proposed method is also applied to human legged motion which is multi-degree of freedom system. The pseudo-reference showed the significant information to realize the walking motion.

The pseudo-references are designed as movie of postures which contain significant information. In the future, motion instruction for beginner will be performed by using the movie of expert's pseudo-references.

Acknowledgements

This research is supported by the "Research on Macro / Micro Modeling of Human Behavior in the Swarm and Its Control" under the Core Research for Evolutional Science and Technology (CREST) Program (Research area : Advanced Integrated Sensing Technologies), Japan Science and Technology Agency (JST).

References

[1] B.D.O.Anderson and Y.Liu. Controller reduction : Concepts and approaches. *IEEE Trans. on Automation and Control*, 34(8):802–812, 1989.

[2] C. de Villemagne and R. E. Skelton. Controller reduction using canonical interactions. *IEEE Transactions on Automatic Control*, 33(8):740 – 751, 1988.

[3] G.S.Dordevic, M.Rasic, D.Kostic, and V.Potkonjak. Representation of robot motion control skill. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS - PART C*, 30(2):219–238, May 2000.

[4] I.Nonaka and H.Takenaka. *The knowledge-creating company: how Japanese companies create the dynamics of innovation*. Oxford University Press US, 1995.

[5] Y. Kawamura and Y. Sankai. Humanoid control method based on human knack for human care service. In *IEEE International Conference on Systems, Man and Cybernetics 2002(SMC'02)*, pages TP1B4(CD–ROM), 2002.

[6] M.Okada and M.Watanabe. Pseudo-reference for motion transfer based on autonomous control system with an orbit attractor. In *Proc. of The IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems(IROS2010)*, pages pp.1297–1302, 2010.

[7] B. C. Moore. Principal component analysis in linear systems - controllability, observability, and model reduction. *IEEE Transactions on Automatic Control*, AC-26(1):17 – 32, 1981.

[8] M. Okada and K. Murakami. Robot communication principal by motion synchronization using orbit attractor. In *Proc. of the IEEE International Conference on Robotics and Automation (ICRA'07)*, pages 2564–2569(CD–ROM), 2007.

[9] M. Okada, K. Osato, and Y. Nakamura. Motion emergence of humanoid robots by an attractor design of a nonlinear dynamics. In *Proc. of the IEEE International Conference on Robotics and Automation (ICRA'05)*, pages 18–23, 2005.

[10] M. Polanyi. *The Tacit Dimension*. University of Chicago Press, 1967.

[11] R.Cortesao, R.Koeppe, U.Nunes, and G.Hirzinger. Data fusion for robotic assembly tasks based on human skills. *IEEE Transactions on Robotics*, 20(6):941–952, Dec. 2004.

[12] A. Safonova, J. K. Hodgins, and N. S. Pollard. Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. *ACM Trans. Graph.*, 23:514–521, August 2004.

[13] T.Hasegawa, T.Suehiro, and K.Takase. A model-based manipulation system with skill-based execution. *IEEE Transactions on Robotics and Automation*, 8(5):535–544, Oct. 1992.

[14] T.Sugie and T.Yoshikawa. General solution of robust tracking problem in two-degree-of-freedom control system. *IEEE Transaction on Automatic Control*, AC-31(6):552–554, June 1986.

[15] Y.Kuniyoshi, Y.Ohmura, K.Terada, A.Nagakubo, S.Eitoku, and T.Yamamoto. Embodied basis of invariant features in execution and perception of whole body dynamic actions — knacks and focuses of roll-and-rise motion. *Robotics and Autonomous Systems*, 48(4):189–201, 2004.

Modeling of Pedestrian's Unintentional Guide Using Vection and Body Sway

Norifumi Watanabe* Tamagawa University
Hiroaki Mikado† Tamagawa University
Takashi Omori‡ Tamagawa University

ABSTRACT

In daily life, our behavior is guided by various visual stimuli such as the information on direction signs. However, our environmentally-based perceptual capacity is often challenged in crowded circumstances, or more so, in emergency evacuation circumstances. In those situations, we often fail to pay attention to important signs. In order to achieve more effective direction guidance, we considered the use of unconscious reflexes in human walking action. In this study, we experimented with vision-guided walking direction control by inducing subjects ' gaze direction shift using a vection stimulus combined with body sway. In this paper, we confirm a shift in subjects ' walking direction and body sway, and discuss a possible mechanism.

Keywords: Pedestrian Navigation, Gaze Control, Vection, Body Sway

1 INTRODUCTION

In the daily act of walking, we take in large amounts of sensory information through visual, auditory, tactile, and olfactory channels, (among others), and decide how to act. An important information source in these action decisions is explicit visual information such as signs or arrows. However, in crowded situations or when avoiding danger, it is difficult to recognize relevant signs, and it may become difficult to take appropriate action[2].

The situation is similar in the artificial environment of augmented reality (AR). Although the awareness of the environment ' s artificiality may tend to keep us from visual attention, this artificial environment nonetheless stresses proper attentional allocation to signs, as compared to the more routine environment of daily life. In order to more effectively guide ambulatory behavior using AR, we considered the use of an unconscious or reflex based guidance method in addition to usual visual action signs. Galvanic Vestibular Stimulation (GVS) is reported as a method of causing unconscious walking direction guidance[1][6][3]. GVS is a produces the illusion on the sense of equilibrium by throwing a slight current on the vestibular organ. When GVS is presented to a subjective pedestrian, the subject perceives that his or her own movement is different from that intended and tries to correct for the difference unconsciously and the reflexive correction produces a change in walking direction. However, the use of GVS in human behavior guidance requires continuous application of electrical current to the vestibular organ, and may not be entirely safe.

Thus, we experimented with body sway perception using a vibration device and vection, instead of GVS. When vibration is delivered to the leg, our sense of equilibrium transfers dependency on body sway to the available visual input. Accordingly, we should expect that the self-body motion illusion by vection to be sufficient to cause a reflexive change in gaze direction and an unconscious modification of body motion direction.

It is thought that these reflexes can safely deliver the level of sensation that can cause unconscious guidance of walking direction. In this study, we examine the behavioral dynamic in a case of walking direction guidance.

2 WALKING INDUCEMENT BY GAZE CONTROL

2.1 Self-motion Sensation

The self-motion sensation is an effective method to control a pedestrian's autonomous motion. The self-motion sensation combines information from multiple sensors such as the vestibular, visual, auditory and tactile systems, and yields perception of own ' s balance, direction and motion. The vestibular and visual senses are the most basic among all those of the sensory systems.

The vestibular system perceives gravity and acceleration, and influences the faculty of equilibrium. However, the sense of self-motion by the vestibular system is temporal, in that it responds to changes in speed and acceleration. Therefore, the perception of self motion due to the vestibular sense disappears when we move at a constant speed. However, the sense of constant speed is necessary for walking as is the sense of acceleration. The sense that plays the principal role in continuous self-motion perception is the vision.

The sensation of visual self-motion is referred to as vection[5]. Vection is the perception of self-motion without actual motion produced by optical flow. When vection occurs, we correct our posture to compensate for the perceived self-motion. Yoshida et al. reported that when a standing subject perceived vection, the center of gravity inclined unconsciously to the opposite side of the perceived self-motion[8]. However, visual information alone was not sufficient to change the walking orbit but at most allowed the body incline to change with respect to the center of gravity.

It was previously thought that information from other sensations such as somatic sensation was contradictory and thereby suppressed by reflexes, just as by vision.

2.2 Self-motion Sensation Using Vision and Body Sway

Suzuki et al. reported that they had achieved changes in body postural change toward the eye-movement direction by applying vection and body sway[7]. In their experiment, the body sway was produced by applying "neck dorsal muscle stimulation" (NS) using tibialis anterior stimulation, (called TAS), and "gastrocnemius stimulation, called "GAS", using a vibration device while a subject was standing. Applying a visual stimulus, they evaluated the amount of postural change quantitatively (See Fig.1). Experimental results with TAS, NS and GAS together were able to induce the postural change towards the gaze direction. But they only evaluated the effect for the standing state and did not try the walking state.

Then, in this paper, we examine a method for inducing a change in walking direction by showing an optic flow stimulus to control the gaze direction and self-motion sensation, together with a vibratory stimulus to the body. Specifically, the vibration device attached to a subject's leg destabilizes somatic sensation and causes the body to sway. The presentation of an optic-flow stimulus causes the illusion that the body is moving in a direction opposite to that of the

*e-mail: norifumi@lab.tamagawa.ac.jp
†e-mail:mkdhi6ec@engs.tamagawa.ac.jp
‡e-mail:omori@lab.tamagawa.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

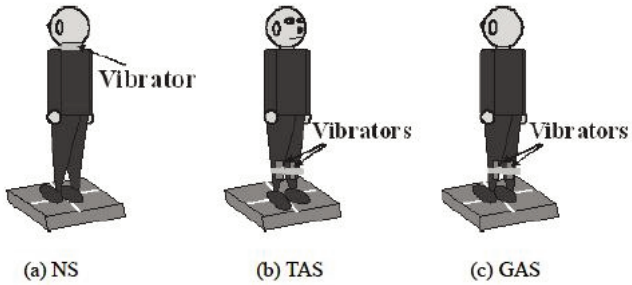


Figure 1: Vibration stimulation presentation part to induce body sway. ”NS: neck dorsal muscles stimulation”, ”TAS: tibialis anterior stimulation”, and ”GAS: gastrocnemius stimulation”. (Ueno, 2007[7])

flow, and the illusion induces the reflex of the subject's body moving in the direction of the stimulus.

3 Pedestrian Inducement Experiment Using Self-motion Sensation

To evaluate the effect of vection stimulation and body sway during walking, we used a motion capture system (Motion Analysis MAC3D). We used 12 cameras, and subjects wore 19 markers on their head (3 points), their shoulders (2 points), their elbows (2 points), their wrists (4 points), their waists (2 points), their knees (2 points), their ankles (2 points), and their large toes (2 points), as a means of measurement.

We used a handy massager ”Slive MD-01” as the vibration device and presented a high frequency (100Hz) and a low frequency (90Hz) vibration. The vibration was applied at left and right GAS, where it would not greatly affect their walking ability. (Fig.2).

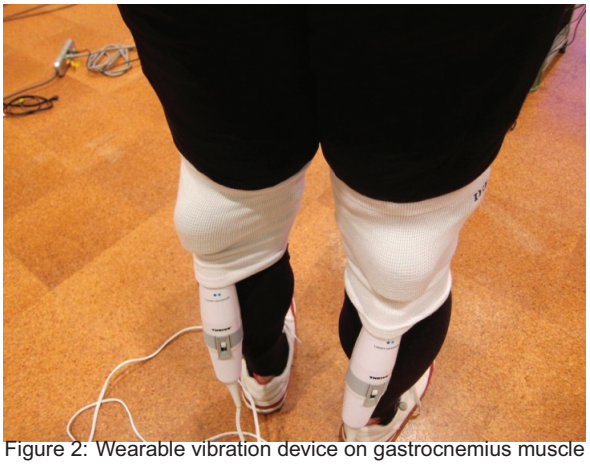


Figure 2: Wearable vibration device on gastrocnemius muscle

For the evaluation of gaze direction guidance by optical flow, we used an eye tracking system (NAC EMR-8B), and measured the subject ' s gaze from the starting position of the walking task. We used a sequence of points aligned in the side and moved it to left or right by a speed of 160mm/sec as the optical flow stimulus. The size of one point is a circle 6 cm in radius, and all the points move at constant speed. The size of screen was 2m height by 3m width (Fig.3).

Six subjects participated in our experiment. In 10 trials of the high frequency (100Hz) and low frequency (90Hz) vibration conditions, 5 right- and 5 left-direction optical flow images were presented by a PC on a screen by a projector.

In the experiment, subjects stood facing the screen, looked at a fixation point on the screen and start walking straight from a starting position. The optical flow stimulus was projected when the subject reached 1.8m from the starting position, and walked an additional 2.5m after the start of projection, while watching the screen (Fig.4). The vibration stimulus was presented simultaneously with the commencement of walking.

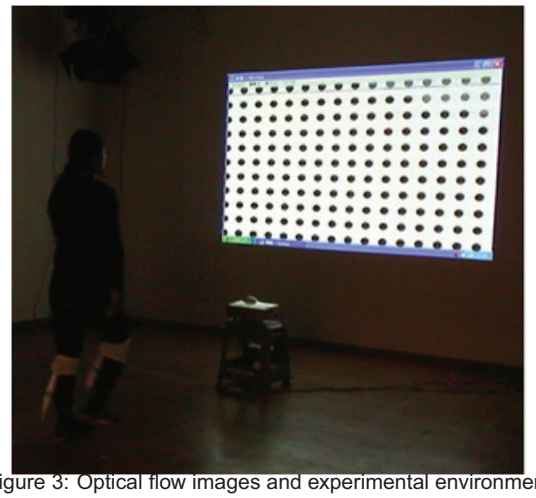


Figure 3: Optical flow images and experimental environment

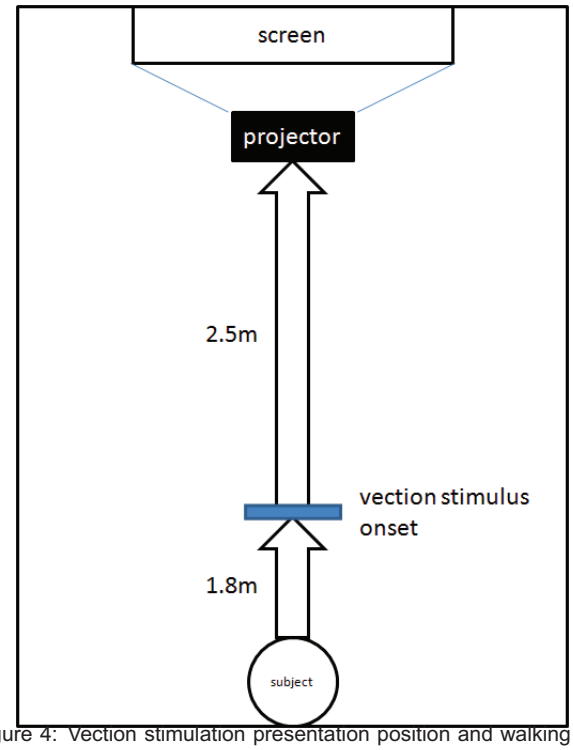


Figure 4: Vection stimulation presentation position and walking distance

3.1 Result

3.1.1 Eye Movement by Vection

Fig.5 shows the eye movement of a subject when they were presented the Optical-Flow stimulus. The X-axis shows the number of frames (30frame/sec) and the Y-axis shows the direction of the eye (upper is right direction in degrees). In the measurements taken, the eye moved to the right by right flowing stimulation and moved to the left by left-flowing stimulation. Namely, the gaze moved in the

direction of the optical flow, and it was confirmed the results that the gaze movement was induced by the vection stimulation in our experiment.

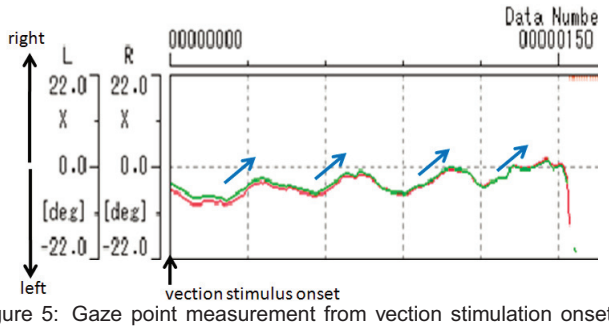


Figure 5: Gaze point measurement from vection stimulation onset to the right side of 5 second. X-axis shows the number of frames (30frame/sec) and Y-axis shows the gaze point movement (deg). Green line shows left eye trajectory and red line shows right eye trajectory.

3.1.2 Body Movement by Vection

Fig.6 shows the probability of body movement when the vibration and the optical-flow stimuli are presented. The Y-axis represents the probability that either left or right movement have occurred. It shows that the body moved with high probability towards the vection stimulation direction, independent of the vibration frequency.

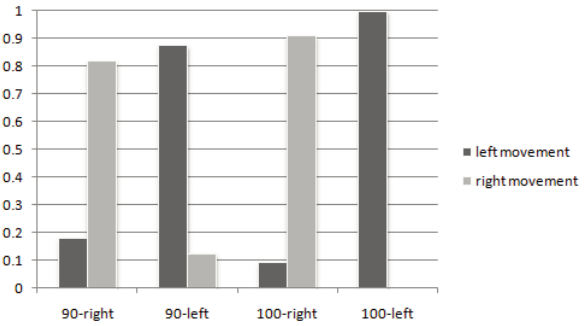


Figure 6: Body movement direction probability after vection stimulation. The X-axis shows vibration frequency and the vection stimulation direction. The Y-axis shows the ratio of left and right movement.

Fig.7 shows the trajectory of a left leg ankle for one step after vection stimulation to the left. The upper graph is a right or left shift of the ankle to forward direction and lower one shows height of the ankle. From this graph, it is confirmed that the inducement of leg movement by vection stimulation occurs at a phase of the leg lifted up in a walking cycle. A feature of the motion was that the ankle moved in a direction opposite to direction of stimulation at a first half of the leg lifting step and moved to the stimulus direction afterwards. It is thought that this opposite direction movement is a correction reaction for the illusion caused by this stimulation.

Next, figure8 shows latency in in the beginning of ankle motion from the vection stimulation onset. The X- axis shows the conditions of the vection direction and the foot from which movement first appeared. Average latency was 1.35sec, 1.4m distant from the stimulation where the walking speed was 1.05m/s. We can say that the inducement effect appeared about one step after the stimulation. There was no significant difference in the latency between the conditions of the vection direction or motion induction of the leg.

Fig.9 shows the duration of ankle movement to the opposite direction of a vection stimulus compared to the conditions of the vection direction and the foot with which movement first appeared. The average duration was 0.38 sec, and the average distance was 0.4m (half step). As a result, it was confirmed that the movement induced by vection stimulation is rather limited in time, when compared to walking behavior.

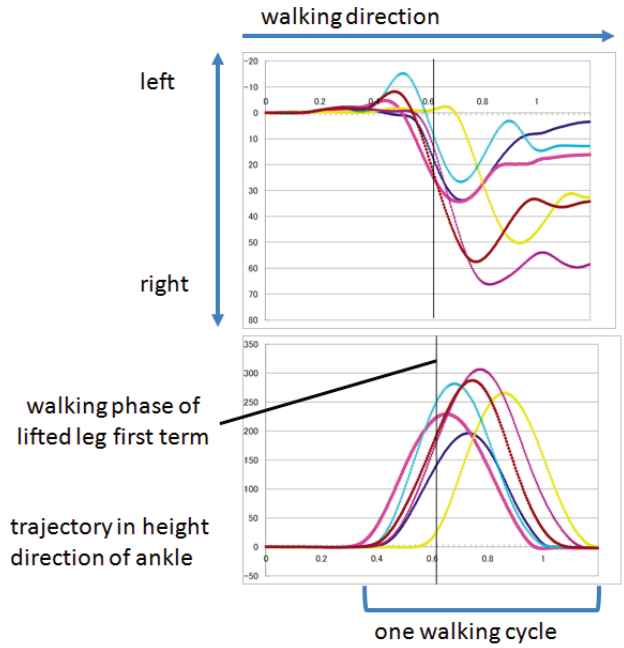


Figure 7: Trajectory of the left ankle for 1 step after vection stimulation. Above graph shows tracks of a right and left ankle to walking direction and below shows tracks in height direction.

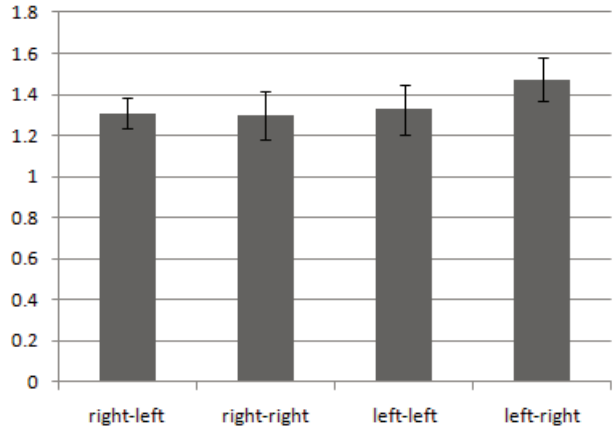


Figure 8: Latency from vection stimulation onset to inducement timing. X axis shows [vection stimulation direction] - [first movement appeared feet], and Y axis is latency time [sec]

4 Conclusion

Based on our experiments, we conclude that gaze-point movement and changes in walking direction are caused by optical-flow stimulation under the body sway proprioceptive sensed, and in this case, produced by a vibratory device. It was also confirmed that the reflexive leg movement by the vection stimulation first occurred in the walking phase of leg lifting, and next, the change in walking direction was induced.

Given these results, we consider the following mechanism for the walking guidance phenomenon that we found: Due to the vi-

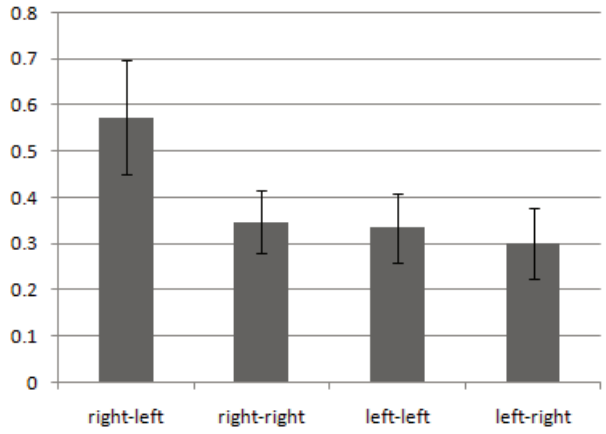


Figure 9: Inducement time to right and left of ankle. X axis shows [vection stimulation direction] - [first movement appeared feet], and Y axis is inducement time [sec]

bration stimulus given to the gastrocnemius, the signal gain within somatosensory the somatosensory channel, (namely, the vestibular system), is lowered. Yet, when an optic-flow stimulus is is visually delivered, the subject perceives his or her body as being shifted in a direction opposite to that of the flow, and thereby generates a correction reflex towards the vection direction.

In an AR environment, it was known that one feels acceleration when watching a moving object in a wide angle of view. On the other hand, this feeling produced a discrepancy between vision and the vestibular senses, and that incongruence consequently induced the what has been called VR sickness[4]. Although devices that display visual and vestibular stimuli simultaneously have been developed to relieve this conflict of inputs, they have until now been large and complicated systems which were inappropriate for daily use. Since the vection stimulus and the body sway in our study are easier to present using an HMD device and a vibration device, we can envision it as a practical inducement device in an actual AR environment.

While GVS was known as a method for inducing unconscious body movement using visual illusion, here in this paper, we demonstrate another method that does not require GVS. And, although we presented a vibratory stimulus continuously within this study, we do not believe it is essential. In the future, we plan to analyze the dynamics of sensory signal usage in the walking phase, and obtain effective walking action induced by only brief vibration during the walking cycle.

Although GVS was thought to be an attractive method for inducing the unconscious body movement using visual stimuli, it eventually did not become popular due to the possible side effect of dizziness when it is applied at length, or continuously. In summary, we here present a method that does not require GVS in order to generate adequate vection by further analyzing the dynamics of sensory signal usage during the walking phase, and inducing effective walking using only minimal vibratory stimulation during the walking cycle.

Acknowledgements

This research is supported by Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology agency (JST). Professor Taro Maeda gave us a lot of important advices. We thank their efforts.

References

[1] R. C. Fitzpatrick, D. L. Wardman, and J. L. Taylor. Effect of galvanic vestibular stimulation during human walking. *The Journal of Physiology*, (517):931–939, 1999.

[2] N. Fridman and G. A. Kaminka. Towards a cognitive model of crowd behavior based on social comparison theory. In *AAAI 2007*, pages 731–737, 2007.

[3] F. Hlavacka and F. B. Horak. Somatosensory influence on postural response to galvanic vestibular stimulation. *Physiol. Res.*, 55(1):121–127, June 2006.

[4] J. Reason and J. Brand. *Motion sicknes*. London ; New York : Academic Press, 1975.

[5] R. Snowden, P. Thompson, and T. Troscianko. *Basic vision: an introduction to visual perception*. Oxford University Press, 2006.

[6] M. Sugimoto, J. Watanabe, H. Ando, and T. Maeda. Inducement of walking direction using vestibular stimulation-the study of parasitic humanoid (xvii)-. *Virtual Reality Society of Japan*, 8:339–342, Sept. 2003.

[7] T. Suzuki, A. Ueno, H. Hoshino, and Y. Fukuoka. Effect of gaze and auditory stimulation on body sway direction during standing. *IEEJ Transactions on Electronics, Information and Systems*, 127(10):1800–1805, 2007.

[8] T. Yoshida, T. Takenaka, M. Ito, K. Ueda, and T. Tobishima. Guidance of human locomotion using vection induced optical flow information. *IPSJ SIG Technical Reports*, 2006(5(CVIM-152)):125–128, 2006.

A Pedestrian Dynamics Simulator for Wearable Navigation

Kosuke Shinoda
RIKEN

Itsuki Noda
AIST

Eimei Oyama*
AIST

Abstract

There are designated routes that should be taken by a crowd of people during emergency situations. When a disaster happens, some of these routes might not be available because of structural problems caused by the disaster itself. A more important factor is the distribution of congestion of the people spread over the area. The flow speed of people (pedestrians) depends on the density of the area. Therefore, when many pedestrians select the same route, or follow others, they may encounter heavy congestion at various places. Therefore, it is important to assist in the navigation of pedestrians by providing them with useful information. We have designed and developed a pedestrian dynamics simulator to perform navigation system studies, and have validated the system. Moreover, we have performed simulation experiments with a novel wearable navigation device that can guide the evacuation process. In one simulated case, we have verified that the navigation system can handle crowd control when the wearable navigation device is used by over 30% of the crowd.

Index Terms: K.6.1 [Management of Computing and Information Systems]: Project and People Management—Life Cycle; K.7.m [The Computing Profession]: Miscellaneous—Ethics

1 Introduction

The issue of how to evacuate people efficiently from crowded spaces (e.g., station's platform, concert venue) is an important problem; and a suitable solution can result in saving lives. A emergency situations can happen anywhere, and it is necessary to develop some countermeasures in order to keep people safe in such events. We can categorize the situations of how to shield the general public against danger, as "emergency precautions" and "hazard mitigation". In the former, we could consider how to design buildings, analyze walk flows, allocate emergency sign boards and perform emergency escape drills for efficient escape behavior. A number of researchers have studied this problem [3][5][7], which is called "*Pedestrian Dynamics*" Those researches aim to analyze the characteristic of pedestrian flow for the effective design of human-usable space, or to be used as educational tools to train general public in preparation for disaster. In the disaster, nobody knows what will happen, when and where a congestion occurs, and so on. Furthermore, the dynamics of crowd walking in emergency evacuation is not steady and it has large uncertainty. We should be able to respond to changing the situations in a flexible and dynamic manner.

As wearable technology and ubiquitous technology develop, we can provide some countermeasures against emergency congestion, i.e., evacuation guidance using some types of evacuation methods. We assume that a number of people can use a personal wearable device, such as a mobile phone, UMPC, and Head Mounted Display (HMD), under such situations in order to support the evacuation by a navigation system. These devices are called "Wearable Devices" [6][8]. We have developed the wearable robotics system, entitled the "Parasitic Humanoid (PH)" systems [10]. These novel devices provide some explicit and implicit support and advice for a specific person, who is equipped with one of the devices. Since these navigation systems have a potential to deal with the changing situations, the utilization of the devices can be a novel evacuation measure.

In order to evaluate the effectiveness of the devices, we developed a pedestrian dynamics simulator to perform evacuation experiments in virtual spaces and we validated our simulator as a pedestrian dynamics simulator. We conducted a number of preliminary evacuation simulations and obtained the results of the simulations to show the effectiveness of the devices.

In this paper, we propose the utilization of the wearable devices with the personal navigation function in an emergency and illustrate the results of the evacuation simulations to illustrate the effectiveness of the devices. First, we discuss the potential of the wearable navigation devices as a crowd navigation system in an emergency. Then, we introduce the pedestrian dynamics simulator and show the properties of the simulator. We illustrate the results of the simulation experiments to evaluate the utility of the wearable navigation devices to decrease congestion (i.e., avoid heavy congestion). Finally, we summarize the main contributions of this paper.

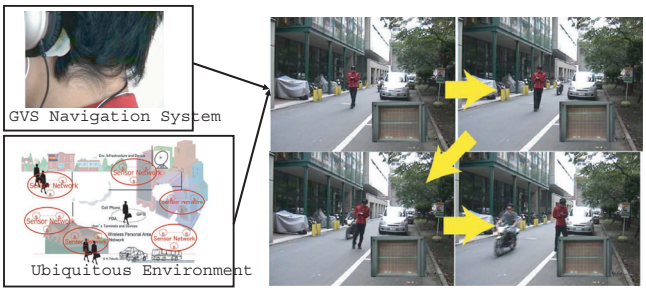


Figure 1: Human Behavior Induction by GVS. By using GVS and the environmental information provided by sensor networks, the wearer of PH can avoid a motorcycle, unconsciously

Figure 2: Prototype Wearable Navigation System using AR Technology. Left: The rapid prototype of the HMD of the PH. Right: The image for the wearer of the PH.

*e-mail: eimei.oyama@aist.go.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

2 PEDESTRIAN DYNAMICS MODEL

2.1 Related Work

There are a number of simulators for simulation of crowd behavior [5][13][18]. Some related work use the particle model or the cellular automaton as an agent model [5]. These approaches are simple for modeling agent actions, and easy to observe crowd behavior. However, it is difficult for these models to describe more complex (and realistic) characteristic properties. Therefore, recent efforts have introduced the multi-agent based model [13][18]. Specifically for disaster evacuation, the RoboCup Rescue Simulator [15][17] and FreeWalk [11] are well-known simulators. RoboCup Rescue aims to promote research and development for disaster risk management, using computer simulations or physical robotic agents. FreeWalk is platform where human participants and autonomous characters can socially interact with one another in virtual city space. The crowd simulation by computers is one of the approaches to construction of a secure and safe society for performing experiments. There are some crowd simulations that include the situation of disaster evacuation [3][17][19]. These simulations aim to design environment in order to realize effective escape behavior. Generally, the method for improving the entire evacuation behavior is effective installation of evacuation directive boards, or participating in an emergency escape drills.

We have developed a simple simulator, and have conducted simulations of situations in order to study the advantages and disadvantages of the proposed devices.

2.2 Wearable Navigation Devices

The target devices provide some explicit [8] and implicit support [9] and advice for a specific person, who is equipped with the devices. We will call these devices "Wearable Navigation Devices" or "Wearable Devices".

The concept of wearable computing has been actively investigated, as the physical sizes of components such as computers, sensors, actuators, etc. are getting smaller, and the wearable VR and wearable robotics have become popular research topics, similar to the popularity of wearable computers [6][8]. The cellular phones with the Global Positioning System (GPS) navigation function are popular now; although, the function does not work inside buildings. Kurata et al. [8] has developed a personal navigation system, which combines the self-contained sensors (accelerometers, gyro sensors and magnetometers), the GPS, and an active Radio Frequency Identification (RFID) tag system for both outdoor and indoor use [8][12].

Our target devices is the wearable robotics device entitled the "Parasitic Humanoid (PH)" system. Maeda et al. developed the wearable robotics system called the PH systems [10]. Although the PH does not have any actuators, the PH can guide the wearer of the PH to walk in desirable course by the galvanic vestibular stimulation (GVS), the wearable moment display, and the Augmented Reality (AR) technology using the HMD with cameras [2][9][1], as shown in Fig.1. The vestibular system is sensitive to GVS intensity changes, and responds by altering the magnitude of the response accordingly.

When GVS is delivered to the mastoid through electrodes during human walking, the wearer of PH responds by deviating towards the anode, as shown in Fig.2. For the motion induction in PH, GVS is available to induce the desirable movement of the direction during human walking. We envision being able to obtain a variety of information, whenever and wherever, desired using these wearable technologies in the ubiquitous societies of near future. In the case of an emergency, the wearable navigation devices directly guide the wearers to follow a safe path. The availability of our wearable navigation devices make a difference between the existing crowd simulations and our simulation at the point of being able to manage individual behavior, rather than the unspecific population behavior.

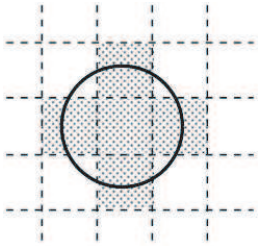


Figure 3: Agent Behavior Decision Model. The circle shows agent's body, and the cells occupied by only one agent. The agent selects its own direction of movement from five cells (forward, back, left, right, stay here). These cells are smaller than the agents' scale. Then, the agent moves with no constraint except that if the target cell is closed or is a wall.

2.3 The Simulation Model

Our target simulation is called "Pedestrian Dynamics Simulation" or "Crowd Simulation". Some researcher model this simulation as cellular automaton model [4] and multi-agent model [13][18]. We should chose our own model to fit the purposes of our experiments. The cellular automate model is based on discrete spatial representations. A pedestrian is expressed as a particle, and the cell is the same as the particle size. Each cell is defined as the potential fields that can use the reference value as the local effect of obstacles or moving pedestrians. The second model, the multi-agent model, focuses on the simulation of a pedestrian's action as an entity, and the interactions among agents. This model is able to deal with individual cognitive capabilities.

In our simulations, some agents obtain a guidance provided by the wearable navigation system and the other obtain no guidance. Since the device is a personal device, we selected the multi-agent model for our simulations. The social force model proposed and developed by Helbing et al.[5] is the representative model, which describes the interaction between the agents. The direction and speed of a pedestrian is assumed to be computed by the combination of the virtual forces between the agents. The description of the behavior of the agents is based on the social force model in our simulations.

2.3.1 The Environment Model

We describe the simulated world as a typical grid world (similar to a cell-based), where the pedestrian is described as a particle. Each cell is smaller than an agents' scale. Our goals are to model the complex interaction between agents in crowded space, and to reduce the amount of required calculations. As the size of the cell decreases, the computational requirements increases. We set the cell size to about 0.1m x 0.1m, with human width of 0.5m. Our simulation breaks the time into discrete steps, where in each step, an agent decides its own behavior. The simulator performs collision detection.

2.3.2 The Agent Model

The pedestrian dynamics simulation aims to reproduce complex behavior patterns using interactions among pedestrians. And our proposed device targets personal user, hence we modeled pedestrians using the multi-agent model, and implemented the multi-agent simulator for simulating pedestrian dynamics. We used the boid model [16] for crowd behavior. The model is regarded as a kind of the social force model.

We developed our simulator based on C. Parker's pseudocode of Boids [14]. Let $x(t)$ be the 2D position vector of an agent i at time t. According to the pseudocode, $x(t)$ is updated as follows:

$$v(t) = v(t-1) + a(t) \quad (1)$$

$$x(t+1) \quad = \quad x(t)+v(t) \tag{2}$$

$v(t)$ is the velocity of the agent and $a(t)$ is the acceleration of the agent. Let $F(t)$ be the real and virtual force affects the agent. We assume the following equation is established.

$$a(t) = F(t) \tag{3}$$

$F(t)$ consists of the following three forces:

$$F(t) = F_{drag}(t) + F_{goal}(t) + F_{boids}(t) \tag{4}$$

$F_{drag}(t)$ is the drag force affecting the agent, which is proportional to $v(t)$. $F_{goal}(t)$ is the force driving the agent to follow the path to the goal. $F_{boids}(t)$ is the social force, which describing the flocking behavior of the agents based on Reynolds' Boids model. The force consists of three simple steering behaviors which describe how an individual boid maneuvers, based on the positions and velocities of its nearby flock-mates: *Separation* ($F^S(t)$) steer to avoid crowding local flock-mates, *Alignment*($F^A(t)$) steer towards the average heading of local flock-mates, *Cohesion*($F^C(t)$) steer to move toward the average position of local flock-mates.

The agent i receives the social force F_{boids}, as follows:

$$F_{boids}(t) = F^S(t) + F^A(t) + F^C(t) \tag{5}$$

An agent i decides its own direction of movement in the grid with roulette selection based on F_i. At each moment, a selection probability calculates a component of the resultant force between boid model and target direction. When the selection probability of each direction is lower than any threshold value, the agent increases a probability to stay there. The property of wearable navigation device is expressed as an internal parameter of agents.

3 Simulator Properties

In this section, we check the basic properties of the simulator described in 2.

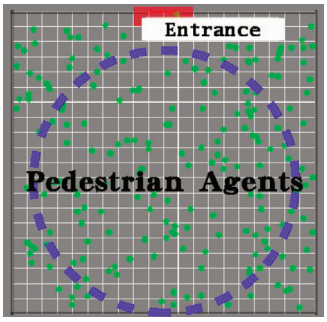


Figure 4: Pedestrians evacuating from a room with one door: Distribution of pedestrians, who can see the door.

3.1 Volume Effect

Since the simulator integrates excluded volume effect [5], it can represent an arch-like blocking (bridging) at an exit. The phenomenon is important for pedestrian simulation. We checked the volume effect of the pedestrian flow.

The room size is 10m x 10m and the exit size is 1m as shown in Fig.4. N pedestrians attempting to leave a one-door square room. Pedestrians are initially placed randomly in the room. As the pedestrians move toward the exit, the arching behavior near the exit can be observed as shown in Fig.5.

The calculation results of the proposed model reflects the volume effect of the social force model.

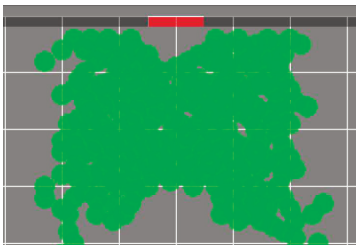


Figure 5: Arching and clogging at exit are observed: People stack near exit, because physical interactions in jams can build up in high density.

3.2 The relationship with crowd flow and entrance width

Since the speed of crowd flow decreases with an increase in density around the exit, the evacuation time decreases as the exit width increases. We checked the characteristics by the evacuation simulations. In the simulations, the number of the pedestrians N changes from 10 to 300 and the entrance width W changes from 1 m to 3 m. Fig.6 shows that the average time changes according to both N and W. The average time increases as W decreases. The narrow entrance, which width is 1 m suddenly rises the average time when more than 70 pedestrians in the room. The arching phenomenon can explain the rapid rise of the average time. The simulator qualitatively emulated the volume effect, which causes this result. The validity of the simulator was partially confirmed by this result.

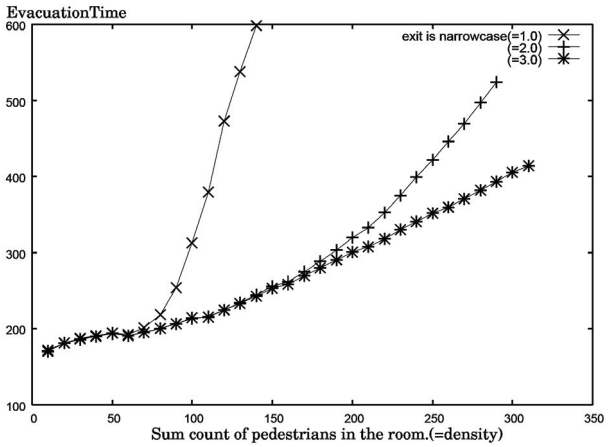


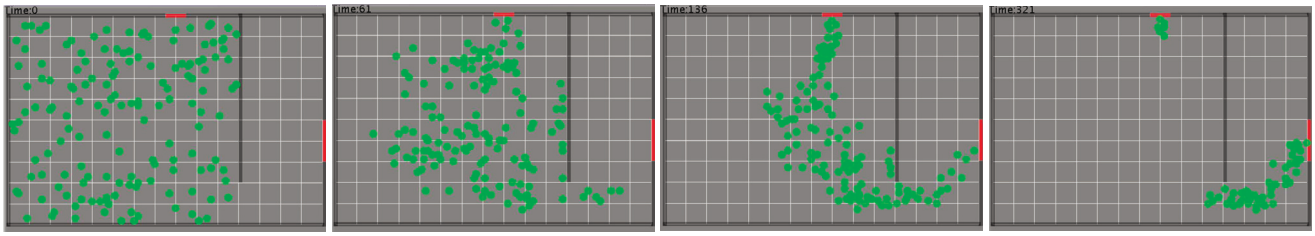
Figure 6: The average time of evacuation depending on entrance width: There are entrances, which are two types of width W. the narrow(W = 1.0) and the wide (W = 2.0, 3.0). The narrow line increased exponentially earlier than the wide line.

4 Preliminary Evacuation Simulations

In order to evaluate the effectiveness of the wearable navigation devices for personal navigation, two preliminary simulations were conducted in this paper.

4.1 Effect of Exit Selection

The first simulation aims to evaluate the effect of instruction in the evacuation. The experimental environment is shown in Fig.7. Usually, there are some designated escape routes at any place (buildings, parks, etc.). There are two exit, Exit A and Exit B in the room. All pedestrians in the room can find Exit A, which is narrow, and some pedestrians can see Exit B, which is wide. Exit A has a



a) initial state b) 61 simulation steps c) 136 simulation steps d) 321 simulation steps

Figure 8: Screen-shoots of evacuation simulation

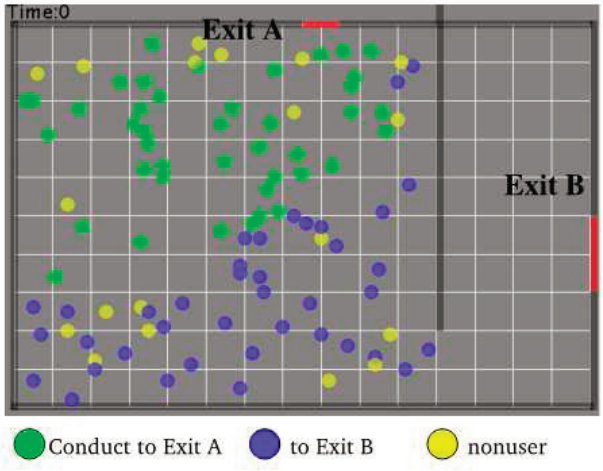


Figure 7: Simulation Field: Field size is 10m x 15m. This situation has two possible escapes: Exit A and B. Exit A is narrow, but is known by all pedestrians. Exit B is wide, but not all pedestrians may know its existence.

tendency to be clogged. All pedestrians receive an instruction that shows an escape pathway. The escape pathway, which is given to one pedestrian approximates the shortest pathway from him/her to one exit. The pedestrian who has shorter escape pathway to Exit B has a tendency to be instructed to move to Exit B. In the simulations, we changed the ratio of the pedestrians, who is instructed to use Exit B. Let α be the ratio. Fig.8 shows the screenshot of one evacuation trial.

All the following experimental results are the average of 1000 trials. 50 initial status of the pedestrians were generated randomly. 20 simulation trials are conducted from each initial status of the pedestrians.

Fig.9 shows the change of the average time of evacuation according to α. The vertical bar of the graph represents the simulation steps corresponding to the time of evacuation. It can be seen that there is an optimal point of efficient evacuation time around $\alpha = 0.5$. Fig.10 shows the ratio of escape completion until a certainly time (= 1500 simulation steps). The clogging happens with large α as well as small α. However, the clogging with large α is relatively small since Exit B is wider than Exit A. Therefore, the escape pathway to Exit B is a tolerant pathway. When N is smaller than 250 and α satisfies $0.4 < \alpha < 0.7$, the ratio of the successful evacuation nearly equals 1. Almost all the pedestrian successfully escape from the room in the emergency. The best value of α exists between 0.5 and 0.6.

4.2 Effectiveness of Wearable Navigation Devices

It is not realistic to assume that all people would have their own wearable navigation devices in case of an emergency. Therefore, we conducted the second simulation to investigate the relationship between the average time of evacuation and the percentage of the pedestrians, who have the wearable navigation devices.

Initially, N pedestrians are distributed randomly. All pedestrians try to leave a room. There are two groups of agent: (1) first type of pedestrians, who have the devices and (2) second type of pedestrians, who have no wearable navigation device. The first type of the pedestrians have support from the navigation system. They will leave the room through the instructed pathways. The other pedestrians, who have no wearable device will leave the room through the nearest door or follow other pedestrians as described in 2.3.2.

As the ratio of the wearable navigation device users increases, the wearable device users who are instructed to move to Exit B increases. Green circles in Fig.7 represent one group of the wearable device users, who are instructed to move to Exit A. Blue circles in Fig.7 represent the other group of the wearable device users, who are instructed to move to Exit B. Yellow circles in Fig.7 represent the other pedestrians, who have no wearble device. Snapshots of one simulation traial is shown in Fig.11.

Fig.12, and Fig.13 show the result of simulation experiment, where the horizontal axis is the ratio of the pedestrians, who use the wearable devices. In Fig.12, the vertical axis is the average time of evacuation. The vertical is the ratio of escapes completed in Fig.13. All the experimental results are the average of 1000 trials as described in the previous section. The average time of evacuation decreases according to increase of the users of the wearable navigation devices. And the ratio of escapes complete is 100% when the percentage of the users is over 30%. When 30% of pedestrians use the navigation system, the secure evacuation is realized in the simulation case.

These results indicate that the wearable navigation devices are effective, when certain percentage of pedestrians use the devices. It should be noted that the devices are effective even when the devices is used by some pedestrians, not all the pedestrians. The nature of the boid model, which encourages pedestrians to flock caused the result.

5 Conclusion

We proposed and developed wearable navigation systems for management of crowd behavior (crowd control) in case of an emergency. In order to perform navigation system studies, we have designed and developed a pedestrian dynamics simulator. In this paper, the simulator is illustrated. By using the simulator, the effectiveness of the wearable devices are evaluated. In one simulated case, the evacuation is successful when the wearable device is used by over 30% of the crowd. We verified that the wearable navigation devices has a potential to handle the crowd control.

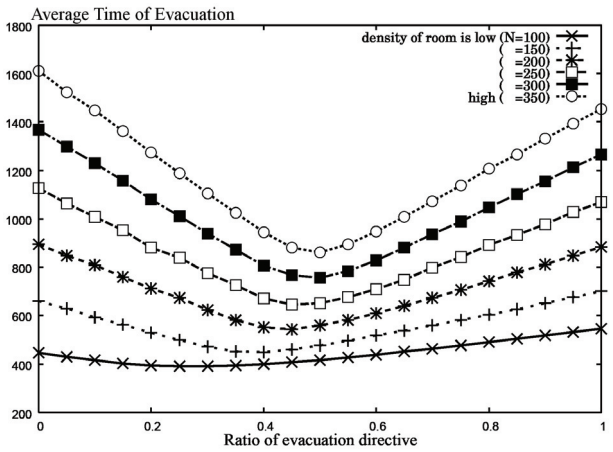


Figure 9: The average time of evacuation: Each line is different density of agents $N = \{100, 150, 200, 250, 300, 350\}$. The x-axis is the ratio of pedestrians α, who leave via Exit B. The y-axis is the average time of evacuation of all pedestrians. When α is around 0.5, all pedestrians escape efficiently.

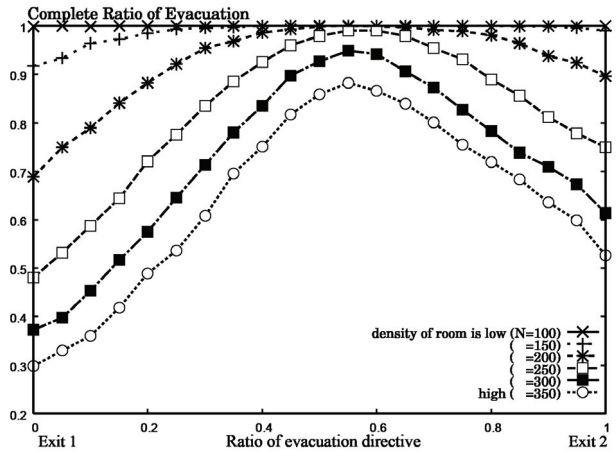
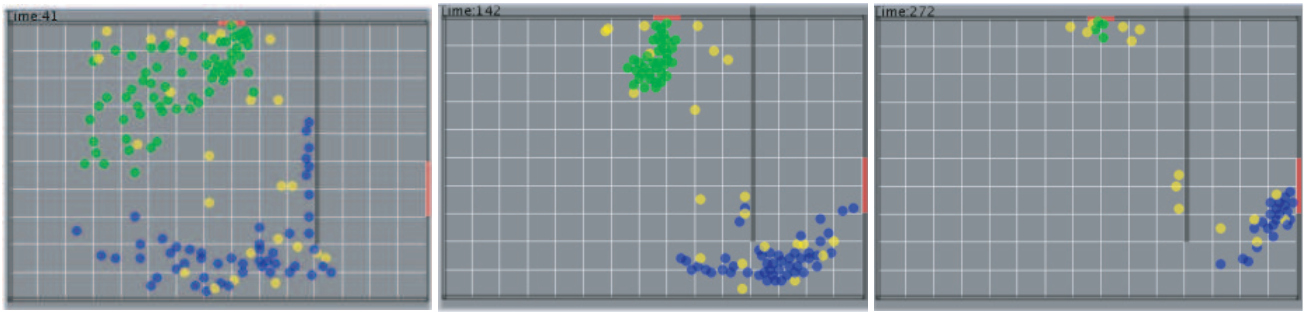


Figure 10: The ratio of escape completion: Each line is different density of agents $N = \{100, 150, 200, 250, 300, 350\}$. The x-axis is the ratio of pedestrians α, who leave via Exit B. The y-axis is the ratio of pedestrians who complete the evacuation by certain counts(= 2000 steps).



a) after simulation started b) Pedestrians via separate exit c) before simulation finish

Figure 11: Screen-shots of the pedestrian dynamic simulation

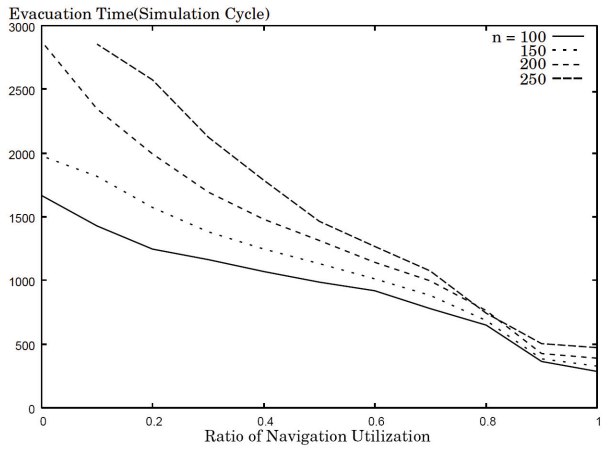


Figure 12: The average time of evacuation. Each line is different density of agents $N = \{100, 150, 200, 250\}$. The x-axis is the ratio of navigation utilization. When all pedestrians are provided the support from navigation system, the average time of evacuation decreases.

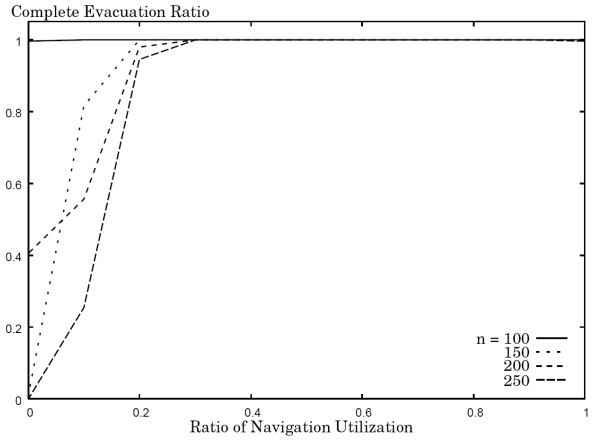


Figure 13: The ratio of complete evacuation. Each line is different density of agents $N = \{100, 150, 200, 250\}$. The x-axis is the ratio of navigation utilization. When N is 100, all pedestrians finish the escape action completely. In Other conditions, when the system utilization is over 30%, all pedestrians may finish escape action completely.

ACKNOWLEDGEMENT

This work was supported by CREST of JST.

REFERENCES

[1] T. Amemiya and T. Maeda. Asymmetric oscillation distorts the perceived heaviness of handheld objects. *IEEE Transactions on Haptics*

archive, 1(1):9–18, 2008.

[2] H. Ando, M. Sugimoto, and T. Maeda. Wearable moment display device for nonverbal communications. *IEICE Trans. Info. and Sys.*, E87-D(6):1354–1360, 2004.

[3] A. Braun, S. R. Musse, L. P. L. de Oliveira, and B. E. J. Bodmann. Modeling individual behaviors in crowd simulation. In *International Conference on Computer Animation and Social Agents (CASA) 2003*, pages 143–148, July 1987.

[4] C. Burstedde, K. Klauck, A. Schadschneider, and J. Zittartz. Simulation of pedestrian dynamics using a two-dimensional cellular automaton. *Physica A: Statistical Mechanics and its Applications*, 295:507–525, 2001.

[5] D. Helbing, I. Farkas, and T. Vicsek. Simulating dynamical features of escape panic. *Letters to Nature*, 407:487–490, June 2000.

[6] M. Hirose, K. Hirota, T. Ogi, H. Yano, N. Kakehi, M. Saito, and M. Nakashige. Hapticgear: The development of a wearable force display system. In *IEEE Virtual Reality (IEEE VR) 2001*, pages 123–129, 2001.

[7] F. Kl`ugl and G. Rindsf`user. Large-scale agent-based pedestrian simulation. In *Lecture Notes in Computer Science*, volume 4687/2007, pages 145–156, 2007.

[8] T. Kurata, M. Kourogi, N. Sakata, U. Kawamoto, and T. Okuma. Recent progress on augmented-reality interaction in aist. In *Proc. of The 2nd International Digital Image Forum: The Future Direction and Current Development of User-centered Digital Imaging Technology and Art*, 2007.

[9] T. Maeda, H. Ando, and M. Sugimoto. Virtual acceleration with galvanic vestibular stimulation in a virtual reality environment. In *Proceedings of IEEE VR 2005*, pages 289–290, 2005.

[10] T. Maeda, H. Ando, M. Sugimoto, J. Watanabe, and T. Miki. Wearable robotics as a behavioral interface -the study of the parasitic humanoid. In *Proc of 6th International Symposium on Wearable Computers*, pages 145–151, 2002.

[11] H. Nakanishi. Freewalk: A social interaction platform for group behavior in a virtual space. *International Journal of Human Computer Studies*, 60(4):421–454, 2004.

[12] T. Okuma, M. Kourogi, N. Sakata, and T. Kurata. 3-d user interfaces for indoor exhibits navigation and reliving experiences on-and-off the spot. In *Proc. International Workshop on Ubiquitous Virtual Reality (IWUVR2008)*, 2008.

[13] T. Osaragi. Modeling of pedestrian behavior and its applications to spatial evaluation. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 834–841, 2004.

[14] C. Parker. Boids pseudocode. http://www.kfish.org/boids/pseudocode.html.

[15] R. Rescue. http://www.robocuprescue.org/.

[16] C. W. Reynolds. Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics*, 21:25–34, 1987.

[17] S. Tadokoro, H. Kitano, T. Takahashi, I. Noda, H. Matsubara, A. Shinjoh, T. Koto, I. Takeuchi, H. Takahashi, F. Matsuno, M. Hatayama, M. Ohta, M. Tayama, T. Matsui, T. Kaneda, R. Chiba, K. Takeuchi, J. Nobe, K. Noguchi, and Y. Kuwata. Robocup rescue project. *Advanced Robotics*, 14(5):423–425, 2000.

[18] M. C. Toyama, A. L. C. Bazaan, and R. da Silva. An agent- based simulation of pedestrian dynamics lane formation to auditorium evacuation. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 108–110, 2006.

[19] J. xun Chu, J. jing Li, M. Xu, and L. Zhao. Simulating escape panic based on the mechanism of asymmetric information distribution. Complex Systems Summer School Final Project Papers, 2005.

Improvement of Wearable View Sharing System for Skill Training

Yuki HASHIMOTO, Daisuke KONDO, Tomoko YONEMURA, Hiroyuki IIZUKA, Hideyuki ANDO and Taro MAEDA

Osaka University / JST CREST

ABSTRACT

We have proposed a view sharing system where two distant people can share their views by mixing or exchanging images captured by head mounted cameras with HMD. Users can share what another user is seeing, and furthermore these users can make their their spatial perception, motion and head movements correspond. Our goal is to transmit non-verbal skills from a skilled person to a non-skilled person by using this view-sharing system. To realize this goal, we have improved this system. This system is light weight, has a wide viewing angle, is stand-alone, and does not require calibration of intraocular distance. These features can facilitate more efficient skill training and expand the sphere of activity for using view sharing systems.

KEYWORDS: view sharing, skill training, parasitic humanoid, wearable system

INDEX TERMS: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems Artificial, augmented, and virtual realities

1 INTRODUCTION

By extending robot-human telexistence[1] technology to human-human situations, we are developing an environment where a skilled person, who actually exists at a different place, can work with high efficacy on the ground instead of non-skilled person. The skilled person feels as if he exists at the place and can work there. The non-skilled person can perform with high quality with the skilled person's help. In order to realize such a telexistence environment in human interactions, we are developing remote communication technologies exploiting sense-motion sharing. In this project, we have developed a view sharing system to share first person perspectives between remote two people[2][3]. The system consists of a head mounted display and cameras, which makes possible a video-see-though head mounted display (VST-HMD). The user wearing the VST-HMD can see his own view and the partner's view, and also send his own view to the partner. The view sharing system can be applied to skill transmission and learning tasks. Previously, we considered the effectiveness of our view sharing system in some skill training scenarios [4][5][6][7][8].

Based on previous work, we developed a new view sharing system to improve effectiveness and expand its applications. In this paper, we describe our existing view sharing system. We also show the design and implementation of our new view sharing system.

2-1 Yamadaoka, Suita, Osaka, Japan
{y.hashimoto, kondo, yonemura, iizuka, hide, t_maeda}@ist.osaka-u.ac.jp

2 VIEW SHARING SYSTEM

To transmit non-verbal skills, body position movement is one of the most important pieces of information. The view sharing system has a potential to transmit body position movement by visual information. Thus, we attempted to transmit three elements of body position movement–visual image, hand position and head motion–in our system. In this chapter, we describe our established system and the method of transmission of these three elements.

2.1 System Construction

The system arrangement and connection of devices are shown in Figure 1 and Table 1. The image taken by the cameras is sent to a PC through IEEE1394 cables, and the PC renders the output images by mixing the input images. The overall latency is 66 milliseconds.

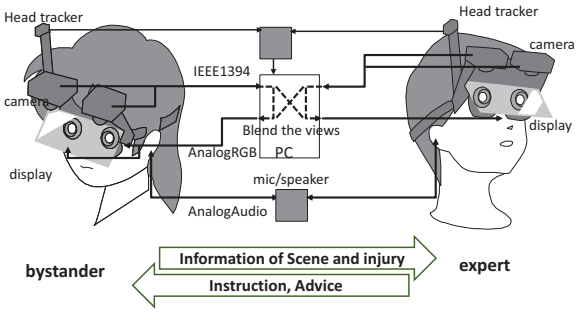


Figure 1. System construction

Table 1. Specification of the system

Display	Model	eMagin Z800 3D Visor
(HMD)	Resolution	800 x 600
	FOV	32° (horizontal) x 24° (vertical)
Camera	Model	Point Grey FireFly MV
	Resolution	752 x 480 (native)
		464 x 348 (effective resolution)
	Connection	IEEE1394a
Motion	Model	Polhemus LIVERTY
Tracker		(electromagnetic
		posision/rotation sensor)
	Update rate	240Hz
	Connection	USB2.0
PC	Model	Hewlett-Packard Z800
	CPU	Intel Xeon E5520
	OS	Windows XP SP3 32bit
	Graphics	nVidia QUADRO FX 4700X2

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

2.2 Visual image transmission

To get an exact first person viewpoint and improve spatial perception, we developed a video see-through HMD (VST-HMD). This VST-HMD has optical conjugation cameras (Figure 2), the user wearing the VST-HMD can see the image taken by the cameras attached on the user's head in real time with little delay. Basically, the user can see his own view as though he were seeing the view directly.

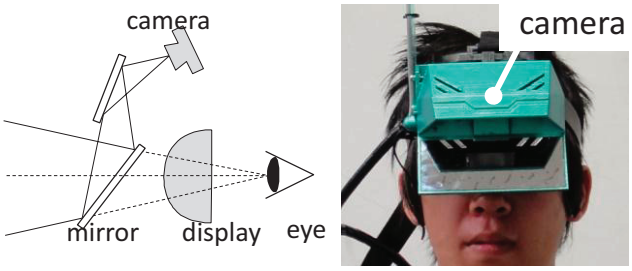


Figure 2. Orthoscopic video see-through HMD

2.3 Hand position transmission

To transmit hand position, image blending is used. In this case, the PC composes the view images from the user and the partner, and the user sees a blended view of him/her and the partner (Figure 3). The user can see the situation around the partner, and furthermore the user can follow the work performed by the other's hands. Because the user can see the partner's hands from a first-person viewpoint, the user can trace easily the action with his hands in real time to mimic the partner's motion.

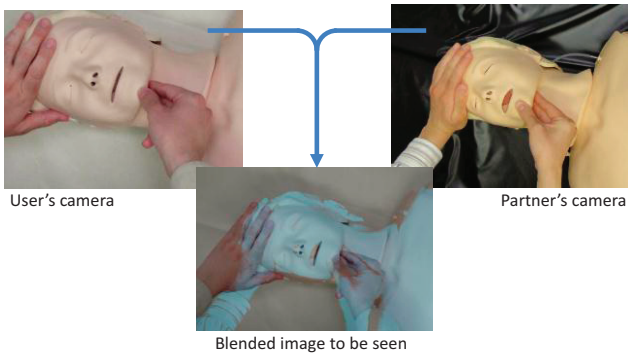


Figure 3. Image Blending

2.4 Head motion transmission

In order to follow the other user's head motions, a target marker is introduced. As shown in Figure 4, the user can see two markers, the green marker (A) is the center marker that is fixed in front of the subject, which indicates the orientation of the user's head. The other blue one (B) is the target marker, which indicates the other person's current head orientation. The marker is in the shape of a ring, with a diameter of 20 cm and a distance from head to the center of the marker of 70 cm. Therefore, the positional relationship between the center marker (A) and the target marker (B) reflects the positional relationship between the user and the partner. The user has to follow the target marker by moving his head. When both markers are aligned, the user and partner's motion are successfully corresponding.

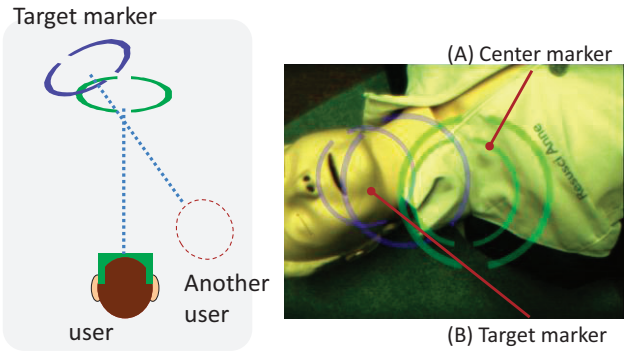


Figure 4. Spatial Relationships of Marker

3 Previous Trials and Problems of our System

3.1 Previous Trials

We have tried to use our system for skill training in the following four kinds of tasks.

1. *Rope work (Figure 5)*
2. *Playing the theremin (Figure 6)*
3. *Juggling (Figure 7)*
4. *Cardiopulmonary resuscitation (CPR) (Figure 8)*

Rope work and playing the theremin require adjustment of hand position. Juggling requires following arm motion over time. CPR requires following position and motion of the body, head and hand.

As a result of these experiments, we confirmed that our view sharing system is effective for skill training[4][5][6][7].

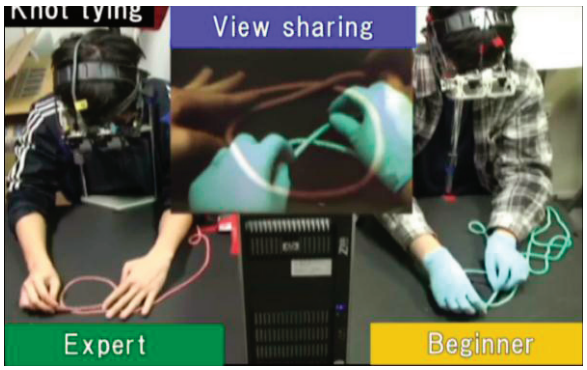


Figure 5. Rope work

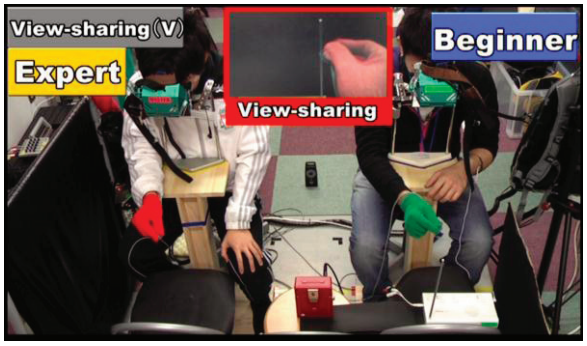


Figure 6. Playing the theremin

Figure 7. Juggling

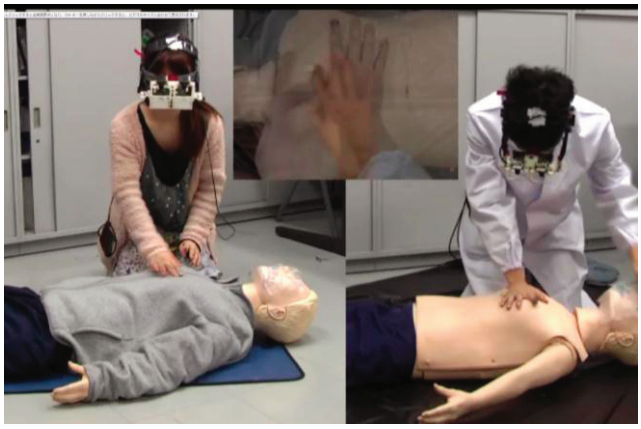


Figure 8. Cardiopulmonary resuscitation

3.2 Purpose

In order to improve the effectiveness of the view sharing system and expand its applications, we proposed a new view sharing system. In this system, we focused on two points. One was to improve the architecture of the VST-HMD to make it easier to wear and to widen the viewing angle. Another one was to make it a stand-alone system to extend users' range of possible actions.

4 Improvements in the New View Sharing System

4.1 System Construction

The system arrangement and connection of devices are shown in Figure 9 and Table 2. In order to realize a stand-alone system, a sensor was changed from an external sensor (Polhemus LIBERTY) to an internal motion sensor (Tokin MDP-A3U9S). The PC was also changed from a desktop PC to a laptop PC. In addition, this system has a battery in its control unit. We simplified connections between the VST-HMD and control unit. Images from cameras and data from motion sensors are sent to the PC through USB2.0. USB2.0 also supplies power to the devices. Therefore, the number of connecting cable between VST-HMD and control unit has been reduced to two (previous system = 7).

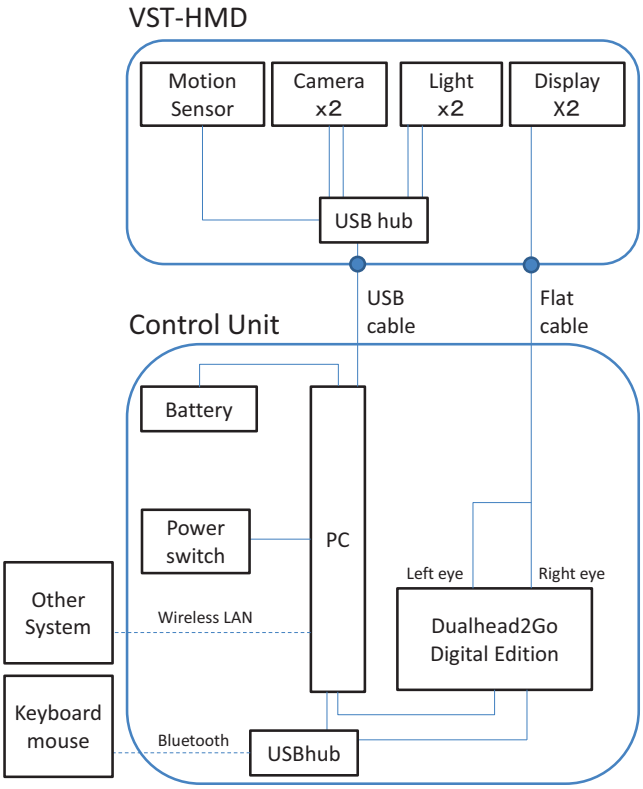


Figure 9. System Construction

Table 2. Specifications of the system

Display	Model	Daeyang FX603
(HMD)	Resolution	800 x 600
	FOV	42° (diagonal)
Camera	Model	Point Grey FireFly MV
	Resolution	752 x 480 (native)
		464 x 348 (effective resolution)
	Connection	USB2.0
Motion	Model	NEC/Tokin MDP-A3U9S
Tracker	Update rate	125Hz
	Connection	USB2.0
PC	Model	Apple Macbook air 11inch
	CPU	Intel Corei7 1.8GHz
	OS	Windows 7 64bit
	Graphics	Intel HD Graphics 3000

4.2 Implementation

4.2.1 Overview

The appearance of the proposed View Sharing System is shown in Figure 10. Weight is drastically lower than the existing system, and there are only two components; the VST-HMD and a vest including the battery, circuits, and laptop PC (Figure 11). Battery life is approximately three hours. For communication, we mainly use wireless networking so that the area of user's actions is not

limited by wires. However, wired networking may be needed in situations which require rapid following.
Therefore, a user can wear the system easily, and expand his/her range with a View Sharing System.

Figure 10. Proposed View Sharing System

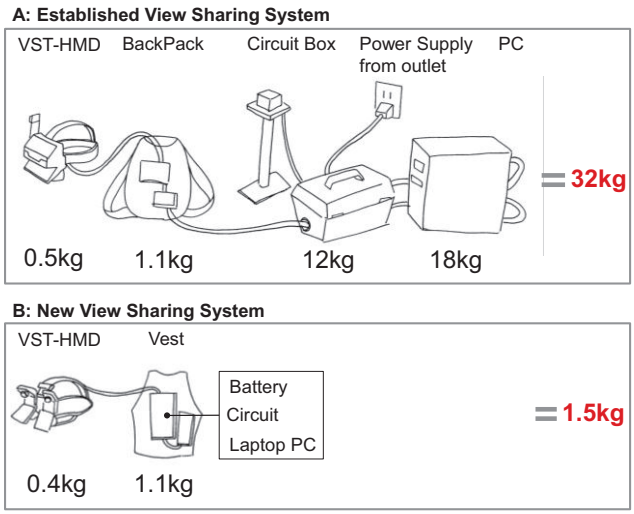


Figure 11. Components and weight

4.2.2 VST-HMD

The VST-HMD part is composed of two cameras (PointGrey FireFly MV), two displays (Daeyang FX603), two mirrors, two lamps and a motion sensor (NEC/Tokin MDP-A3U9S) (Figure 12). The viewpoint of camera and eye maintains an optically conjugated situation. The motion sensor is attached to the right side of the display unit. The display is bonded to goggles. The edges of the goggles are covered with soft rubber so that the eyes and goggles maintain close contact (Figure 13).
This VST-HMD has three advantages compared to the established system. The first is simplicity in wearing the system. The user can wear the proposed system like a pair of goggles, and will not need to adjust the fit after wearing them. The second is that there is no need to adjust the distance between the two displays to a user's interocular distance. Because a display contacts each eye, the viewpoint of each eye and its display is maintained without adjustment (Figure 14, Figure 15). However, the angle of convergence must be adjusted. To solve this problem, we have used software calibration with ARtoolkit. The user sees a marker with the VST-HMD, and we acquire each camera's position and rotation from the marker. The angle of convergence is calculated from this information.

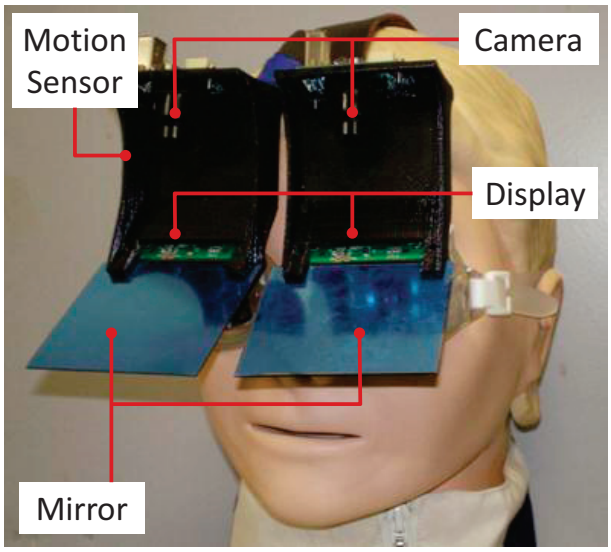


Figure 12. Proposed View Sharing System

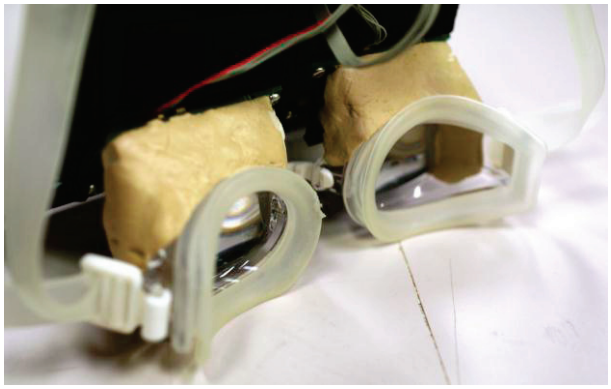


Figure 13. Contact part to eyes

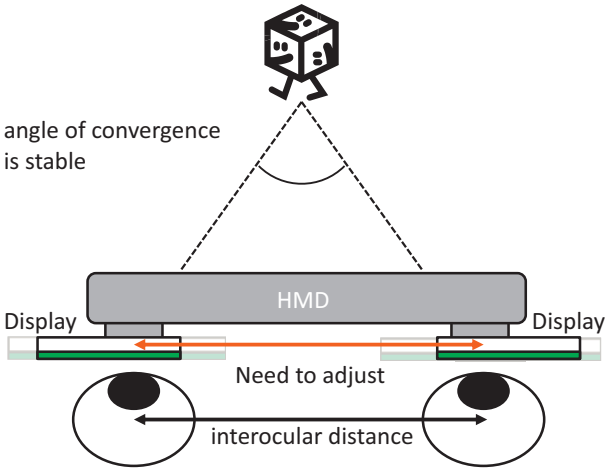


Figure 14. Relationship between interocular distance and angle of convergence in established the VST-HMD

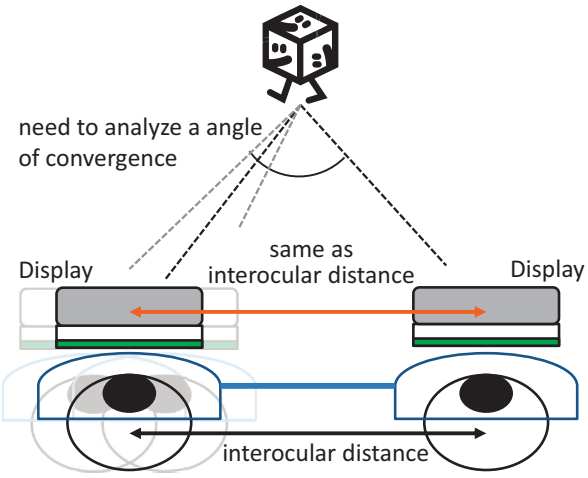


Figure 15. Relationship between a Interocular distance and an angle of convergence in proposed VST-HMD

The third is a wide view angle. The field of view is expanded because the display unit is fitted to the curvature of the face (Figure 16). The field of binocular vision (FOBV) of the proposed VST-HMD has been measured to be about 50° (horizontal) by 24° (vertical). In contrast, the FOBV of established the VST-HMD was approximately 35°(horizontal) by 24°(vertical) (Figure 17). However, the overlapping area between the two views was reduced (approximately 30° horizontal in the established system versus approximately 15° horizontal in the proposed system, distance = 50cm). One of our next steps will be optimization of the overlapping area and viewing angle.

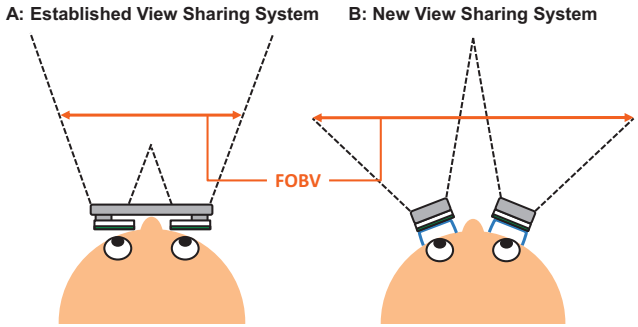


Figure 16. FOBV of VST-HMD

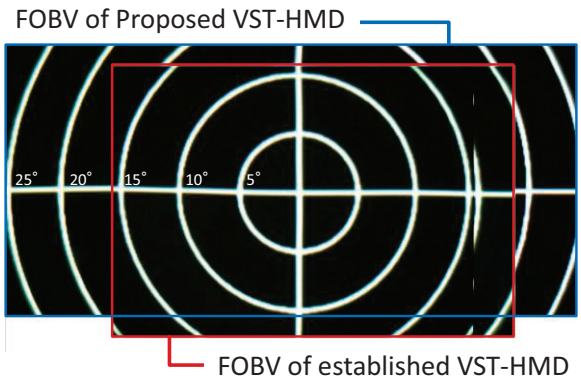


Figure 17. Measurement of FOBV

4.2.3 Control Unit

The control unit is composed of a PC board (Apple Macbook Air 11-inch), a circuit (Matrox Dualhead2Go and USBHub) and a battery (Japan Trust Technology Energizer XP8000). To trim weight, we deconstructed and rebuilt these components. For the laptop PC in particular, we removed the display, internal battery, a keyboard and cut the chassis. As a result, the weight of the PC board was reduced to 280g, making the total weight 800g (without cables) (Figure 18). The control unit is stored in the back pocket of a vest (Figure 19).

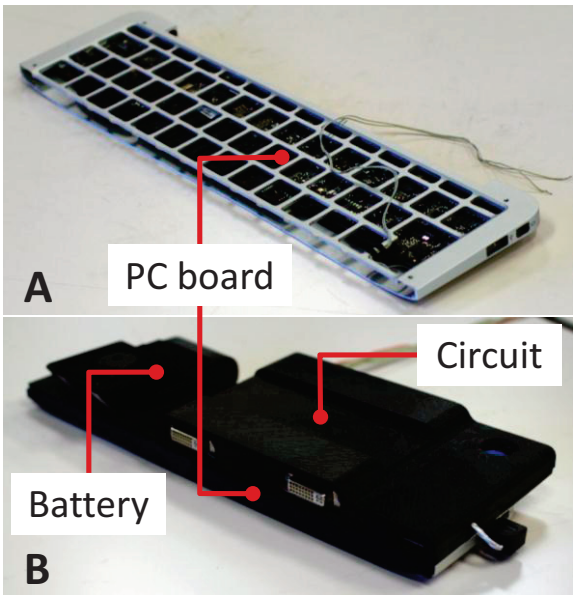


Figure 18. A: PC board, B: Control unit

Figure 19. Circuit unit is stored by back pocket of vest

5 Conclusion

In this paper, we propose a new view sharing system to facilitate more effective skill training and an expanded sphere of activity.

After describing our established view sharing system and previous trials, we explained the concept for a proposed view sharing system and described a prototype.
This system is expected to enable a user to work more efficiently and comfortably. In addition, the system will be able to support tasks which require larger physical movements, such as dance, baseball, or complicated emergency medical procedures.
As a next step, we will solve some issues of the proposed view sharing system, and compare its effectiveness in skill training with the old system. We will also use this system in new applications which require larger body movements.

6 ACKNOWLEDGMENTS

This research was supported by JST, CREST.

REFERENCES

[1] Susumu Tachi: Tele-existence - Toward Virtual Existence in Real and/or Virtual Worlds, Proceedings of the International Conference on Artificial Reality and Tele-existence (ICAT '91), pp.85-94, 1991.

[2] T. Maeda, H. Ando, H. Iizuka, T. Yonemura, D. Kondo and M. Niwa : Parasitic Humanoid: The Wearable Robotics as a Behavioral Assist Interface like Oneness between Horse and Rider, 3rd Augmented Human International Conference, 2011.

[3] D. Kondo, K. Hattori, K. Kurosaki, H. Kawasaki, Y. Hashimoto, T. Yonemura, H. Iizuka, H. Ando and T. Maeda : Effect of Wide FOV and Image Stabilization on Spatial Perception for View Sharing System, Proceedings of 20th International Conference on Artificial Reality and Telexistence, 2010.

[4] K. Kurosaki, K.. Hamada, D. Kondo, H. Iizuka, H. Ando, T. Maeda : Development and Evaluation ofa Zero Eye offset and Orthoscopic Video See-Through Head-Mounted Display, VRSJ the 15th Annual Conference, 2009. (in Japanese)

[5] K. Kurosaki, H. Kawasaki, D. Kondo, H. Iizuka, H. Ando and T. Maeda : Skill Transmission for Hand Positioning Task through View-sharing System, 3rd Augmented Human International Conference, 2011.

[6] H. Kawasaki, H. Iizuka, S. Okamoto, H. Ando, T. Maeda : Collaboration and Skill Transmission by First-person Perspective View Sharing System, 19th IEEE International Symposium in Robot and Human Interactive Communication, 2010.

[7] D. Kondo, K. Kurosaki, H. Iizuka, H. Ando and T. Maeda : View Sharing System for Motion Transmission, 3rd Augmented Human International Conference, 2011.

[8] H. Iizuka, H. Ando and T. Maeda : The Anticipation of Human Behavior Using Parasitic Humanoid, Proceedings of the 13th International Conference on Human-Computer Interaction. Part III: Ubiquitous and Intelligent Interaction, 2009.

Adaptive Substrate for Enhanced Spatial Augmented Reality Contrast and Resolution

Markus Broecker* Ross T. Smith† Bruce H. Thomas‡

Wearable Computer Lab
University of South Australia

ABSTRACT

This paper presents the concept of combining two display technologies to enhance graphics effects in spatial augmented reality environments. The appearance of the projected light images and text are enhanced by using an ePaper display as the substrate. The ePaper display employed does not emit light but provides a high resolution greyscale display surface that can dynamically change the appearance of the projected light pixels. We demonstrate graphics techniques that leverage this novel approach to provide an improved spatial augmented reality appearance. Our results are an improved black level that results in greater contrast and several image and text enhancement methods.

Index Terms: H.5.2 [Information interfaces and Presentation]: Graphical User interfaces—Input Devices and Strategies; I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction Techniques

1 INTRODUCTION

This paper explores the concept of enhancing the displayed appearance of Spatial Augmented Reality (SAR) graphics by combining two display technologies, a digital projector and an ePaper[1] display. Our technique enhances both the resolution and the contrast ratio of projected objects. Using this method, we have explored a range of techniques such as image enhancement, text enhancement and texture effects that leverage both display technologies working in synergy to provide an optimized appearance.

Augmented Reality (AR) systems commonly employ hand-held displays or head-mounted displays with optical or video see through technologies to present computer generated graphics. SAR employs projected light to present computer graphics that directly illuminate physical objects to enhance their appearance [7]. To achieve this, a simple substrate is constructed with the desired shape, for example a small white rectangular box for a mobile phone mock-up, with the appearance and interactive functionality provided by the SAR system.

There are two limitations with this approach that can be improved with a composite display. Firstly the black appearance is limited to the performance of the projector and the underlying substrate. With projection technologies, black is achieved with the absence of projected light and the colour of the underlying substrate. Achieving a "true black" representation using projected light is impossible with a white substrate. Secondly, the resolution of SAR objects is limited to the number of pixels provided by the projector. Both these aspects are limitations for the appearance fidelity since with these technologies we are unable to develop fine grained details on SAR objects such as small text and intricate surface textures.

Our solution to this problem is is to exchange the traditional white substrate with an ePaper display to provide a controllable, high-resolution display surface. The combination of display technologies allows us to leverage their different capabilities simultaneously. In particular, we are interested in display technologies that have different dynamic ranges, resolutions, colour spectrum, and refresh rates. The use of ePaper combine with projectors has been previously explored [5] for the express purpose of constructing an HDR display device. The research presented in this paper explores this combination for the purpose of darker blacks, finer resolution displays, augmenting ePaper with colour, and augmenting ePaper with animation effects.

Using our display method, we envision physical substrates will be covered with ePaper displays to provide regions of the user interface with the high resolution functionality. The adaptive surface regions will be used to simulate lighting effects, surface structures and compensation for overlapping areas of projector frustum. The models our SAR system is illuminating are tracked by a 6 DOF tracking system. While support for fully dynamic and constantly moving objects is possible solely in SAR, we support for our combined display technique the common case that an object gets moved to a new place, examined and moved again. This accommodates and compensates for the low refresh rate of current ePaper displays, as the new lighting or compensation masks have to be recalculated and displayed only after a change in position.

This scenario describes how ePaper displays will be used as adaptive substrates in SAR environments and the advantages they provide. Contrary to TFT and other digital displays, they are lighter, smaller and flexible, so they are able to be fixed to even limited curved surfaces. As they are requiring power only on state change, only simple electronic control hardware is required, and new surface descriptions would have to be "uploaded" to the device only on demand. As the ePaper display is still a display, it allows changes to the displayed content during runtime, contrary to paper paper printouts. Finally, the ePaper surface substrate is not light-emitting and has reflection properties close to the white paint we used earlier for our models in our SAR system.

The following paper describes a hybrid display using a projector and an eInk display for the implementation. Section 2 discusses related works including existing composite displays, spatial augmented reality systems and previous projector-based display systems. Section 3 describes the theoretical framework for our concept where the theory and properties of this hybrid display are discussed. The following section presents a variety of appearance techniques based on the theory such as image enhancement and image contrast improvement. Section 5 discusses the implementation, using an LCD projector and a Kindle ebook reader[2] as a technology demonstrator. This section also discusses techniques for registration and calibration of the two displays. Finally, Section 6 discusses the limitations of the current systems and Section 7 describes possible

*e-mail:Markus.Broecker@unisa.edu.au
†e-mail:ross@r-smith.net
‡e-mail:Bruce.Thomas@unisa.edu.au

[1]We refer to ePaper as generic electronic paper, and eInk as the product from www.eink.com.

[2]http://amazon.com/Kindle

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

future work, as well as new research directions as the capabilities of ePaper technologies advance.

2 Related Work

SAR enhances the physical world with perspectively correct computer generated graphics using digital projectors [19]. This is in contrast to other AR display technologies, such as Head Mounted Displays (HMD) which place augmentations on an image plane in front of the user's eyes, and hand-held devices which show augmentations on a hand-held display [2]. SAR requires physical surfaces to project onto. These surfaces can consist of any objects in the environment that are of interest to the user; projections are not limited to walls or purpose built screens.

Unlike projection-based CAD displays and other AR display technologies, SAR allows users to physically touch the virtual information. The surfaces provide passive haptic feedback and all stereoscopic depth cues are naturally provided by the physical substrate. Previous virtual reality research has shown that the ability to touch virtual objects and information enhances user experience [14], and can improve users' performance [21]. This physical nature of SAR makes it a compelling choice for industrial design applications since the designers can physically touch the mock-up prototypes and leverage the flexible computer controlled appearance. Hare et al. [13] describe the importance of physical prototypes in the design process. Using SAR, designers can naturally interact with design mock-ups, without having to hold or wear display equipment. As SAR places computer generated information directly onto objects in the real world, groups can view and interact with the system. This makes SAR an ideal choice for collaborative tasks.

Objects that are augmented with projected imagery are either custom built props with ideal projection properties or they are existing entities. For the second case, research has been undertaken to investigate how to best project onto non-optimal surfaces taking into consideration their colour and geometry. Grossberg et. al describe a camera-projector method in which the colour response of the surface is taken into account [12]. A compensation image is created which allows the projection on any kind of coloured surface without any degradation in image quality. Bimber et al. also compensated for the irregular shape of the projection surface as well as inter-reflection properties of non-planar geometries and light-transport differences [3, 6, 22].

SAR systems are increasingly used in museums and other public spaces. Implementing a projector-based system instead of regular displays allows the seamless integration into existing exhibitions. In paintings, for example, certain details can be highlighted, explanations can be projected directly in place and previous restoration and cleaning processes can be displayed on the object of interest [4]. Aliaga et al. [1] describes a camera-projector system that is able to fix damages to a sculptures' surface, as well as relight it under virtual lighting conditions.

The lack of dynamic range of projectors is a well known challenge and a few attempts have been made to create a high-dynamic range display device using projectors. Stürzlinger and Pavlovych combined projectors with LCD displays, in effect replacing the light source of a TFT display with a projector or a custom-built, addressable LED display [18, 20]. They describe image splitting functions for both created displays, that split a high-dynamic range (HDR) input image into a subimage for the light-providing display (the projector or the LED wall) and the modulating display (the LCD screen). Bimber [5] created a composite high-dynamic range display using a projector and a projection surface as well. The projection surface could be any kind of display, from a TFT display, to a printed paper and also an ePaper display. His work extends Stürzlingers in that it extends and generalises the HDR image splitting function. The focus of both papers was to create a HDR capable composite display. Special care was taken to transform the input HDR image into the increased dynamic range of the new display. Our work however works with low-dynamic range data and does not focus on creating a HDR device, but rather on preserving a darker black.

While modern LCD projectors offer a high resolution, these projected images suffer from different problems up close. As the distance between projector and projection surface increases, the local resolution or pixel density of the projected image decreases. Additionally, fine darker lines between individual pixels become visible, the so-called "screen-door effect". This effect can be compensated if multiple projectors are used to illuminate a single surface. Due to small spatial differences in the projections, images don't overlap perfectly. If the rendering pipeline takes into account these differences, and compensates for it, local *Super-resolution* can be achieved [9, 10]. These overlaid images enhance each other and provide a natural anti-aliased image.

Another solution for the low local resolution of projection-based displays are composite displays. Olwal [17] et al. described different methods on how to interact with mobile, tracked, high-resolution devices to enhance local areas of the overall projection, thereby *replacing* one display locally with another, more suitable one. Contrary to this concept of employing two separate displays, we are seeking to combine the properties of two display technologies into one composite display.

Electronic paper displays are commonly used for ebook readers and low power application, such as supermarket price displays. eInk is a company and line of display products using their proprietary implementation of an electronic-paper display. The workings of electronic paper is described by Comiskey et. al [8]; the *sheet* of electronic paper consists of microcapsules, filled with electronically charged, white and black particles. A current can be applied to either the upper or lower side of this microcapsule, thus separating the particles and changing the apparent colour of the microcapsule to either black or white. To drive these sheets of paper like a display, a matrix display controller is used [11], and groups of microcapsules are addressed together, so that a pixel raster is created. One of the fundamental differences to regular displays is that electronic paper displays are passive and only need control commands (or applied voltage) when a display change is required, thus making them very energy efficient when compared to LCD technologies. Finally, the high resolution and contrast offer a legibility similar to traditional paper printouts [16].

3 Combining Display Technologies

This section provides a description of the properties of interest for the two display technologies employed in this paper, digital projectors and ePaper. A discussion on the current technology challenges is presented followed by the theory of how a hybrid display surface can be employed to improve display performance for SAR environments.

3.1 Technology Challenges

Currently the image contrast in projection-based environments is not sufficient to provide realistic blacks compared to a physical black object. When projecting black on a large area this problem is not so significant, however blacks are very important for defining details in images to highlight edges and require a good implementation to achieve this well. The current limitation is a technical challenge with the method used to create black with current projector systems. Black content is achieved with the absence of projected light, where the darkest perceived colours are produced by the underlying material colour alone and whites are achieved with the maximum light from the projector. One technique used to improve the perceived darkness of projected blacks is to use a grey screen in a darkened room. Using this technique blacks appear to

be darker compared to when a white screen is used. However, the range between the darkest and lightest appearance (or contrast) remains unaffected since with a grey screen the projected white appears to be darker. This poses a problem with SAR systems since they are used to support presentations in rooms with ambient light. For example designers developing SAR prototypes for clients. This task requires high contrast images to provide optimal realism so as to maintain the fidelity when compared to physical prototypes and presented in a room with enough ambient light to allow a meeting to be held.

Ambient light significantly affects the appearance of projected light displays. Even if a room could be made completely light-absorbent, inter-reflections of the projected light between projection surfaces on models or "bounced" light from the observer's physical body will illuminate the scene and reduce the accurate representation of black. Achieving darker colours and black allows the operation of the projection system in much brighter ambient light environments without losing details of the projected textures. One solution to this challenge is to use a custom built dark room that prevents all ambient light from entering the room. This approach does improve the projected appearance but still does not overcome the limitation of the image contrast that is fixed to the performance of the projectors specifications.

The *effective resolution* is determined by the distance between the projector and the projected surface, the surface shape and the area used. Unlike simple planar displays, SAR systems are often projecting on objects that are not employing the entire projector frustum area. This leads to a lower effective resolution on the models' surface. Although the projectors can be moved closer to the surface to increase the resolution, this is limited by the minimum focal distance and the size of the objects to be projected on. These requirements decrease the effective resolution which reduces the detail of the projected information. Artefacts such as aliasing and texture filtering appear, and this effects the visual outcome such as reducing the readability of text.

3.2 Hybrid Display Theory

This paper presents our investigations into overcoming these challenges by replacing the single coloured projection surface of SAR objects with an adaptive substrate. By incorporating a projected display and a surface display, both technologies will work in synergy to improve the contrast and resolution.

3.2.1 Adaptive Substrate

Projection surfaces for traditional planar displays and SAR systems are usually a uniform colour and have an evenly reflective surface, such as a timber substrate painted white. We propose the concept of an adaptive adaptive substrate that is a projection surface capable of changing its base colour to influence the appearance of projected images. In our implementation we are using an ePaper display to provide the functionality of the adaptive substrate although other technologies could be put in its place.

One supporting argument of this method is the increasingly flexible nature of ePaper displays. Currently ePaper substrates are either rigid, as seen in ebook readers, or they are flexible allowing some limited bending around a surface or rolling up on itself. We envisage that future ePaper displays will go beyond flexible substrates that can be wrapped around simple shapes such as a cylinder, to allow elastic properties that will allow them to be wrapped around almost any organic shape. However, until this technology is invented and made commercially available, the hybrid display surface that uses both a projector and ePaper technology can provide much of the future functionality. For example consider a car dashboard where the entire surface is a display. Using our approach, the majority of the dash area will be textured using the projected light to provide simple colour details. While the instrument panel would use ePaper in conjunction with the projector to allow the fine annotations and details to be displayed.

The refresh rate of ePaper displays is currently much slower ($\sim 1Hz$) compared to projection technologies. Although this is a limiting factor, there are two reasons our approach is viable. Firstly, animated details can be provided by the projection system for animation while the ePaper is only used for static and non-time critical details. Secondly, we have already seen ePaper displays improve their refresh rates and it is likely they will continue to improve as the technology advances.

3.2.2 Contrast

Display contrast is defined as the ratio between the darkest and the brightest achievable colour. In projection displays, black is achieved by projecting no light at all. However, this leaves the pixel at the surface colour, which is usually a white surface, and is then lit by ambient light.

The adaptive substrate is able to turn individual pixels or regions to darker shades of grey and finally black. This reduces the amount of reflected light and makes the surfaces appear darker. By keeping other parts of the surface at a neutral "white" colour the maximum amount of light is reflected there, therefore increasing the contrast ratio of this display.

Image 2 highlights the difference between contrasts. On the left, the projector alone is displaying the colour black, and as a result the darkest achievable colour is the colour of the projection surface. Image 2b shows the contrast of the combined display. The ePaper display darkens the pixels that should appear black, while keeping the white pixels in the checkerboard at a neutral setting.

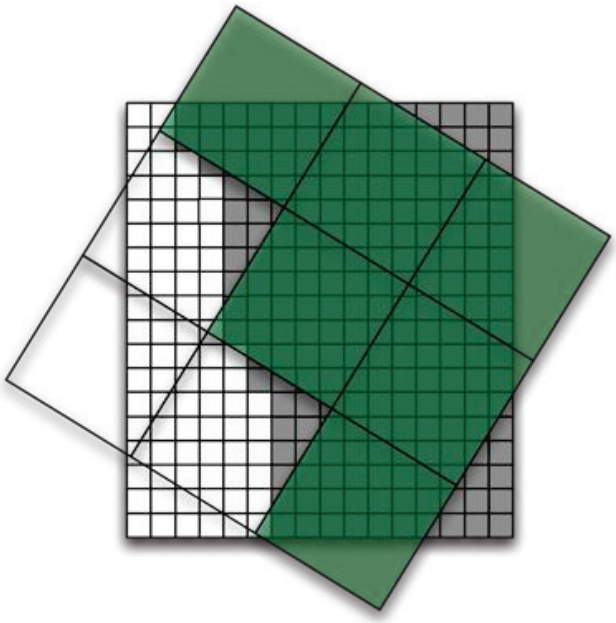


Figure 1: Comparison of different display resolutions. Let the big grid represent the projected pixel size of projectors, while the smaller grid represents the resolution of the substrate. The shaded projector pixel highlights how many smaller pixels are covered and can, on the other hand, influence the appearance of the projector pixel.

3.2.3 Local Resolution

In a SAR context, projectors usually use only parts of their display to illuminate surfaces, that may also be oriented at an oblique, non-optimal angle to the projector. With increasing distance between projector and projection surface, these pixels are also growing bigger in size (and dimmer in light). With a fixed projection surface, an increased distance means less projected pixels on this surface. We can therefore define the effective resolution of a projection surface as the actual count of pixels that are illuminating this surface. While the effective resolution decreases, the information we want to display on the same surface patch, a texture for example, stays the same.

Compared to the projector, the adaptive substrate has a much higher effective resolution on the same surface patch. It will also never decrease since the adaptive substrate is the projection surface itself. The much higher resolution allows modulation of projected light on this surface on a much finer scale than the resolution of the projected image. Controlling such fine details allows to preserve, enhance or even simulate image features and details that would have either been lost or invisible due to the low resolution of the projected image. Finally, the resolution and display of the ePaper device is projection and perspective independent. Where a projected texture would have to be interpolated to project correctly, and therefore is subjected to texture interpolation, the ePaper displays its information flat and perspective independent on the substrate.

A particularly useful aspect of the combined display technologies for SAR systems is the concept of employing an adaptive substrate for specific areas on an object. Using this approach, sections on an object that only require a low resolution may use only the projected technology for most of their appearance while areas such as interactive controls may employ the hybrid display surface to provide optimum performance and detail. Areas of interest can also be defined in this manner, drawing the user naturally to certain surface areas.

3.3 Discussion

Table 1: Comparison of display systems

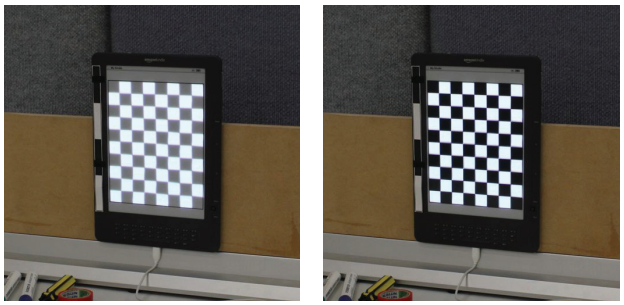
	Advantages	Shortcomings
ePaper	High contrast "True" black High local resolution	No colour Slow refresh rate Limited flexibility
Projector	Colour output High refresh rate Project on any surfaces	High black level Low effective res. Requires line-of-sight
Projector+ ePaper Composite	Colour output High dynamic range Local superresolution True black	Requires registration Inhomogenous refresh rate

By creating a composite display of two very different displays, we are utilising the strengths of each system and at the same time compensating the weaknesses (See Table 1). As can be seen, the ePaper compensates for the projector display in areas of high contrast, true black, and resolution. Project on the other hand compensates for colour range, refresh rate, and ability to conform to complex shapes. The combined effect is an overall improved visual outcome.

4 Appearance Techniques

This section describes techniques that demonstrate how the theory of using an active ePaper substrate can be employed to enhance the contrast and sharpness of projected imagery. All photographs presented in this paper were taken with a high-resolution digital camera at room ambient light. We measured the ambient light in the room with a Digitech QM1586 light meter to be about 200 lx. Preprocessing of the photographs was restricted to cropping and slight distorting, so that they would fit a rectangle. They were not colour processed.

4.1 Improved Contrast and Black Level



(a) Checkerboard projected only (b) Checkerboard on the composite display

Figure 2: Comparing the black level of a projector alone (2a) with the black level of our composite display (2b).

In Figure 2, the black levels of a projector and our composite display are contrasted. On the left, the projector alone is displaying black; the darkest achievable black is therefore determined by the ePaper colour and the ambient light at that point. On the right, the ePaper is also displaying black at the required pixels, achieving a much darker shade of black. Both images were taken at ambient light levels.

4.2 Image Enhancement

In Figure 3, we display the "Lena" image on both the projector and the ePaper substrate. The left column shows the projector-only output, while the right column shows the same output on the composite display. A close-up of the eye in the second row highlights that details like the exact shape and details of the iris or the eye lashes that have been preserved using the composite display. The last row shows an interpretation on how the various pixel sizes and colours are interacting to form the final image. Details, such as fine details in the iris, are preserved in the composite image.

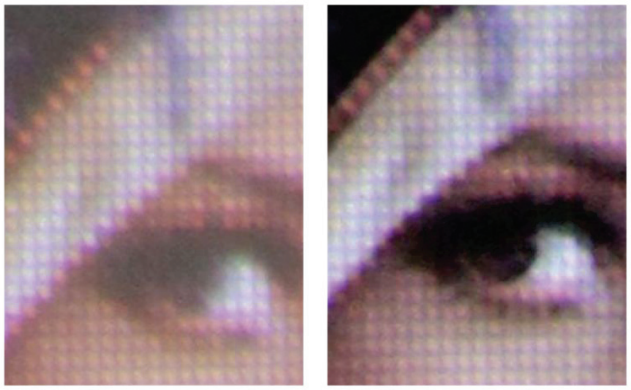
4.3 Text Enhancement

In many SAR applications, there is a need to display small legible text on a surface. This is usually done by displaying a texture which contains the text. However the density of projected pixels is often not adequate to display characters sharply. Additionally, texture filtering limits the details we can effectively reproduce. The ePaper substrate overcomes this problem by providing a very high local effective resolution. Rather than depending solely on the substrate for displaying text, we used a combined rendering techniques to display both black and coloured text using the projection system for colour and the ePaper display for improved sharpness and contrast.

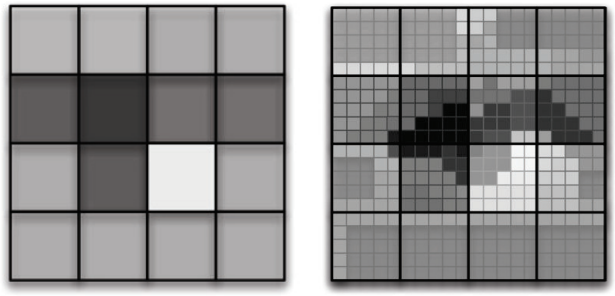
For black text on white and coloured background, high-quality images of the text are drawn on the ePaper substrate. A higher resolution on the substrate allows for smaller and sharper text. Depending on the text size, the projector might "fill in" the letters of the text as well, although it usually is concerned with providing the colour of the text background.

Figure 4 demonstrates this approach in practice. The left column shows the text texture, as it is displayed by the projector. The text is too small to be effectively rendered at that resolution and parts of letters are missing. By displaying the same texture on the ePaper

(a) Lena projected (b) Lena on the composite display



(c) Close-up of the eye in the photo (d) Close-up of the eye in the photo



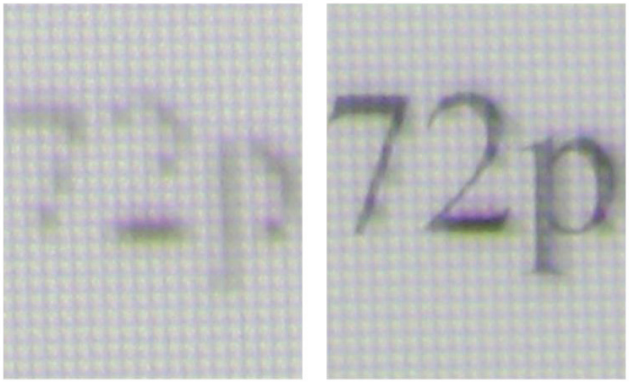
(e) Projected pixel sizes (f) ePaper pixel size

Figure 3: Contrast and detail enhancement for images, demonstrated on Lena.

underneath it, the higher resolution of the ePaper display provides crisp outlines and makes the text readable with much smaller font sizes.

4.4 Animated and Static Displays

The concept of augmenting static images can be further developed using the composite display. Static parts are augmented with the ePaper substrate and dynamic parts are displayed using the projector alone. For example Figure5 demonstrates a simulated instrument display that employs the hybrid display to create an animated instrument gauge. The frame and most details are present in the projected details and also in the ePaper substrate, thus increasing detail and sharpness. Animated parts, such as the needle, are displayed with the projector alone.



(a) Projector close-up (b) Composite close-up

Figure 4: Text enhancement.

With the composite display a compromise must be reached between lower image quality for some parts of the display and the ability to animate such a simulated instrument. However, only the currently very low refresh rate of an ePaper display prohibits us from implementing a fully augmented dynamic instrument.

4.5 Texture effects

The final appearance of the augmented surface can be altered by displaying static textures on the ePaper substrate. The combination of this static texture and the projected image provides an enhanced version of the final image output. This section describes two techniques, detail textures and detail shading.

4.5.1 Detail Textures

An effective example of the hybrid display technology is texture effects that leverage the high resolution of the substrate display to provide the fine-grained details. A projector illuminates the surface with a very low effective or "local" resolution, as most of the projected image falls on other surfaces, the background and so on. Small details, such as fine surface structures and high resolution textures can not be projected as their detail gets lost during due to the low resolution and texture filtering.

A solution is to use the high-resolution substrate display to present the *detail texture* information. This provides detail information of the surface at close distances. Traditionally, these textures were multiplied with the original texture. In our case, we provide this texture on the ePaper substrate surface, while the projector is illuminating with the original image. When the substrate is viewed up close, these details add an extra layer of surface information.

4.5.2 Detail Shading

The idea of detail textures can be extended to include variable lighting. Surface details on the texture are visible because a light source illuminates and shades irregularities on this surface. If the scale of these irregularities get larger, we are able to simulate the general shading of surfaces using bump maps.

In Figure 7 this idea is explored using prerendered bump map images. A virtual light source is moved across the display, and the shading thereof changes accordingly. The ePaper display is responsible for showing the shading, while the projector is displaying an orange tone, simulating the colour of an orange. If specular reflections are required, the projector would display them as well.

(a) ePaper only

(b) Projector only

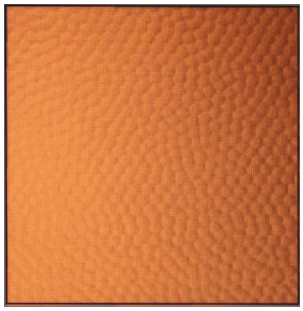
(c) Composite display

Figure 5: An example of an animated display. (5a) shows the ePaper displays output, (5b) the projector output and (5c) the final image on the composite display.

(a) Concrete

(b) Canvas

Figure 6: Demonstration of concrete and canvas detail-textures employing the high resolution ePaper display for details and projector for colour information.



(a) Frame 1

(b) Frame 2

Figure 7: Two bump mapping frames using the composite display. The light source moves from left to right.

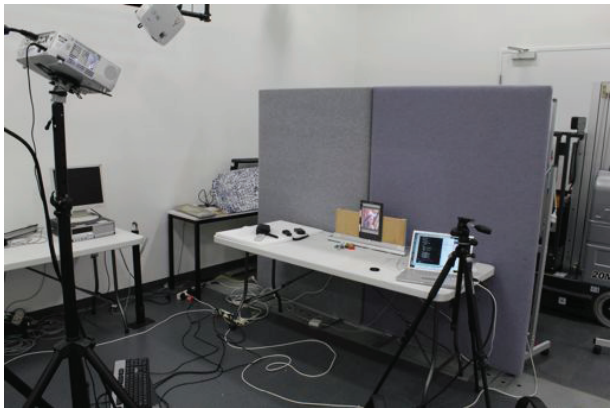


Figure 8: The experimental setup showing a projector on the left mounted at a distance of 2.5-3 metres from the ePaper display.

5 Implementation

We implemented a prototype by using an off-the-shelf ebook reader as one possible implementation of an adaptive surface substrate. A python script, responsible for aligning a displaying a textured quad, was running on a computer, attached to a projector. Projected images were generated with OpenGL.

5.1 Prototype Display

We used an Amazon Kindle DX with an eInk "Pearl" display. Its display has a physical size of 10.4" by 7.2" and a physical reso-

lution of 825x1200 pixels. Each pixel can display 16 shades of grey; intermediate values are displayed using dithering. The Kindle's operating system displays PDF documents rasterised and has a built-in image viewer. The display itself has has very low reflectiveness, similar to painted wood we use for the construction of our other prototypes for SAR objects. The projector we used is a NEC 501W LCD projector with a resolution of 1280 x 800 pixels. It is driven by a computer over a DVI connection.

This setup is used as a technology demonstrator and early prototype and we intend to update as more feature rich hardware is available. In this setup the Kindle has some shortcomings, for example it is not a directly controllable display. The image and PDF viewer are able to display data on the screen although direct access to the display would be more optimal. Using this technique, images get stretched if they are not the correct resolutions. The refresh rate for full screen images was lower then 1 Hz.

5.2 Setup

We built a solid stand to which the Kindle was firmly attached using double-sided tape. The stand and Kindle was placed on a desk at roughly 2.5 metres distance from a projector (see Figure 8). This represents a typical case of *Desktop SAR*, in which a model is placed on a desktop and lit on all sides by projectors mounted on a frame or gantry around it. The projector was controlled by an attached computer, while the Kindle display was controlled by uploading generated images into an "ebook" representation and viewed in the Kindle's image viewer.

To create a usable composite display, the projectors image requires registration with the eInk display to ensure that the pixels of the different output devices overlay each other. As described in Section 5.1, using the image viewer is problematic, as we found it to be unreliable in clearing and refreshing the screen and displaying series of images with similar grey tones. Through experimentation we found out that only some images at a certain resolution are displayed correctly. We therefore defined a "valid" area inside the Kindle's display, and drew a border around it. This area's resolution was 800x1000 pixels, and it was framed by a 5 pixel black border. All of our generated images had this black border, allowing easy alignment between the eInk and the projector image.

To align a projected quad to this valid frame a quad was projected. Its four corner vertices were then aligned using mouse selection and keyboard commands and their positions can be stored in a file. As long as neither the projector nor the Kindle moves, the coordinates can be read from the file and the alignment step therefore has to be performed only once.

Images are displayed by texturing this aligned quad. For animated displays, the texture is created by rendering the animation into a framebuffer object and binding it as a texture. It must be noted that the effective resolution of the projector in this valid area usually quite low and even textures of the size 512x512 pixels have to be filtered.

5.3 Contrast

We measured contrast in an ambient lit room, with a ambient light of roughly 250 lx. For contrast measurements, we were displaying the ANSI contrast pattern, a 4x4 black-and-white checkerboard pattern. The illuminance of all the white and all the black areas was measured and averaged using a calibrated spectrometer in absolute irradiance measurement mode. Based on those values, the final display contrast is calculated by dividing the average white by the average dark illuminance. All the measured values were rounded to the next integer. Table 2 shows the results of the calculations.

In the first instance, we measured the contrast of the projector alone projecting on the empty Kindle surface, in the second we projected the ANSI pattern on a displayed ANSI pattern. The comparatively low contrast ratio in the first case can be explained by the display surface of the Kindle. It is not perfectly diffuse white but rather grey, this lowers the amount of reflected light.

Table 2: Contrast measurements for a single projector and a combined adaptive substrate display.

	Avg. black	Avg. white	Contrast
Projector alone	269 lx	3722 lx	$\sim 14:1$
Kindle + Projector	50 lx	3720 lx	$\sim 74:1$

6 Limitations

One shortcoming is the ePaper we employed does not provide direct access to the display. We are using an eBook reader (with an unsupported image viewer) to display our images. However, we have little control over how the images are being displayed. Image scaling, cropping and dithering were the results of incompatible images. As previously mentioned, we have no direct control, displaying a series of images meant creating an image series in advance and manually flipping to the next entry in this series. Manual operation might move the device however, which could invalidate the projector - eInk display registration.

Secondly, our current prototype has a very limited refresh rate of about 1Hz. It is not possible to display moving or animated images. While the sixteen grey levels are appropriate for many applications, they lack the finer control needed to compensate for complex lighting problems, like overlapping projection areas.

One of the major advantages of using SAR for early prototyping is the ease and low cost to build models to project on. Incorporating eInk devices as surface replacements would possibly increase the complexity and price of such prototypes. A solution would then be to use eInk displays only on selected, important parts of such a model. Finally, all these displays would be wired up to a central controller, requiring additional cable connections and control or video outputs. Fortunately, ePaper displays have a very low power requirement, so expensive and heavy power equipment is not required.

7 Future Work

7.1 Adaptive Projection Screens

Spatial Augmented Reality and multi-projector display walls have areas, in which multiple projections are overlapping. In these overlapping areas, the brightness is increased as multiple projectors, instead of one, are illuminating a surface. Using adaptive substrate, one could compensate for this over-illumination, by lowering the brightness of the surface in this intersection area only. These overlapping areas can be determined through matrix decomposition (if the projector's extrinsic and intrinsic matrices are known) or through camera-projector systems. The current limitation of only 16 shades of grey prevents our system from creating effective compensation levels. Additionally, as we can only control the blackness for now, we cannot compensate for different projector white balances, as each projector might contribute a different colour to this overlapping area.

7.2 Subpixel Antialiasing

A composite display could have an inter-display look-up-table, that maps pixel coordinates from one device to another (for example through automatic, structured light registration techniques). Having now subpixel-correct control of the higher-resolution display, we are able to perform antialiasing by decreasing the darkness level of single pixels. Text, lines or silhouettes can be smoothed using this method. This also requires direct control of the ePaper display.

7.3 Improved eInk

Our description of the composite display is a RGBK display system. We were using a greyscale ePaper display for our prototype. However future versions of eInk displays will be coloured and possible provide RGB control.

Using RGB control to describe the surface of an object allows us to reformulate the surface description of the rendering system. While we now split between colour and brightness (RGB and black), it will be possible to split the rendering in surface description (colour + black) and a lighting description. The surface description will contain the diffuse texturing and the black detail shading, while the coloured lighting by the projector will describe the lighting, the specular effects and so on. Part of the rendering equation[15] will take place by shining coloured light onto the coloured surface substrate.

Finally, an eInk display with a higher refresh rate does not change the techniques we presented so far. It might be interesting for displaying animations, as the surface change of the display can be synchronised with the refresh rate of the projection display, thereby creating a high-dynamic range composite display.

7.4 Adaptive Substrate Props

In our SAR environment, we are exploring methods of projecting on simple props and how add detail to these models by intricate surface simulation. For example, instead of creating detailed physical models of control panels, we build simple, almost box-shaped models and use SAR to enhance their surface so that the end result looks like a real control panel.

Although these models are currently painted matte white for best projection properties, we envision that bendable eInk displays will be used as surface coating for future models. They have many advantages compared to traditional displays, like their reduced power consumption, lower weight and their flexibility. For a future SAR system, a number of geometrically simple shapes, like boxes, cylinders or cones could be built using eInk surfaces as adaptive substrates instead of simply painting them white.

8 Conclusion

This paper described a new form of hybrid display using both a digital projector and eInk together to create a single composite display surface for use in SAR systems. Replacing the commonly employed uniform white projection surface of SAR objects with our composite display has allowed us to develop a series of theoretical approaches that define how this form of display can be employed. This approach significantly enhances the contrast ration and sharpness appearance of SAR objects. We have implemented image enhancement techniques that are built around two core concepts provided by this new method: effective resolution and a lowered black level. The results we achieved by controlling the adaptive substrate to work in synergy with a digital projector have significantly improved the SAR object appearance.

References

[1] D. G. Aliaga, A. J. Law, and Y. H. Yeung. A virtual restoration stage for real-world objects. In *ACM SIGGRAPH Asia 2008 papers*, SIGGRAPH Asia '08, pages 149:1–149:10, New York, NY, USA, 2008. ACM.

[2] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. *Computer Graphics and Applications, IEEE*, 21(6):34–47, 2001.

[3] O. Bimber. Multi-projector techniques for real-time visualizations in everyday environments. In *ACM SIGGRAPH 2006 Courses*, SIGGRAPH '06, New York, NY, USA, 2006. ACM.

[4] O. Bimber, F. Coriand, A. Kleppe, E. Bruns, S. Zollmann, and T. Langlotz. Superimposing pictorial artwork with projected imagery. In *ACM SIGGRAPH 2005 Courses*, SIGGRAPH '05, New York, NY, USA, 2005. ACM.

[5] O. Bimber and D. Iwai. Superimposing dynamic range. In *SIGGRAPH Asia '08: ACM SIGGRAPH Asia 2008 papers*, pages 1–8, New York, NY, USA, 2008. ACM.

[6] O. Bimber, D. Iwai, G. Wetzstein, and A. Grundhöfer. The visual computing of projector-camera systems. In *SIGGRAPH '08: ACM SIGGRAPH 2008 classes*, pages 1–25, New York, NY, USA, 2008. ACM.

[7] O. Bimber and R. Raskar. *Spatial augmented reality: Merging real and virtual worlds*. AK Peters Ltd, 2005.

[8] B. Comiskey, J. D. Albert, and J. M. J. H. Yoshizawa. An electrophoretic ink for all-printed reflective electronic displays. *Nature*, 394:253–255, July 1998.

[9] N. Damera-Venkata, N. Chang, and J. DiCarlo. A unified paradigm for scalable multi-projector displays. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1360 –1367, nov.-dec. 2007.

[10] N. Damera-venkata and N. L. Chang. On the resolution limits of superimposed projection. In *In Proc. IEEE International Conference on Image Processing (ICIP*, 2007.

[11] H. Gates, R. Zehner, H. Doshi, and J. Au. 31.2: A5 sized electronic paper display for document viewing. *SID Symposium Digest of Technical Papers*, 36(1):1214–1217, 2005.

[12] M. D. Grossberg, H. Peri, S. K. Nayar, and P. N. Belhumeur. Making one object look like another: Controlling appearance using a Projector-Camera system. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–452–I–459 Vol.1, 2004.

[13] J. Hare, S. Gill, G. Loudon, D. Ramduny-Ellis, and A. Dix. Physical fidelity: Exploring the importance of physicality on Physical-Digital conceptual prototyping. In *Human-Computer Interaction – INTERACT 2009*, pages 217–230. 2009.

[14] H. Hoffman, A. Hollander, K. Schroder, S. Rousseau, and T. Furness. Physically touching and tasting virtual objects enhances the realism of virtual experiences. *Virtual Reality*, 3(4):226–234, 1998. 10.1007/BF01408703.

[15] J. T. Kajiya. The rendering equation. *SIGGRAPH Comput. Graph.*, 20(4):143–150, 1986.

[16] F. M. Meyer, T. L. Trissell, D. L. Aleva, S. J. Longo, and D. G. Hopper. Readability evaluation of an active matrix electrophoric ink display. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 6225 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, June 2006.

[17] A. Olwal and S. Feiner. Spatially aware handhelds for high-precision tangible interaction with large displays. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction*, TEI '09, pages 181–188, New York, NY, USA, 2009. ACM.

[18] A. Pavlovych and W. Stuerzlinger. A high-dynamic range projection system. In *Photonic Applications in Biosensing and Imaging, Proceedings of SPIE 5969*. SPIE—The International Society for Optical Engineering, 2005.

[19] R. Raskar and K. Low. Interacting with spatially augmented reality. In *Proceedings of the 1st international conference on Computer graphics, virtual reality and visualisation*, pages 101–108, Camps Bay, Cape Town, South Africa, 2001. ACM.

[20] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs. High dynamic range display systems. *ACM Transactions on Graphics*, 23:760–768, 2004.

[21] C. Ware and J. Rose. Rotating virtual objects with real handles. *ACM Trans. Comput.-Hum. Interact.*, 6(2):162–180, 1999. 319102.

[22] G. Wetzstein and O. Bimber. Radiometric compensation of global illumination effects with projector-camera systems. In *ACM SIGGRAPH 2006 Research posters*, SIGGRAPH '06, New York, NY, USA, 2006. ACM.

Occlusion Handling and Image-based Lighting using Sliced Images in 3D Photo Collections

Frank Nagl* Konrad Kölzer† Paul Grimm‡

Fulda University of Applied Sciences, Germany

Figure 1: (a) 3D Photo Collection with embedded object, (b) Close-up of unlit object with wrong occlusion. (c) Photo divided into 3D segments (Sliced Image). (d) Convincing augmentation with our occlusion handling and image-based lighting features.

ABSTRACT

This paper presents novel methods to handle potential occlusion problems and to render realistically the lighting of embedded virtual 3D objects in photo-based 3D worlds. These 3D worlds are called 3D Photo Collections. The first focus of this work handles potential occlusion problems between image parts of photos and the embedded virtual objects by dividing photos into 3D segments, which we call *slices*. The segmentation of the photos is done by a gradient-based contour tracing algorithm, which divides a photo into several contours. These contours are used in the 3D world as slices. Each slice will be moved to a depth position estimated by the provided spatial interdependency between photos of the 3D Photo Collection. This results in several slices at different depth positions per photo. We call the divided photo *SlicedImage*. The second focus of this paper is concentrated on the realistic rendering of the illumination of embedded virtual objects by extracting the environmental lighting of the real scene from the generated Sliced Images.

For fully automated generating of image-based 3D worlds, 3D Photo Collections (generated by software like Microsoft Photosynth or Google Street View) are used. These 3D Photo Collections provide spatial interdependency between photos, which will be used as input data for our algorithms to compute the Sliced Images.

Keywords: 3D Photo Collection, Occlusion, Sliced Image, Depth Map, Image-based Lighting, Augmented Reality

Index Terms: H.5.1 [Information Systems]: Multimedia Information Systems—Artificial, augmented, and virtual realities

1 INTRODUCTION

Augmented Reality can be used in applications of urban planning and interior design to combine virtual products and real environments for a convincing impression. A well-established way of Augmented Reality is the augmentation of photos. In contrast to the usage of a single photo, an image-based 3D world can be built by shaping the real environment and matching a set of photos into it. This assumes the scene geometry measures and good experience with Photoshop or 3D modeling tools as well as a high effort to build an image-based Augmented Reality world. Another way is the usage of a structure-from-motion software for unordered image collections like Photosynth [20] or Google Street View [15]. A Structure-from-motion tool processes a set of unordered images and automatically provides intrinsic and extrinsic camera parameters as well as a sparse 3D model of the real scene as a key point cloud. The outcome of this is an image-based 3D world, which we call 3D Photo Collection. Information about the scene geometry or good experiences with special 3D modeling tools are not necessary. Also, arbitrary cameras are usable instead of a specialized hardware (e.g. see-through displays or tracking devices).

Owing to these advantages, our work uses 3D Photo Collections as base AR environment. In these 3D Photo Collections virtual 3D objects will be embedded. In this connection the visual appearance of virtual portions needs to be very authentic for the user to ensure a convincing AR impression. This requires to deal with a number of typical open issues in AR environments like

- occlusion problems of virtual parts by real parts, and
- illumination of virtual parts consistent with real parts.

This paper is focused on handling potential occlusion problems between image parts of the photos and the embedded virtual objects by dividing photos into 3D segments, which we call slices. Each slice will be moved to a depth position estimated by the provided spatial registration between photos. This results in several slices at different depth positions per photo. We call the divided photo Sliced Image. The second focus of this paper is concentrated on the realistic rendering of the illumination of embedded virtual objects by extracting the environmental lighting of the real scene from the generated Sliced Images. For lighting purposes, the images of the 3D Photo Collection should be shot with varying light exposure.

This paper is organized into 5 chapters. The next chapter is engaged with related works to this paper. Chapter 3 describes the concept of occlusion handling and image-based lighting of virtual products in 3D Photo Collections. After that, chapter 4 discusses the implementation as well as our results. The last chapter summarizes the content of this paper and gives some information about potential future work topics.

*e-mail: frank.nagl@hs-fulda.de
†e-mail: konrad.koelzer@hs-fulda.de
‡e-mail: paul.grimm@hs-fulda.de

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

2 RELATED WORK

This chapter is separated into the three issues

- 3D Photo Collections,
- occlusion handling, and
- image-based lighting.

Generating **3D Photo Collections** from unordered image sets is automatically performed by structure-from-motion algorithm tools like Bundler[29] or Photosynth [20]. Bundler first calculates image features for each photo using the SIFT algorithm[19] and then performs a pairwise image matching. After that, the intrinsic and extrinsic camera parameters are reconstructed for each image by using a modified version of the Sparse Bundle Adjustment package of Lourakis and Argyros[18]. The output of Bundler also contains sparse scene geometry as a set of 3D key points. Additionally, for each key point Bundler determines a color and the set of cameras that use this point as image feature. Based on Bundler, our IP3D framework was presented in [22]. This framework provides a generic architecture for fast and robust development of applications using 3D Photo Collections to build authentic augmented 3D worlds. Also in this work, the IP3D framework is used as the underlining 3D Photo Collection viewer.

In [14] it is shown how the spatial data structures of 3D Photo Collections will be used to realize a ambient view interpolation of foreground objects shown in several photos [14].

Several approaches for **handling occlusion** issues in the context of Augmented Reality with 3D Photo Collections exist. The patch-based multi view stereo algorithm of [12] and the 3D reconstruction algorithm of [30] show how 3D Photo Collections can be used to reconstruct 3D geometry from objects, which are depicted on many photos. These 3D objects handle possible occlusions of other virtual objects in an augmented 3D Photo Collection. The requirements for these works are a large quantity of photos, the foreground objects have to be emphasized considerably from the background in the photos and the algorithms do not accept arbitrary photo scenes. In conclusion, these approaches are too restrictive for our proposed ideas.

A depth map can be used to extract foreground and background objects in photos. Further work dealing with estimating depth maps from photos is discussed in [3]. A foreground object in single photos is extracted by estimating a depth map. Therefore, a special lens aperture with an integrated RGB filter is used to encode three different grayscale images into one three-channel photo. Another work [31] estimates the depth positions of foreground objects in a video stream by using a stereo view algorithm. Two exact calibrated cameras are needed for the stereo effect. Both approaches do not work with arbitrary photos and require special hardware. As a consequence, they cannot be used in our approach. In our poster [23], we presented a first approach for generating Sliced Images by using 3D Photo Collections. Every 3D key point is projected into the photo and the distance between camera position and key point is used for the color value of the pixel. Every 2D projected key point is used as cell nucleus for a Voronoi Diagram [1]. A depth map will be generated by applying the Voronoi diagram. In contrast to this first approach, our new approach in this paper provides an improved version of generating the depth map for building the slices. Instead a Voronoi diagram, a contour tracing algorithm divides the photos into segments. 3D key points which belong to a segment are used for estimation the depth of this segment (see section 3.1).

The next approaches deal with the issue **image-based lighting**. Fournier et. al. [11] presented differential rendering to embed virtual objects with correct shadows into a real scene. This is done by performing an approximate manual reconstruction of geometry, camera parameters and lighting conditions of the real scene.

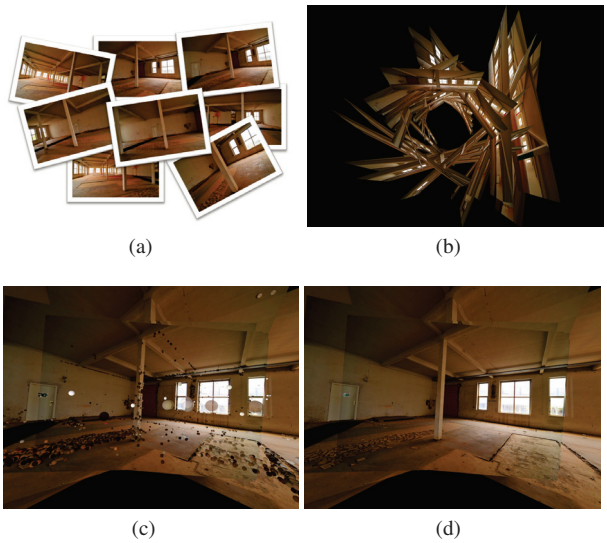


Figure 2: (a) Unsorted photos. (b) Matching photos together to a 3D Photo Collection, seen from top view. (c) 3D Photo Collection with key points. (d) Generated 3D Photo Collection, seen from front view.

Then the scene is rendered once without and once with virtual objects. The difference is then added to the photo to produce realistic shadows. Another approach was presented by Debevec in [9]. Debevec divides the real environment into distant and local scene. Light sources and geometry of the local scene are manually approximated. The lighting of the distant scene is described as a high dynamic range 'omni-directional Radiance-Map', which is provided by a measured light probe. The usage of light probes was later adapted to real-time rendering, by extracting few directional light sources from the environment map [10, 8].

A novel approach of Gibson [13] also uses light probes, which are projected on the coarsely reconstructed geometry of a real scene. The geometry is subdivided into patches and Inverse Radiosity is used to calculate the diffuse factors for each patch. Additionally, an Irradiance-Volume is precalculated that stores the environmental lighting for each point on a 3D grid as spherical harmonics coefficients and is used to perform real-time capable lighting. This was later extended by Grosch in [16] by generating shadows from direct illumination. Both approaches give good results with realistic shadows, however they also share the restriction to panorama scenes: The light of the real scene is measured by a light probe, which only specifies lighting information for scene parts that are not occluded from the viewing position of the light probe.

In [17] and [23], we presented another approach for image-based lighting by using 3D Photo Collections. Each pixel of an image is treated as a ray of light that is either emitted or reflected from a point on a surface of the real scene. In contrast to this paper a much simpler depth reconstruction of the image pixels is used: For each photo a single depth value is estimated and used for all pixels which is a extremely rough reconstruction.

3 APPROACH

The main goal of our work is to increase the visual quality by addressing occlusion problems and image-based illumination in augmented 3D Photo Collections for convincing urban planning and interior design. A 3D Photo Collections viewer is used to visualize the reconstructed real environment (see fig. 2) and 3D geometries are given for embedding virtual objects.

To address the problems of our main goal, the viewer has to provide features for occlusion handling and for image-based lighting.

These features will be described in the following subsections.

3.1 Occlusion Handling

Using augmented photo collections some image parts of photos may occlude parts of the embedded virtual objects and vice versa. Since occlusions are essential for human visual perception [4][2] and a faulty visualization or an incorrect occlusion handling destroys the Augmented Reality experience [25], occlusion handling in AR applications is a fundamental aspect. Therefore, a photo has to be divided in separated regions. Next, the correct rendering order of foreground image parts, virtual objects and background image parts has to be considered. Thus, three issues are important:

- Separation of foreground and background image parts,
- estimation of 3D depth position of separated image parts, and
- occlusion of embedded virtual objects by image parts.

Separation of foreground and background image parts The photos have to be divided and segmented. For this purpose we use a novel contour tracing mechanism. In contrast to other contour tracing and edge detection algorithms our contour tracer results in closed contours with the knowledge which pixel belongs to which contour. We presented in [24] a previous version of this contour detector. For better clarity of the overall process, we describe the contour detector in detail.

By applying an edge detection to the image (converted in gray scale), the gradient and consequentially the orientation are known for each pixel [7]. Based on this knowledge a three-stage tracing algorithm is applied. For each pixel its neighbor pixels will be analyzed in a specified radius, if they are

1. edge pixels, when this returns no results then
2. own contour pixels, when this returns no results then
3. other contour pixels.

The analysis algorithm for all three steps searches for the neighbor with the lowermost match value:

```
if (Neighbor has same orientation)
  match = Max(distanceX, distanceY)
else
  match = Max(distanceX, distanceY) + weight
```

The lower the *match* value is the higher the significance of the neighbor for the contour is. The parameters *distanceX* and *distanceY* represent the distances between the pixel and its neighbor on the x- and y-axis in pixel coordinates. For calculating *match* the bigger value of the distances in x- and y-direction is used instead the euclidean distance. This is done to fill the match value matrix with integers and these computations are more efficient with the same result. Parameter *weight* is a specified constant value to privilege neighbors with same orientation. The higher this value is set the higher is privilege of neighbors with same orientation. This value is not set automatically, because different requirements at the content of photos are controllable by setting this parameter. For example a photos of buildings with many squared objects should get a high *weight* value in contrast to a photo with rolling contours. So the user has an influence on the algorithm by controlling this parameter manually.

Following example illustrates the analysis results for a red line in an image. The pixel p belongs to a diagonal red line and is the current pixel to analyze. In this example the *weight* constant is set to 5 to privilege very strong neighbor pixels with same orientation.

The resulting match value matrix shows that the direct neighbor top right has the lowermost match value ($match = 1$) and hence the

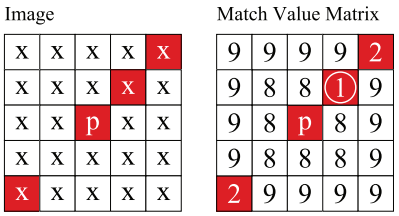


Figure 3: Illustration of calculated match values: Initial image with a diagonal red line and the corresponding match value matrix.

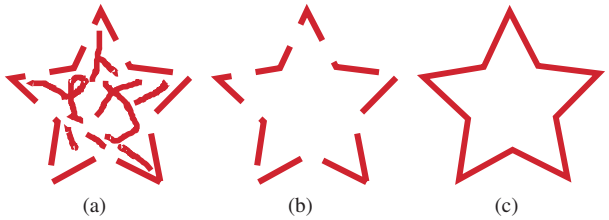


Figure 4: Getting a closed contour: (a) All pixels of one contour after analysis algorithm. (b) Detected borderline of the contour. (c) All borderline pixels connected by a polyline to a closed contour.

biggest significance. Consequential this neighbor becomes part of the contour of pixel p and the analysis starts from the neighbor's position again. This recursive process is done until all edge pixels belong to a contour.

The next step removes all contour pixels which do not belong to the contour's borderline (see fig. 4(b)). This is done by detecting the first and the last contour pixel per line as well as per column of the image. Finally, all detected pixels will be connected and drawn as a polyline [6] (see fig. 4(c)). This results in closed contours with the knowledge which pixel belongs to which contour. Fig. 4 shows the several steps to get a closed contour of a star-shaped object.

Estimation of 3D depth position of separated image parts After contour tracing, photos divided into image parts (contour segments) are given. The next step estimates the correct 3D depth position for every segment. The 3D depth position of a segment is defined by the distance between the segment and the 3D camera position of the segmented photo. The resulting values will be stored and visualized in a depth map.

Our approach for estimating the depth map uses all 3D key points which belong to a segment and calculates their *distances* to the corresponding 3D camera position. For evaluating, which 3D key points belong to a 2D segment, all 3D key points have to be projected into the 2D photos. Fig. 5 visualizes this projection and shows their *distances* to the 3D camera position.

With the knowledge of 2D image coordinates of the projected key points combined with the knowledge of each key point image coordinate belongs to which contour segment, the depth values of a segment can be estimated. Starting from each key point image coordinate a 2D flood-fill process is started until the fill area touches the segment's borderline or another fill area of a further key point. Thus, every key point produces a flood-fill area inside a contour.

Fig. 6 shows a draft with all key points related to a segment for the distance calculation.

Occlusion of embedded virtual objects by image parts The depth position of every image part (contour segment) is known. To answer the question, how image parts occlude the virtual objects, the image parts have to be moved to their calculated depth. This results in many slices per image, which we call **Sliced Image**. Fig. 7 shows the construction of a Sliced Image.

After considering occlusions, the next section deals with the image-based lighting of virtual objects in 3D Photo Collections.

3.2 Image-based Lighting

The generated slices of each image are a very sparse representation of the real scene's geometry. In this section we discuss how they are used to estimate the lighting conditions of the real scene.

Looking at the given data, image pixels can be identified as the smallest piece of available information. As shown in fig. 7, each image pixel provides an RGB color value, an incident direction (derived from its known camera view frustum) and a roughly estimated depth value. The challenge is to combine all this information from all image pixels to a meaningful representation of light conditions of the real scene.

Collect lighting information From now on collecting the environmental lighting of the scene shall be simplified to collect lighting that is received by a point of interest in the scene. For small sized virtual objects, the environmental lighting can be collected only for a single point (usually the object center). For larger objects, light can be collected from multiple points around the virtual objects. The latter case gives better results, but the extracted lighting has to be interpolated between points for final rendering.

The color value of an image pixel represents a certain amount of radiance that was reflected by a surface of the real scene onto the CCD chip of a digital camera. With the known direction and the estimated depth of a pixel, this point on the surface can be located roughly.

The scene surfaces are considered as diffuse (which is true for most surfaces in indoor scenes), so each pixel of an image does not only provide information for its camera position, but also for other positions in the scene, as long as it is not occluded by other surfaces of the scene.

From the above observations a colored ray can be constructed for each located surface point that is visible for the point of interest. Each ray gets the RGB color of its associated image pixel. The set of colored rays represent all available environmental lighting information for the point of interest.

Recovering relative radiance The RGB color of each ray cannot be used for lighting, because it is subject to clamping errors – high radiance is clipped to 100% white – and image noise – low radiance results in a more or less visible grain depending on the quality and the settings of the digital camera. Also the RGB intensities are not linear to the original radiance that was received by the camera lens due to camera specific non-linear transformations.

However, the color information can be linearized and extended to a high dynamic range relative radiance value, because the same point of a surface is usually shown on multiple images, which are

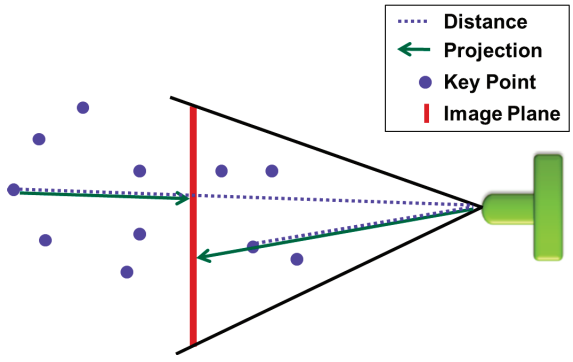


Figure 5: Projection of 3D key points to 2D image coordinates with their distances to the corresponding 3D camera position as pixel values.

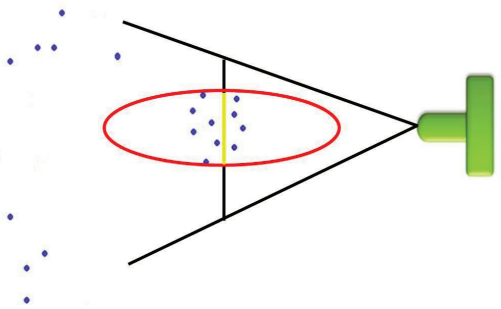


Figure 6: Red encircled 3D key points belong to the yellow 2D segment.

shot with varying light exposure. In other words the ray set contains many rays with identical direction, but varying RGB color.

The linearization is done by estimating the camera response curve using Robertson's algorithm [26]. Extending multiple rays to a ray with a relative radiance value can be calculated by adapting Robertson's HDR image estimate from image pixels to colored rays.

It is calculated as the weighted average of each ray r that shares the same direction. Each ray has the color value y_r the exposure time t_i originating from its associated image. The weighting function w returns the significance of color value and the camera response curve I linearizes the color value:

$$r_p = \frac{\sum_r w_{y_r} t_i I_{y_r}}{\sum_r w_{y_r} t_i^2} \quad (1)$$

Using this equation, a new set of rays is calculated that represents the known incident relative radiance for the point of interest.

4 REALIZATION AND RESULTS

To implement a 3D Photo Collection viewer, which provides mechanisms for occlusion handling and image-based lighting of virtual objects several toolkits are used. The structure-from-motion software Bundler [27] is used to extract the photos viewpoint (intrinsic and extrinsic camera parameters) of all images in the photo collection. Furthermore, a point cloud with all 3D key points is extracted, which is used to calculate the distance from camera to the image plane. The IP3D framework [22] renders the photos and key points in correct 3D orientation and position. For processing images the SBIP framework [21] is used for GPU-based performance improvements.

The solutions of our concept for a convincing augmentation are implemented in our 3D Photo Collection viewer. Fig. 8 shows

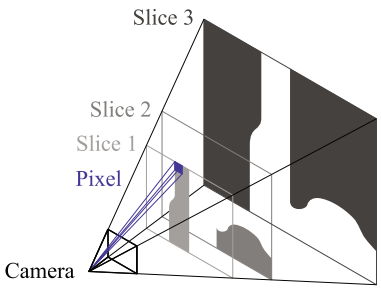


Figure 7: Pixel of a slice with known color, depth and incident direction.

(a)

(b)

Figure 8: A 3D Photo Collection with an embedded virtual chair: (a) Default augmentation without increasing the visual quality. (b) Augmentation with our occlusion handling and image-based lighting features.

the comparison of a 3D Photo Collection with an embedded virtual chair by standard augmentation as well as augmentation with our features. Fig. 9 shows the same scene with the virtual chair from a closer range to emphasize our results. The following subsections describe the implementation features and discuss the results.

4.1 Occlusion Handling Feature

The occlusion feature realizes a correct occlusion handling for every camera view (see comparison in fig. 9). For separating foreground and background image parts, a novel contour tracing algorithm is implemented, as described in section 3.1. The canny edge detector [5] is used for edge detection. The resulting contour segments are stored with all necessary information like pixel coordinates of their borderlines. Fig. 11 shows the result of our contour tracer.

The next step calculates the depth map (see fig. 12). The distances between the 3D key points related to segments and the 3D camera position are stored as depth values for the segments. For determining the relation of key points to segments, all 3D key points are projected on the 2D photos. For each camera the focal length (f), the rotation matrix (R) and the translation vector (t) is given from the IP3D framework. The following equations projects a 3D

(a)

(b)

Figure 9: Picture detail of fig. 8 with the virtual chair from a closer range: Comparison again (a) without and (b) with our occlusion handling and image-based lighting features.

key Point (X) into 2D pixel coordinate (p'):

$$P = R \cdot X + t \tag{2}$$

$$p = \frac{-P}{P.z} \tag{3}$$

$$p' = f \cdot p \tag{4}$$

(2) is to convert from world to camera coordinates, (3) represents the perspective division and (4) is the conversion to pixel coordinates [28]. In the current implementation a reviewing process for potential consolidation of segments by several conditions is not realized yet.

Finally, for occluding parts of the embedded virtual objects by image parts of photos and so handling the correct occlusion, the

Figure 10: Sliced Image occludes partial a virtual chair by a pillar in a 3D Photo Collection (seen from top view). Every slice is drawn with a white border to emphasize the multiple image planes.

(a) (b)

Figure 11: Our contour tracer: (a) Original image. (b) Contour image with colored segments to emphasize the result.

Sliced Image will be generated and rendered. This process is realized by using shader programs to ensure a real-time capable rendering of the scene. After contour tracing and generating the corresponding depth map of an image, a collection of contour segments and the correct depth position of every segment is given. For rendering the Sliced Image, a new image (slice) for every segment is generated. In this slice only the corresponding segment is opaque, all other image parts are transparent. Fig. 10 shows a Sliced Image and a partial occluded virtual chair in a 3D Photo Collection.

The next subsection describes the implemented image-based lighting using the Sliced Images and discusses the results.

4.2 Image-Based Lighting Feature

The image-based lighting feature applies a realistic illumination on embedded virtual object (see comparison in fig. 8 and 9). As well as the rendering of Sliced Images, the collection of environmental lighting is implemented as shader programs that are available on modern graphics hardware and lead to a shorter processing time.

In our render setup floating point cube maps are used as environment maps to represent the set of rays for the point of interest. Each texel represents a solid angle and stores the relative radiance value that is accumulated from the colors of all rays within the texel's solid angle. Each of the six cube map sides is handled as the image plane of a camera that is located at the point of interest. Depending on the cube map side, the camera is facing in the +X, -X, +Y, -Y, +Z or -Z direction.

All rays lying inside the texel's solid angle will be projected on the same texel. This is accomplished by simply rendering each Sliced Image with a three-pass shader into the cube map side.

The first pass evaluates for each pixel of the cube map side the denominator of equation 1 and the second pass evaluates the numerator. Additive blending is used to implement the sum of weighted color values. Results of the first and second pass are rendered into a temporary texture and the final pass calculates the quotient of both for each pixel.

As seen in fig. 13, the result is a cube map that contains all available environmental lighting for the point of interest as relative radiance values, although the cube map contains artifacts due to the very rough estimation of depth values.

(a) (b)

Figure 12: Estimating the depth map: (a) Original image with 3D key points. (b) The corresponding depth map as gray-scale image.

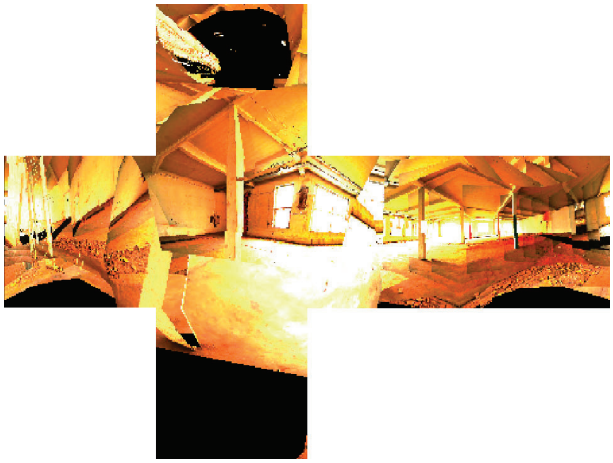


Figure 13: Rendered Cube Map describing the environmental lighting for the point of interest

Once the cube map has been rendered it can be used for various rendering techniques. For instance, it can be used to render specular surface by using reflection mapping. By applying a blur filter on the cube map contents, specular materials with different levels of shininess can be simulated.

In order to perform real-time capable illumination of 3D objects, the cube map's lighting information needs to be transformed to a simpler representation. As seen in fig. 14, our implementation extracts 16 directional light sources from the environment map using Debevec's approach[10]. These directional lights are used to illuminate the embedded virtual object. Additionally, the brightest light source is used to render the shadow of the virtual object.

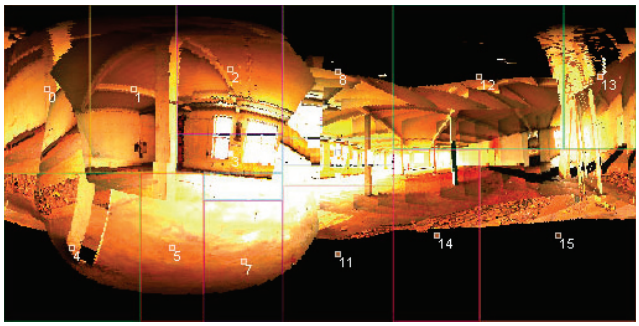


Figure 14: Cylinder projection of cube map with extracted directional lights

4.3 Run-time behavior

This subsection discusses the processing times for two exemplary 3D Photo Collections with 27 as well as 47 JPEG images. The format of the images is 24 bpp and their resolution is 2144x1424 pixel. A machine with an Intel Core i7 CPU 2.67 GHz, 6.00 GB RAM and two NVIDIA GeForce GTX 285 was used. The generation of the depth maps takes 397.2 seconds for the scene with 27 images and 676.579 seconds for the scene with 47 images. The rendering of sliced images is done in real-time. Due to its massive parallelization the generation of cube-maps takes not more than 0.358 seconds for the scene with 47 images. The run-time for the extraction of directional lights from the cube-map does not depend on the scene complexity, only on the resolution of the cube-map. In our setup

	Scene 1	Scene 2
Scene Complexity	27 images	47 images
Depth Map Generation	397.2 sec.	675.6 sec.
Cube Map Generation	0.188 sec.	0.358 sec.
Directional Light Extraction	0.321 sec.	0.323 sec.
Overall preprocessing time	397.709 sec.	676.281 sec.
Display of lit model and scene with occlusion handling	44.3 fps	35.1 fps

Table 1: Processing times of our occlusion handling and image-based lighting feature.

a cube-map with a resolution of $6 \times 128 \times 128$ pixels was used, which took less than 0.323 seconds (see table 1).

5 SUMMARY AND FUTURE WORK

5.1 Summary

A novel approach has been presented for dealing with occlusion problems and image-based lighting of embedded virtual objects in image-based 3D worlds. This approach relies on 3D Photo Collections that can be automatically constructed from unordered images with varying exposure.

Occlusion problems have been addressed by generating Sliced Images. These are created by segmenting the photos with a gradient-based contour tracing algorithm. The depth of these contour segments (called slices) are estimated by related 3D key points, which represents a sparse model of the spatial scene structure of the photos.

The generated Sliced Images are also used to accumulate the environmental lighting for a point of interest. The result is a HDR environment map. In our implementation, 16 directional light sources are extracted from the environment map and used to illuminate virtual objects.

5.2 Future Work

In future work, we will enhance our contour tracer for detecting polygons directly. This avoids potential loss of objects with less gradients.

The generation of the depth map does not contain potential merging of segments. In future work, all segments may be reviewed for potential consolidating. Therefore, several combinations of conditions are possible. Neighboring segments can be merged, when segments have

- similar depth and similar hue,
- similar depth and similar light intensity,
- similar depth, similar hue and similar saturation, or
- combinations of all described similarities.

Concluding from these observations, the merging is parameterizable. The larger the parameter spaces are, the more segments grow together. The more segments are merged, the less the (merged) segment depth value is exact.

In the current approach, the depth reconstruction of an image does not depend on other images. This means the advantage of having multiple views of the real scene is not exploited. This could be achieved by fitting small scale geometrical objects (e.g. spheres) into the point cloud. After this step they could be used to project the images of the scene on them. A gain in depth precision would also increase the quality of the presented image-based-lighting approach, since the origin of the light is modeled with higher accuracy.

In addition, we plan to implement stereoscopic rendering in our viewer. This requires the modify of our rendering routines of Sliced Images. Based on the 3D key points and the contour-based segmentation, which is done for the occlusion handling, we will add stereoscopic depth into all photos.

ACKNOWLEDGEMENTS

This work was funded by BMBF (Federal Ministry of Education and Research, project no.: 17N0909). Furthermore, special thanks to Ekkehard Beier (EasternGraphics GmbH) for scientifical support to this project. Also, we thank Bastian Birnbach, Tobias Bindel and Stephan Rothe (Erfurt University of Applied Sciences) for their technical support.

REFERENCES

[1] F. Aurenhammer. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Comput. Surv.*, 23(3):345–405, 1991.

[2] R. T. Azuma. A Survey of Augmented Reality. In *Presence: Teleoperators and Virtual Environments 6*, 1997.

[3] Y. Bando, B.-Y. Chen, and T. Nishita. Extracting depth and matte using a color-filtered aperture. In *SIGGRAPH Asia '08 papers*, pages 1–9, 2008.

[4] R. Brinkmann. *The Art and Science of Digital Compositing*. Morgan Kaufmann, 1999.

[5] J. Canny. A computational approach to edge detection. *Readings in computer vision*, vol.184, 1987.

[6] B. Chanda and D. D. Majumder. Boundary-based Description. In *Digital Image Processing and Analysis*, pages 312–314. Prentice-Hall of India Pvt.Ltd, 2004.

[7] B. Chanda and D. D. Majumder. Edge and Line Detection. In *Digital Image Processing and Analysis*, pages 239–277. Prentice-Hall of India Pvt.Ltd, 2004.

[8] N. Dachuri, S. M. Kim, and K. H. Lee. Estimation of few light sources from environment maps for fast realistic rendering. In *ICAT '05 Proceedings*, pages 265–266, 2005.

[9] P. Debevec. Rendering Synthetic Objects into Real Scenes: Bridging Traditional and Image-Based Graphics with Global Illumination and High Dynamic Range Photography. In *SIGGRAPH98*, pages 189–198. ACM, 1998.

[10] P. Debevec. A Median Cut Algorithm for Light Probe Sampling. In *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*, 2005.

[11] A. Fournier, A. S. Gunawan, and C. Romanzin. Common Illumination between Real and Computer Generated Scenes. Technical report, University of British Columbia, Vancouver, BC, Canada, Canada, 1992.

[12] Y. Furukawa and J. Ponce. Carved visual hulls for high-accuracy image-based modeling. In *SIGGRAPH '05 Proceedings*, page 146. ACM, 2005.

[13] S. Gibson, J. Cook, T. Howard, and R. J. Hubbold. Rapid Shadow Generation in Real-World Lighting Environments. In *Rendering Techniques*, pages 219–229, 2003.

[14] M. Goesele, J. Ackermann, S. Fuhrmann, C. Haubold, R. Klowsky, and T. Darmstadt. Ambient point clouds for view interpolation. *ACM Transactions on Graphics (TOG)*, 29(4):1–6, 2010.

[15] Google. Street View. `http://maps.google.com/`, 2007.

[16] T. Grosch. PanoAR: Interactive augmentation of omnidirectional images with consistent lighting. *Mirage 2005, Computer Vision / Computer Graphics Collaboration Techniques and Application*, pages 25–34, 2005.

[17] K. Kölzer, F. Nagl, B. Birnbach, and P. Grimm. Rendering Virtual Objects with High Dynamic Range Lighting Extracted Automatically from Unordered Photo Collections. In *ISVC 2009 Proceedings: Part II*, pages 992–1001, 2009.

[18] M. I. Lourakis and A. A. Argyros. The Design and Implementation of a Generic Sparse Bundle Adjustment Software Package Based on the

Levenberg-Marquardt Algorithm. Technical report, Heraklion, Crete, Greece, 2004.

[19] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[20] Microsoft. Photosynth. `http://photosynth.net/`, 2008.

[21] F. Nagl, M. Friedl, A. Schäfer, and A. Tschentscher. Shader-Based-Image-Processor. In *GI Seminars 9. Informatiktage 2010*, pages 223–226, 2010.

[22] F. Nagl, P. Grimm, and D. Abawi. IP3D - A Component-based Architecture for Image-based 3D Applications. In *SEARIS@IEEEVR2010 Proceedings, IEEE VR 2010 Workshop*, pages 47–52, 2010.

[23] F. Nagl, P. Grimm, B. Birnbach, and D. Abawi. PoP-EYE environment: Mixed Reality using 3D Photo Collections. In *ISMAR 2010 Poster Proceedings*, pages 255–256, 2010.

[24] F. Nagl, K. Kölzer, P. Grimm, T. Bindel, and S. Rothe. Congrap – contour detection based on gradient map of images. In G. S. Hamid R. Arabnia, Leonidas Deligiannidis, editor, *Proceedings of the 2011 International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV 2011)*, volume II, pages 870–875. CSREA Press, USA, 2010.

[25] H. Regenbrecht. *Faktoren für Präsenz in virtueller Architektur*. PhD thesis, Bauhaus-Universität Weimar, 2000.

[26] M. Robertson, S. Borman, and R. Stevenson. Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *Journal of Electronic Imaging*, 12:219, 2003.

[27] N. Snavely. Bundler: Structure from Motion for Unordered Image Collections. `http://phototour.cs.washington.edu/bundler/`, 2010.

[28] N. Snavely. Bundler v0.4 User's Manual. `http://phototour.cs.washington.edu/bundler/bundler-v0.4-manual.html`, 2011.

[29] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3D. In *SIGGRAPH '06 Proceedings*, pages 835–846, 2006.

[30] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the World from Internet Photo Collections. *Int. J. Comput. Vision*, 80(2):189–210, 2008.

[31] J. Zhu and Z. Pan. Occlusion registration in video-based augmented reality. In *VRCAI '08 Proceedings*, pages 1–6, 2008.

Depth-assisted Real-time 3D Object Detection for Augmented Reality *

Wonwoo Lee† GIST U-VR Lab.
Nohyoung Park‡ GIST U-VR Lab.
Woontack Woo§ GIST U-VR Lab.

ABSTRACT

In this paper, we propose a novel method of real-time object detection that can recognize three-dimensional (3D) target objects, regardless of their texture and lighting condition changes. Our method computes a set of reference templates of a target object from both RGB and depth images, which describes the texture and geometry of the object, and fuses them for robust detection. Combining both pieces of information has advantages over the sole use of RGB images: 1) the capability of detecting 3D objects with insufficient textures and complex shapes; 2) robust detection under varying lighting conditions; 3) better identification of a target based on its size. Our approach is inspired by a recent work on template-based detection, and we show how to extend it with depth information, which results in better detection performance under varying lighting conditions. Intensive computations are parallelized on a GPU to achieve real-time speed, and it takes only about 33 milliseconds for detection and pose estimation. The proposed method can be used for marker-less AR applications using real-world 3D objects, beyond conventional planar target objects.

1 INTRODUCTION

Computer vision-based target detection techniques have been studied extensively and have been applied successfully to markerless augmented reality (AR) applications. A planar object has often been used as a tracking target due to its simple geometry [4, 13]. Results of recent work have shown that 3D objects can be used in AR applications with primitive-based modeling [12].

Local feature descriptors have been shown to have good performance for target detection [2, 9, 16] and hence they have been widely used for target detection. The descriptors are usually computed from local patches centered at keypoints; therefore, these methods can barely handle a textureless object with few keypoints on its surface. On the other hand, depth-based object recognition approaches have focused on a target's geometrical properties, such as normals and curvatures. A target object is identified from depth images by building local feature histograms from those properties [7, 10]. Local feature descriptors have also been introduced to depth-based object recognition [3, 15, 8, 1]. As RGB-D cameras, which capture color and depth images, become widespread recently, the RGB and depth information have been considered together for object recognition and pose estimation. In [5], depth information was employed to reduce influences from background and occlusion, but it was not explicitly used in object recognition. [14] reported that combination of visual features and shapes achieved higher performance and individual cues.

Figure 1: Detection and pose estimation of a 3D object. Our method can detect and estimate the pose of a 3D object. It also provides occlusion between the real and the virtual object based on depth information.

In this paper, we propose a novel method of object detection that can handle 3D target objects, regardless of their texture and lighting condition changes. As shown in Figure 1, our method can deal with a 3D object with a complex shape and insufficient textures in real-time. To do that, we combined information from both RGB and depth images to exploit both the texture and the geometry of a target. We modified a recent template-based detection method [11]. A set of reference templates of a target is computed not only from RGB images but also from depth images to reflect the target's textures and geometrical properties. In runtime, the reference templates are compared with incoming RGB and depth images, to identify a target object. Once the target is detected, its pose is estimated by aligning the 3D points of the current depth image with those of the reference templates.

By combining the texture and geometry information, a 3D object can be detected robustly under changing lighting conditions, and the two objects that have similar shapes and textures but different sizes can also be distinguished from each other. Both detection and pose estimation are computationally expensive and can barely run in real-time on a CPU. We adapted them for the use on a graphics processing unit (GPU) to achieve real-time speed. In our implementation, the overall detection and pose estimation take approximately 33 milliseconds.

In the remainder of the paper, we provide background information in Section 2 and describe our approach to target detection and pose estimation in Section 3. Experimental results are presented in Section 4. Finally, we offer our conclusions in Section 5.

2 BACKGROUND

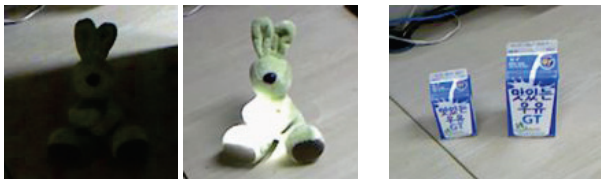
2.1 Overview

Figure 2, shows typical examples of when the sole use of an RGB image fails to detect a specific target. In Figure 2(a), an object with

*This work was supported in part by the Global Frontier R&D Program on <Human-centered Interaction for Coexistence> funded by the National Research Foundation of Korea grant funded by the Korean Government(MEST) (NRF-M1AXA003-20100029751), and in part by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research & Development Program 2011.

†e-mail: wlee@gist.ac.kr
‡e-mail: npark@gist.ac.kr
§e-mail: wwoo@gist.ac.kr

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278



(a) (b)

Figure 2: Examples where the sole use of an RGB image is unsuccessful in detection: (a) a target object is under different lighting conditions; (b) two objects have almost the same shapes and textures, but their sizes are different.

insufficient texture is under very different lighting conditions. It is difficult to handle these cases using RGB images only because the target's appearances look very different. The objects shown in Figure 2(b) have almost the same shapes and textures, while they have different sizes. It is also difficult to identify one of them solely using RGB images.

Thus, we attempted to overcome these limitations in 3D target detection by combining the shape and texture information. We modified a recent template-based detection method [11] to consider the texture and shape of a 3D object for robust detection We compute a set of reference templates from both RGB and depth images taken from different viewpoints. The gradients in both images are considered for template computation; the reference templates are built from the gradients with large magnitudes or frequent appearances in local patches. In runtime, the same gradient features are computed from incoming RGB and depth images and compared with the reference templates. As a result, a target's ID and 2D location in both images are retrieved, and a reference template with the greatest similarity is chosen as a match. Then, the detected target's pose is estimated by aligning the 3D points that are computed from the incoming depth image and those of the reference template chosen as a match. After point registration, the target's identity is finally verified from the registration error. Computationally expensive procedures in detection and pose estimation are parallelized by GPU programming for real-time speed. Figure 3 shows the overall procedure of our method.

2.2 RGB-Depth Calibration

The RGB-D camera we used consists of two cameras: an RGB camera for capturing color images and a depth camera for capturing depth images. The depth camera provides depth information as discretized values in a certain range, rather than actual distances. Depth-to-distance calibration is required to convert raw depth values into real distances before using depth information. On the other hand, the RGB and depth cameras have different characteristics, such as field of views and focal lengths; consequently, two pixels at the same location in RGB and depth images do not correspond to the same location in a scene. Thus, the RGB and depth images should be aligned to determine the depth of pixels in the RGB image. We first calibrated the RGB and depth cameras through a typical camera calibration method [19] using a chessboard pattern, and as a result, intrinsic parameter matrices of RGB and depth cameras, K_c and K_d, were estimated. Then, we performed the depth-to-distance calibration and the RGB-depth alignment.

Depth-to-distance calibration: We captured RGB and depth images of the chessboard pattern from different viewpoints, changing the distance between the pattern and the camera from 50 centimeters to 2.5 meters. Then, the actual depth values of the corners of the chessboard pattern were computed from extrinsic parameters. The raw depth values at the corner locations in the depth images were also collected. Finally, we estimated a polynomial of the 6-th degree from the collected data to represent a mapping between the raw depth values and the real distances.

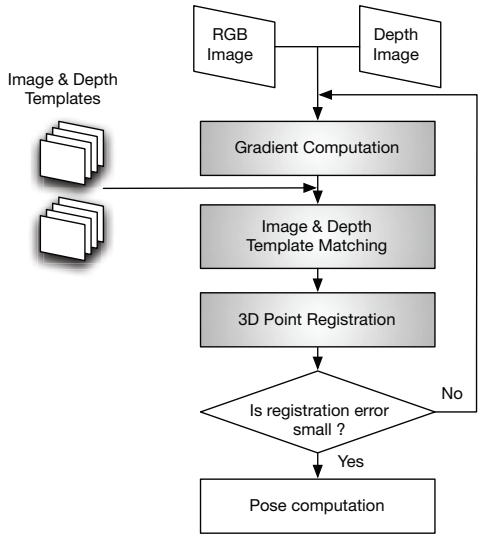


Figure 3: Overall procedure of the proposed method. The steps marked in shade runs in parallel on a GPU.

RGB-depth alignment: Let us denote by $M_{d\to c}$ the relative transformation between the RGB camera and the depth camera. The depth of a pixel in an RGB image is computed as follows. First, a pixel, $\mathbf{p}_d$, in the depth image is back-projected based on its depth value, z_p, to calculate a corresponding 3D point, $\mathbf{P}_d$.

$$\mathbf{P}_d = z_p K_d^{-1} \mathbf{p}_d \tag{1}$$

Then, P_d is transformed to the RGB camera's coordinate frame and projected onto the RGB image plane.

$$\mathbf{P}_c = M_{d\to c} \mathbf{P}_d \tag{2}$$

$$\mathbf{p}_c = K_c \mathbf{P}_c \tag{3}$$

Finally, the depth of a pixel at $\mathbf{p}_c$ in the RGB image is determined as the depth of $\mathbf{P}_c$. If two or more $\mathbf{P}_d$s correspond to the same $\mathbf{p}_c$, the closest one is chosen. As a result of the RGB-depth alignment, we obtain a depth image whose pixels correspond to the RGB image's [1].

3 Depth-assisted 3D Object Detection and Pose Estimation

3.1 Definitions

We define $\mathcal{O}_I$ and $\mathcal{O}_D$ as a target's reference patches taken from an RGB image and a depth image, respectively [2]. We denote by $\mathcal{T}_I$ and $\mathcal{T}_D$ the templates computed from $\mathcal{O}_I$ and $\mathcal{O}_D$, respectively. We call $\mathcal{T}_I$ 'image template' and $\mathcal{T}_D$ 'depth template'.

When computing a template, $\mathcal{T}$, related to a specific viewpoint, we also store its pose, $\mathcal{H}$, and depth. $\mathcal{T}$ is defined as

$$\mathcal{T} = \{\mathcal{T}_I, \mathcal{T}_D, \mathcal{H}, \mathcal{O}_D\}, \tag{4}$$

[1] In the remainder of the paper, 'depth image' means an aligned depth image instead of a raw depth image.

[2] Although we use grayscale images instead of color images, we keep calling it 'RGB image' to distinguish it from a depth image.

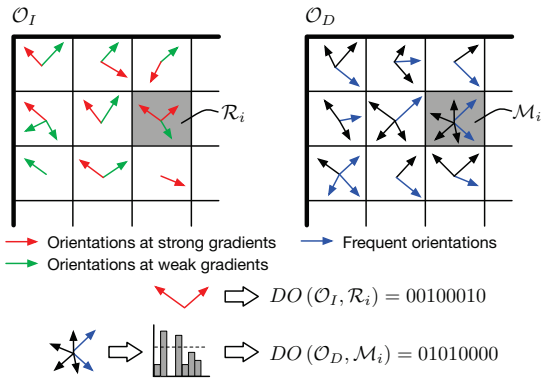


Figure 4: Building the image and depth templates from gradients

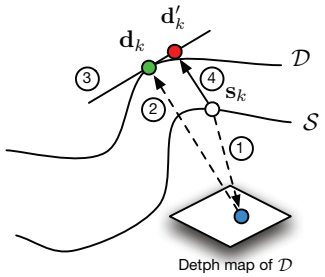


Figure 5: Point-to-plane registration: 1) a source point, $\mathbf{s}_k$ is projected onto the destination depth map; 2) a point on $\mathcal{D}$, $\mathbf{d}_k$, is determined by back-projecting the projection of $\mathbf{s}_k$; 3) a plane tangent to $\mathcal{D}$ on $\mathbf{d}_k$ is computed; 4) a match, $\mathbf{d}'_k$, is determined by projecting $\mathbf{s}_k$ onto the tangent plane.

3.2 Building Templates

Our image and depth templates are inspired by *Dominant Orientation Template* (DOT) algorithm [11]. The image and depth templates are computed from gradients in $\mathcal{O}_I$ and $\mathcal{O}_D$ but they represent different properties of the target object. Figure 4 depicts how image and depth templates are built from gradients.

Image templates are computed in the same way in [11]. The region of $\mathcal{O}_I$ is divided into $m \times n$ subregions, $\mathcal{R}_i$, and the orientations of strong gradients, whose magnitude is larger than a threshold, are collected from each $\mathcal{R}_i$. When gathering orientations, $\mathcal{R}_i$ is translated in a certain range to improve the robustness of detection to small deformations [11]. The collected orientations in $\mathcal{R}_i$ are denoted by $DO(\mathcal{O}_I, \mathcal{R}_i)$. The orientations of gradients are discretized to 7 bins to represent $DO(\mathcal{O}_I, \mathcal{R}_i)$ as an 8-dimensional binary vector. Each element of the vector indicates that strong gradients exist in a specific direction. The remaining 8-th element is set to 1 if there are no strong gradients in $\mathcal{R}_i$. Finally, an image template, $\mathcal{T}_I$, is represented as:

$$\mathcal{T}_I = \{DO(\mathcal{O}_I, \mathcal{R}_i) \,|\mathcal{R}_i \in \mathcal{O}_I\} \tag{5}$$

A depth template $\mathcal{T}_D$ is computed in a similar way as $\mathcal{T}_I$, but we consider the orientations that appear frequently in $\mathcal{O}_D$, rather than those of strong gradients because we are interested in the geometrical changes on the target's surface. In a depth image, strong gradients usually appear on the boundary between a target and the background, and thus, strong gradients barely provide useful information about the target's surface.

Given a depth patch, $\mathcal{O}_D$, and its subregions, $\mathcal{M}_i$, we build a histogram by accumulating orientations of gradients in $\mathcal{M}_i$. Orientations are discretized to 7 bins, and $\mathcal{M}_i$ is also translated to improve the robustness to small deformations, as we do for an image template. The histogram is then binarized by applying a threshold δ_h. Typically, δ_h is set to 30% of the number of subregions in $\mathcal{O}_D$. $\mathcal{T}_D$ is represented as an 8-dimensional vector from the resulting binary values. Each element indicates that a specific orientation appears frequently in $\mathcal{M}_i$. In case of the depth template, the remaining 8-th element represents the existence of a strong gradient in $\mathcal{M}_i$, where depth information is unreliable. A depth template, $\mathcal{T}_D$, is defined as:

$$\mathcal{T}_D = \{FO(\mathcal{O}_D, \mathcal{M}_i) \,|\mathcal{M}_i \in \mathcal{O}_D\}\,, \tag{6}$$

where $FO(\mathcal{O}_D, \mathcal{M}_i)$ represents orientations that frequently appear in $\mathcal{M}_i$.

3.3 Template matching

When there is an incoming RGB image patch, $\mathcal{P}_I$, the similarity between an image template, $\mathcal{T}_I$, and $\mathcal{P}_I$ is defined as

$$Sim(\mathcal{P}_I, \mathcal{T}_I) = \sum_{\mathcal{R}i} \delta\left(do(\mathcal{P}_I, \mathcal{R}_i), DO(\mathcal{O}_I, \mathcal{R}_i)\right)\,, \tag{7}$$

where $do(\mathcal{P}_I, \mathcal{R}_i)$ computes only the orientation of the strongest gradient in $\mathcal{R}_i$. The function $\delta(\cdot)$ is an element-wise *AND* operation between two 8-dimensional vectors.

The similarity between a depth image patch, $\mathcal{P}_D$, and a depth template, $\mathcal{T}_D$, is defined as

$$Sim(\mathcal{P}_D, \mathcal{T}_D) = \sum_{\mathcal{M}_i} \delta\left(fo(\mathcal{P}_D, \mathcal{M}_i), FO(\mathcal{O}_D, \mathcal{M}_i)\right), \tag{8}$$

where $fo(\mathcal{P}_D, \mathcal{M}_i)$ computes the most frequent orientation in a region $\mathcal{M}_i$.

From both similarity measures, the similarity function between a 2-tuple patch $\mathcal{P} = (\mathcal{P}_I, \mathcal{P}_D)$ and a template, $\mathcal{T}$, is defined as

$$Sim(\mathcal{P},) = (1-\alpha)\, Sim(\mathcal{P}_I, \mathcal{T}_I) + \alpha Sim(\mathcal{P}_D, \mathcal{T}_D)\,, \tag{9}$$

where α controls the weights of similarity values. In practice, we set $\alpha = 0.4$, which was experimentally determined.

All possible image regions are compared with the reference templates by changing the location of $\mathcal{P}$, and the most similar template is chosen as a match:

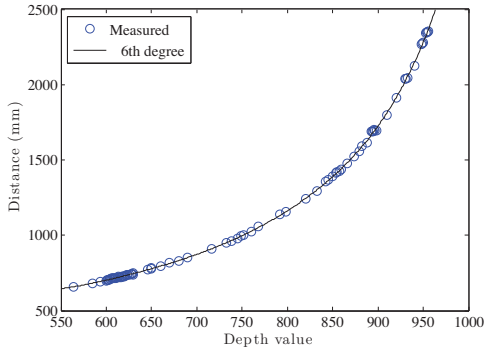
$$(\mathcal{P}^*, \mathcal{T}^*) = \underset{i,j}{\text{argmax}} Sim\left(\mathcal{P}^i, \mathcal{T}^j\right)\,, \tag{10}$$

where $\mathcal{P}^i$ and $\mathcal{T}^j$ represent a 2-tuple patch at i-th locations of RGB and depth images and the reference template corresponding to the j-th viewpoint, respectively.

As a result of the template matching, the location of a target, $\mathcal{P}^* = (\mathcal{P}^*_I, \mathcal{P}^*_D)$, and a matching reference template $\mathcal{T}^*$ are obtained.

3.4 Pose Estimation of a 3D Object

We assume the target's pose is initially identical to the pose of $\mathcal{T}^*$. The initial pose is refined by aligning 3D points computed from the current depth region, $\mathcal{P}^*_D$, with those of the reference template, $\mathcal{T}^*$. We adopt the point-to-projection approach for point registration because it is faster than the other iterative closest points (ICP) methods [18]. Let us denote by $\mathcal{S}$ the 3D points on a source surface, $\mathcal{D}$ a 3D points on the destination surface, and $\Delta W = [\Delta R|\Delta t]$ an incremental transformation between the source surface and the destination surface. In our problem, $\mathcal{S}$ and $\mathcal{D}$ correspond to the 3D



(a)

(b)

Figure 6: RGB-depth calibration: (a) Mapping between the raw depth values and real distances through the 6-th order polynomial; (b) RGB-depth alignment result (from left to right: RGB image, raw depth image, and aligned depth image).

points of $\mathcal{T}^*$ and those of $\mathcal{P}_D^*$, respectively. The initial pose is $\mathcal{H}^*$ of $\mathcal{T}^*$ and ΔW is initialized as $[I|0]$.

In the point-to-projection algorithm, a point $\mathbf{s}_k$ of $\mathcal{S}$ is projected onto the depth map of $\mathcal{D}$. The projection of $\mathbf{s}_k$ is back-projected, and a 3D point, $\mathbf{d}_k$, on $\mathcal{D}$ is determined from the depth map. Then, a plane that is tangent to $\mathcal{D}$ on $\mathbf{d}_k$ is computed from $\mathbf{d}_k$'s normal vector. Finally, $\mathbf{s}_k$ is projected onto the tangent plane, and its projection, $\mathbf{d}'_k$, is considered as a match. Figure 5 illustrates this procedure. After matches for all $\mathbf{s}_k$s are found, we obtain a new surface $\mathcal{D}'$ consisting of $\mathbf{d}'_k$s.

A correlation matrix, H, relating $\mathcal{S}$ and $\mathcal{D}'$ is defined as

$$H = \sum_k \left(\mathbf{d}'_k - \mathbf{c}_{d'}\right)\left(\mathbf{s}_k - \mathbf{c}_s\right)^T \ , \tag{11}$$

where $\mathbf{c}_s$ and $\mathbf{c}_{d'}$ represent the centroids of $\mathcal{S}$ and $\mathcal{D}'$, respectively.

ΔW between $\mathcal{S}$ and $\mathcal{D}'$ is computed from H:

$$\Delta R = VU^\top \tag{12}$$

$$\Delta t = \mathbf{c}_s - R\mathbf{c}_{d'} \ , \tag{13}$$

where V and U are determined by applying singular value decomposition (SVD) to H ($H = U\Sigma V^T$).

By iteratively estimating and accumulating the incremental transformation, ΔW, the refined pose after the i-th iteration, W_i, is computed as

$$W_i = \Delta R_i \Delta R_{i-1} + \Delta R_i \Delta t_{i-1} + \Delta t_i \ . \tag{14}$$

The iteration is terminated if the difference between W_{i-1} and W_i is small.

The target's identity is verified based on the registration error defined as the average distance between points on $\mathcal{S}$ and $\mathcal{D}'$. If the registration error is smaller than a threshold, detection is considered successful. Typically, the threshold is set to 7 millimeters in our experiments.

3.5 Parallelization on a GPU

To increase the speed of detection, we parallelized the gradient computation and template comparison steps. As pixel- or patch-wise operations, they are suitable for running in parallel on a GPU.

Gradient computation consists of two steps. First, the magnitude and the orientation of each pixel's gradient is computed from both RGB and depth images. In the GPU, a thread is assigned to a pixel location, and gradients are computed by a 3×3 Sobel mask in parallel. Then, $do(\cdot)$ and $fo(\cdot)$ are computed from the gradients in RGB and depth patches that would be compared with the reference templates. In the second step, each region is independently processed in a thread. The resulting $do(\cdot)$ and $fo(\cdot)$ are stored as an array of 2D vectors in the GPU's memory to use them in template matching on the GPU.

When conducting template matching on the GPU, the reference templates of a target are copied to the GPU's memory as a 2D image block, where each row corresponds to a template related to a viewpoint. A thread is assigned to an image region and compares the previously computed orientations, $do(\cdot)$ and $fo(\cdot)$, with the reference templates. The threads in the same thread block are synchronized and they access the same reference template concurrently. After all of the reference templates are compared, the index of the most similar reference template and the similarity value are stored as a 2D vector in the memory for each image region.

Pose estimation step was also implemented on the GPU for real-time speed because point registration is computationally expensive and can barely run in real-time on a CPU. We parallelized two steps of the registration procedure, finding matches and computing the correlation matrix, H.

When computing point matches, each match is computed independently; therefore, it is also good to be parallelized. The source points converted to 3-channels image and the depth map of the destination surface are copied to the GPU's memory. In the GPU, a thread is assigned to each source point, $\mathbf{s}_k$, and the matches to $\mathbf{s}_k$s are computed in parallel. The resulting matches, $\mathbf{s}_k$s and $\mathbf{d}'_k$s, are kept in the GPU for computing the correlation matrix, H. Summation is a major operation in computing the centroids, $\mathbf{c}_{d'}$ and $\mathbf{c}_s$, and in calculating the correlation matrix H from the centroids as well. Summations in those computations were accelerated by a memory reduction technique [17]. The reduction consists of two steps. In the first step, we launch 100 thread blocks, each containing 256 threads. The summation of the 256 elements is computed in each block. Then, in the second step, 100 threads are launched in a block to compute the final sum from the 100 partial sums.

4 Experimental Results

4.1 Setup

We performed experiments on a PC with a 2.93GHz CPU and an NVIDIA Geforce GTX580 GPU. For GPU programming, we used NVIDIA's CUDA. The Microsoft KINECT sensor was employed to capture RGB and depth images (in 640×480 at 30 Hz). The size of a reference template was 154×154 and the number of total templates was 256. The size of a subregion was 7×7 as it was reported that good performance was achieved with it [11]. The maximum number of iterations in the point registration step was 50.

Reference template acquisition When building reference templates from a 3D object, we placed the target object on a flat surface. A simple background subtraction method was applied to identify the image regions that belonged to the object. We took image patches from different viewpoints and computed the image and depth templates from them. The poses of the reference templates were computed from a known planar tracking target, which was located beside the object and used as a reference coordinate frame.

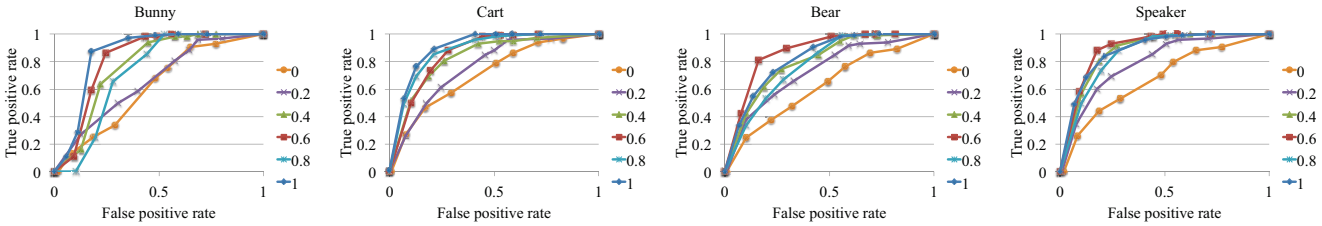


Figure 7: ROC curve with the varying scale factor α.

RGB-Depth calibration Figure 6(a) shows the relationship between the depth values of a raw depth image and those computed from extrinsic parameters. The 6-th order polynomial function estimated from them represents the relationship between raw depth values and actual distances quite well. Note that the estimated mapping function is effective only in the distance range where the data were obtained. In Figure 6(b), the result of RGB-depth alignment is depicted. The raw depth image was misaligned with the RGB image due to the parallax and different characteristics of the cameras. After the alignment, we acquired a new depth image that was correctly aligned with the RGB image.

4.2 Results

Fig 7 shows detection performance with varying α values. We recorded a video sequences under a varying lighting conditions and measured the performance. The α varied from 0 to 1 by the step of 0.2. When using depth only ($\alpha = 0$), detection performance was worse than using RGB only, and many false detection cases were observed. Thus, the depth template is not much discriminate. However, when both information is fused, it was comparable or better than single cue cases, i.e., using RGB or depth only. Depending on the target object, the best result is achieved with differnt alpha values ranging from 0.4 to 0.6.

We show the target detection results under changing lighting conditions in Figure 9. Detection was unsuccessful when using RGB images only, because gradients of the target's textures largely changed due to spotlights and strong shadows. In case of the RGB-only detection, the similarity values decreased when the spotlights were projected or the shadows were drawn. In contrast, our method detected the targets successfully by taking advantages from the depth information, which is not disturbed by changing lighting conditions. The similarity values were maintained high when the depth information was combined with the RGB information.

Thanks to the depth information, the two objects that had almost the same shape and textures but different sizes could also be distinguished successfully as depicted in Figure 8. RGB-only detection failed to identify one of them due to their similar textures, and hence, false detection and wrong pose estimation occurred. More target detection and augmentation results are shown in Figure 10. Most of targets have poor textures (e.g., *Speaker*, *Building*, *Cart*, and *Bunny*). Our method successfully detected them and estimated their poses. It also worked well with a target with sufficient textures as shown in row 4 (*Harubang*). The last row demonstrates that the capability of multiple target detection.

Table 1 shows the overall speed of template matching running on both a CPU and a GPU. In the case of CPU implementation, template clustering was applied to increase the speed of template matching as in [11]. As we can see, template matching ran much faster on the GPU than on the CPU. In the template comparison step, the GPU version was approximately 4 times faster than the CPU version. Without template clustering, template matching took about 80 ms on the CPU. In the 3D registration, the GPU implementation was approximately 25 times faster than the CPU version

Figure 8: Detection of targets that have almost the same textures and shapes: (top row) False detection occurs when using RGB images only, and hence the estimated poses are wrong; (bottom row) two objects are identified correctly by our method.

(see Table 2). A major speed improvement was achieved in the point match step. The speed of correlation matrix computation was also improved on the GPU, but it became slower than the point match step due to multiple launches of GPU kernels for reduction. Note that the $[R|t]$ computation was conducted on the CPU, and consequently, its speed was the same on both cases. The overall speed of single target detection and pose estimation was 33.5 ms, which was adequate for real-time AR applications. In the case of multiple targets, our method was able to handle up to 3 targets in real-time.

The accuracy of point registration was measured under smooth camera motions by computing the average distance between points of the source and destination surfaces. As we can see in Table 3, the registration quality was good and the pose estimation result could be used for overlaying virtual objects on video sequences. In our experiments, 3D point registration tended to be more accurate with the objects having planar geometries (e.g., *Cart* and *Building*), than those with more complex shapes (e.g., *Bunny*, *Harubang*, and *Bear*).

The RGB-D camera used in our experiments was unable to compute the depth of a location that was excessively close to the camera. The distance between the camera and a target object should be more than approximately 60 centimeters. Depth information provided by the camera was noisy, which caused jittering in the estimated pose. Applying noise reduction filters [6] can be a possible solution to reduce the noise. On the other hand, the depth information became inaccurate and unstable if a target had thin structures; therefore, it was difficult to estimate the pose of such a target.

5 Conclusions and Future Works

We proposed a novel 3D target detection method that exploits both the target's texture and geometry information. Our method can handle 3D objects with complex shapes and insufficient textures, and detects targets under changing lighting conditions by combining information from RGB and depth images. Real-time speed

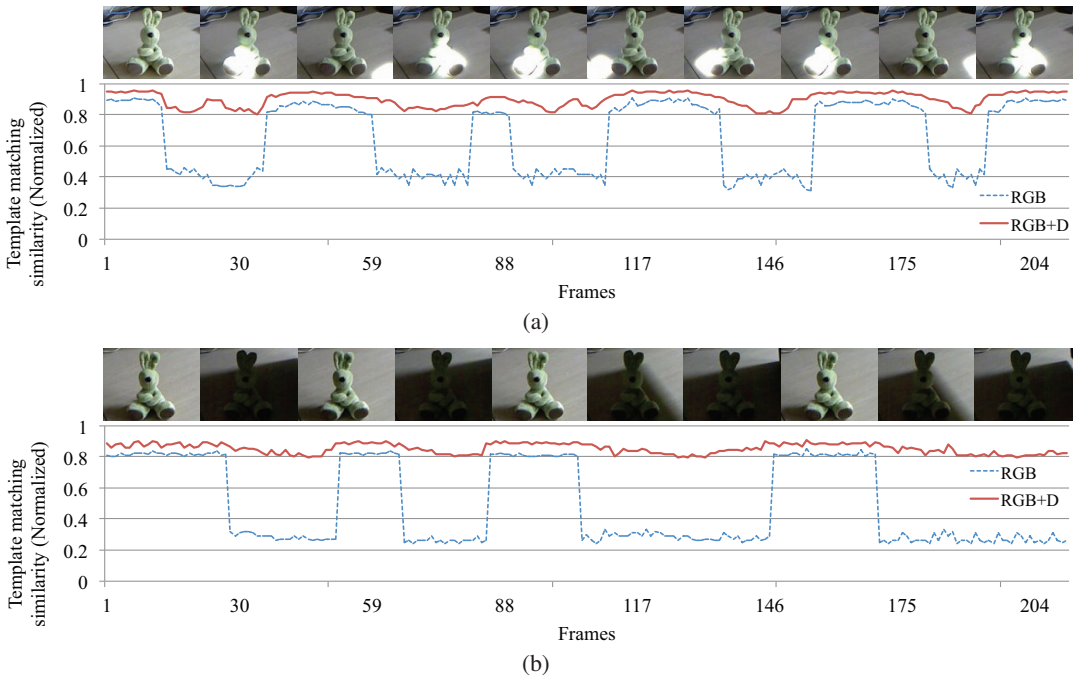


Figure 9: Template matching results under changing lighting conditions: (a) the target is under a moving spotlight; (b) the target is under shadow.

is achieved by parallelizing computationally expensive steps on a GPU. In the current method, jitters occur in the estimate pose when a target's surface is occluded by other objects. Consequently, we will focus on stable pose estimation under partial occlusion in the future.

Table 1: Template matching speed on a CPU and a GPU (unit: ms)

Procedure	CPU	GPU
Gradient computation	10.7	0.5
Template comparison	39	10.2
Total	49.7	10.7

Table 2: 3D Registration speed on a CPU and a GPU (unit: ms)

Procedure	CPU	GPU
Point match	422.8	6.3
Correlation matrix computation	160	15.6
$[R\|t]$ computation	0.9	0.9
Total	583.7	22.8

Table 3: Point registration error (unit: millimeters)

Object	Mean	Stdev
Speaker	3.75	0.43
Building	3.13	0.44
Cart	2.15	0.54
Bunny	4.49	0.55
Harubang	3.89	0.32
Bear	4.21	0.29

REFERENCES

[1] P. Bariya and K. Nishino. Scale-hierarchical 3d object recognition in cluttered scenes. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 1657–1664, 2010.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. VanGool. Speeded-Up Robust Features. *Computer Vision and Image Understanding*, 110(3):346–359, June 2008.

[3] N. Bayramoglu and A. A. Alatan. Shape index sift: Range image recognition using local features. In *Proceedings of the International Conference on Pattern Recognition*, pages 352–355, 2010.

[4] S. Benhimane and E. Malis. Homography-Based 2d Visual Tracking and Servoing. *International Journal of Robotics Research*, 26(7):661–676, 2007.

[5] N. Burrus, M. Abderrahim, J. Garcia, and L. Moreno. Object reconstruction and recognition leveraging an rgb-d camera. In *The 12th IAPR Conference on Machine Vision Applications*, June 2011 (in press).

[6] D. Chan, H. Buisman, C. Theobalt, and S. Thrun. A Noise-Aware Filter for Real-Time Depth Upsampling. In *Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications 2008*, 2008.

[7] H. Chen and B. Bhanu. 3d free-form object recognition in range images using local surface patches. *Pattern Recogn. Lett.*, 28:1252–1262, July 2007.

[8] B. Drost, M. Ulrich, N. Navab, and S. Llic. Model globally, match locally: Efficient and robust 3d object recognition. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 998–1005, 2010.

[9] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[10] G. Hetzel, B. Leibe, P. Levi, and B. Schiele. 3d object recognition from range images using local feature histograms. In *Proceedings of CVPR 2001*, pages 394–399, 2001.

[11] S. Hinterstoisser, V. Lepetit, S. Ilic, P. Fua, and N. Navab. Dominant Orientation Templates for Real-Time Detection of Texture-Less Objects. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 2010.

[12] K. Kim, V. Lepetit, and W. Woo. Keyframe-based Modeling and Tracking of Multiple 3D Objects. In *Proceedings of the International Symposium on Mixed and Augmented Reality*, 2010.

[13] K. Kim, V. Lepetit, and W. Woo. Scalable real-time planar targets tracking for digilog books. *Vis. Comput.*, 26:1145–1154, June 2010.

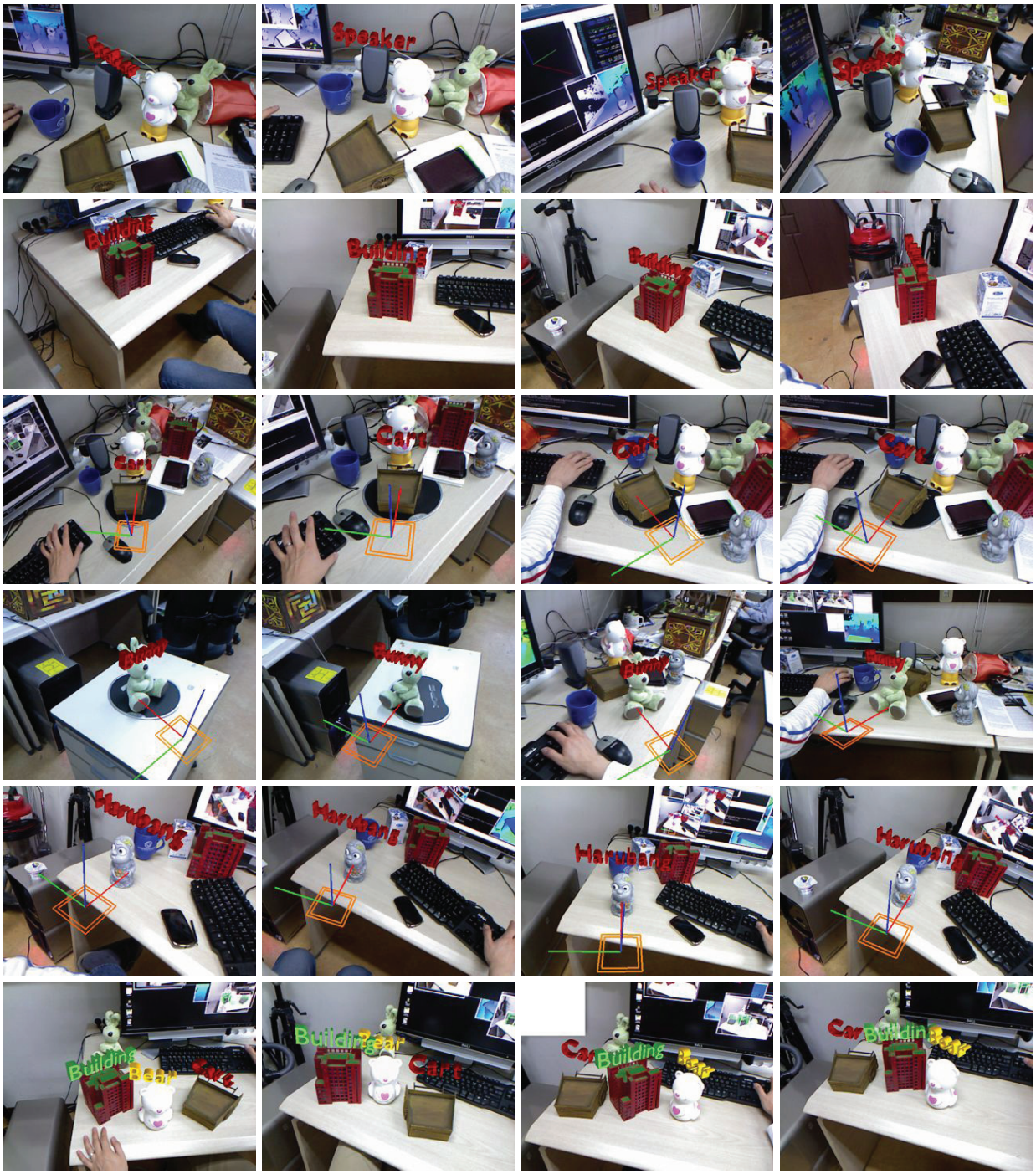


Figure 10: Target detection examples: (from rows 1 to 5) *Speaker*, *Building*, *Cart*, *Bunny*, *Harubang*; (row 6) multiple target detection. Our method handles the targets with complex shapes and insufficient textures successfully.

[14] K. Lai, L. Bo, X. Ren, and D. Fox. Sparse distance learning for object recognition combining rgb and depth information. In *IEEE International Conference on Robotics and Automation*, 2011.

[15] T.-W. R. Lo and J. P. Siebert. Local feature extraction and matching on range images: 2.5d sift. *Comput. Vis. Image Underst.*, 113:1235–1250, December 2009.

[16] M. Ozuysal, M. Calonder, V. Lepetit, and P. Fua. Fast Keypoint Recognition Using Random Ferns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3):448–461, March 2010.

[17] D. Roger, U. Assarsson, and N. Holzschuch. Efficient stream reduction on the GPU. In *1st Workshop on General Purpose Processing on Graphics Processing Units (GPGPU)*, 2007.

[18] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *Third International Conference on 3D Digital Imaging and Modeling (3DIM)*, pages 145–152, June 2001.

[19] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *Computer Vision, IEEE International Conference on*, volume 1, page 666, 1999.

Remote Ikebana with Olfactory and Haptic Media in Virtual Environments

Pingguo Huang*, Yutaka Ishibashi†, Norishige Fukushima‡, and Shinji Sugawara§
Department of Scientific and Engineering Simulation
Nagoya Institute of Technology

Abstract

In this paper, we handle a remote ikebana (i.e., flower arrangement) system with olfactory and haptic media. In the system, a teacher or a student can hold a flower, adjust the length of the held flower's stem with a pair of scissors, and impale the flower on a flower pinholder in a 3-D virtual space. We investigate the influence of the size of smell space (defined as a sphere in which we can perceive the smell of flower) on QoE (Quality of Experience), and we illustrate that there exists the optimum value of the smell space size.

Keywords: Ikebana, Olfactory media, Haptic media, QoE, Virtual environment

1 Introduction

Recently a number of researchers have been paying attention to multi-sensory communications, in which we treat vision, auditory sensation, gustation, olfaction, and tactile sensation [1]. By handling multiple sensations together, we can improve realistic sensations [2] and immerse ourselves in various applications such as ikebana (i.e., flower arrangement) [3], cooking [4], and harvesting fruit [5]. However, there is few papers which study networked applications using vision, olfactory media and haptic media together.

In this paper, we deal with a remote ikebana system with vision, olfaction media, and haptic media. Since the output timing of olfactory media affects the experience of realistic sensations, it is very important to clarify the influence of the output timing of olfactory media. Thus, we assess the influence of the output timing of olfactory media on QoE (Quality of Experiment).

2 Remote Ikebana System

In the remote ikebana system, by manipulating a haptic interface device, a teacher or a student can hold a flower, adjust the length of the held flower's stem with a pair of scissors, and impale the flower on a flower pinholder (see Fig. 1, in which the student is going to cut the held rose's stem). The teacher is able to teach the student at a remote place how to arrange flowers, and the teacher is also able to change the arrangement designed by the student. Moreover, when the viewpoint of the teacher or student enters the *smell space* (defined as a sphere in which we can perceive the smell of flower. It is called "aroma aura" in [2]) of a flower, the smell of the flower is diffused by using an olfactory display, and he/she can perceive the smell of the flower. We employ PHANToM Omni as a haptic interface device and use SyP@D2 as an olfactory display.

3 Assessment Method and Results

We carried out QoE assessment to investigate the influence of the output timing of olfactory media by changing the smell space size (i.e., the radius of each sphere). In the assessment, each subject was asked to select a rose from among flowers on the table, hold and move the rose at a constant speed toward his/her viewpoint until he/she starts to perceive the smell of the rose. Then, he/she moved it away from the viewpoint until he/she became insensitive to the smell at the same speed as the speed when he/she moved the flower toward his/her viewpoint. Each subject was asked to judge how good the output timing of olfactory media is. The subject gave a score from 1 (bad) through 5 (excellent) to each test to obtain the mean opinion score (MOS).

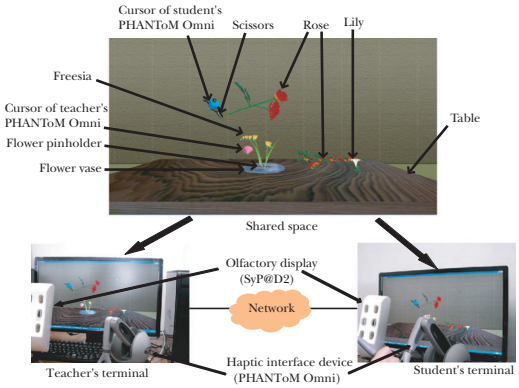


Figure 1: Configuration of remote ikebana system.

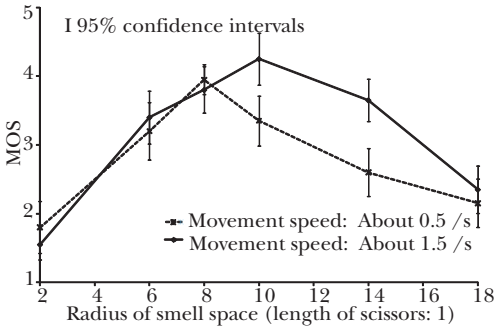


Figure 2: MOS versus radius of smell space.

We show MOS for two movement speeds in Fig. 2. We see in the figure that there exists the optimum value of the smell space size, and the optimum value depends on the movement speed.

4 Conclusions

In this paper, we handled a remote ikebana system with olfactory and haptic media. We also investigated the influence of the smell space size on QoE. As a result, we saw that there exists the optimum value of the smell space size, and the optimum value depends on the movement speed. As the next step of our research, we plan to investigate the influences of network delay, delay jitter, and packet loss on QoE in the remote ikebana system.

Acknowledgements

The authors thank an ikebana's professional Ms. Junko Ando for her valuable comments on the remote ikebana system. This work was supported by the Grant-In-Aid for Scientific Research (C) of Japan Society for the Promotion of Science under Grant 22560368 and the HORI SCIENCES AND ARTS FOUNDATION. The work was also done in cooperation with TSUJI KOSAN CO., LTD. Olfactory Multimedia Labo., Exhalia, and SHIONOKORYO KAISHA, LTD.

References

[1] T. Nakamoto *et al.*, *Proc. ICAT*, Dec. 2008.
[2] K. Tominaga *et al.*, *Proc. VSMM*, Oct. 2001.
[3] N. Mukai *et al.*, *Proc. World IMACS/MODSIM Congress*, July 2009.
[4] I. Siio *et al.*, *Lecture Notes in Computer Science*, vol. 4551, 2007.
[5] S. Hoshino *et al.*, *Proc. IEEE CQR*, May 2011.

*e-mail: huang@mcl.nitech.ac.jp
†e-mail: ishibasi@nitech.ac.jp
‡e-mail: fukushima@nitech.ac.jp
§e-mail: shinji@nitech.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

QoE Comparison of Competition Avoidance Methods for Management of Shared Object in Networked Real-Time Game with Haptic Media

Yuji Kusunose* Yutaka Ishibashi† Norishige Fukushima‡ Shinji Sugawara§

Department of Scientific and Engineering Simulation
Nagoya Institute of Technology

Abstract

In this paper, we investigate competition avoidance methods for management of a shared object in a networked real-time game with haptic media. For competition avoidance, we deal with a priority method and a combined method of AtoZ (Allocated Topographical Zone) and CDP (Count Down Protocol). We also clarify the influence of network delay on QoE (Quality of Experience) for the two methods.

Keywords: Networked real-time game, Haptic media, Competition avoidance, Network delay, QoE

1 Introduction

It is expected that using haptic media in networked real-time games gives players a higher sense of immersion. However, the consistency and causality may be disturbed owing to network delay, delay jitter, and packet loss in a QoS (Quality of Service) non-guaranteed network like the Internet.

The authors investigated the influence of network delay on QoE (Quality of Experience) in a networked air hockey game with haptic media [1]. For consistency and causality, they employed the adaptive Δ-causality control scheme with adaptive dead-reckoning (referred to as Adaptive DR + Adaptive Δ). However, they demonstrated that disagreement of the owner of a shared object occurs when the network delay is large. To solve this problem, we need a competition avoidance method.

In this paper, we treat a priority method and a combined method of AtoZ (Allocated Topographical Zone) and CDP (Count Down Protocol) [2], [3] (called AtoZ + CDP) for competition avoidance. Then, we compare the two methods by QoE assessment in the networked air hockey game with haptic media. We also investigate the influence of network delay on QoE.

2 Networked Air Hockey Game with Haptic Media

In the game, two users fight against each other. Each user operates his/her mallet with a haptic interface device, and he/she hits a puck toward his/her opponent's goal. Each terminal uses PHANToM Omni (just called PHANToM) as the haptic interface device. When a mallet touches the puck, a player of the mallet feels force feedback. The game is based on a peer-to-peer (P2P) model. The owner of the puck, who has hit the puck last, calculates the position and velocity of the puck. A terminal which is not the owner of the puck outputs the puck at a position which has been received from the owner. When the owner of the puck is different between the terminals, we use the priority method or AtoZ + CDP for competition avoidance. For consistency and causality, we use Adaptive DR + Adaptive Δ in each method. In this scheme, the output time of position information is given by the generation time of the information plus Δ (> 0) ms. The value of Δ is dynamically changed according to the network delay and satisfies the following relation: $0 < \Delta_L \leq \Delta \leq \Delta_H$.

3 Competition Avoidance Methods

3.1 Priority Method

In this method, if the owner of the puck is different between the two terminals for a fixed time (set to $2\Delta_H$ ms in this paper), one terminal (called the *priority terminal*) which is determined in advance becomes the owner of the puck, and the other terminal (the *non-priority terminal*) conforms the position of the puck to that of the priority terminal. Then, a warp (momentary and large movement) immediately after a pause of the puck occurs at the non-priority terminal; thus, the output quality of the puck deteriorates.

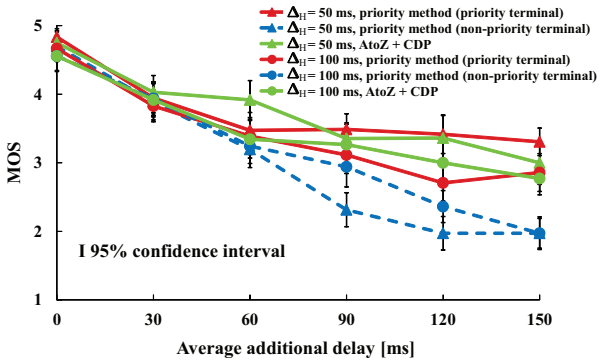


Figure 1: MOS of comprehensive quality.

3.2 AtoZ + CDP

AtoZ is used to determine which terminal can access to the puck most quickly by taking account of the positions and velocities of the mallets. The determined terminal acquires the ownership of the puck when the owner of the puck is different between the two terminals.

CDP is a protocol used when the puck exists in a field called Dead Zone, where the owner of the puck cannot be uniquely determined owing to the influence of network delay in AtoZ. The basic idea of CDP is that a terminal resigns the ownership of the puck until the terminal receives the information that the other terminal manages the puck. If the owner of the puck is different between the terminals after using AtoZ + CDP, we use the priority method. The reader is referred to [2] and [3] for details of AtoZ and CDP.

4 Assessment Method and Results

We show the mean opinion score (MOS) of comprehensive quality [1] in Fig. 1, where the average MOS value is plotted in AtoZ + CDP since there was almost no difference in MOS between the two terminals. From this figure, we see that the fairness between the terminals is ruined when the network delay is large in the priority method, while we can keep the fairness and MOS high in AtoZ + CDP.

5 Conclusion

In this paper, we made a comparison between the priority method and AtoZ + CDP in a networked real-time game with haptic media. As a result, we demonstrated that the fairness between terminals is ruined when the network delay is large in the priority method, while we can keep the fairness and MOS high in AtoZ + CDP.

References

[1] Y. Kusunose *et al.* in *Proc. NetGames'10*, Nov. 2010.

[2] Y. Kawano *et al.* (in Japanese). *Trans. of the Virtual Reality Society of Japan*, 9(2):141-150, June 2004.

[3] Y. Kawano and T. Yonekura. *IEICE Trans. on Inf. and Syst. (Japanese Edition)*, J89-D(10):2219-2228, Oct. 2006.

*e-mail: kusunose-y@mcl.nitech.ac.jp
†e-mail: ishibasi@nitech.ac.jp
‡e-mail: fukushima@nitech.ac.jp
§e-mail: shinji@nitech.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Development of Inner Strings Haptic Interface SPIDAR-I

*Yan ZHU, Tatsuya KOYAMA, Tatsuro IGARASHI, Katsuhito AKAHANE, Makoto SATO

Precision and Intelligence Laboratory, Tokyo Institute of Technology

ABSTRACT

This paper describes the design and implementation of an inner strings haptic device in 6-DOF, called SPIDAR-I. This device is developed to improve the calculational fidelity of position, orientation and force feedback for higher performance. Moreover, it promotes a new structure that strings and frame are inside the grip, so making it particularly compact in size for wider use.

KEYWORDS: SPIDAR, haptic device, user interface.

1 INTRODUCTION

As 3D virtual environment has been widely used, haptic interfaces are also drawing widespread attention. To all haptic devices, it is necessary to make further improvement on the calculational fidelity of position, orientation and force feedback for more real user experience.

This paper is based on SPIDAR[1], which is a wire-driven haptic device. Because its configuration of frame, grip and strings have a great effect on force display fidelity[2], we have discussed the optimization of the structure and developed the new SPIDAR-I with the optimal result, for a higher calculational fidelity of position, orientation and force feedback.

At the same time, the new type has been compacted into a smaller size and is expected for a wider application.

2 DESIGN AND IMPLEMENTATION OF SPIDAR-I

2.1 Derivation of SPIDAR-I

For the optimization of structure, we first modeled the SPIDAR into 3D space as shown in figure 1(A) and conducted the equations for position, orientation calculation and force feedback calculation. Because there is a relation between the two equations that when one` fidelity is improved, so is the other, we chose to discuss the former one. Then using Least Squares Method, it is simplified into (1).

$$\mathrm{M}^{\mathrm{T}}\Delta \mathrm{l} = \mathrm{M}^{\mathrm{T}}\mathrm{M}\Delta r \tag{1}$$

Here if Δr decreases, the calculational fidelity will be improved. Therefore, we defined the evaluation function as (2) to increase the eigenvalues of the coefficient matrix $\mathrm{M}^{\mathrm{T}}\mathrm{M}$ on average.

$$\mathrm{J} = \left|\mathrm{M}^{T}\mathrm{M}\right| = \lambda_1\lambda_2\lambda_3\lambda_4\lambda_5\lambda_6 \tag{2}$$

After maximizing (2), we got the optimization result as follows.

$$\mathrm{D} = \frac{1}{2}\mathrm{R},\ \mathrm{D}_{\mathrm{Z}} = \frac{\sqrt{2}}{2}\mathrm{R} \tag{3}$$

2.2 Design and implementation of SPIDAR-I

Depending on the optimal result of structure, we found it possible that strings and frame could be compacted inside the grip, so the overall size of the device has been much reduced. And the final design is shown in the figure 1(B).

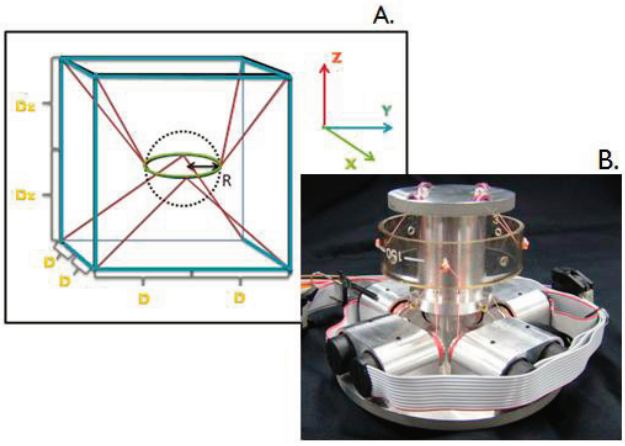


Figure 1. A is the model of SPIDAR and B is a picture of SPIDAR-I

3 EVALUATION EXPERIMENTS

The evaluation experiments have been executed as two parts, one is to measure the position and orientation by OPTOTRAK, the other is to measure the force feedback by load cell. And both of them compared SPIDAR-I to a classic type SPIDAR-G[3]. As a result, even the overall size is much reduced, SPIDAR-I still has a sufficient haptic perform in a high calculational fidelity.

4 CONCLUSION

A new type of haptic device called SPIDAR-I was proposed with a higher calculational fidelity of position, orientation and force feedback. At the same time, it has been designed as a particularly compact style in size, so making it possible for a wider use.

As the future work, the friction between strings and other components, which has a quite effect on force display, should be reduced. Moreover, in this paper, only home position of the grip was taken into account, it is necessary to consider the situations of other positions.

REFERENCES

[1] M. Sato, Y. Hirata and H. Kawarada. Space Interface Device for Artificial Reality, SPIDAR. In *The Transactions of the Institute of Electronics, Information and Communication Engineers.* 74(7), p887-894, July 1991.

[2] M. Inoue, S. Hasegawa, S. Kim and M. Sato. Tension calculation algorithm for the wire driven force display using the quadratic programming method. In *Proceedings of the Virtual Reality Society of Japan*, 6, 91-94, September 2001.

[3] S. Kim, S. Hasegawa, Y. Koike and M. Sato. A Proposal of 7 DOF Force Display: SPIDR-G. In *Transactions of the Virtual Reality Society of Japan*, 7(3), 403-412, September 2002.

*email: {zhu.y.ab@m, t_koyama@hi.pi, t-igarashi@hi.pi, kakahane@hi.pi, msato@pi}.titech.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

The effects of using a modified motorcycle simulator training for the spinal cord injury patients

Siao-Ying Wu[1] — The University of Tokyo

Wen-Hsu Sung[2] — National Yang-Ming University

Yun-An Tsai[3] — Taipei Veterans General Hospital

Henrich Cheng[4] — Taipei Veterans General Hospital

Jin-Jong Chen[5] — National Yang-Ming University

ABSTRACT

This is a first step study to investigate the effects of the training with virtual reality (VR) system in different situations for Spinal cord injuries (SCI) patients and evaluate the riding performance and the balance ability of patients with SCI to deal with different road conditions. And is also a very early stage study to investigate whether the training effects can transferred to the real road. SCI patients increase by 1200 people/ year in Taiwan. Some studies found that SCI patients who have jobs and transportation had more positive self-concept. Modified motorcycles are the most popular transportation tools used by SCI patients in Taiwan. The purpose of this study is to investigate the training effects of a motorcycle riding training program with modified motorcycle virtual reality simulator (MCVRA) for the spinal cord injury patients. In this study, five SCI subjects were included in this study. They received ten 30-minutes training sections in one month, and the riding performance, balance ability and questionnaires were measured before and after 5 and 10 training sections to evaluate the program effects. However, only three subjects completed 10 training sections due to discharge or being transferred. Results revealed that the riding performance and balance ability under VR environment and on road test were improved after the MCVRA. The subject's enjoyment, confidence and motivation of motorcycle riding also increased dramatically after the MCVRA. No cyber-sickness or other side-effects was noted during the training program. We found that the riding performance and balance ability would be improved after VR training and the training effects seem able to be transferred to the real world road. Trained and assessed with MCVRA designed for SCI should be feasible and useful.

KEYWORDS: Virtual reality, spinal cord injury, physical therapy, modified motorcycle.

1 INTRODUCTION

Spinal cord injuries usually damage to the central or peripheral nervous systems that cause problems of sensory and motor functions. The key functional deficit of SCI is poor balance and it usually makes the patients' performance of activities of daily living decrease and influences their quality of life [1]. VR is a new high technology developed in recent years and has been used in many fields, such as aviation, military, or medical field. Now it also has been attention on rehabilitation because training with VR has many advantages, such as more safety, interesting, users have less discouragement, and low cost [1-2]. Although there are many studies about the simulator designing and development, there is no research documenting the applications on rehabilitation. There were also only very few studies related to the connection of training effects between the VR environment and the real world. We will investigate those topics in this study by analyzing the performance of riding and balance ability in MCVRA and the real world.

2 EXPERIMENT EQUIPMENT

The MCVRA, showed in Fig.1, was designed for both assessment and for training. All the scenarios were designed as urban and followed the transportation laws. One of the scenarios was designed as the assessment route for evaluating subjects in the real world to compare the performance.

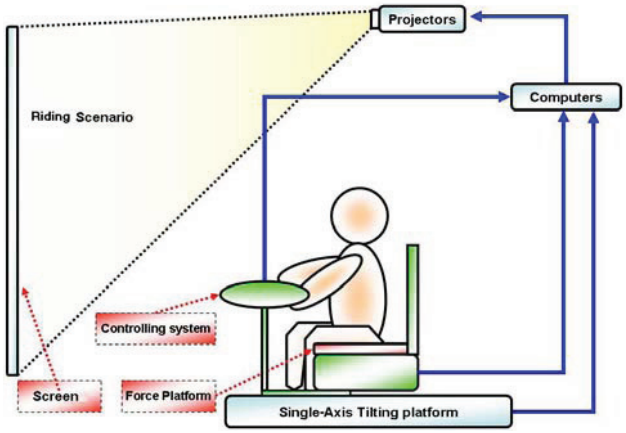


Figure 1. Modified motorcycle virtual reality simulator system

3 RESULTS AND CONCLUSION

Although the VR systems can't 100% simulate the real world, it could induce similar human body reaction and movements. It means we may assessment and training by a safer way with a less impact to participants in VR. There is a potential of VR for some patients who need assessment and training but have higher risk. Through the VR system, we can get many detail objective and precise data which hardly to get from the real world to evaluate the performance. Moreover, the same results of improvement tendency on riding performance and balance ability performance were found in the VR system and in the real world. That reveals the training effects can be transferred to the real world. Therefore, it seems using MCVRA to do assessment and training riding modified motorcycle and balance ability is feasible. Because the relative researches are still limited, more studies are needed.

REFERENCES

[1] R Kizony, L Raz, N Katz, H Weingarden, PL Weiss. Video-capture virtual reality system for patients with paraplegic spinal cord injury. *J Rehabil Res Dev*, 42(5):595-608, Sep-Oct 2005.

[2] YS Lam, DW Man, SF Tam, PL Weiss. Virtual reality training for stroke rehabilitation, *NeuroRehabilitation.,* 21(3):245-53, 2006.

[1]e-mail: d097640@h.k.u-tokyo.ac.jp
[2]e-mail: whsung@ym.edu.tw
[3]e-mail: yatsai@vghtpe.gov.tw
[4]e-mail: hc_cheng@vghtpe.gov.tw
[5]e-mail: jjchen@ym.edu.tw

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Feasibility Study on the Estimation of Photo Shoot Position and Direction Based on Virtualized Reality Environment Models

Koji MAKITA* AIST†
Jun NISHIDA University of Tsukuba, Japan
Tomoya ISHIKAWA AIST
Takashi OKUMA AIST
Jun YAMASHITA University of Tsukuba, Japan
Hideaki KUZUOKA University of Tsukuba, Japan
Takeshi KURATA AIST / University of Tsukuba, Japan

Abstract

This paper provides a feasibility study on the estimation of photo shoot positions and directions in a modeled environment for augmented reality applications. This study focuses on the ability to compare a photo and images generated from virtualized reality environment models.

Keywords : Localization, Virtualized reality environment model, Model based matching, Mobile augmented reality

1 Introduction

Mobile augmented reality (AR) has possibility to be used in various places, conditions, and contexts. But it is difficult to achieve robust and global localization in wide areas in a single method, because each method has its own merits and demerits. Therefore, a combination of pedestrian dead reckoning and initialization methods has been attracting attention as one of the localization method in reasonable accuracy and cost [1]. For efficiently constructing an initialization method, we focus on the estimation of photo shoot position and direction by comparing the photo and images generated from virtualized reality environment models. In this study, a coarse-to-fine framework is introduced for reducing computational cost. In this framework, we evaluated three types of image similarities and an interest point matching.

2 Estimation procedure

The coarse estimation phase (A-1,2,3) is introduced in order to select some adequate images. In the fine estimation phase (B-1,2), photo shoot position and direction are estimated. Detailed descriptions of the two phases are given below.

(A-1) Generation of generated images from models

Images are generated from virtualized reality environment models. Each image is generated using one camera parameter in the model's coordinate system. In this study, discretely distributed positions and directions are introduced to determine the camera parameters.

(A-2) Calculation of similarities

Similarities between a photo and generated images are calculated. In this study, three different types of similarities are introduced for comparison purposes. They are based on the correlation of hue-saturation histogram, SSD (Sum of Squared Differences), and ZNCC (Zero-mean Normalized Cross-Correlation).

(A-3) Selection of generated images for fine estimation

Some images that have high similarity are selected to be used in the fine estimation phase.

(B-1) Matching of corresponding feature

The photo and each generated image selected in **(A-3)** are matched according to corresponding interest points. The SIFT [2] is introduced as a method to detect interest points in the photo and generated images.

(B-2) Estimation of photo shoot position and direction

The photo shoot position and direction are estimated. The algorithm must be developed considering characteristics of the matching result of **(B-1)**. Therefore, the development of the algorithm is treated as a future work in this study.

3 Experimental results

Experiments were conducted at three points (Point A, B, C) in an indoor office, and photos taken at three points by an iPhone 4 phone were used. For each photo, approximately the same image (corresponding image) was generated using the corresponding camera parameter in the model's coordinate system. Figure 1 shows an example of a photo and corresponding image. Next, to generate images from models, discretely distributed positions and directions were introduced. To set the position, the height of the camera is fixed and positions are discretely set every one meter on a 2D plane. To set the direction, the roll angle and the pitch angle are fixed and yaw angles are discretely set every 5 degrees.

For the coarse estimation phase, we set maximum of the threshold by which corresponding images are included in selected images. In hue-saturation histogram based similarity, the percentage of the narrow range at point A is 43%, at point B is 69%, and at point C is 40%. In SSD-based similarity, the percentage at point A is 24%, at point B is 67%, and at point C is 97%. In ZNCC-based similarity, the percentage at point A is 10%, at point B is 7%, and at point C is 96%. In the result, ZNCC based method was dominant at point A and B. The main reason why a large region was selected at point C is a reflection of window glass. In future, characteristic features of models should be considered for similarity calculations.

For the fine estimation phase, we counted the number of correct matching of interest points. In the result, the maximum number of correct matching of interest points was 14. Therefore, new methods to determine correctness of matching must be introduced in future.

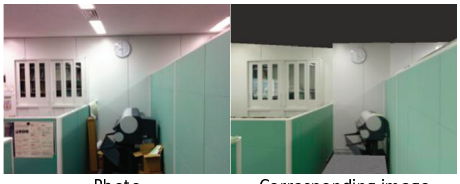


Figure 1: An example of a photo and corresponding image.

4 Conclusion

In this study, for the feasibility study of estimation of photo shoot position and direction based on virtualized reality environment models, some experimental results are described. In the coarse estimation phase, similarities based on the correlation of hue-saturation histogram, SSD, and ZNCC have been evaluated. In the fine estimation phase, matching results of interest points have been evaluated. In future studies, we need to research characteristics of the matching of interest points to develop the algorithm for fine estimation phase. For the research, we will introduce a variety of models.

Acknowledgements

This work was supported by ANR in France and JST in Japan.

References

[1] T. Ishikawa, M. Kourogi, and T. Kurata: Economic and Synergistic Pedestrian Tracking System with Service Cooperation for Indoor Environments, Int. Journal of Organizational and Collective Intelligence, Vol.2, No.1, pp.1–20, 2011.
[2] D. G. Lowe: Distinctive image features from scale-invariant keypoints, Journal of Computer Vision, 60, 2 , pp. 91–110, 2004.

*e-mail:k.makita@aist.go.jp
†Center for Service Research, National Institute of Advanced Industrial Science and Technology, Japan

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Obstacle sensation augmented by enhancing low frequency component for horror game sound

Shuyang Zhao[1] *Taku Hachisu*[1] *Asuka Ishii*[1]
Yuuki Kuniyasu[1] *Hiroyuki Kajimoto*[1]

[1]The University of Electro-Communications
1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan

[1]{zsy, hachisu, asuka, kuniyasu, kajimoto}@kaji-lab.jp

ABSTRACT

Horror computer games provide users with a mental stimulation that the real world cannot. Current horror games can provide the user with a visible ghost and stereo background sound to thrill the user. Inspired by obstacle sense- blind people localizing only with hearing, a novel method to augment existence is proposed. Obstacle sense is caused mainly by coloration by reflected sound and the attenuation by shielding. By focusing on the attenuation, we found an effective sense can be created by decreasing high frequency component and increasing low frequency component simultaneously. Experiments were conducted to evaluate our proposal.

KEYWORDS: Augmented reality, game background sound, horror game, obstacle sensation.

1 INTRODUCTION

Computer games combine the aesthetic and the social aspects in a way the old mass media, such as movies, and novels do not. Horror computer games are one of the most popular categories.

To elicit the emotional reactions of the upcoming frightening events and a more euphoric atmosphere, the fact of obstacle sense inspired us. Human can perceive the existence and the position of non-sound object aurally without visual information. This ability is known as “obstacle sense”. The factors of this perception may include the impression due to the change of acoustic field caused by the reflected sound [1]; the reduction in volume due to absorption- attenuation is another factor [2] (Figure 1).

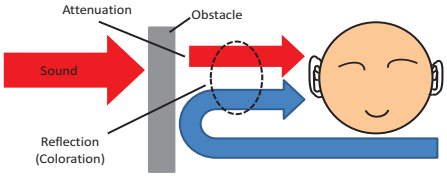


Figure 1. Schematic explanation of obstacle sense.

Our purpose is to “augment” the euphoric atmosphere of horror game background sound based on the principles of obstacle sense. Keeping with our proposal, attenuation caused by relatively smaller objective is set as a fast candidate. The attenuation is mainly observed by high frequency component since it tends to transmit straight and reflect, while the low frequency sound tends to diffract [3]. Two possibilities were assumed. First, drastic decrease of the high frequency component might create a more effective obstacle sense. Second, if human sense the obstacle by comparing the high and low frequencies, a more effective sense can be created by not only decreasing high frequency component, but also “increasing” low frequency component. By conducting the first preliminary trial, the result showed that the second possibility created a more effective obstacle sense.

2 EXPERIMENTS AND RESULTS

Two specific experiments were conducted. The first experiment compared the situation under different cut-off frequencies of low pass filter without volume difference. The sound stimulus was pink noise. The direction and subjective certainty of an obstacle (1~5) were asked to evaluate the effect. The results showed that as the cut-off frequency of low pass filter decreased, the obstacle sense perceived by subjects became stronger (Table 1).

Table 1. Results of experiment 1

Cut-off frequency(Hz)	Rates of correctness	Obstacle sensation certainty
400	91%	3.54
800	88%	3.32
1600	88%	2.92
3200	87%	2.27
6400	51%	1.37
10000	48%	1.30

The second experiment was under the same situation of the first experiment but with volume difference and classical music as sound stimulus. The results showed that when volume was decreased or increased, the subject answered the direction where volume was smaller (Table 2). The purpose of the demonstration is to demonstrate the experiments.

Table 2. Results of experiment 2

Cut-off frequency(Hz) \ Volume(dB)	-10	Constant	10
400	27% 3.46	56% 3.33	63% 3.23
800	33% 3.90	42% 3.07	70% 3.13
1600	35% 3.37	43% 2.33	63% 3.33
3200	39% 3.47	27% 1.90	80% 3.50
6400	30% 3.47	47% 1.57	70% 3.73
10000	40% 3.47	53% 1.57	63% 3.60

3 CONCLUSION

The two experiments’ results showed that human judge the obstacle existence according to the comparative difference in volume.

REFERENCES

[1] Y. Seki and K. Ito. Coloration perception depending on sound direction. IEEE Trans. Speech Audio Processing, 11:817–825, 2003.

[2] Y. Seki, T. Ifukube, and Y. Tanaka. The influence of sound insulation effect on obstacle sense of the blind. J. Acoust. Soc. Jpn. (J), 50(5):382–385, 1994.

[3] M. A. Price, K. Attenborough, W. Nicholas, and J. Heap. Sound attenuation through trees: measurements and models. J. Acoust. Soc. Am. 84(5):1836-1844, 1988.

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Intra-expo: Augmented Emotion By Superimposing Comic Book Images

Sho Sakurai, Shigeo Yoshida, Takuji Narumi, Tomohiro Tanikawa and Michitaka Hirose

The University of Tokyo

ABSTRACT

This paper proposes a method for conveying emotions in face-to-face communication by superimposing comic book images on real world using augmented reality technology. We implemented a system named "Intra-expo" that superimposes comic book images which influence our estimation of emotional state around the user. In order to build the system, we analyzed how subjects estimated the emotional state of a person when the person's image was superimposed on various comic book images. We then constructed a prototype of Intra-expo, which detects the user using the Microsoft Kinect system and projects different comic book images around the user, who selects their emotional state using an Apple iPhone as a mobile interface.

KEYWORDS: Conveying Emotion, Augmented Reality, Comic Book Images, Face-to-face Communication.

1 INTRODUCTION

Research on conveying emotion by using computer technology has become an active topic in recent years [1]. There are many approaches for conveying emotional states by using abstractly symbols, such as color or sound tone [2]. The practices of these approaches have been advanced in study to conveying emotion via online communication, but are rarely used in face-to-face communication. We propose a system named "Intra-expo" to convey emotion in face-to-face communication by augmenting emotion on real world. We focused on comic book images as a means of describing emotional states, because they are effectively used to describe the emotional states of characters in comics [3].

2 "INTRA-EXPO": AUGMENTED EMOTION BY SUPERIMPOSING COMIC BOOK IMAGES ON THE REAL WORLD

Firstly, we experimented to analyze how each of some comic book images has been generally-regarded as to be able to describe a certain emotional state and whether there are some comic book images which can describe a certain emotional state effectively. We analyzed 25 comic book images and some emotional states. The results are shown in Figure 1.

Next, we implemented a system named "Intra-expo." The system superimposes an appropriate comic book image around the user by projection the image, which describes the emotional state selected by the user, while remaining focus on the user's partner in face-to-face communication. "Intra-expo" has three components: a mobile interface to manually select the own emotional state, a section for user detection, and a section for projection of comic book images around the user (Figure 2). The users of "Intra-expo" said that the system enable them to convey own emotion properly.

7-3-1, Hongo, Bunkyo-ku, Tokyo, Japan

[sho | shigeodayo | narumi | tani | hirose]@cyber.t.u-tokyo.ac.jp

Figure 1. Correspondences of Comic book images used in our experiment and emotional state.

Figure 2. Scenes of superimposing comic book images describing some emotional states.

3 CONCLUSION

We proposed a method for conveying emotion in face-to-face communication using comic book images. In the future, we will establish the methods to support conveying emotional information in face-to-face communication. Especially we try to examine ways of conveying emotion in real world without user's operation of devices.

REFERENCES

[1] R.W. Picard. 2007. Affective Computing. *MIT Press*.

[2] Shimura, S.. Hirano, Y., Kajita, S. and Mase, K. 2005. Experiment of Recalling Emotions in Wearable Experience Recordings. In *Proceedings of the 3rd International Conference on Pervasive Computing*(May). 19-22.

[3] McCloud, S. 1990. Understanding Comics: The invisible Art. *Harper Parerbacks*.

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Pan-Tilt Projector Path Planning for Adaptive Resolution Display

Kei Kodama* Osaka University
Daisuke Iwai† Osaka University
Kosuke Sato‡ Osaka University

Abstract

We present a novel multi-projection system whose spatial resolution is not uniform.Recently, projection interactive systems which you can operate by touching with fingers or pens are studied. These systems' projection images need to have high resolution because they are often watched at close range. However, projectors which are used in these systems do not satisfy this condition. And also, the number of pixels of image sensors have more than 10 megapixel, and you can get a billion pixels by panoramcomposition. On the other hand, number of pixels of projectors on the market have only about 2 million pixels of Full HD, and even Super-HD projectors have only about 33 million. So it is out of doubt that they are not enough to display pictures photographically. In contrast, the resolution can be increased by using multiple projectors. Human recognizes in high resolution only by central field, and in low resolution by peripheral vision. So only to array projectors in the shape of a tile makes the region which does not need high resolution has high resolution.Therefore, in this study, we propose the system to realize dynamically reconfigurable pixels.

Keywords: multiple pan-tilt projectors, focus+context display, spatially varying resolutions, dynamically reconfigurable pixels.

1 Focus+Context Screen

Recently, focus+context display has been ploposed as an example of the technique to arrange the limited number of pixels effectively.

Human visual system performs detailed recognition in high resolution in the central field and in low resolution in the peripheral field. Focus+context display is the system that displays the picture by high resolution only in the region which is recognized by central field and by low resolution in the peripheral region.

There are a lot of study examples using this method. For example, Baudisch et al. proposes the method to display the whole image by a projector and display only the region where high resolution is needed by LCD[1]. In contrast, the system which can change high resolution region is proposed. Cotting et al. proposed the system which can display by high resolution a part of region by mobile projector and display by low resolution by a stationary projector[2].

These systems focus on the method for moving projection regions by manual operation, so to control the projective regions automatically has not been realized.

2 Proposed System

We apply multiple pan-tilt projectors to realize dynamically reconfigurable pixels. Each pan-tilt projector has different spatial resolution, and consequently the proposed system displays high resolution image contents with spatially varying resolutions. In this paper, particularly, we refer the case where the movie contents are displayed by the proposed system. In this case, because the region which should be displayed by high resolution is different in each frame, projection image should be moved while being moved. However, image degradation may occur when projection image moves. For example, a motion blur occurs if projection image is moved while displaying something, and the resolution decreases if projection image is moved while not displaying. So if the projection image is moved to a region of each frame simply, image deterioration occurs every time when optimal regions change. Therefore, we propose the optimal method for moving the projection image when movie contents are displayed using the information of user's gazed domain in each frame. In this study, we propose two methods for moving while displaying the projection image(method 1), and while not displaying(method 2) as described in figure1.

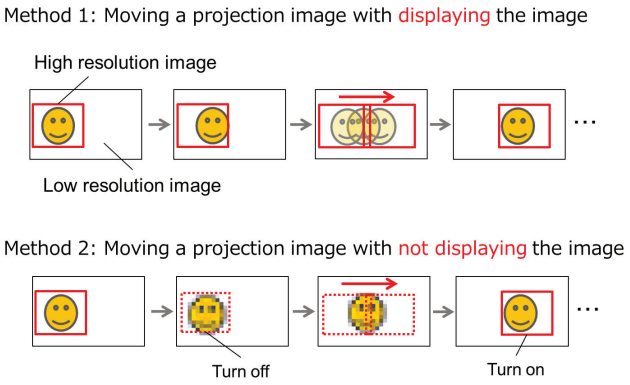


Figure 1: Concept of the proposed method

3 Image Quality Evaluation

We performed an experiment to evaluate a image quality of proposed system. In this experiment, we compare the image quality of 3 methods; One of these is that the projection image is moved to optimal region of each frame(method 0), and others are the method 1 and 2. In the simulation experiment by a monitor, we could get the result that the method 2 was the best, 1 was the second, and 0 was the worst. And in the experiment by the prototype, we compared the method 2 with 0, and we could get the result that the method 2 was the better in 3 contents, and in the other, the method 0 was the better.

References

[1] P. Baudisch, N. Good, V. Bellotti, and P. Schraedley. Keeping things in context: a comparative evaluation of focus plus context screens, overviews, and zooming. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 259–266, 2002.

[2] D. Cotting and M. Gross. Interactive Visual Workspaces with Dynamic Foveal Areas and Adaptive Composite Interfaces. *Computer Graphics Forum*, Vol. 26, No. 3, pp. 685–694, 2007.

*e-mail: kodama@sens.sys.es.osaka-u.ac.jp
†e-mail:daisuke.iwai@sys.es.osaka-u.ac.jp
‡e-mail:sato@sys.es.osaka-u.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Re-PITASu Concept: Touch-based Interaction Using Range Image Sensor with Image Projected onto Wall Surface

Yuki Uranishi, Goshiro Yamamoto, Hirokazu Kato*
Nara Institute of Science and Technology, Japan

Petri Pulli†
University of Oulu, Finland

Abstract

This paper proposes a concept of a system using a range image sensor for interacting with an image projected onto wall surface. A target surface and human hands are observed, and the finger motion is detected by the range image sensor. The user can interact with projected images by using the finger. The experimental results show that the prototype can detect the finger motion, and the contents shown by the projector can be interact with the user's finger.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, Augmented and Virtual Realities; H.5.2 [Information Interfaces and Presentation]: User Interfaces—Input Devices and Strategies

1 Introduction

Touch-based interaction is intuitive for interacting, and several methods have been proposed for interacting with the images projected on surfaces using a camera. PALMbit [3] has been proposed for projecting images onto the user's hand and operating the projected images by another hand using an Infrared camera and a projector. The method can detect the motions stably, however, the shape of the target surface is limited to hand shape. PiTaSu (Picture based Input Method Using Tapping on Wall Surface) [2] can detect the hand motion accurately. However, detectable motions are limited due to a property of the acceleration sensor.

We propose a system concept for interacting with surfaces called Re-PITASu (Rangeimage-Projector-based Interaction Tool for Arbitrary Surfaces). The proposed system consists of a range image sensor and a projector in use. The range image sensor is used to detect hands and a surface to interact with the images projected by the projector. The range image sensor is robust over colors of the scene. The scene of projection is generally dark and affected by the light from the projector. The Re-PITASu concept is aiming at calibrating between the range image sensor and the projector precisely compared to OmniTouch [1]. Experimental results show the calibration method of the Re-PITASu and the touch-based interaction by the Re-PITASu.

2 Overview of Proposed System

The Re-PITASu consists of a range image sensor and a projector. Figure 1 shows an overview of the Re-PITASu concept. To calibrate between the range sensor and the projector precisely, a camera is introduced to calibrate between the range image sensor and the projector indirectly. Firstly, a finger region to tap is extracted from an image taken by the range image sensor. Secondly, a distance between the finger and the surface is evaluated in the range image coordinate system. Lastly, the tapping action is detected by observing the distance between the finger and the surface.

*e-mail:{uranishi, goshiro, kato}@is.naist.jp
†e-mail:petri.pulli@oulu.fi

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

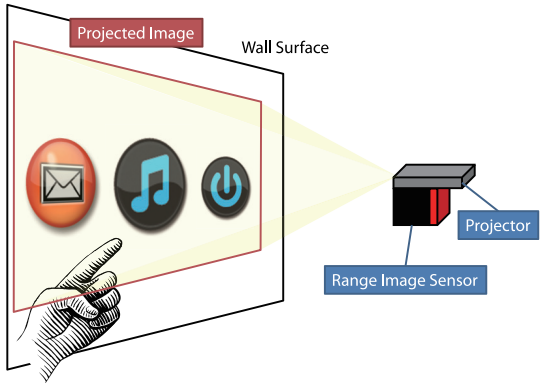


Figure 1: An overview of Re-PITASu.

(a) (b)

Figure 2: An example of the experimental results. (a) The finger is off the surface. (b) The finger is on the surface.

3 Experimental Results and Conclusions

Figures 2 (a) and (b) show an example of the experimental results. The prototype could detect the tapping and the label is displayed by the projector according to the position of the estimated finger position by the proposed calibration method.

Future work will aim at lifting restrictions on camera positions. We have used the background subtractions for the prototype. It is desirable that the position of the camera is estimated simultaneously without preliminary-taken images. In addition, the accuracy of the calibration should be improved. The indirect calibration in the prototype is so simple that the prototype has not been applicable for the practical use yet.

References

[1] C. Harrison, H. Benko, and A. D. Wilson. Omnitouch: Wearable multitouch interaction everywhere. *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, Oct 2011.

[2] G. Yamamoto, T. Kuroda, D. Yoshitake, S. Hickey, J. Hyry, K. Chihara, and P. Pulli. Pitasu: Wearable interface for assisting senior citizens with memory problems. *International Journal on Disability and Human Development*, Jul 2011.

[3] G. Yamamoto and K. Sato. PALMbit: A PALM interface with projector-camera system. In *Proceedings of the 9th International Conference on Ubiquitous Computing 2007*, pages 276–279, Sep 2007.

Localization with Microsoft Kinect using Natural Features and Depth Data

*Yuki Takabatake[a], Yuichi Tamura[b], Naoya Kashima[b] and Tomohiro Umetani[b]

[a]*Graduate School of Natural Science, Konan University*
8-9-1 Okamoto, Higashinada-ku, Kobe 658-8501, Japan
[b]*Department of Intelligence and Informatics, Konan University*
8-9-1 Okamoto, Higashinada-ku, Kobe 658-8501, Japan

ABSTRACT

This paper proposes a localization method using the Microsoft Kinect sensor and natural feature tracking with a red-green-blue (RGB) camera. It is difficult to measure self-position accurately because measurement errors in depth tend to be greater when using only an RGB camera. The Kinect sensor has both an RGB image camera and a depth camera; therefore it can overcome some of these problems. Finally, we provide examples of 3D reconstruction.

KEYWORDS: Kinect, 3D reconstruction, natural feature tracking.

1 INTRODUCTION

It is useful to provide current location information to a user. The Global Positioning System (GPS) is generally used to obtain such information; however, it cannot be used inside buildings. To overcome this problem, many studies have considered using machine vision with cameras. Vision-based localization methods with [1] and without artificial markers [2] have been proposed. However, measuring depth value with only a camera is generally inaccurate. We propose a localization system using Microsoft Kinect, which has both a red-green-blue (RGB) camera and a depth camera.

2 PROCESSING PROCEDURE

Figure 1 shows the procedure of the proposed system. First, the RGB and depth images are calibrated. Then natural feature points are explored, and base points, which are feature points with 3D location data, are selected and stored. Finally, the 3D location of the system (Kinect) is calculated from the base points. Base points are tracked with template matching and their locations are updated.

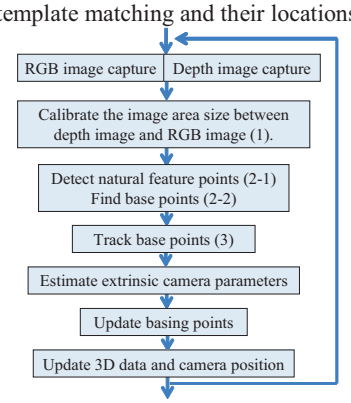


Figure 1. Processing procedure

*email: mn024010@center.konan-u.ac.jp

3 RESULT

We used this system to estimate the dimensions of a room and a corridor. The room size is about 7.5 m × 6.5 m. The Kinect was located on the marked point in Fig. 2 and rotated 360 degrees on a turntable. Figure 3 shows the result of estimation in the corridor. The width of the corridor is about 1.7 m.

The estimates were repeated 10 times in the room. The average measurement error was 31 cm (± 6 cm standard deviation) in the x-direction, 17 cm (± 3 cm) in the y-direction, and 41 cm (± 12 cm) in the z-direction.

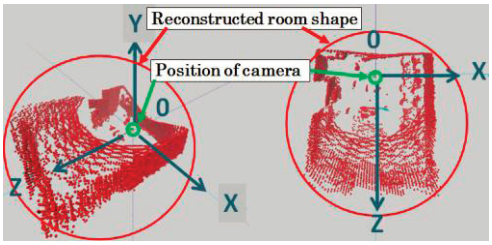


Figure 2. 3D reconstruction result in the room.

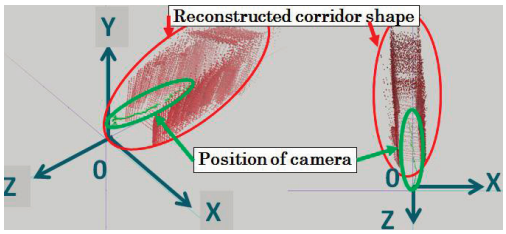


Figure 3. 3D reconstruction result in the corridor.

4 CONCLUSION

This paper proposes a localization method using the Microsoft Kinect sensor. We applied the system to a room and a corridor. The estimation results had a horizontal error of around 40 cm and a vertical error of about 20 cm.

REFERENCES

[1] L. Naimark and E. Foxlin. Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker. *Proc. IEEE/ACM Int. Symp on Mixed and Augmented Reality*, pp. 27-36, 2002.

[2] M. Oe, T. Sato and N. Yokoya. Estimating camera position and posture by using feature landmark database. *Proc. 14th Scandinavian Conf on Image Analysis (SCIA2005),* volume 13, pp. 171–181, June 2005.

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Real-Time Diminished Reality using Multiple Smartphones

Toshihiro Honda* Keio University
Takuya Inoue† Keio University
Hideo Saito‡ Keio University

Abstract

In this paper, we present a system for real-time Diminished Reality with multiple smartphones. In this system, we assume multiple smartphones capture the same scene that is occluded by obstacles. Areas of the obstacles are extracted from each camera image replaced with image of the hidden areas which are captured using different viewpoint camera. In the proposed method, we suppose that the target scene can be approximated as a plane. Therefore, we compute homography matrices between each camera image by using natural features. Then, obstacle area which is not approximated as a plane can be removed by synthesizing the image warped with the homography matrix and the user viewpoint image. We can perform real-time processing because we send each camera image to PC which returns obstacle-removed images at every frame. We experimentally demonstrate the effectiveness of the proposed method using three viewpoint images.

Keywords: Diminished Reality, Multiple Smartphones, Real-Time, Homography

1 Introduction

Diminished Reality(DR) is a technique for removing obstacles and replacing the area with a proper target scene image. Enomoto et al. proposed DR using multiple web cameras[1]. In their system, they allow each web camera connected each PC to share data through a network, and compute homography matrices between each camera using ARTag, then remove obstacles by blending process.

In the proposed method, no ARTag is required because homography matrices are computed in real-time by matching natural feature points. We also apply median process in order to determine the color of the replaced pixels, which is more simple algorithm than blending process. Users can move freely by using smartphones as a device for capturing camera image.

2 Proposed System

Figure 1 shows our proposed system. We use multiple smartphones in our system. In this paper, we introduce our method using three smartphones.

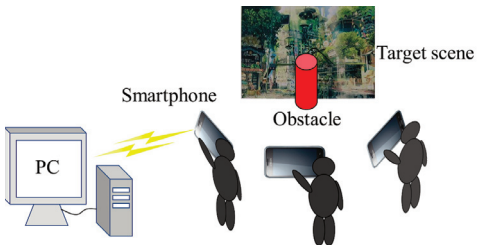


Figure 1: Experimental system.

*e-mail:t-honda@hvrl.ics.keio.ac.jp
†e-mail:inoue@hvrl.ics.keio.ac.jp
‡e-mail:saito@hvrl.ics.keio.ac.jp

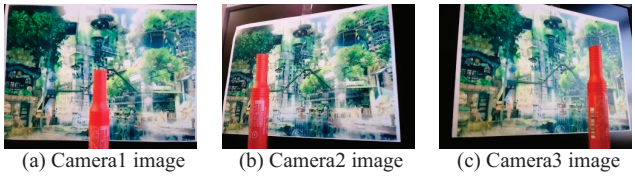
The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

C_1, C_2, C_3 are camera images captured by the three smartphones. We detect natural feature points of C_1, C_2, C_3 using CenSurE[2], then describe feature values of the points using BRIEF[3]. The detected feature points in each camera are mathched with the feature points deteceted in the other cameras based on the descriptors. According to the correspondenced feature points, we can compute homography matrices $\mathbf{H}_{21}$ (betweeen C_1 and C_2) and $\mathbf{H}_{31}$ (between C_1 and C_3). In this case, C_1 is a basis image that relates C_2 and C_3, so that the obstacles in C_1 can be replaced with the object scene images captured in C_2 and C_3.

C_2 is warped to an image which is seen from the smartphone capturing C_1 by $\mathbf{H}_{21}$. Similarly, C_3 is warped by $\mathbf{H}_{31}$. At this time, locations of obstacles are different among the three images because obstacles are not approximated as a plane. Therefore, pixel values of the other two images in the location of obstacles which are projected on one of three images are pixel values of the target scene. Because of this, we can get the image removing obstacles if we use pixel values in the middle of ones of three images for every pixel.

3 Experimental Result

Figure 2 shows the result of removing an obstacle by using this system. The experimental platform was implemented on GALAXY S and 3.4GHz Intel Core i7-2600 desktop with 3.49GB RAM. The image size is 320x240. We can see that the obstacle is removed and the entire background poster can be seen. In addition, proposed system runs in approximately 5-6 fps, which can be approximately considered as real-time.



(a) Camera1 image (b) Camera2 image (c) Camera3 image

(d) DR image of (a) (e) DR image of (b) (f) DR image of (c)

Figure 2: Removing an obstacle.

4 Conclusion

Through experiments aimed at the real scene, we confirmed that we could remove obstacles in real-time.

References

[1] Akihito Enomoto and Hideo Saito. Diminished Reality using Multiple Handheld Cameras. ACCV'07 Workshop on Multi-dimensional and Multi-view Image Processing, pp. 130-135, 2007.

[2] Motilal Agrawal, Kurt Konolige and Morten Rufus Blas. CenSurE: Center Surround Extremas for Realtime Feature Detection and Matching. ECCV, volume 5305, pages 102-115, 2008.

[3] Michael Calonder, Vincent Lepetit, Christoph Strecha and Pascal Fua. BRIEF: Binary Robust Independent Elementary Features. ECCV, volume 6314, pages 778-792, 2010.

Adaptive Annotation Layout in Projection-Based Mixed Reality by Considering Its Readability

Tatsunori Yabiki* Daisuke Iwai† Kosuke Sato‡

Graduate School of Engineering Science, Osaka University

Abstract

Superimposing annotations on physical objects such as the 3-D model of a human body in projection-based mixed reality can help our understanding of the objects. For the projection-based annotation, it must be carefully considered that the readability of the annotation varies according to the shape and texture of the object's surface. In this paper, we propose a method which applies a genetic algorithm (GA) to compute the adaptive layout of superimposing annotation on an arbitrary surface so that the readability is not much degraded. This paper also shows the result of a psychophysical test which was carried out to investigate the issue.

Keywords: Projection-based mixed reality, readability, optimal layout, superimposing annotation

1 Introduction

Annotations help our understanding of objects. In recent years, much research has been studied that automatically calculates an appropriate layout [1]. For superimposing annotations for real 3-D objects, mixed reality (MR) has been used. This paper aims at superimposing annotations to the real object surface using projection-based MR. In that case, there is a problem that the readability of a projected annotation significantly degrades when the shape and texture of a projection surface spatially vary.

Therefore we propose an adaptive annotation layout technique which computes the readability of each annotation projected on a non-planar and textured surface.

2 Proposed Technique

The flow of the proposed technique is shown in Figure 1. The input is the shape and texture data of object, labeling problem (annotations and their regions) and position of projectors. An energy-minimizing optimization using the genetic algorithm then generates the adaptive layout of annotations by considering the readability of projected letter.

3 Evaluation of the Readability of Projected Letter

It is thought that distortion and the shadow of a projected character, and the contrast of a background color to a projected letter affect the readability of a projected image. So we conducted a psychophysical test to investigate the issue. In the experiment, the projected result of an alphabet on a non-planar surface was simulated. The gauss noise was added to the generated image. Examinees observed the image displayed on the screen and answered which character was projected.

*e-mail: yabiki@sens.sys.es.osaka-u.ac.jp
†e-mail: daisuke.iwai@sys.es.osaka-u.ac.jp
‡e-mail: sato@sys.es.osaka-u.ac.jp

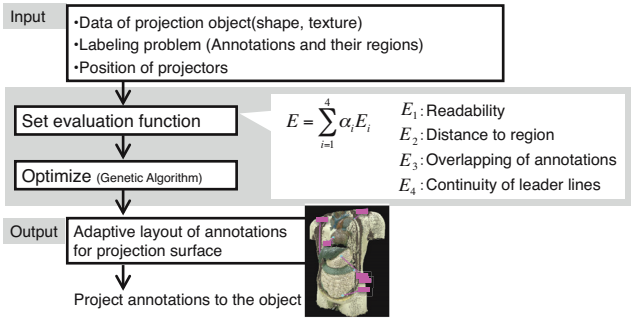


Figure 1: Flow of the proposed technique

The index of readability R is defined as the maximum quantity of the gauss noise in that the examinee could answer correctly. The result was fitted to a plane by the least-square method as follows:

$$R(D,O,C) = a_1(C) + a_2(C)D + a_3(C)O \quad (1)$$

where D is an average of the distance of each pixel of the character domain from a position in simulation image to the position in case that it is not distorted, O is a rate of the portion that is invisible from a virtual viewpoint or a virtual projector position, and C is the contrast of a background color to a projected letter. a_1, a_2, a_3 are represented by the continuous function using C as follows: $a_1(C) = 395.368 - 753.654C + 390.324C^2$, $a_2(C) = -9.11365 + 11.2027C$, $a_3(C) = -100.728 - 273.116exp(-7.259C)$.

In order to evaluate that the function R can estimate the readability of projected characters in real environments, we conducted the subject experiment. Subjects observed alphabets projected on real 3-D objects and ranked them according their readability. As a result, the correlation coefficient between the average value of the ranking which all the subjects answered and the value of readability calculated by R was -0.829. Therefore, it is thought that the readability of the projection character in real environment is computable using the proposed function R.

4 Conclusion

We proposed a technique that computes the adaptive layouts of superimposing annotation by considering the readability of projected letter. We conducted a psychophysical test to evaluate the readability of projected letter and defined the model function of the readability.

We plan to construct a layout system of annotation by using this result of the experiment.

References

[1] I. Vollick, D. Vogel, M. Agrawala, and A. Hertzmann. Specifying label layout style by example. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*, UIST '07, pages 221–230, New York, NY, USA, 2007. ACM.

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Model based tracking of rigid curved objects using sparse polygonal meshes

Marina Atsumi Oikawa*
Nara Institute of Science and Technology

Goshiro Yamamoto†
Nara Institute of Science and Technology

Makoto Fujisawa‡
University of Tsukuba

Toshiyuki Amano§
Yamagata University

Jun Miyazaki¶
Nara Institute of Science and Technology

Hirokazu Kato‖
Nara Institute of Science and Technology

ABSTRACT

In this paper a framework to improve model based tracking of rigid curved objects using polygonal meshes is presented. Previous approaches deal with curved objects treating them as polyhedral objects but requiring the use of dense meshes which may be computationally inefficient. Furthermore, when considering, for instance, applications targeting mobile devices, the data size of this model can become an inconvenience to the final user. However, reducing the quality of the object mesh creates a trade-off between the computational time and tracking accuracy. In order to solve this problem, we suggest the use of quadrics calculated for each patch in the mesh to give local approximations of the object shape. Then, curves representing the quadrics projection in the current viewpoint are used for distance evaluation. This representation allowed us to considerably reduce the level of detail of the polygonal mesh and keep an accurate tracking.

Keywords: Apparent contour, quadrics, model-based tracking, camera pose estimation, sparse polygonal meshes.

1 INTRODUCTION

Tracking the 3D pose of a known object is a common task in computer vision and many approaches to achieve real-time tracking have been developed in order to attend different applications and scenarios as can be seen in the crescent number of systems in Augmented Reality (AR).

This work focus on model-based tracking that considers the object edges while doing tracking, similar to the approaches presented in [3] and [2]. A polygonal mesh of the target object is used for matching with the edge information found on the video image. Given an initial estimation of the pose, edge normal search of projected edges in the image is performed and the final pose is obtained after an optimization process. However, the approaches mentioned above are applied mainly to polyhedral objects with flat faces.

When dealing with curved objects, the tracking becomes more challenging because dense meshes are required to accurately recover the object shape, creating a trade-off between the computational time and tracking accuracy as exemplified in Fig.1(a): the number of patches in the mesh was reduced in order to improve the system efficiency, but a larger distance d between projected and detected edges is used for error evaluation, affecting accuracy.

*e-mail: marina-o@is.naist.jp
†e-mail: goshiro@is.naist.jp
‡e-mail: fujis@slis.tsukuba.ac.jp
§e-mail: amano@yz.yamagata-u.ac.jp
¶e-mail: miyazaki@is.naist.jp
‖e-mail: kato@is.naist.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

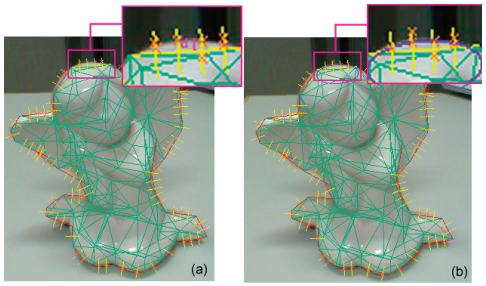


Figure 1: A sparse polygonal mesh is overlaid on the target object and a comparison between the distance evaluation of (a) the standard tracking and (b) our proposed approach using conics is showed.

2 PROPOSED FRAMEWORK

In our framework, a general quadric equation is calculated for each patch in the mesh and curves representing the quadrics projection in the current viewpoint are used for matching with detected edge points in the video image (conics represented by the blue lines). In Fig.1(b), it is possible to notice it approximates better the object contour and the error is clearly smaller when compared to Fig.1(a) - part of the ellipse passes exactly on the detected edges.

Quadrics were chosen because they have simple contour generators and their apparent contour can be easily obtained by using the theory provided by differential geometry [1]. They are represented by conics and using them instead of the original edges from the mesh makes the tracking more robust when dealing with sparse meshes because more correct point correspondences can be found and more accurate because it is able to approximate better the local shape of the object.

Our main contribution is the creation of a simple representation that can be easily constructed and at the same time efficient when dealing with curved objects having different shapes. This representation also allowed to considerably reduce the data size of the polygonal mesh (in some cases, the number of patches can be reduced to 10% of the number of patches from the dense mesh), making it a good option for AR applications in mobile devices, for instance. Experimental results comparing the use of dense and sparse meshes will be presented using both simulated and real image data.

REFERENCES

[1] R. Cipolla and P. Giblin. *Visual Motion of Curves and Surfaces*. Cambridge University Press, 2000.

[2] A. I. Comport, E. Marchand, M. Pressigout, and F. Chaumette. Real-time markerless tracking for augmented reality: The virtual visual servoing framework. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):615–628, 2006.

[3] T. Drummond and R. Cipolla. Real-time visual tracking of complex structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):932–946, 2002.

AR based Co-located Meeting Support System

*Igor de Souza Almeida ¶ Jun Miyazaki † Goshiro Yamamoto ‡ Makoto Fujisawa § Toshiyuki Amano // Hirokazu Kato

* ¶ † // Nara Institute of Science and Technology ‡ University of Tsukuba § Yamagata University

ABSTRACT

The use of Augmented Reality (AR) for the purpose of meeting support has been explored in several experimental and commercial solutions however most of these works focus on remote communication. This work targets the use of AR for co-located (face-to-face) meeting support in order to enhance interaction among and between meeting participants and provide collaboration awareness to all. A prototype system called *Meetsu* was developed and it is currently under experimentation.

KEYWORDS: meeting support system, augmented reality, co-located human-human communication.

1 INTRODUCTION

Meeting support systems can also be referred to as group support system (GSS). Even though GSS stands for a general term, its focus is primarily to promote social interactions and enhance meeting performance. In other words, GSS presents important factors which can be used to examine the social interaction between meeting attendants.

One work aiming co-located meeting support is [1] where a prototype for a face-to-face meeting support system (called *HEMS*) based exclusively on the use of handhelds. This system allows people to meet in any place where the handheld connection is able to support the various tasks and processes that arise over the life-cycle of a meeting.

[2] describes an informal communication tool called *MoCHA*. It is aimed to support co-located hospital workers. The prototype system's sharing service consists of displaying the contents of any device in the vicinity, such as a PDA, a PC or a public display, on a handheld computer, and being able to remotely share the control of the device with its owner and/or other users.

Our prototype system, Meetsu, intends to provide the intrinsic benefits of GSS by means of AR to a co-located meeting situation. The exploration of the possibilities given by AR to the aforementioned context constitutes the basis of this work. The study case being used for the experimentation is that of a weekly research meeting in which the participants include students and professors.

2 PROTOTYPE SYSTEM

Meetsu is a web system mainly developed using PHP and JQuery. Its features were decided over iterative discussions on the frequent scenarios for research meeting context as well as general meeting support features.

The main idea is to create a non-intrusive smooth solution for participants to express actions in the context of a research meeting where students present their research progress to their lab members. Users interact through a web interface which controls the icons being overlaid on top of the live video image of the meeting room with all participants.

In our first attempt to explore AR, the supported actions include making a question and expressing agreement/disagreement through the use of interactive icons (Figure 1) which will be displayed on top of the user's head (Figure 2).

Figure 1. Interactive icons that are available in Meetsu's web interface.

Figure 2. Icons being overlaid on top of a live camera feed.

Questions can also be sent at anytime through the Meeting phase using a submission form, also resulting in the question mark icon being shown in real-time.

2.1 Meeting setup

For this study, the system considers a presentation setting where two screens are visible to the attendants. One screen displays the presenter's slides (a Powerpoint presentation) and the second screen contains Meetsu's AR feature which shows a live camera feed of the audience. A webcam is positioned on top of the front screen capturing images of the meeting room.

3 CONCLUSION

We believe the simple idea of using icons to visually represent one's action (for example, making a question) stimulates participation, provide awareness of everyone's contribution and facilitates management of turn-taking during Q&A.

REFERENCES

[1] Zurita, Gustave; Baloian, Nelson. Handheld-Based Electronic Meeting Support. In Proceedings of Collaboration Researcher's International Workshops on Groupware, p.341-350, 2005.

[2] Mejia, David A. Supporting Informal Co-located Collaboration in Hospital Work. Collaboration Researchers International Working Group (CRIWG), p.255-270, 2007.

*igor-a@is.naist.jp

† goshiro@is.naist.jp

‡ fujis@slis.tsukuba.ac.jp

§ amano@yz.yamagata-u.ac.jp

¶ miyazaki@is.naist.jp

// kato@is.naist.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

SURF-based Line Marker for Augmented Reality

***Hiroki Yoshinaga**[1] **Yoichi Muraoka**[2]

[1 2] Waseda University

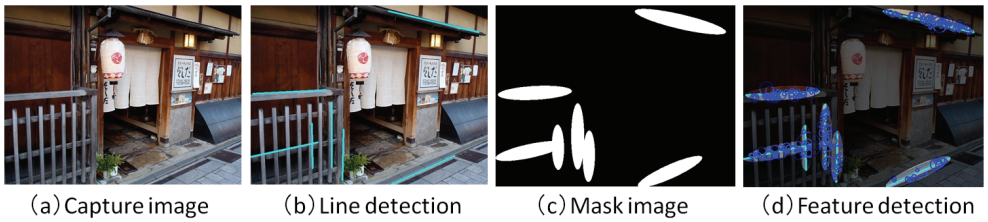


Figure 1. Line-based Marker

ABSTRACT

A SURF-based Line marker for Augmented Reality (AR) and AR tourist guide system for an iPhone is presented in this paper. In some tourist spots like Kyoto, putting up a sign is prohibited for landscape policy. However the AR marker may be put freely without ruining the landscape. In this system we display information on structures like temples, shrines, and shops. Most of them are composed of straight lines.

Hence detecting feature points around the lines, using the Hough transform, enables to reduce computational cost and memory usage.

KEYWORDS: natural features, Augmented Reality, object recognition.

INDEX TERMS:

H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – Artificial, augmented, and virtual realities; I.4.8 [Image Processing and Computer Vision]: Scene Analysis – Object recognition

1 BACKGROUND

When using features as a marker, it is not matching step but feature detection step which takes most of the computation time. By limiting target area to domain around the straight lines, we can reduce the area by 90%. Therefore detecting features time also can be reduced dramatically.

This approach is not improving on feature detection itself like [2] but restrict on target area. Then it enables to use both other approaches like [2] and this one at the same time.

2 OVERVIEW

This system is composed of two steps: feature detection, feature matching. In the detection step, the straight lines (b) are detected using Hough transform to an image (a) (see Figure 1&2). Next a mask image (c) which is drawn ellipses around each straight lines is created and feature detection (d) is done using it at last. In this approach, we use this set of feature points contained in one ellipse as a marker and store in the database.

*email:hirokiy@muraoka.info.waseda.ac.jp

In the matching step, this system uses Nearest Neighbor Search and recognition rate is about 70%. On the other hand it is effective approach to reduce memory usage since the number of detected feature points are reduced by 90%. It is also effective to reduce computational cost and reduces detecting time by 60%.

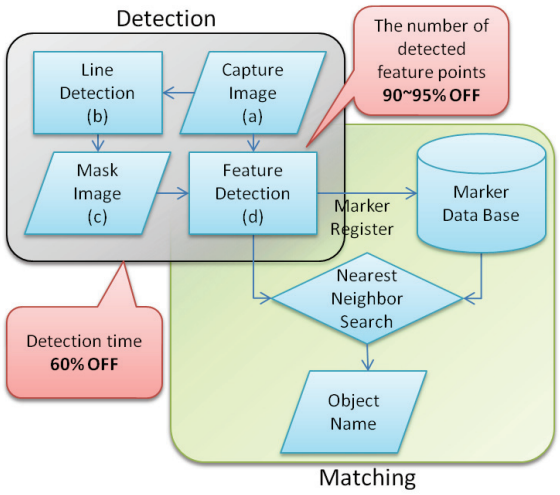


Figure 2. System Flow

3 DEMONSTRATION

In our demonstration, this system is actually run on an iPhone. Although it is originally a guide system for tourist, some substitutes are used instead of buildings this time. We register those data as markers in advance and display the name and information when you take a picture of them.

REFERENCES

[1] H. Bay, T. Tuytelaars, and L. V. Gool: Speeded up robust features, In Proc. ECCV 2006, 2006.

[2] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond, D. Schmalstieg: Pose Tracking from Natural Features on Mobile Phones . ISMAR '08: Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality.

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

CG Image Generation of Developmental Origami Model of Hypercube

Haruki Chiba* Keimei Kaino** Kuniaki Yajima† Takatoshi Suenaga ‡

Sendai National College of Technology, Hirose Campus

ABSTRACT

A four-dimensional space is a space whose fourth axis is perpendicular to a three-dimensional space. In a four-dimensional origami we fold a solid material along flat planes in a four-dimensional space. We will show a developmental wire-frame model of hypercube and construct it from its development. Its development is regarded as a four-dimensional origami with a front and a back side. Defining the four-dimensional CG methods as shading and painter's algorithm in the three-dimensional CG methods, we will show CG images of constructing developmental origami of hypercube from the development.

INDEX TERMS: I.3.3 [Computer Graphics]: Picture/Image Generation—Display Algorism

1 INTRODUCTION

An origami has much to offer as an instrument for experimenting with scientific and educational ideas. A four-dimensional origami is an analogue of the well-known origami folded in a three-dimensional space [1]. This is folded in a four-dimensional space where an additional fourth u-axis is perpendicular to a usual three-dimensional xyz-space (from now on, we will abbreviate "n-dimensional" by "n-D").

Figure 1 shows a model of this 4-D space where the basal plane presents the 3-D space. We call this 3-D space the u=0 hyper-plane. Miyazaki showed some figures of constructing a four-dimensional cube from its development which consists of eight congruent cubes [1]. This 4-D cube has not both front and back sides of a material for the sake of simplicity. First we will present a developmental wire-frame model of 4-D cube. When the side cells around the base cell are folded along their fold planes, a projection of the wire-frame model on the u=0 hyper-plane gives the same figure as shown by Miyazaki [1]. Secondly we will define a front and a back of an origami material in a 4-D space [2]. For CG image generation of 4-D objects, we use 4-D painter's algorithm. We may construct a view space by using the stereogram.

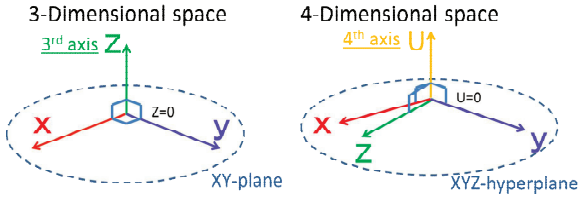


Figure 1. 3-D space and 4-D space

2 DEVEVOPMENTAL ORIGAMI MODEL OF 4-D CUBE

Figure 2 shows a developmental wire-frame model of 4-D cube. Note that a desk on which the development is laid corresponds to the u=0 hyper-plane. The 4-D cube consist of eight cubes, where both top and bottom bases are red and the other cubes are around the bottom base. A boundary between two cubes is a fold plane. When the side cubes around the base are folded along their fold planes, its projection of the wire-frame model on the u=0 hyper-plane is the same as one of the figures shown by Miyazaki [1].

Let us take a solid in a u=0 hyper-plane and move this solid a bit downward in the direction of the u axis. In this pair of solids, we regard the upper solid as a front and the lower one as a back and define its normal vector n as a direction of the relative movement (0,0,0,1). We will call such a pair of solids a 4-D origami [2].

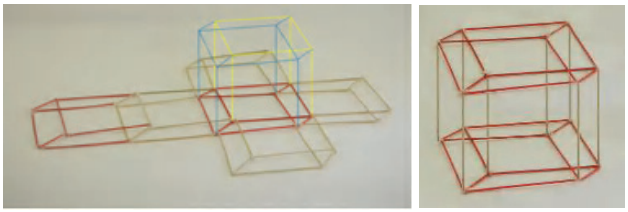


Figure 2. Developmental wire-frame model of 4-D cube

4-D painter's algorithm is a method of projecting 4-D objects on a view space which is 4-D analogue of a view plane. Shading and shadowing methods are applicable to 4D objects. Note that there is only one point of intersection where a line of light and a hyper-plane meet. Figure 3 shows CG images of during a folding process of 3-D origami cube which are analogue of ones of 3-D origami cube.

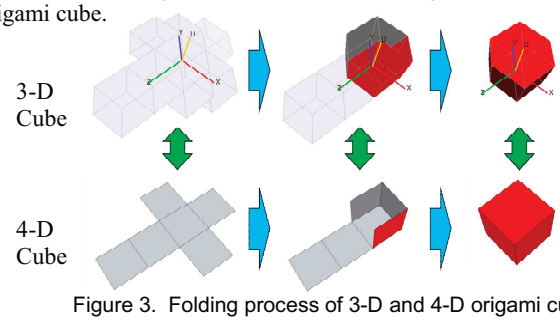


Figure 3. Folding process of 3-D and 4-D origami cube

3 CONCLUSION

CG images of a construction of 4-D cube from the development are intuitively understood by using the developmental wire-frame model of 4-D cube. Using 4-D CG methods we have generated CG images of a developmental origami model of 4-D cube. Those give us a good understanding of 4-D object and space.

REFERENCES

[1]K. Miyazaki. Four-dimensional origami. In *Proceedings of the Second International Meetings of Origami Science and Scientific Origami* (Shiga, Japan, Nov 29–Dec 2, 1994), pp.51–61. Seian University of Art and Design, March 1997.

[2]A. Inoue et al. CG image generation of four-dimensional origami. In *The Journal of the Society of Art and Science*, volume 4, pp.151–158. December 2005.

4-16-1 Ayashi-chuo,Aoba-ku,Sendai,9893128,Japan
* a1102029@sendai-nct.jp,**kaino@sendai-nct.ac.jp,
† yajima@sendai-nct.ac.jp, ‡ sue@sendai-nct.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Data adjustment methods of a low-priced data glove

Shinichi Hamaguchi* Sanshiro Yamamoto Kenji Funahashi† Hidenori Kanazawa
Department of Computer Science and Engineering, Nagoya Institute of Technology NTT COMWARE Tokai Co.

1 Introduction

A data glove is one of devices which are used in the field of virtual reality. We must use a data glove which has many sensors to capture a variety of human hand motions. However it is expensive and a low-priced data glove does not have enough sensors to capture hand data correctly. There are researches about a data adjustment method with low-priced glove [1], it used finger angle correlations only for the grip motion. In this paper, we propose two data adjustment methods. One is based on object shape knowledge which is held [2], another is based on a hand motion pattern estimation [3]. In our experiment system we chose three representative motions to hold; **grip** for a cube/cylinder type object, **nip** for a thin object, and **pinch** for a small object held by a thumb and an index finger. Then we calculate all finger joint angles from each hand motion pattern which is surveyed in advance. Using our new methods, we can adjust finger joint angles from just five sensors of a glove.

2 Object knowledge method

When we hold an object, our hand motion pattern is usually affect with a shape of the object. Thus we can estimate a hand motion pattern from an object shape knowledge. We supposed an object shape to be a rectangular solid in an experiment system. From a preliminary survey we decided basic sizes of an object and made six appropriate hand motion patterns for each size object. When the object is not any basic dimension, an interpolated motion is made as shown in figure 1. Although these patterns are assumed that a hand confronts the object directly, a hand does not always confront directly it. In this case we also make an interpolated hand motion pattern. Then we obtain equations for the pattern which is suitable for an unknown size object and a hand motion. Using this equations, we calculate proper angles of all finger joints.

3 Finger relation method

The hand motions can be expressed with difference of finger angles. It means that each hand motion has each relations among angles of fingers during operation. Therefore we can estimate a hand motion pattern using these relations. For a preliminary survey, we sampled the angles of five finger sensors and all joints for the representative motions, then we matched sensor value relation to each motion which can obtain all angles. However a hand does not behave according to exact representative motions actually. So an interpolated relation is made for the motion among the representative motions. Then we obtain the equation for the present motion and calculate proper angles of all finger joints.

4 Conclusion

Figure 2 shows differences of grip and pinch operation on the system based on our methods. Although sensor angles of index finger are the same, calculated angles of all joints are different as shown in CG images. As a result of evaluation experiments, both methods are effective to capture different hand motions. In the future, we should improve these methods to adopt other hold styles, and to integrate each other. We have also researched about VBDG (vision based data glove) which capture a hand motion from camera image of the hand. We are going to combine these adjustment method and VBDG together.

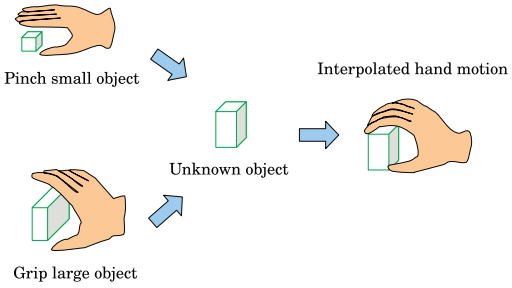


Figure 1: Interpolate hand motion

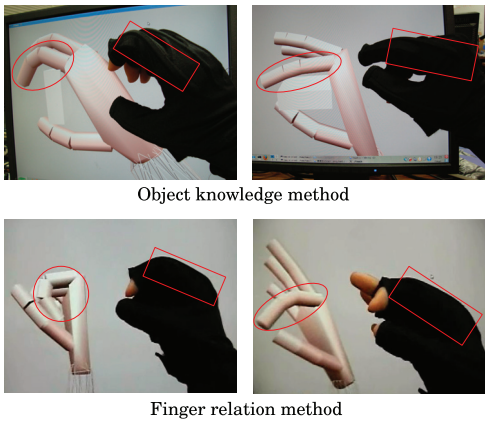


Figure 2: Difference of two hand motions

References

[1] D. G. Kamper and E. G. Cruz, and M. P. Siegel. Stereotypical fingertip trajectories during grasp. In *J Neurophysiol* 90, pages 3702-3710, 2003.

[2] S. Yamamoto, K. Funahashi and H. Kanazawa. A data adjustment method of low-priced data-grove based on object shape knowledge. In *Proceedings of the* 16*th Annual Conference of the Virtual Reality Society Japan*, 33D-5 (DVD-ROM, in Japanese), 2011.

[3] S. Hamaguchi and K. Funahashi. A data adjustment method of low-priced data-grove based on hand motion pattern estimation. In *Proceedings of the* 16*th Annual Conference of the Virtual Reality Society Japan*, 33D-6 (DVD-ROM, in Japanese), 2011.

*e-mail: hama@center.nitech.ac.jp
†e-mail: kenji@nitech.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Addition of 3D sound based on the position and the area of an object in a silent video

Miwa Nishimura[1)], Tsuyoshi Kobayashi[2)], Jun Murayama[3)], Yukihiro Hirata[4)], Makoto Sato [5)] and Tetsuya Harada [3)]
1) 3)Tokyo University of Science 2)5) Tokyo Institute of Technology 4) Tokyo University of Science, Suwa

KEYWORDS: Auditory sense, Sound generation, Motion vector

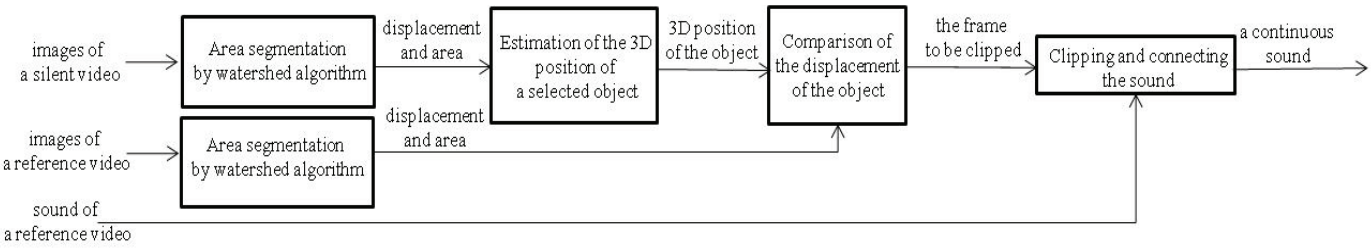


Figure 1. Processing procedure

1 INTRODUCTION

Recently, it has become possible to save and use old valuable videos easily through video sharing services. However, those videos often lose important information, such as sound or color.

This paper proposes a system for adding appropriate sound to silent videos. The proposed system gives more reality to the scenes, by adding sounds - it generates appropriate sounds based on the position and the area of objects in the videos. These are obtained by applying a watershed algorithm. Appropriate sounds were selected from a database prepared previously, and added to the silent video.

2 METHOD OF ADDING SOUNDS

A reference video is filmed prior to the addition, and is a record of an object which makes sounds similar to the object in a silent video. Figure 1 shows a block diagram of the processing procedure for adding sounds to silent videos.The system computes the position and the area of the object in each frame in the silent video using the watershed algorithm. The system also computes the displacement and the area of the object in the reference video. Next, the system clips sounds corresponding to the displacement in the reference video which has the nearest value to that in the silent video in every 0.5[s]. Figure 2 shows an example in which continuous sound corresponding to the silent video is clipped. The continuous sound that is added to the silent video is generated by linking the fragments of those sounds in sequence. The three dimentional position of the object is given by (1).

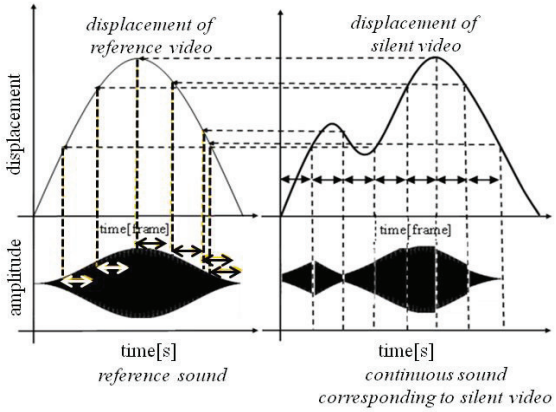


Figure 2. The rink of the *reference sound to the silent*

$$\begin{cases} x = \sin\left(\tan^{-1}\left(\frac{a_s}{a}\tan\phi_x\right)\right)\sqrt{\frac{S}{S_S}}r_s \\ y = \sin\left(\tan^{-1}\left(\frac{a_s}{a}\tan\phi_y\right)\right)\sqrt{\frac{S}{S_S}}r_s \\ z = \cos\left(\tan^{-1}\left(\frac{a_s}{a}\tan\phi_x\right)\right)\sqrt{\frac{S}{S_S}}r_s \end{cases} \quad (1)$$

The graphical representation of (1) is shown in Figure 3. Finally, the continuous sound at the three dimensional position of the object is added.

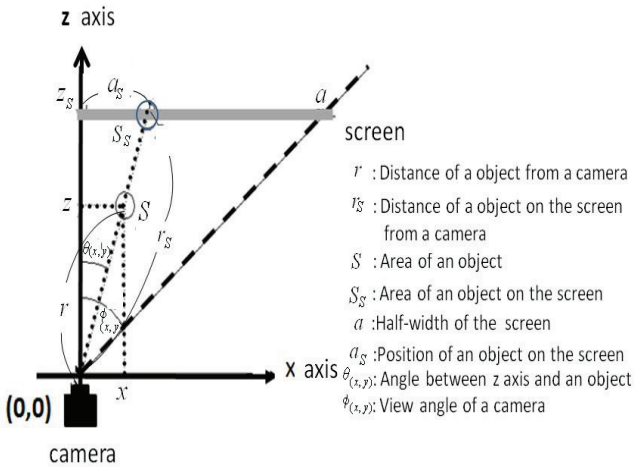


Figure3. Presumption of three dimensional position

1)email: j8110645@ed.noda.tus.ac.jp
2)email: tkobayashi@hi.pi.titech.ac.jp
3)email: {harada,murayama}@te.noda.tus.ac.jp
4)email: yhirata@rs.suwa.tus.ac.jp
5)email: msato@pi.titech.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

A Web Application for an Interior-Design Simulator using Augmented Reality

Tomoki Tanaka[1)]
Graduate Student,
Graduate School of Engineering
Chiba University

Takuma Nakabayashi[2)]
Graduate Student,
Graduate School of Engineering
Chiba University

Keita Kado[3)]
JSPS Research Fellow,
Building Department National Institute for Land and Infrastructure Management

Gakuhito Hirasawa[4)]
Associate Professor,
Graduate School of Engineering
Chiba University

ABSTRACT

We describe an interior-design simulator implemented as an augmented reality (AR) web application. The system is freely available over the internet and open for use by anonymous users. The purpose of this research is to evaluate the effectiveness of the AR system in the architectural field; therefore in the near future we plan to interview users.

KEYWORDS: Augmented Reality, Interior Design Simulator, Web Application.

1 INTRODUCTION

As described in the paper by TAMURA[1], it is difficult to evaluate the practicality of AR systems where various factors complicate the evaluation. The gap, jitter and occlusion of overlapping 3D graphical representations on a back-plate image are typical factors. Completeness of a virtual 3D model, reproducibility of light environments, and video frame rates can also be factors. Which of these factors is important depends on the particular evaluator.

After considering practicalities in evaluating AR systems, we noticed that third-person evaluations offer reasonable solutions. We developed the AR system as a web application and made it available to a large number of anonymous evaluators. Interviews have as yet to be performed but will be attempted soon.

2 SYSTEM ARCHITECTURE AND IMPLEMENTATION

The system is composed of client and server programs written in PHP and C/C++ languages. Dynamic HTML programs reside on the server; the client loads and executes the HTML including Java Scripts if necessary. The maker-tracking program in C/C++ resides also on the server and calculates the camera and model space parameters of a digital image when uploaded to the server.

2.1 Marker Tracking

The maker-tracking program employs ARToolkitPlus. If tracking is successful, the output data is stored in the PostgreSQL database. The server retrieves the data from the database when requested by a client.

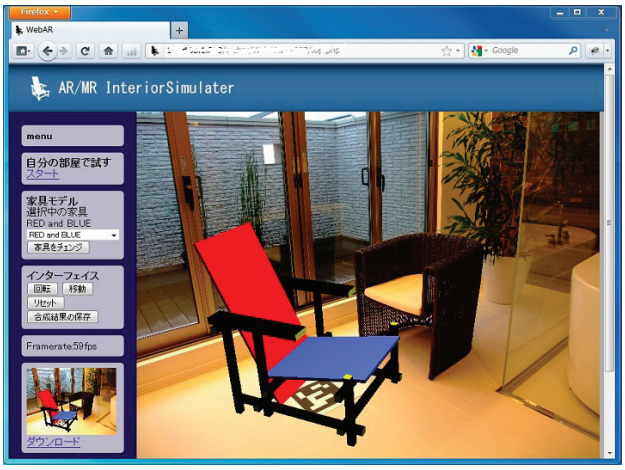


Figure 1. Augmented Reality on a Web Browser

2.2 Superimposing

The client superimposes a 3D model of the item of furniture on the digital image recovered from the server's database. The HTML file includes WebGL descriptions controlled by JavaScript code running in a web browser.

2.3 Furniture Model

We adopted Blender as our 3D modeler to create representations of furniture items because the exporter of an x3d file is provided by default. The parser that we developed in C/C++ reads the x3d file and outputs modeling data of geometry and appearance as a JavaScript code.

3 CONCLUSION

We described the development of an AR system as a web application. In the near future, with improvements made, we will conduct an open experiment involving users of the system as anonymous respondents to a survey. Through this experiment, we will analyze the practicality in evaluating an AR system.

REFERENCES

[1] Hideyuki Tamura, Hirokazu Kato, The TrakMark Working Group: "Proposal of International Voluntary Activities on Establishing Benchmark Test Schemes for AR/MR Geometric Registration and Tracking Methods". Proceedings on International Symposium on Mixed and Augmented Reality (ISMAR2009), October 2009.

[1)] email: TANAKA_Tomoki@graduate.chiba-u.jp
[2)] email: takuma_nakabayashi@chiba-u.jp
[3)] email: keita_kado@graduate.chiba-u.jp
[4)] email: hirasawa@faculty.chiba-u.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Landscape Simulation in Outdoor Settings using Stereoscopic Augmented Reality

Takuma Nakabayashi[1)]
Graduate Student,
Graduate School of Engineering,
Chiba University

Keita Kado[2)]
JSPS Research Fellow, Building
Department National Institute for
Land and Infrastructure Management

Gakuhito Hirasawa[3)]
Associate Professor,
Graduate School of Engineering,
Chiba University

ABSTRACT

This paper reports on an augmented reality (AR) system that handles virtual realizations of buildings and civil engineering structures at real scale in outdoor settings. This system consists of a real-time kinematic (RTK) GPS and a 3DOF inertial measurement unit (Sensor). With additional software, these devices improve the precision in computing camera position and orientation. Moreover, the system uses a 3D head-mounted display (HMD) rendering shadowing of all virtual buildings to achieve a real-world look. The system enables practical AR landscape simulations for architectural design to be made.

KEYWORDS: Augmented Reality, Outdoor Setting, Real Scale, 3DOF inertial measurement unit, Real Time Kinematic - GPS, 3D Head Mounted Display.

1 INTRODUCTION

Although advanced information technologies have been introduced in architectural design and building construction, more effective simulation methods are currently required. Conventional methods, such as photo montage and virtual reality, are inadequate in giving realistic simulations in exterior settings. With AR technology, we can freely move camera position to borrow real-scene backdrops that can be added to a simulation.

Similar systems from previous research focused primarily on portability [1][2]. In contrast, we emphasis performance over portability aiming for a stereoscopic AR that can handle large and complicated 3D shapes.

2 SYSTEM ARCHITECTURE

2.1 Hardware

We use a RTK-GPS and a 3DOF inertial measurement unit to obtain camera position and orientation. In addition, we employ a 3D-HMD and two cameras to develop stereoscopic views for more realistic simulation. We have packed these devices into a compact portable unit suitable for outdoor use.

2.2 Software

First, we developed a VRML parser to ease loading of 3D models into the system. Second, because the devices generate small errors that betray a real-world appearance, we implemented five functions with the following descriptions:

[1)] email: takuma_nakabayashi@chiba-u.jp
[2)] email: keita_kado@graduate.chiba-u.jp
[3)] email: hirasawa@faculty.chiba-u.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Figure 1. The virtual bridge experiment (Right is the virtual bridge)

1) ignores Sensor outputs if angular velocities fall below a threshold value;
2) ignores GPS outputs if accelerations fall below a threshold;
3) corrects Sensor outputs accelerations due to gravity;
4) corrects Sensor outputs by using a series of locations previously acquired from GPS data; and
5) corrects Sensor outputs through template matching from previously-captured landscape images.

Functions 1) and 2) resolve the slight shaking in the rendering of virtual buildings even if a camera is at rest, whereas functions 3), 4), and 5) improve the precision in computing camera orientation.

Ultimately, high quality rendering is required in architectural design simulations. We achieve a realistic rendering by shadowing the virtual buildings. Furthermore, to preserve real-time performance, our program was developed with multi-threading to ensure the functions run concurrently and smoothly.

3 CONCLUSION

In this research, we developed an AR system that handles complicated virtual buildings on a real scale with outdoor backdrops. Implementing corrective functions improved computations of camera position and orientation and casting of shadows enriched landscape simulations of large virtual buildings. Field settings such as the one illustrated in fig. 1 are possible.

REFERENCES

[1] P. Honkamaa *et al*., "Interactive outdoor mobile augmentation using marker-less tracking and GPS", Proc. of Virtual Reality International Conference, pp. 285-288, 2007.

[2] Michael Bang Nielsen *et al*., "Mobile Augmented Reality Support for Architects Based on Feature Tracking Techniques", International Conference on Computational Science, pp.921-928, 2004.

AR Whiteboard: Handling Written Contents as Digital Information Using Tools for Whiteboards

Yuta Tsukada* Keita Ushida† Satoshi Tsurumi‡

Gunma National College of Technology

Abstract

The purpose of this study is to improve operations on whiteboard environments. To do this, we focus attention on digital copy and paste functions on a whiteboard. Using these functions, the contents and drawing data on a whiteboard can be moved, reproduced, and reused easily. The features of our proposed enhanced whiteboard system include a projective AR (Augmented Reality) and a high affinity for the working styles on a whiteboard with magnets, erasers and pointing sticks.

Index Terms: H.5.1 [Information Interfaces and presentation]: Multimedia Infor mation Systems—Artificial, augmented and virtual realities, Evaluation/methodlogy; H.5.2 [Information Interfaces and presentation]: User Interfaces—Screen design, Input devices and strategies

1 Introduction

Recently the use of electronic whiteboards has spread in school and office. However, their user interface is mostly based on computers, so working styles with them are based on computers rather than blackboards.

In related studies[1][2], the functions of computers are enhanced so that they can be applied to working environments with whiteboards. In such cases, users might feel they use computers while using these whiteboards. On the other hand, we have developed an enhanced whiteboard system which has a user interface based on existing operations on whiteboards. The concept of this system is that it enables users to carry out enhanced operations which are founded on the operations on whiteboards, not on computers.

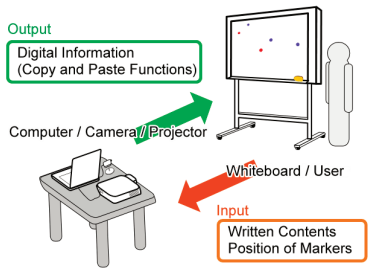


Figure 1: System concept

2 Implemented System

The implemented system is illustrated in Fig. 1. The system is based on projective AR technology: the surface of a whiteboard is captured by a camera, and then digital information on a whiteboard is projected by a projector. Our AR whiteboard system has copy and paste functions for written contents and drawing data on a whiteboard.

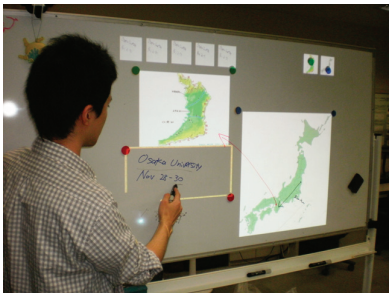


Figure 2: Copy and paste for written contents on a whiteboard

Users can use these functions with magnets. The magnets play a role of markers. Their positions define rectangle regions to be copied and pasted. Other tools like erasers and pointing sticks are used in this system too. They are detected by their colors or shapes.

3 Function

The main function of this system is to copy and paste written contents on a whiteboard. When users put two magnets on a whiteboard, they get an image of a rectangle region formed by two magnets. Then the image is stored on a whiteboard as a thumbnail. Users use three magnets when they paste. To set a paste position, users put two magnets on a whiteboard. By putting them, users define the size and position of a pasted image. Next, users put a magnet on a thumbnail. The magnet serves to select a thumbnail which users want to paste. The selected thumbnail appears on a rectangle region formed by two magnets. In addition to reusing written contents, users can also utilize images which they prepared beforehand. By employing these copy and paste functions, users can save written contents on a whiteboard, and then they can reuse and reproduce the contents on a whiteboard. When thumbnails increase, you may think you want to erase them. This system has a function of data deletion. By erasing thumbnails with an eraser, users can clear the image data of unnecessary thumbnails.

Furthermore, when rooms are large and there are many people, some people may think it is difficult to see what is written on a blackboard or a whiteboard. The pointing sticks have a function of expanding an area which lecturers point at.

References

[1] E. R. Pedersen, K. McCall, T. P. Moran and F. G. Halasz. Tivoli: An Electronic Whiteboard for Informal Workgroup Meetings. In *Proc. of INTERCHI '93*, pp. 391–398, 1993.

[2] K. Kurihara, M. Goto, J. Ogata and T. Igarashi. Speech Pen: New Pen Input Interface Capable of Utilizing Speech Recognition for Digital Writing. *Computer Software*, Vol. 23, No. 4, pp. 60–68, 2006 (In Japanese).

*e-mail: ap10820@ipc.gunma-ct.ac.jp
†e-mail: ushida@ice.gunma-ct.ac.jp
‡e-mail: tsurumi@ice.gunma-ct.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Interactive Cardiovascular Editor Using Echocardiographic Images

*M. Nakao[1], Y. Masuda[2], R. Haraguchi[3], K. Kurosaki[3], K. Kagisaki[3], I. Shiraishi[3], K. Nakazawa[3] and K. Minato[2]

[1]Graduate School of Informatics, Kyoto University
[2]Graduate School of Information Science, Nara Institute of Science and Technology
[3]National Cerebral and Cardiovascular Center

Abstract

We propose a three-dimensional cardiovascular modeling system where medical doctors can interactively construct patient-specific cardiovascular models based on neonatal echocardiographic images, and share the complex topology and the shape information. For the construction of cardiovascular models with a variety of congenital heart diseases, we propose a set of algorithms and interface that enable editing of the topology and shape of the three-dimensional models. The cardiovascular models generated from some patient data confirmed that the developed technique is capable of constructing cardiovascular disease models in a tolerable timeframe.

Keywords: Interactive editing, vascular model, echocardiogram

1 Introduction

In this study, we present a cardiovascular modeling system not only using echocardiogram images but with the simple interaction to the model by the user. Since echocardiogram images do not include complete information of the anatomical structures, conventional segmentation-based approaches have difficulty in representing detailed structures of the vessels. We utilize a deformable template to interpolate anatomical features common to sparsely selected images, and introduce topology-editing interface to easily create a cardiovascular model.

2 Cardiovascular Model and Editing Algorithms

With surface mesh models, mesh subdivisions and collision detection is needed to edit the shape and the topology. Also, the radius of the vessels is not easy to handle. We define a cardiovascular centerline for a skeleton S_i comprising nodes and edges. For the construction of cardiovascular models with a variety of congenital heart diseases, we propose algorithms and interface that enable editing of the topology and shape of the three-dimensional models. In order to facilitate interactivity, the centerline and radius of the vessels are used to edit the surface of the heart vessels. This forms a skeleton where the centerlines of blood vessel serve as the nodes and edges, while the radius of the blood vessel is given as an attribute value to each node. Parent-child relationships are given to each skeleton. They are expressed as the directed acyclic graph (DAG), where the skeletons are viewed as graph nodes and the connecting points are graph edges. (Fig. 1) For the details of our algorithms and interface, see [1].

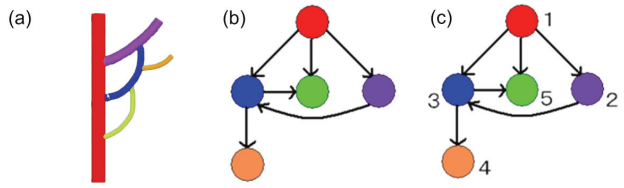


Fig. 1 Topological sort of the skeletons for topology-based editing.

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

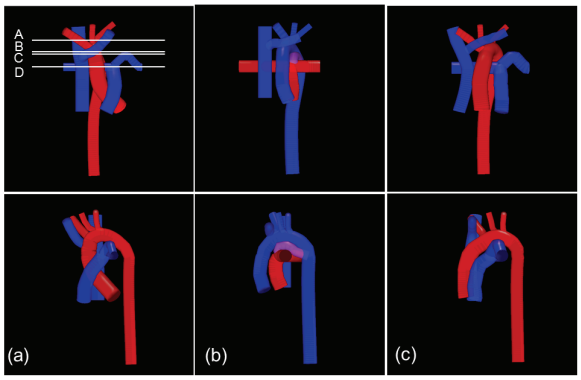


Fig. 2 Cardiovascular modeling results from a template model.

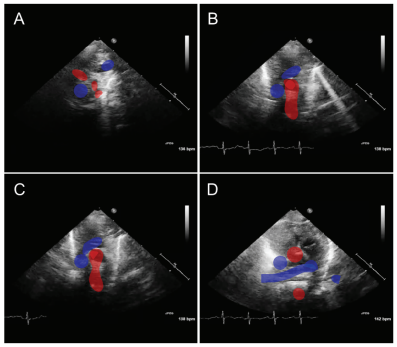


Fig. 3 Editing interface for manipulating cross sections of the model.

3 Results and Conclusion

Fig. 2 shows the interactive editing results from a normal cardiovascular template model. It took approximately 5-7 minutes to complete each model. The aorta and pulmonary artery form a helical shape in the normal heart and the modeling results confirmed that this feature could be expressed. Fig. 3 displays echocardiogram images and cross sections of the cardiovascular model on the four planes A, B, C and D depicted in Fig 2a. We note the cross sections include the branch and small vessel structures that cannot be extracted from the echocardiogram images. The user only indicates the correct 2D position of the vessels based on the information clearly revealed on the reference echocardiogram images. However, our system can generate 3D cardiovascular models that satisfy user's knowledge from the partial or incomplete anatomical information given on the arbitrarily selected 2D planes. Our approach can support representation of the patient-specific heart vessels for better understanding in preoperative planning and tele-diagnosis.

Reference

[1] M. Nakao, K. Maeda, R. Haraguchi *et al*, "Cardiovascular Vessel Modeling of Congenital Heart Disease Based on Neonatal Echocardiographic Images", IEEE Trans. on Information Technology in Biomedicine, 2012. (in press)

Whirling Interfaces: Smartphones & Tablets as Spinnable Affordances

Michael Cohen[*], Rasika Ranaweera, Hayato Ito, & Shun Endo
Spatial Media Group, University of Aizu

Sascha Holesch[†]
Eyes, JAPAN

Julián Villegas[‡]
Language and Speech Laboratory, University of the Basque Country

Abstract

Interfaces featuring smartphones and tablets that use magnetometer-derived orientation sensing can be used to modulate virtual displays. Embedding such devices into a spinnable affordance allows a "spinning plate"-style interface, a novel interaction technique. Either static (pointing) or dynamic (whirled) mode can be used to control multimodal display, including panoramic and turnoramic images, the positions of avatars in virtual environments, and spatial sound.

"Spinning," in which a flatish object is whirled with an extended finger or stick, is a disappearing art. We hope to re-motivate this vanishing skill, modernizing it and opening it up to internet-amplified multimedia. The ubiquity of the modern smartphone makes it an attractive platform for even location-based attractions. We are experimenting with embedding mobile devices into suitable affordances that encourage their spinning. Using azimuthal (yaw) tracking especially allows such devices to control horizontal planar displays such as periphonic spatial sound, as well as avatar heading and (QTVR-style) panoramic and turnoramic imaged-based rendering.

(a) Below (b) Above

Figure 1: Double-headed configuration

We use modern mobile smartphones and tablets (Google Android Samsung Galaxy S and Apple iOS iPhone & iPad), in particular their magnetometers (electronic compasses), to track azimuth, sending such heading to a collocated computer via WiFi. Such sensing affords two modes of operation: static pointing and dynamic whirling [2]. By simply pointing the device in a certain direction, anything can be steered. More innovatively, spinning it yields a whirling controller. An important feature of such design is that mobile devices can display graphically. By compensating for rotation, graphical display can be stabilized. A "double-headed" back-to-back configuration, as shown in Figure 1, allows the display to be seen when the spinning is both below and above eye level. To enable integration with various multimodal displays, including those used for stereographic, panoramic, or turnoramic viewing, we use our own Collaborative Virtual Environment (CVE) to synchronize distributed clients, as seen in Figure 2. Using a software "transmission" to downscale azimuthal displacement allows using even a spinning mode to control other azimuthal displays, such as our "Schaire" (for 'share chair') rotary motion platform [3] or the position of avatars or other objects in virtual environments such as Alice,[1] Open Wonderland,[2] or Second Life.

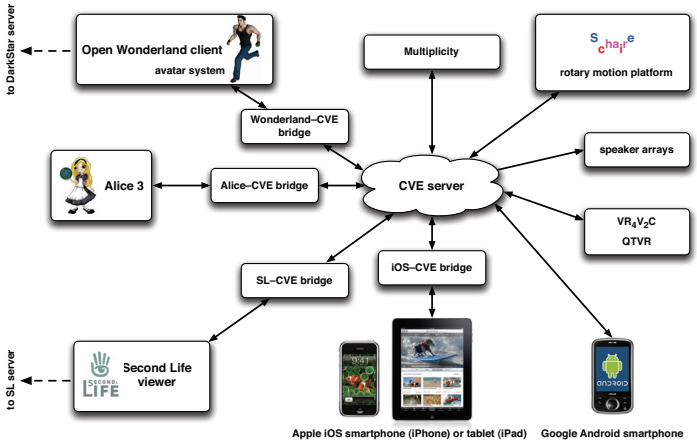


Figure 2: Our CVE provides a shared infrastructure, so that heterogeneous multimodal clients can display data from multiple spinning affordances.

This "exertoy" or "exergame" represents an active interface, a physical interface for whole body interaction. Its groupware capabilities encourage social interaction through physical play [1]. It's a "come as you are" interface, requiring no special markers or clothing. Direct manipulation gives immediate multimodal feedback, in either static (pointing) or dynamic modes (whirling). A video of its operation surveys such applications.[3]

References

[1] T. Bekker, J. Sturm, and E. Barakova. *PUC: Personal and Ubiquitious Computing*, volume 14. Springer, 2010. Design for Social Interaction through Physical Play, ISSN 1617-4909.

[2] M. Cohen. Integration of laptop sudden motion sensor as accelerometric control for virtual environments. In *VRCAI: Proc. ACM SIGGRAPH Int. Conf. on Virtual-Reality Continuum and Its Applications in Industry*, Singapore, Dec. 2008.

[3] N. Koizumi, M. Cohen, and S. Aoki. Japanese patent #3042731: Sound reproduction system, Mar. 2000.

[*]e-mail: {mcohen,d8121104,s1160024,s1160037}@u-aizu.ac.jp
[†]e-mail: sascha@nowhere.co.jp
[‡]e-mail: julovi@yahoo.com

[1] www.alice.org
[2] www.openwonderland.org
[3] sonic.u-aizu.ac.jp/spatial-media/Videos/Twin_Spin.m4v

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Collaboration between Networked Heterogeneous 3D Viewers through a PAC-C3D Modeling of the Shared Virtual Environment

Thierry Duval *
Université de Rennes 1, IRISA UMR CNRS 6074, Rennes, France

Cédric Fleury†
INSA de Rennes, IRISA UMR CNRS 6074, Rennes, France

ABSTRACT

We propose to illustrate how the PAC-C3D software model makes it possible to share networked 3D Virtual Environments (VE) between heterogeneous 3D viewers written in Java3D and jReality.

Keywords: Software Architectural Models for CVE

Index Terms: H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces—Theory and methods; I.3.7 [Computer Graphics]: 3-Dimensional Graphics and Realism—Virtual reality; D.2.11 [Software Engineering]: Software Architectures—Patterns

1 THE PAC-C3D MODEL

We propose to design each object of a Collaborative Virtual Environment (CVE) according to the PAC-C3D model [1] illustrated figure 1. It is an explicit evolution of the PAC model dedicated to 3D CVE. On each user's computer, shared virtual objects of a CVE must be decomposed into three main kinds of components described by three interfaces. The *Abstraction* is in charge of the core data and behavior of the object, the *Presentations* are in charge of the virtual representation of the object to the user, and the *Control* is in charge of the consistency maintenance between *Abstraction* and *Presentations*, and between all the distributed *Controls* of the shared object.

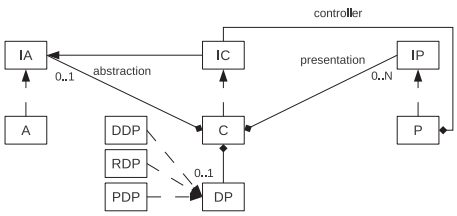


Figure 1: Adaptation of the PAC model for 3D CVE

2 VIEWING THE SAME VE WITH DIFFERENT VIEWERS

PAC makes it possible to design a VE with very small dependency to the 3D graphics API used for the 3D rendering: it proposes to confine all the graphics features of the virtual object in its *Presentation*. PAC-C3D proposes to use explicit interfaces components to strengthen this separation: it makes the *Control* components totally independent of the implementation of the *Presentation* components. The same 2D GUI and external interaction devices allow also to drive in a similar way these 3D viewers, as all the interaction and navigation orders are sent to *Control* and then to *Abstraction* components. We have used this model to design our IIVC [2] and three viewers based on Java3D, jReality and jMonkey.

*e-mail: thierry.duval@irisa.fr
†e-mail: cedric.fleury@irisa.fr

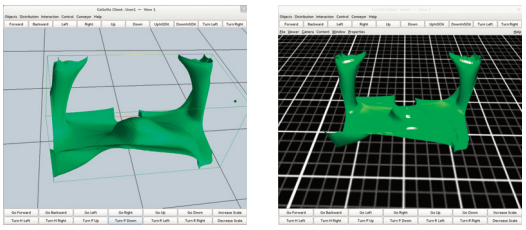


Figure 2: Two different viewers sharing the same virtual environment

3 SHARING THE SAME VE BETWEEN DIFFERENT VIEWERS

PAC-C3D allows these different 3D viewers to share the same 3D VE at run-time (see figure 2) over a network. PAC-C3D proposes to put all the collaborative features (distribution, synchronization, etc.) in the *Control* components, ensuring that each evolution of a virtual object is distributed to the other *Control* components of this virtual object, according to their distribution policy [3].

4 ENRICHING 3D VIEWERS INTER-OPERABILITY

It is possible to enrich the interaction possibilities of a viewer *X* with extra functions provided by an other viewer *Y*. It consists in allowing the viewer *X* to control a virtual object provided by the viewer *Y*. The viewer *X* owns a local *Control* of this object, so each order given to this local *Control* by the viewer *X* will be forwarded to the *Control* of the object on the viewer *Y* that will forward it to its *Abstraction*. To be more efficient, the distribution policy of this virtual object can be changed dynamically at run-time. For example, the *Referent Control* can be migrated to the same process than the viewer *Y*. These exchanges between different viewers strengthen their inter-operability.

5 CONCLUSION

PAC-C3D makes it possible to design a CVE with very small dependency on a 3D graphics API, and it makes it easy to use different 3D graphics API on different remote computers sharing the same collaborative session, providing easy inter-operability between 3D graphics API such as Java3D, jReality and jMonkey.

ACKNOWLEDGEMENTS

This work was partly funded by the French Research National Agency project named Collaviz (ANR-08-COSI-003-01).

REFERENCES

[1] T. Duval and C. Fleury. PAC-C3D: A New Software Architectural Model for Designing 3D Collaborative Virtual Environments. In *Proc. of ICAT*, page to Appear, 2011.

[2] C. Fleury, A. Chauffaut, T. Duval, V. Gouranton, and B. Arnaldi. A Generic Model for Embedding Users' Physical Workspaces into Multi-Scale Collaborative Virtual Environments. In *Proc. of ICAT*, pages 1–8, 2010.

[3] C. Fleury, T. Duval, V. Gouranton, and B. Arnaldi. A New Adaptive Data Distribution Model for Consistency Maintenance in Collaborative Virtual Environments. In *Proc. of JVRC*, pages 29–36, 2010.

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

KINECTing Superheroes in MR Space:
Matching Head-Tracking Coordinates and Gesture-Interaction Coordinates

*Masaki Oda, Le Van Nghia, Katsuyoshi Tomita, Asako Kimura, Fumihisa Shibata, and Hideyuki Tamura
College of Information Science & Engineering, Ritsumeikan University

ABSTRACT

This paper proposes an example of method for integrating a rangefinder-type sensor to a gesture-driven mixed reality (MR) application. The result is an application, "KINECTing Superheroes in MR Space," using the sensor as its space interaction and showed significant improvement of free motion and its limitations of interaction in MR space.
KEYWORDS: Mixed Reality, Interactive System, Gesture

1 INTRODUCTION

Not only visual and auditory mixture but also natural gesture interface is necessary for realizing an immersive MR system (e.g. MR attractions). Nevertheless, it is rare to see such systems because of cost and operative concerns. BLADESHIPS [1] utilizes hand's pose and Hyak-Ki Men[2] utilizes hand motion as its interface but the uses are limited. However circumstances have changed these days because of emergence of cost effective rangefinders such as Kinect sensor and free of charge libraries to manipulate the sensors [3].

This paper proposes an MR application in which a user fights against a menace with super power. The superpower is realized by motions of user's whole body recognized by a rangefinder-type sensor and visual mixture. Possibility of uses of multiple sensors as gesture interfaces in MR applications and method to match their differing coordinate systems is proposed.

2 SETTING MULTIPLE SENSORS

2.1 Selecting Sensors

Followings are criteria for selecting sensors as gesture interface:

Range: The user moves around in space in this system. However, the range is basically limited by length of a cable attached to an HMD. Therefore, the range is limited to about 2 meters.
Occlusion: Iron, mirrors, walls, and lights would affect sensors.
Visibility: It is important for MR applications to place sensors where a user cannot see them not to destroy the scenery.
Burden: Measuring whole body motions requires putting many sensors on the user. However, big sensors and cables attached to them disturb the user's motions. Concerning practical use, getting on and off many sensors on the user are time consuming.

Concerning the criteria, this system adopts a rangefinder-type sensor, Kinect sensor, which can measure bones of the user's body without requiring putting any sensors on the user. The sensor obtains gesture interaction coordinates, but it cannot be used for acquiring head's pose and position (head-tracking coordinates) because of its low accuracy. Then, this system uses optical sensors, Vicon Bonita, for tracking pose and position of the head because the sensor has a wider range.

2.2 Matching Sensors' Coordinate Systems

MR systems using multiple sensors require matching head-tracking and gesture-interaction coordinates systems. Let M_{wp}, M_s, M_p, and s be pose and position of a measuring part in a world coordinates system, those of a sensor in a world coordinates system, those of a measuring part in the sensor's coordinates system (gesture-interaction coordinates), and scale factor, since

$$M_{wp} = s(M_s M_p)$$

M_s can be easily obtained by using a sensor defining world coordinates, in this case, Vicon Bonita.

*email:oda@rm.is.ritsumei.ac.jp

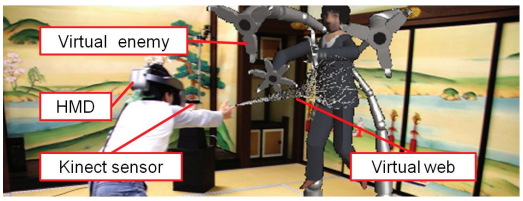


Figure 1. Concept image: Spiderman mode

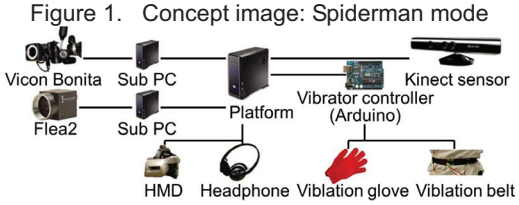


Figure 2. System configuration

3 SYSTEM OVERVIEW

This application has two modes: Spiderman mode and Incredible Hulk mode (See **Figure 1**). Details are shown below.

Spiderman mode: In this mode, the user can shoot a virtual web to attack the enemy. The web comes out from the user's wrist by swinging the arms. The power varies depending on the user's pose.
Incredible Hulk mode: In this mode, the user can cause virtual cracks by trampling down the feet to attack the enemy. The power also varies depending on the user's pose.

System configuration is depicted in **Figure 2**. This system manipulates Kinect sensor through OpenNI library [3] and has a spectator camera. Flow of experience of application is followings: Mode select, tutorial, battle, and game over.

4 CONCLUSION

This paper proposed a gesture-driven MR application, "KINECTing Superheroes in MR Space", method for selecting sensors and matching multiple their coordinates systems, and demonstrated uses of a rangefinder-type sensor in an MR application. The sensor presented a significant improvement of space interaction in MR and showed a limitation; it should be placed in front of a user and that is visually obstacle for MR applications. Therefore, future work will include visual removal of the sensor using diminished reality technologies.

REFERENCES

[1] M. Takemura, *et al.* An interactive attraction in mixed reality: BLADESHIPS. Trans. VRSJ, Vol. 10, No.1, pp. 119 - 128, 2005.
[2] K. Kikuya, *et al.* Hyak-Ki Men: Development of a mixed reality attraction with gestural user interface. Proc. Symp. IPSJ,vol. 2011, No, 3 pp. 469 - 472, 2011.
[3] OpenNI Official Web-page. http://75.98.78.94/default.aspx

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Direct Volume Manipulation for Navigating Liver Resection

*Y. Oda[1], M. Nakao[2], K. Imanishi[3], K. Taura[4], K. Minato[1]

[1]Graduate School of Information Science, Nara Institute of Science and Technology
[2]Graduate School of Informatics, Kyoto University
[3]E-growth.inc
[4]Graduate School of Medicine, Kyoto University

ABSTRACT

We propose a set of algorithms and interface on direct volume manipulation to navigate liver resection surgery. Our navigation system creates the incision surface on a three-dimensional liver model created from patient's CT images, and visualizes a change of the incision surface by deformation of the liver model. Visualization of the incision surface assuming intraoperative deformation can help a surgeon resect precisely in the surgery.

KEYWORDS: Direct volume manipulation, Liver resection surgery, Surgery planning

1 INTRODUCTION

In liver resection surgery, surgeons guess the direction to resect from the image of preoperative planning, operative field and intraoperative ultrasounography and gradually resect the organ.
Existing liver resection surgery simulation system can plan the incision surface against a three-dimensional model created from CT images. However, surgeons deform the liver due to physical constraints in the actual surgery. This makes resection procedure difficult because the incision surface and positional relationship of the blood vessels totally change from the initial state. In this study, we aim to visualize the incision surface deformed by surgical procedure and to support preoperative planning of liver resection surgery.

2 GENERATING AND VISUALIZATION THE INCISION SURFACE

Our navigation system has three steps: (1) Input of the incision surface, (2) Visualization of the incision surface, (3) Deformation of the organ. (Fig.1)

(1) Input of the incision surface: The user inputs a free curve with the mouse, and two parameters: width and depth. Cylindrical objects defined by the free curve, width and depth create open space between incision surfaces.
(2) Visualization of the incision surface: By using different LUT for each of the hepatic parenchyma part and the incision surface part, the incision surface is visualized. LUT (Look Up Table) is a user-defined color map information that transfers CT values to RGBA values.
(3) Deformation of the organ: The tetrahedral mesh created from CT images is deformed, and deformation of the incision surface is visualized using volume rendering based on deformation of the mesh.

*email: yuya-o@is.naist.jp

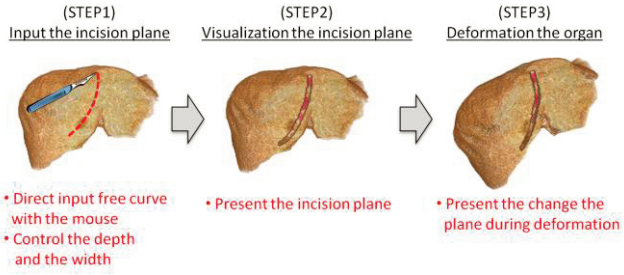


Figure 1. Our navigation system flow

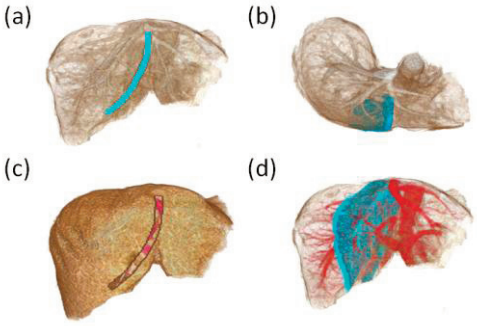


Figure 2. Input results of the incision surface

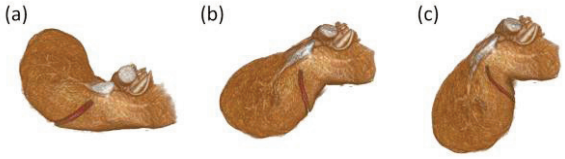


Figure 3. Results of deformation

3 RESULTS AND CONCLUSION

Fig.2 shows results of the incision surface, which is given by the user on a liver object modeled from patient's CT images. (a) and (b) show a planned incision surface visualized by different viewpoints. (c) and (d) are other visualization results of incision surface using different LUT sets. This result makes it easy to understand the relationship between the incision surface and others. Fig.3 shows results of deformation of an organ. This performs with high-quality rendering and user can operate interactively (13-15FPS). The surgeons used the developed system commented that direct deformation of liver with incision surface could contribute to more precise resection in the surgery.

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Experiencing Shape-COG Illusion in Mixed-Reality Space

Hiroki Omosako*, Asako Kimura*, Fumihisa Shibata*, and Hideyuki Tamura*

*Graduate School of Science and Engineering, Ritsumeikan University

ABSTRACT

Mixed reality (MR) is a technology that merges real and virtual worlds in real time. In MR space, visual appearance of a real object can be changed by superimposing a virtual object on it. Because it is well known that the sense of weight can be changed intentionally by providing the appropriate visual stimulation, we believe it has a similar effect in the case of presenting MR visual stimulation. If the behavior and extent of MR visual influence is well investigated, real objects can be differently perceived. In this study, we focus on the center-of-gravity (COG) and verify the influence of MR visual stimulation on the COG in MR environment. In this paper, we conducted experiments to examine the influence of superimposing virtual objects having different COG positions onto real objects. As a result, we confirmed that (1) the presence of COG can be changed by MR visual stimulation; (2) although COG differs in vision and force, the presence of COG can be represented by MR visual stimulation under certain conditions; (3) the influence of MR visual stimulation reduces when positions of COG of a real object and a virtual object are distant; and (4) COG perception can also be changed by varying the mass of the real object in MR space. We named this illusion the "Shape-COG Illusion."

KEYWORDS: Mixed Reality, Center-of-Gravity, Illusion, Psychophysical Influence, Visual Stimulation.

1 INTRODUCTION

This paper describes the influence of visual stimulation on center-of-gravity sense in mixed reality (MR) environment; We conducted experiments to analyze the influence of visual stimulation on COG perception in an MR environment. Specifically, we superimposed virtual objects of different shapes onto real objects having the same mass and volume in order to verify whether COG perception can be changed using MR virtual stimulation. In the experiments, we examined the changing aspect ratios of the virtual object, as well as the mass of the real object.

This study is inspired by the industrial application of MR technology in [1]. This can be summarized as **Fig. 1**. In [2] it is reported that visual stimulation affects tactual sense. Similar to our study, the influence in MR environment is investigated in [3].

2 EXPERIMENTAL ENVIRONMENT

In the following experiments, we adopted an MR system with video see-through mechanism that merges real and virtual worlds visually. Wearing an HMD and touching real objects, the user can see the computer-generated images (CGI) onto the objects with high geometric precision. As the real object used in the experiments, we employed a plastic case with the handle of a real attaché case (**Fig. 2(a)**). As virtual objects used in the experiments, we employed CG models such as the attaché case shown in Fig. 2 (b), which are available in many sizes and shapes.

*1-1-1 Noji-Higashi, Kusatsu 525-8577, Shiga, Japan

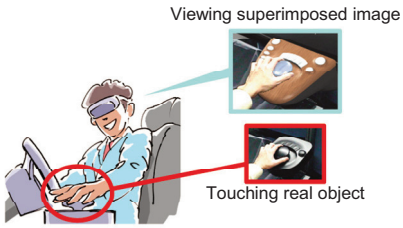


Fig. 1 Mixed reality presentation

(a) In Real space (b) In MR space

Fig. 2 Experimental scene

3 PRELIMINARY EXPERIMENT AND RESULT

In the preliminary experiment, to verify whether COG perception can be changed by superimposing virtual objects on real objects, we superimposed landscape- and portrait-oriented virtual objects onto real objects (Fig. 2(b)). Then, the subjects were asked in which position was the COG farther from their hand. As a result, the COG of real object was perceived in a different place than actual COG position by superimposing virtual objects. Therefore, we considered that COG perception can be influenced by superimposing virtual objects. For the results, we named this illusion the "Shape-COG Illusion."

4 EXPERIMENTS AND RESULTS

In experiment1, virtual objects with different aspect ratios were sequentially superimposed onto the same real object. As a result, the difference in the COG tended to be perceived by changing the aspect ratio of virtual object.

In experiment 2, the same virtual object was superimposed onto the real objects with different mass in order to study the influence of variation in mass. As the result, the perceived COG was changed by changing mass of the real object. Therefore, COG perception can also be changed by varying the mass of the real object in MR space.

REFERENCES

[1] T. Ohshima, *et al.*: "A mixed reality system with visual and tangible interaction capability," Proc. 2nd ISMAR, pp. 284 - 285, 2003.

[2] J. Kim, *et al.*: "Visual touch in virtual environments: An exploratory study of presence, multimodal interfaces, and crossmodal sensory illusions," *Presence*, Vol. 10, pp. 247 - 265, 2001.

[3] M. Nakahara, *et al.*: "Sensory property in fusion of visual/haptic cues by using mixed reality," Proc. World Haptics 2007, pp. 565 - 566.

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Walk-in-Place Locomotion Interface using Footprint Images

Hidetoshi Kiyofuji[1)], Katsuhide Nagasaki[1)], Jun Murayama[2)], Tetsuya Harada[2)]
Tokyo University of Science

KEYWORDS: Walk-in-Place, Locomotion Interface, Footprint images, Image analysis, OpenCV

1 INTRODUCTION

Many locomotion interfaces which have been proposed need sensors attached to the body or expensive special equipments. This paper proposes a simple and low cost locomotion interface using a USB camera. This interface only requires a USB camera, a stepping stage and a PC. This locomotion interface allows users to move in VR space using walk-in-place. Purpose of this demonstration is to experience effect of this system.

2 OVERVIEW

2.1 System configuration

The system configuration of this locomotion interface is shown in Figure 1. This system consists of a PC, a stage for walk-in-place, and a USB camera which is placed at the rear of the stage on the floor. The stage uses an acrylic board on it to project a shadow on the back of the board.

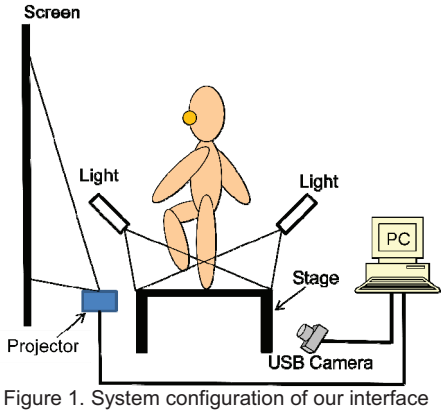


Figure 1. System configuration of our interface

2.2 Footprint images processing

Footprint images are obtained by the USB camera. Image processing below uses OpenCV. Footprint raw images are given projection transformation and background differencing. The image noises are removed by opening-filter. The footprint raw images are binarized as Figure 2. In this system the locomotion direction and the velocity are obtained by the images changing with time.

The determination of direction is shown in Figure 2. The direction of foot print (right or left) is decided by comparing the number of white pixels in every corner of a circumscribed oblong surrounded with lines parallel to x-axis and y-axis. If the number of white pixels at the upper right and lower left corner is larger than that at the upper left and lower right, the rotation direction is assumed right. If the number is reverse, the direction is assumed left. Then this system approximates the rotation angle (θ) as shown in equation (1),

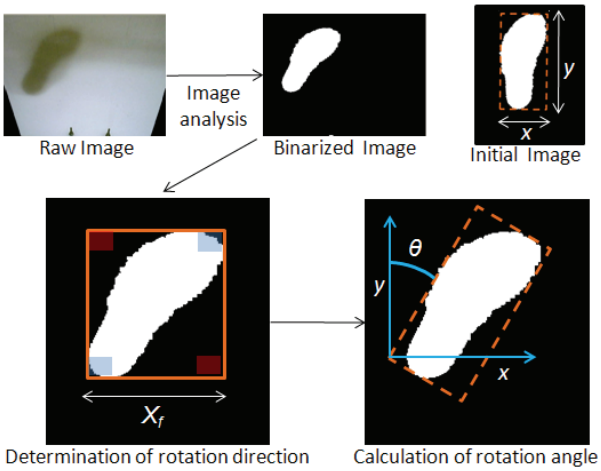


Figure 2. Determination of direction process

$$\theta \approx \sin^{-1}\left(\frac{X_f}{\sqrt{x^2+y^2}}\right) - \tan^{-1}\left(\frac{x}{y}\right) \quad (1)$$

where X_f shows width of the circumscribed oblong, and x and y show width and height of foot initially obtained respectively . If a user steps with rotating one's foot, the projected image on the screen moves with rotating by calculated angle.

The relation between footprint patterns and locomotion velocity is shown in Figure 3. If one footprint image is detected, the locomotion velocity and the direction are applied. To move continuously if two footprint images are detected, the locomotion velocity is decayed with time not to be 0 soon.

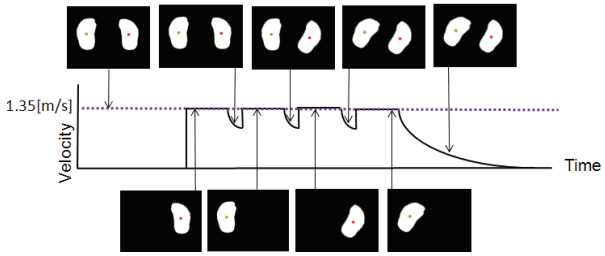


Figure 3. The relation between footprint patterns and locomotion velocity

1) e-mail:{j8110622, j8111633}@ed.noda.tus.ac.jp
2) e-mail:{murayama, harada}@te.noda.tus.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Skill Transmission by Using Parasitic Humanoid System

* Yuki HASHIMOTO, Daisuke KONDO, Tomoko YONEMURA, Hiroyuki IIZUKA, Hideyuki ANDO and Taro MAEDA

Osaka University / JST CREST

ABSTRACT

We have proposed a Parasitic Humanoid (PH) system where two distant people can share their views, audio and tactile sensation by mixing or exchanging them by our devices. The users can share what the other user is seeing, hearing, and touching. Our goal is to transmit non-verbal skills from skilled person to the non-skilled person by using PH system. To realize this goal, we improved this system. The system has the feature of a light weight, a wide viewing angle, stand-alone and a calibration of intraocular distance is unnecessary. These features can make more efficient of skill training and expand sphere of activity.

KEYWORDS: view sharing, skill training, parasitic humanoid, wearable system

1 PARASITIC HUMANOID SYSTEM

By extending the robot-human telexistence[1] technology to human-human situation, we are developing an environment where a skilled person, who actually exists at a different place, can work with high quality on the ground instead of non-skilled person. The skilled person feels as if he exists at the place and work there. The non-skilled person can show high-quality performances with the skilled person's help. In order to realize such a telexistence environment in human interactions, we are developing remote communication technologies exploiting sense-motion sharing is. In this project, we have developed the system to share the first person perspectives, environmental sound and touch feeling between remote two people[2]. Based on previous works, we developed a new PH system for improvement of effect and expansion of application (Figure 1).

The new PH system has three advantages. One is light weight. Weight of the system is drastically lighter than established one (established one = about 32kg, new one=about1.5kg). The second is stand-alone system. In this system, the user can wear all components. In addition, there is a battery to operate without external power supply. Therefore, it is possible to extend user's behavior range. The third is wide field of vision. We developed a new video see-through HMD (VST-HMD) that the viewpoint of eyes and display continues suiting without adjustment because of sticking displays on each eye. This design effects a wide field of binocular vision at horizontal axis; about 50° (horizontal) x 24° (vertical).

2 DEMONSTRATION

We have tried to use our system for skill transmission about the following for several tasks [3][4] in real time. And we are also trying to skill transmission by represent a recorded data of expert. For example, a user learns about Cardiopulmonary resuscitation (CPR) by reliving of expert's movement in first person stand point. The demo allows some remote cooperating works and reliving of behavior of an expert by using our system.

*email:{y.hashimoto, kondo, yonemura, iizuka, hide, t_maeda}@ist.osaka-u.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

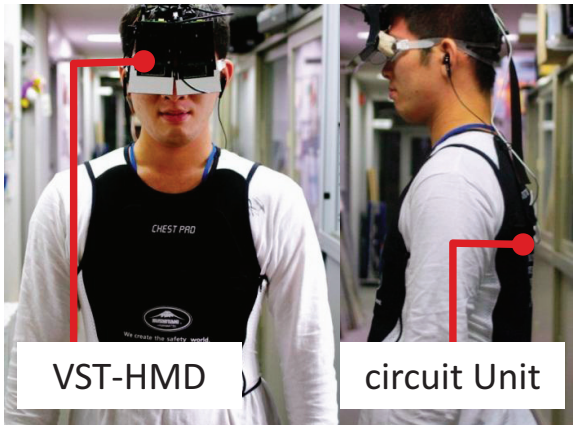


Figure 1. Implementation of parasitic humanoid system

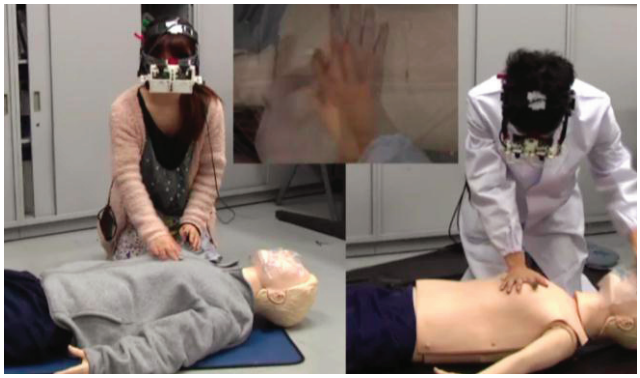


Figure 2. Cardiopulmonary resuscitation

3 ACKNOWLEDGMENTS

This research was supported by JST, CREST.

REFERENCES

[1] Susumu Tachi: Tele-existence - Toward Virtual Existence in Real and/or Virtual Worlds, Proceedings of the International Conference on Artificial Reality and Tele-existence (ICAT '91), pp.85-94, 1991.

[2] T. Maeda, H. Ando, H. Iizuka, T. Yonemura, D. Kondo and M. Niwa : Parasitic Humanoid: The Wearable Robotics as a Behavioral Assist Interface like Oneness between Horse and Rider, 3rd Augmented Human International Conference, 2011.

[3] K. Kurosaki, H. Kawasaki, D. Kondo, H. Iizuka, H. Ando and T. Maeda : Skill Transmission for Hand Positioning Task through View-sharing System, 3rd Augmented Human International Conference, 2011.

[4] H. Kawasaki, H. Iizuka, S. Okamoto, H. Ando, T. Maeda : Collaboration and Skill Transmission by First-person Perspective View Sharing System, 19th IEEE International Symposium in Robot and Human Interactive Communication, 2010.

Interaction for remote collaboration with tabletop system

Keiji Uemura* Nobuchika Sakata† Shogo Nishida‡
Graduate School of Engineering Science, Osaka University

Abstract

This demonstration shows the tabletop and Projector-Camera(ProCam) system in a remote collaboration. In that case, we propose the method improving the usability with the proposal method. This paper describes the implemented tabletop system and the proposal method.

Index Terms: H.5.1 [Information Interfaces and presentation]: User Interfaces—Screen design, Input devices and strategies;

1 Introduction

Works conducted by a local worker under instructions of a remote instructor is called the remote collaboration. With using telecommunication terminal, the remote instructor and the local worker transmit and receive sounds and videos to accomplish their work since they cannot share voices and views directly. On the other hand, a worker and an instructor sometimes communicate regarding objects and places in real work spaces in local collaborative works. Especially, this study focuses on an interaction to make a communication comfatable in a collaborative work. The goal of our research is to achieve an interaction which allows the worker to realize the situation of the work field, and allows the instructor to instruct the field worker accurately.

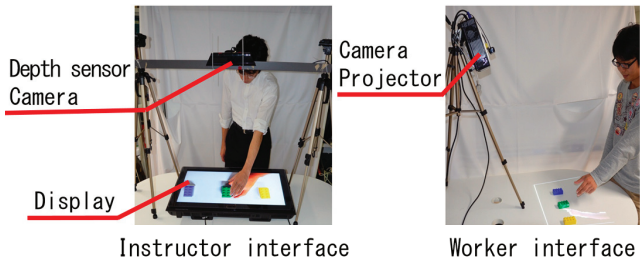


Figure 1: System appearance

2 Configuration

Our implemented system has two interfaces(Figure 1). One is the ProCam system at the work field consisting of a projector and a camera. The other is the tabletop system at the instructor field consisting of a display, a camera and a depth sensor. The process of this system goes as follows. The image captured by the ProCam system on the work field is displayed on the tabletop display at the remote location. Next, the instructor's arm is extracted from the image of the instruction field by the depth sensor. Then, overlapping the image to the work field allows communicating with keeping information of the embodiment which consists of the moving arms and the pointing. Some research have realized the interaction of a tabletop system for remote collaboration [1] [2]. However, these research only consider the person to person situation. In the case of multi-instructor, for example, a worker's watches several arms instructing by pointing or gesture. However, several arms moves at same time, then it is difficult for the worker to realize the instruction. To solve these problem, we propose the "scaling" method. Figure 2 and Figure 3 show this method.

*e-mail: uemura@nishilab.sys.es.osaka-u.ac.jp
†e-mail:sakata@nishilab.sys.es.osaka-u.ac.jp
‡e-mail:nishida@nishilab.sys.es.osaka-u.ac.jp

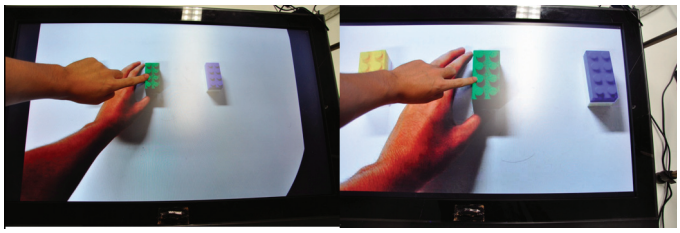


Figure 2: Tabletop display in instructor interface (left:original view, right:magnified view)

The left of Figure 2 shows the display without the method on the instructor interface. As shown in the left of Figure 2, it is difficult for the instructor to point precisely. Therefore, the proposal method magnifies the image(right of Figure 2).

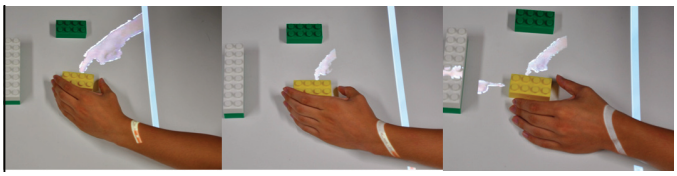


Figure 3: Projected instructor's arm in worker interface (left:original view, center:diminished view, right:multi-instructor)

The left image of Figure 3 shows the projected instructor's arm and worker's arm without the method. In this case, it is difficult for the instructor to point accurately because of the arm's size. The proposal method diminishes the image (center of Figure 3). Therefore, the method allows the worker to realize the instruction. Even in the case of multi-instructor, the projected image is easy to realize (right of Figure 3).

3 Demonstration

The demonstration shows the interaction between the worker interface and the instructor interface. The system allow to communicate each other easily by our proposal method, "scaling" method.

References

[1] S. Izadi, A. Agarwal, A. Criminisi, J. Winn, A. Blake, and A. Fitzgibbon. C-slate: A multi-touch and object recognition system for remote collaboration using horizontal surfaces. In *In IEEE Workshop on Horizontal Interactive Human Computer Systems*, pages 3–10, 2007.

[2] D. Kirk, A. Crabtree, and T. Rodden. Ways of the hands. In *In Proc. 9th European Conference on Computer-Supported Cooperative Work*, pages 1–21, 2005.

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

An Indoor Navigation System using a Wide-view Head Mounted Projective Display with a Semi-transparent Retro-reflective Screen

Duc Nguyen Van[1] Tomohiro Mashita[1,2] Kiyoshi Kiyokawa[1,2] and Haruo Takemura[1,2]

[1] Graduate School of Information Science and Technology, Osaka University
[2] Cybermedia Center, Osaka University

ABSTRACT

We demonstrate a simple indoor navigation system using a wearable Hyperboloidal Head Mounted Projective Display (HHMPD). We have been developing a HHMPD which has unique characteristics such as a wide field-of-view (FOV), a large observational pupil, and optical see-through capability. However, a conventional HHMPD requires a stationary retro-reflective screen in the environment thus is unable to be used in a wearable environment. The wearable HHMPD has been prototyped using a newly developed semi-transparent retro-reflective screen. A user of the demo system is able to observe indoor annotations though the wearable HHMPD.

KEYWORDS: Wide field-of-view, Head mounted projective display, Wearable augmented reality

1 INTRODUCTION

Our research goal is to realize a wearable computing system with a more intuitive and flexible information display by employing a wide field-of-view (FOV) video display. An optical see-through head mounted display (HMD) is commonly used in a wearable augmented reality (AR) system to enjoy a variety of IT services. We have developed a wide FOV optical see-through HMD suitable for wearable AR based on a Hyperboloidal Head Mounted Projective Display [1]. The wearable HHMPD is composed of a pair of custom-made mirrors (see Figure 1) and two pocket projectors (3M MPro110, VGA, 17.7 by 14.4 degrees), and a pupil-division semi-transparent retro-reflective screen [2]. It provides a 109.5-degree horizontal view angle and a 66.6-degree vertical view angle. Note that the HHMPD's optical design is theoretically capable of providing a horizontal field of view wider than 180 degrees [1], if appropriate mirror parameters and wider horizontal projection angles (~50 degrees) are given.

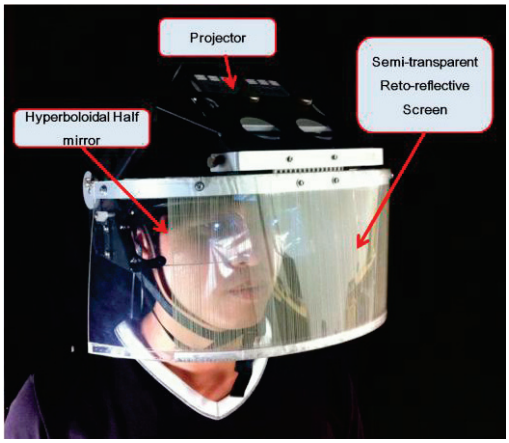


Figure 1. Wearable HHMPD

2 INDOOR NAVIGATION SYSTEM

In this demonstration, we present a simple indoor navigation system using a wearable HHMPD. An optical tracking system OptiTrack is used to track user motion. Thanks to the retro-reflective projection technology and a semi-transparent retro-reflective screen, a user is able to observe annotation information as well as the real objects they refer to at the same time comfortably (see Figure 2).

Figure 2. Implemented AR application using Wearable HHMPD

ACKNOWLEDGEMENT

This research was funded in part by Grant-in-Aid for Scientific Research (B), #22300043 from Japan Society for the Promotion of Science (JSPS), Japan.

REFERENCES

[1] Kiyokawa, K., "A Wide Field-of-view Head Mounted Projective Display using Hyperbolic Half-silvered Mirrors," Proc. IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR), pp. 207-210, 2007.

[2] Nguyen van, D., Mashita, T., Kiyokawa, K. and Takemura, H., "Design Consideration for Semi-transparent Retro-reflective Screen for a Wide-view Head Mounted Projective Display using a Hyperboloidal Half-mirror," IEICE Technical Report, MVE2010-119, 2011. (in Japanese)

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Owens Luis – A Proposal of a Smart Office Chair in an Ambient Environment

Kiyoshi Kiyokawa[1,2] Masahide Hatanaka[2] Kazufumi Hosoda[2] Masashi Okada[2] Hironori Shigeta[2]
Yasunori Ishihara[2] Fukuhito Ooshita[2] Hirotsugu Kakugawa[2] Satoshi Kurihara[2] and Koichi Moriyama[2]

[1] Cybermedia Center, Osaka University
[2] Graduate School of Information Science and Technology, Osaka University

ABSTRACT

This demonstration introduces a smart office chair, Owens Luis, whose pronunciation has a meaning of "an encouraging chair" in Japanese. For most of the people, office environments are the place where they spend the longest time while awake. To improve the quality of life (QoL) in the office, Owens Luis monitors an office worker's mental and physiological states such as sleepiness and concentration, and controls the working environment by multi-modal displays including a motion chair, a variable color temperature LED light and a hypersonic directional speaker.

KEYWORDS: Multi-modal Displays, Ambient Intelligence, Context-aware Systems

1 INTRODUCTION

In an ambient environment, where in our definition ambient intelligence functions and watches people within, a variety of appropriate services are provided depending on the user and environmental contexts in a timely fashion. As an ambient environment, we study on an ambient office, where a variety of sensors are embedded to recognize individual workers' status and to improve the quality of life (QoL) in the office by controlling lighting, air conditions, BGM etc.

More specifically, we aim to develop a rest control system in an ambient office. In Japan, 60 to 80 percent of people are said to feel stressed physically and / or mentally at office. It is often the case an office worker continues to work even when it is actually recommended to take a rest from a point of view of health or work efficiency. In an ambient office, a worker will be suggested to take a rest when necessary, and on the other hand, be encouraged to continue working when appropriate. As a first step, we prototyped a smart rest control system in a form of an office chair, named Owens Luis.

2 OWENS LUIS: SMART OFFICE CHAIR

Figure 1 shows a configuration of Owens Luis, our smart office chair. Owens Luis is composed of a motion chair (Panasonic, EU-JA50), a directional speaker (HSS Japan, H450), a variable color temperature LED light (Color Kinetics Japan, iW Blast Powercore), a high-speed camera (PointGrey Flea3, VGA, 120Hz) and two PCs. Owens Luis estimates sleepiness and concentration of a worker simply from blinking speed and body motion (Figure 2). Then, based on these parameters, Owens Luis's action is determined using an attracter selection model. Attracter selection is a biologically inspired approach found in E. coli cells to self-adaptively react to changes of a nutrient in the environment. In our scenario, Owens Luis is configured to shake the worker up by horse-riding motion when he/she is sleepy, and to slightly and randomly change the inclination of the seat when the concentration level is low. Owens Luis also changes the lighting and BGM settings to cheer up or relax the worker according to the worker's status.

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

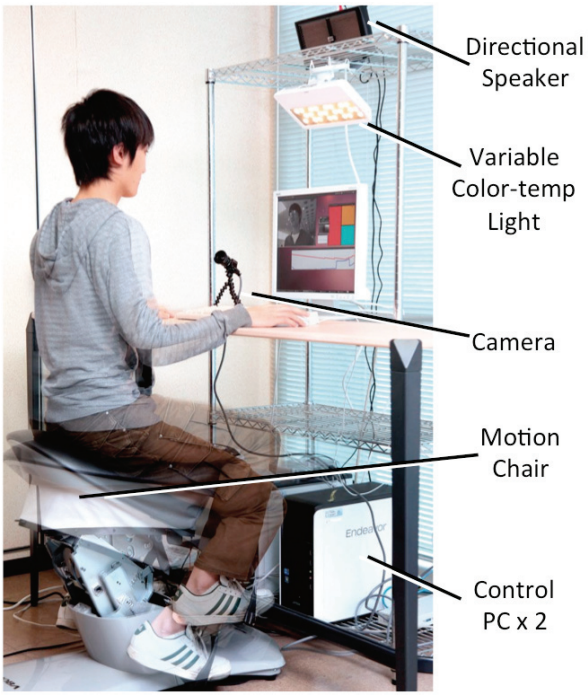


Figure 1. Configuration of Owens Luis, our smart office chair (Courtesy of the Mainichi Newspapers Co., Ltd.)

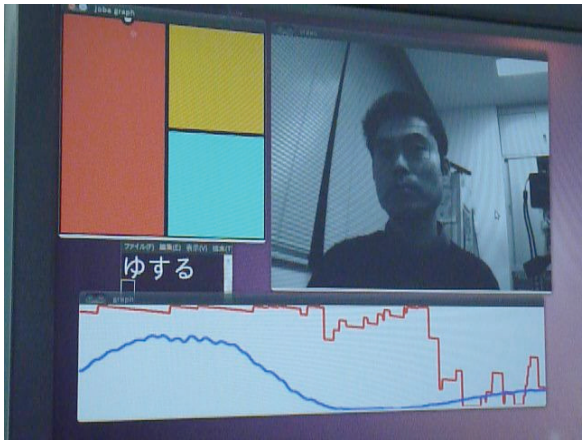


Figure 2. Sleepiness and concentration estimation from camera.

ACKNOWLEDGEMENT

This research was funded in part by "The Global Center of Excellence (GCOE) Program and Grant-in-Aid for Scientific Research on Priority Areas (18049050)" of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

High Dynamic Range 3D Display System with Projector and 3D Color Printer Output

Saeko Shimazu* Osaka University
Daisuke Iwai† Osaka University
Kosuke Sato‡ Osaka University

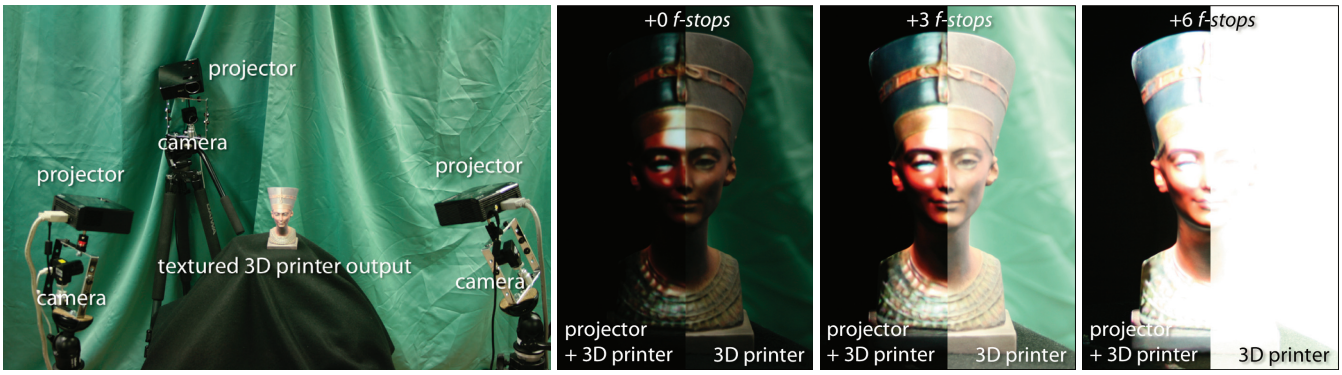


Figure 1: Superimposition of multiple projected images onto a textured 3D printer output leads to multiplicative luminance and chrominance modulation. This consequently leads to contrast enhancement. The leftmost image shows the experimental setup, and the remaining three images show the appearance of the object by the proposed method (left) and 3D printer output (when viewed under environment light) alone (right) captured under different exposures.

Abstract

This demonstration describes a new high dynamic range (HDR) display system that generates a physical 3D HDR image without using stereoscopic methods. To boost contrast beyond that obtained using either a hardcopy or a projector, we employ a multiprojection system to superimpose images onto a textured solid hardcopy that is output by a 3D printer or a rapid prototyping machine. The physical contrast ratio obtained using our method was approximately 5,000:1, while it was 5:1 in the case of viewing the 3D printout under environmental light and 1,000:1 in the case of using the projectors to project the image on regular screens.

1 Introduction

High dynamic range (HDR) display technologies allow the display of images on 2D surfaces with luminance ranging several orders of magnitude. Most of these displays are based on the principle of double light modulation using transmissive spatial light modulators, such as LCD panels [2]. On the other hand, a new approach based on reflective image modulation has been recently introduced for viewing static HDR content [1]. Images are projected onto hardcopies, such as photographs, to boost contrast beyond that obtained using hardcopies or projectors alone.

This demonstration presents a novel HDR display system that allows the generation of a physical 3D HDR image without the use of any stereoscopic methods (Fig.1). Transmissive methods cannot be used for this purpose. Because transmissive modulators are limited to planar surfaces, they display virtual 3D HDR content using a stereoscopic approach. In contrast, reflective approaches generate images by using 3D textured physical objects as hardcopies. A physical 3D HDR display is indispensable in specific applications, such as in industrial design for the assessment of a product prototype or in archeology for the realistic physical reproduction of digitally archived historic objects. In these fields, an enhanced material perception (e.g., specularity) of the displayed 3D information is required. We employ a multiprojection system to superimpose images onto a textured solid hardcopy that is output by a 3D printer or a rapid prototyping machine.

*e-mail: shimazu@sens.sys.es.osaka-u.ac.jp
†e-mail: daisuke.iwai@sys.es.osaka-u.ac.jp
‡e-mail: sato@sys.es.osaka-u.ac.jp

2 Technical Approach

We introduce two basic techniques for our 3D HDR display. The first technique computes an optimal placement of projectors so that projected images cover the hardcopy's entire surface while maximizing image quality. For each combination of projector positions, we evaluate the images projected on a hardcopy surface in terms of both the reflected image quality and the degree of coverage on the basis of the geometric and photometric models of the projectors. The second technique allows a user to place the projectors near the computed optimal position by projecting from each projector images that act as visual guides. Each projector projects an image, which is generated by assuming that the projector is placed at the optimal position. The user adjusts the projector such that the projected image is registered on the hardcopy. Then, we measure the shape of the hardcopy by projecting structured light pattern, in particular gray code, from each projector. The iterative closest point (ICP) algorithm is applied to find out the actual relative position and orientation of each projector to the hardcopy. The calculated geometric information is used to generate a new projected image, which is registered on the hardcopy.

References

[1] O. Bimber and D. Iwai. Superimposing dynamic range. *ACM Transactions on Graphics*, 27(5):150:1–150:8, 2008.

[2] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs. High dynamic range display systems. In *Proceedings of SIGGRAPH*, pages 760–768, 2004.

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Optically Hiding of Information with Polarized Complementary Projection

Mariko Miki *
Osaka University

Daisuke Iwai †
Osaka University

Kosuke Sato ‡
Osaka University

Abstract

We propose the concept and implementation of a graphical information hiding technique for interactive tabletops where users can view the information by simply casting real shadows. We placed three projectors (one in the rear and two in the front) in such a way that the rear one projects graphical information onto a tabletop surface, and the front ones project a complementary image, so that the combined image displayed on the surface becomes uniformly gray, thus hiding the information from the viewer. Users can view the hidden information by blocking the light from the front projector, revealing the complementary image that is being projected onto the occluder. We use the other front projector and polarization filters to make the complementary image projected onto the occluder also uniformly gray. Because the technique completely relies on optical phenomena, users can interact with the system without suffering from any false recognitions or delays.

Keywords: Multi-projection, complementary color, shadow.

1 Introduction

When interactive tabletops are installed in our daily environment, it is necessary to take into account the following two issues: (1) tabletops sometimes react to users' unintentional actions or physical objects that are placed on the tabletops, and (2) sensing and recognition processes inherently cause the delays in the systems' reactions some of which are perceived by users. As soon as these occur, users' natural interactions with tabletops are significantly disturbed. We propose an interaction technique in which users interact with tabletop systems by casting shadows on the surfaces (Fig. 1). The basic idea of hiding information with two projectors and revealing it by casting shadows was proposed by Minomo et al. [1]. However, they did not consider that complementary images would be revealed on anything occluding the projection of images. On the other hand, our technique solves the problem of visual disturbance from objects occluding images.

2 Optical design

Figure 2 shows the proposed system's configuration. A beam splitter is used to ensure that the two front projectors share the same perspective. A polarization filter f_p is placed in front of one of the front projectors, while another polarization filter $f_p^{\perp}$, the polarization direction of which is perpendicular to f_p, is placed in front of the other front projector. We place another filter f_t on the tabletop surface so that the polarization direction is the same as that of f_p. The rear projector p_r projects graphical information i_r on the surface. One of the front projectors p_f (with f_p) projects a complementary image i_f on the surface. The other front projector $p_f^{\perp}$ (with $f_p^{\perp}$) projects an image $i_f^{\perp}$ which is complementary to i_f. Consequently, the projected images i_r and i_f are overlaid on the surface

*e-mail: mariko@sens.sys.es.osaka-u.ac.jp
†e-mail: daisuke.iwai@sys.es.osaka-u.ac.jp
‡e-mail: sato@sys.es.osaka-u.ac.jp

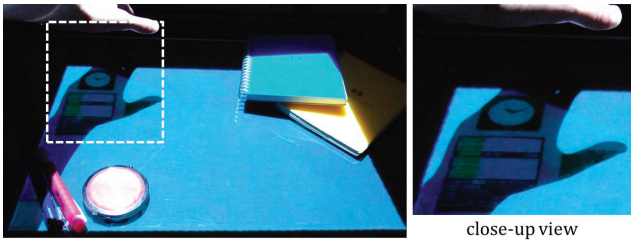


Figure 1: Graphical information is revealed in a real shadow area.

while $i_f^{\perp}$ does not reach it. The combination of images results in a uniform gray image being displayed on the surface. When an object blocks the front projection images i_f and $i_f^{\perp}$, a uniform gray image appears on the object's surface by the overlay of i_f and $i_f^{\perp}$. At the same time, the information is revealed in the shadow area where only i_r is displayed. To compute the projection image for p_f to display i_f, i_r from p_r, and $p_f^{\perp}$ to display $i_f^{\perp}$, we used a camera and a color mixing matrix to estimate the appearance of a projected image on the surface. [2].

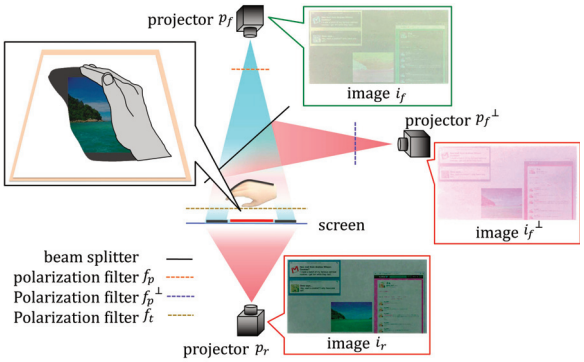


Figure 2: Optical configuration.

3 Result and conclusion

We built a proof-of-concept system consisting of three projectors. And we confirmed that the proposed technique could hide any graphical information projected from the rear projector and the hidden information would be revealed only in shadow areas.

References

[1] Y. Minomo et al.: Transforming your shadow into colorful visual media: multiprojection of complementary colors, ACM CIE, vol. 4, no. 3, article no. 10, 2006.

[2] T. Yoshida et al.: A virtual color reconstruction system for real heritage with light projection, In Proc. of VSMM, pp. 161-168, 2003.

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Interacive Cutting Simulation System of Physics-based Electrosurgery

Yoshihiro Kuroda* Osaka University
Shota Tanaka† Osaka University
Ryosuke Yokohata‡ Osaka University
Masataka Imura§ Osaka University
Osamu Oshiro¶ Osaka University

Abstract

Foregoing studies have never proposed the physics-based real-time electrosurgery simulation, because of the complexity of the phenomena and computational costs. The aim of this study is to construct an interactive surgical simulator with a unified physics-based modeling of electrosurgery. In this paper, we proposed an interactive simulation system of physics-based electrosurgical cutting. Especially, pre-processing independent of contact information reduced real-time calculation time of electrical potential. The proposed system allowed a user to cut a 3D mesh with a virtual electrosurgical tool interactively.

Keywords: virtual reality, medical information systems, finite element methods.

Index Terms: I.6.3 [Simulation and Modeling]: Applications; C.3 [Special-purpose and Application-based Systems]: Real-time and embedded systems—

1 Introduction

Virtual reality-based electrosurgical simulators are demanded for training of the major skills in laparoscopic surgery to reduce complications. However, foregoing electrosurgery simulators ignored underlying physical phenomena, due to their complexity and required update rates for graphics and haptics. Thus, the conventional simulators removed the material when the temperature reached 100 °C [1], although stress concentration is closely-linked to tissue destruction. The temperature-based cutting cannot take into account the effect of the tension given to the tissue by grasping or pulling.

The aim of this study is to construct an interactive surgical simulator with a unified physics-based modeling of electrosurgery. In this study, we simulate interactive physics-based electrosurgery cutting.

2 Interactive electrosurgical cutting simulation

An electrosurgical unit is a device for cutting soft tissue by Joule heat caused by the current of high frequency. The electrical phase determines the current density distribution caused by the contact between the electrosurgical tool and the object. The thermal phase determines the temperature distribution caused by Joule heating and temperature conduction. The governing equations of electric conduction and heat transfer are given by Laplace and heat equations.

$$\nabla^2 V = 0 \tag{1}$$

$$c\rho \frac{\partial T}{\partial t} = \lambda \nabla^2 T + \mathbf{J} \cdot \mathbf{E} \tag{2}$$

where V is voltage, T is absolute temperature, $\mathbf{J}$ is current density, $\mathbf{E}$ is electric field, and c, ρ, λ are specific heat, density, and thermal conductivity, respectively. The structural phase determines the structural change caused by vaporization and stress concentration. The stress is derived from the expansion of the volume caused by water vaporization.

*e-mail: ykuroda@bpe.es.osaka-u.ac.jp
†e-mail: s-tanaka@bpe.es.osaka-u.ac.jp
‡e-mail: yoko@bpe.es.osaka-u.ac.jp
§e-mail: imura@bpe.es.osaka-u.ac.jp
¶e-mail: oshiro@bpe.es.osaka-u.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

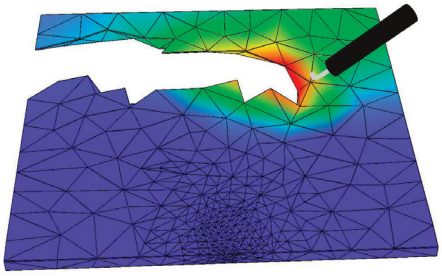


Figure 1: Interactive cutting simulation

Our approach is to reduce the computational cost in real-time processing by pre-processing independent of prior contact information. The proposed system inverts the whole coefficient matrix for linear simultaneous equation of electric potential, before the contact nodes by an electrosurgical tool are given. The conventional method requires the contact information in the high-cost matrix operation, such as the matrix inversion and factorization, while the proposed method does not require the contact information in the high-cost matrix operation. Thus, the proposed method excludes the high-cost matrix operation from real-time processing to pre-processing.

3 System

The simulation system was equipped with Intel CPU (Core i7 3.07GHz), 12GB main memory, a PHANToM Omni haptic device, and an nVidia GeForce GTX 580 graphics board. The object used in the simulation had 1013 nodes (tetrahedral mesh). The object size was 100 mm × 100 mm × 10 mm. Fig. 1 shows the temperature distribution and the cut region by a user's manipulation. The electrical potential was calculated with consideration of interactive manipulation of the electrosurgical tool. The temperature of the nodes was increased by the Joule's heat, while the temperature was diffused in the object as time progressed. The stress increased after the vaporization of water in the elements. The elements whose stress was over the criteria were removed.

4 Conclusion

The simulation results showed that the proposed system enabled an interactive simulation with consideration of a series of physical phases: electrical, thermal, and structural phases.

Acknowledgements

This study was partly supported by the Global COE Program "*in silico* medicine" at Osaka University.

References

[1] A. Maciel and S. De. Physics-based real time laparoscopic electrosurgerysimulation. In *Medicine Meets Virtual Reality 16*, pages 272–274, 2008.

STELET Display: Tactile Augmentation with Handheld Tool

Shunsuke Yoshimoto* Osaka University
Naritoshi Matsuzaki Osaka University
Yoshihiro Kuroda Osaka University
Masataka Imura Osaka University
Osamu Oshiro Osaka University

Abstract

This research introduced a new concept of a tactile display, "Spatial Transparency", which enables tactile augmentation without mediating a device. We developed Spatially Transparent Electrotactile (STELET) display based on tactile illusion caused by electrical stimulus and the anatomical nerve structure. A simple tactile AR system in which the virtual grip sensation is augmented to the real sensation was provided by using the developed device. In the demonstration, users can interact with the virtual object which is superimposed to the real material by using an instrument.

Keywords: Haptics, Tactile augmentation, Electrotactile display.

1 Concept

This research introduces a new concept of a tactile display, "Spatial Transparency", which is defined as tactile augmentation without mediating a device. A spatially transparent tactile display enables users to feel a real environment augmented with synthetic tactile stimuli. Because grip sensation is related to the physical event caused by handheld tool, to augment grip sensation by using a spatially transparent tactile display has potential to modulate the perception such a texture and stiffness as well as to display a virtual object. This research aims at developing a spatially transparent tactile display and investigating modulation of the perception during a tool manipulation with a virtual object superimposed to the real material. The grip force sensation is augmented to present real environment augmented with synthetic tactile stimuli. Previous tactile displays cannot augment tactile sensation without disturbing real sensation because the tactually augmented position corresponds to the stimulated position. We focused on the tactile illusion caused by electrical stimulus to separate the perceived position from the stimulated position and developed Spatially Transparent Electrotactile (STELET) display [1]. We provide a simple tactile Augmented Reality (AR) system which enables users to feel augmented object by using the developed device. In the demonstration, users can interact with the virtual object which is superimposed inside the real material by using an instrument.

2 Innovation

The principle of STELET display is based on tactile illusion caused by electrical stimulus and the anatomical nerve structure. When the nerve is electrically stimulated, users perceive the sensation at the terminal of the nerve. We investigated the optimal stimulus to apply the phenomenon to the tactile augmentation.

Hardware Design: The strategy to achieve the spatial transparency is to stimulate the afferent nerve rather than mechanoreceptors. The two main afferent nerves which run along with the side of a finger are stimulated. The device consists of an electrical pulse generator, two cathodes and a ground electrode. In order to augment tactile sensation without mediating device, the cathodes are attached on the each sides of the media of a finger. The device

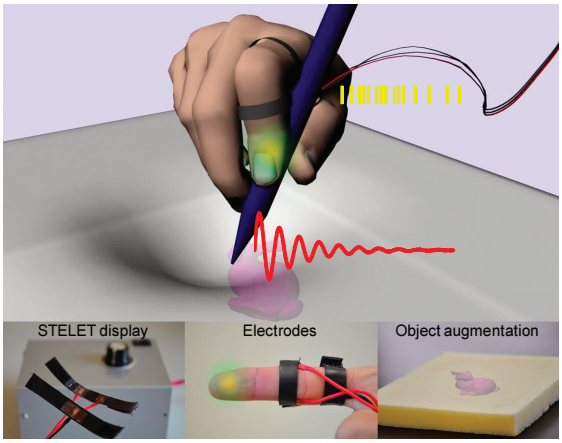


Figure 1: Tactile augmentation with STELET display

can control the pulse rate up to 100 pulses per second (pps) and the pulse height up to 2 mA. The pulse width is fixed at 200 μs.

Stimulus Design: In order to interact with a virtual object, transient acceleration and static force are calculated and displayed. Both acceleration and force are presented by the electrical pulse stimulus and the amounts are linearly controlled by the pulse rate (pulse rate modulation). The acceleration of the instrument according to the physical events such a tapping and dragging are modeled by sinusoidal wave. The static force according to the object deformation is also calculated by using Finite Element Method. The calculated acceleration and force are presented as grip force to integrate with the real sensation.

3 Demonstration

The purpose of demonstration is to investigate a potential of a tactile AR system with the developed spatially transparent tactile display. Especially, modulation of the perception during a tool manipulation with a virtual object superimposed to the real material will be evaluated. The system overview is illustrated in Fig. 1. Users attach the electrodes on their index finger, and calibrate the amount of current. Users touch the real object with holding an instrument and feel both real and virtual object which is visually augmented on the real material. We utilize silicon as a real material and project the virtual object from the back of the material. The position of the instrument is measured by using a camera, and the stimulus is calculated according to the position and velocity of the instrument. The developed tactile display is utilized to present the calculated tactile stimulus. Users feel the integrated grip force according to the physical interaction with an instrument and perceive the augmented virtual object.

References

[1] S. Yoshimoto, Y. Kuroda, M. Imura, and O. Oshiro. Development of a spatially transparent electrotactile display and its performance in grip force control. In *Proceeding of 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3463–3466, Boston, August/September 2011.

*e-mail: yoshimoto@bpe.es.osaka-u.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

A Compact Pseudo-Force Glove Using Shape Memory Alloys

Yu Shigeta* Osaka University
Yoshihiro Kuroda† Osaka University
Masataka Imura‡ Osaka University
Osamu Oshiro§ Osaka University

ABSTRACT

We propose a pseudo-force glove using SMAs(SMA: Shape Memory Alloy), which can present a sensation of grasping a virtual object. SMAs stimulate a fingertip and finger's joints to induce cutaneous and proprioceptive sensations, respectively. Downsizing and large force were achieved by setting the fulcrum of SMAs wire at finger's joints. The developed system enabled to present a sensation of grasping a virtual object.

Keywords: Pseudo-force, Shape memory alloy, Proprioceptive sensation

1 INTRODUCTION

MR(Mixed Reality), which integrates real and virtual spaces, is intensively studied. The wearable haptic devices that can be operated in a wide space draw the attention of MR researches. As a haptic device operable in a wide space, a pseudo-haptic display is noted[1], because the pseudo-haptic display is compact and lightweight. The pseudo-force presents original or different stimuli to stimulate a part of mechanoreceptors or proprioceptors effectively with a compact and lightwight device. Foregoing studies stimulate only finger mechanoreceptors[2]. We propose a pseudo-force glove using SMAs(SMA: Shape Memory Alloy), which can present a sensation of grasping a virtual object. SMAs stimulate a fingertip and finger's joints to induce cutaneous and proprioceptive sensations, respectively.

2 DEVICE CONFIGURATION

A pseudo-force device is composed of four parts per finger. Fingertip part is presented cutaneous sensations by compressing fingertips. Proprioceptive sensations are presented by stretching distal interphalangeal(DIP) joint, proximal interphalangeal(PIP) joint, and metacarpophalangeal(MP) joint parts to provide the joint torque.

SMAs shrinks by only 5% of full length. Therefore, if the length of SMAs changes when the finger is flexed, it can not get enough displacement to give the torque. In this study, in the DIP and PIP joints, the fulcrum is placed in the center of rotation of the fingers, passing through the fulcrum in SMAs, to reduce the change of SMAs length due to the finger flexion and to generate the torque effectively.

Assuming that the force $\vec{f}$ is applied to the fingertip, joint torque $\vec{\tau}$ is given by the formula (1).

$$\vec{\tau} = J^T \vec{f} \tag{1}$$

where $J = \dfrac{\partial \vec{r}}{\partial \vec{\theta}}$ is Jacobian matrix, $\vec{r}$ is position vector, and $\vec{\theta}$ is joint angle vector.

*e-mail:shigeta@bpe.es.osaka-u.ac.jp
†e-mail:ykuroda@bpe.es.osaka-u.ac.jp
‡e-mail:imura@bpe.es.osaka-u.ac.jp
§e-mail:oshiro@bpe.es.osaka-u.ac.jp

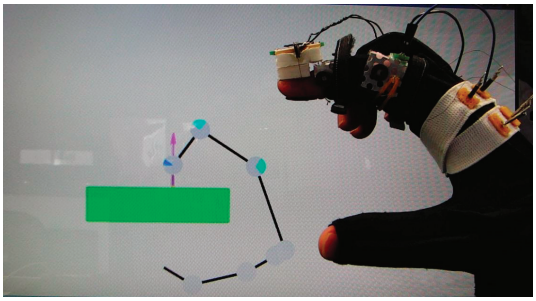


Figure 1: Prototype device.

Table 1: Specification of the prototype device.

Name	Value
Size (figertip)	15[mm] × 20[mm] × 20[mm]
Size (DIP joint)	15[mm] × 20[mm] × 10[mm]
Size (PIP joint)	15[mm] × 25[mm] × 15[mm]
Mass (per finger)	12[g]

3 SYSTEM

The system was equipped with Intel CPU(Core i5 2.26GHz), 4GB main meory, Intel HD graphics board, and the data glove 5DT 14 Ultra by Fifth Dimension Technologies. Specification of the prototype device is shown in Table1. Fig.1 shows a prototype of the pseudo-force gloves in this study. Joint angles and positions of the finger were measured using the data glove. Finger forces and finger joint torques were calculated from the penetration of the finger to the virtual object. Displacements of the SMAs were determined from the joint torque. As a result, the force sensation of grasping a virtual object was presented to the user.

4 CONCLUSION

A pseudo-force glove which stimulated a fingertip and finger's joints to induce cutaneous and proprioceptive sensations, respectively was developed. The result showed that the developed system enabled to present a sensation of grasping a virtual object with the lightweight device.

ACKNOWLEDGEMENTS

This study was partly supported by the Globel COE Program "*in silico* medicine" at Osaka University.

REFERENCES

[1] Y. Kuroda, M. Nakatani, S. Hasegawa, K. Fujita: Research Trend of Display and Calculation of Pseudo-Force Induced by Physical Stimuli, Transactions of the Virtual Reality Society of Japan, Vol. 16, No. 3, pp. 379-390, 2011 (in Japanese).
[2] K. Minamizawa, S. Fukamachi, H. Kajimoto, N. Kawakami, S. Tachi: Wearable Haptic Display to Present Mass and Internal Dynamics of Virtual Objects, Transactions of the Virtual Reality Society of Japan, Vol. 13, No. 1, pp. 15-24, 2008 (in Japanese).

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278

Multi-Viewpoint Interactive Fog Display

Masataka Imura* Asuka Yagi Yoshihiro Kuroda Osamu Oshiro

Osaka University

Abstract

We propose a 360-degree observable fog display which provides different images according to observers' position. The proposed display utilizes directional light scattering of fog so that multiple images which are projected from different directions on one cylindrical fog screen can be transmitted to appropriate observers. The fog display brings motion parallax to observers that can recognize a 3D structure of the presented objects.

Keywords: fog display, Mie scattering, multiple projection, motion parallax

1 Introduction

Fog displays are one of immaterial display systems. Foregoing systems (e.g. [1]) provide only 2D images on a flat screen. We propose a novel fog display system which uses multiple projectors can bring observers recognition of 3D shape of the presented object by motion parallax. 360-degrees viewable 3D displays which utilize projection of multiple images, such as Hitachi's Transpost[2] and Sony's RayModeler[3], have been developed. The advantage of the proposed fog display is that the proposed display enables direct operation to the virtual objects by hands.

2 Methods

The proposed display projects multiple images of a virtual object from different viewpoints onto one cylindrical fog screen (fig. 1). Since the water drops constituting the fog screen show the Mie scattering that is a strong forward directional scattering of lights, observers see only one image projected from the frontal projector at a time. In this way, the 3D shape of the object can be recognized from the motion parallax when the observer walks around the cylindrical fog screen.

Most of 3D displays cannot realize to touch and to grasp the virtual object because the projection system is covered and sometimes dangerous. One of the advantages of the fog display is that the observer can interact with the presented virtual objects by his or her hands. The proposed 3D fog display utilizes an infrared LED illumination and an infrared camera to detect the motion of observer's hand around the fog screen. This infrared vision system enables observers to handle the virtual object directly without any markers.

3 Results

Figure 2 shows the prototype system of the proposed fog display. A diameter of the cylindrical fog screen is 80 mm. We used three projectors and projected three slightly different images. The resolution of each image is 800×600. The projected images from projectors and the images on the fog screen are shown in figure 3. The result reveals that the projected images are not mixed and observers can see the virtual object as it were inside the fog.

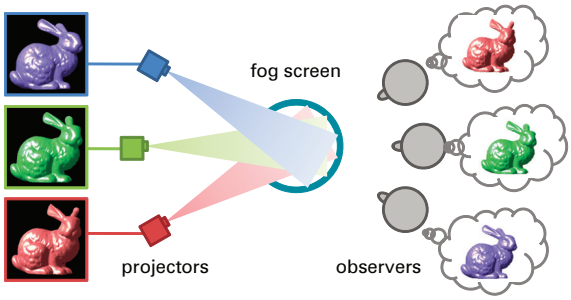


Figure 1: Concept of the display.

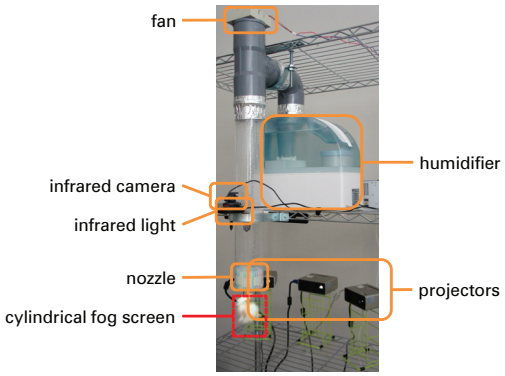


Figure 2: Prototype system.

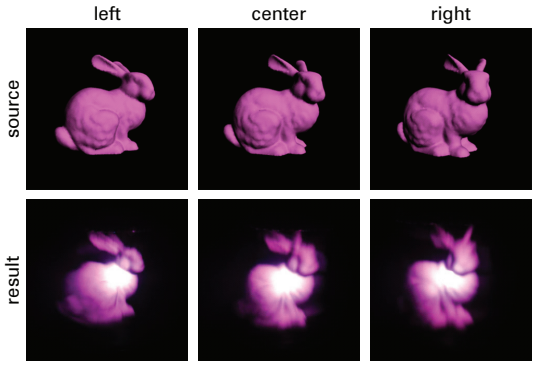


Figure 3: Results of projection onto fog screen.

References

[1] I. Rakkolainen, S. DiVerdi, A. Olwal, N. Candussi, T. Hüllerer, M. Laitinen, M. Piirto, and K. Palovuori. The interactive fogscreen. In *ACM SIGGRAPH 2005 Emerging technologies*, article 8, 2005.

[2] R. Otsuka, T. Hoshino, and Y. Horry. Transpost: A novel approach to the display and transmission of 360 degrees-viewable 3D solid images. *IEEE Transactions on Visualization and Computer Graphics*, 12(2):178–185, 2006.

[3] K. Ito, H. Kikuchi, H. Sakurai, I. Kobayashi, H. Yasunaga, H. Mori, K. Tokuyama, H. Ishikawa, K. Hayasaka, and H. Yanagisawa. 360-degree autostereoscopic display. In *ACM SIGGRAPH 2010 Emerging Technologies*, article 1, 2010.

*e-mail: imura@bpe.es.osaka-u.ac.jp

The 21st International Conference on Artificial Reality and Telexistence
November 28-30, 2011, Osaka, Japan
ISSN: 1345-1278