AMBIENT 2014

The Fourth International Conference on Ambient Computing, Applications, Services and Technologies

ISBN: 978-1-61208-356-8

August 24 - 28, 2014

Rome, Italy

AMBIENT 2014 Editors

Maarten Weyn, University of Antwerp, Belgium

Ivan Evgeniev, TU Sofia, Bulgaria

AMBIENT 2014

Forward

The Fourth International Conference on Ambient Computing, Applications, Services and Technologies (AMBIENT 2014), held on August 24 - 28, 2014 - Rome, Italy, was devoted for a global view on ambient computing, services, applications, technologies and their integration.

On the way for a full digital society, ambient, sentient and ubiquitous paradigms lead the torch. There is a need for behavioral changes for users to understand, accept, handle, and feel helped within the surrounding digital environments. Ambient comes as a digital storm bringing new facets of computing, services and applications. Smart phones and sentient offices, wearable devices, domotics, and ambient interfaces are only a few of such personalized aspects. The advent of social and mobile networks along with context-driven tracking and localization paved the way for ambient assisted living, intelligent homes, social games, and telemedicine.

The conference provided a forum where researchers were able to present recent research results and new research problems and directions related to them. We welcomed technical papers presenting research and practical results, position papers addressing the pros and cons of specific proposals, such as those being discussed in the standard forums or in industry consortiums, survey papers addressing the key problems and solutions on any of the above topics, short papers on work in progress, and panel proposals.

We take here the opportunity to warmly thank all the members of the AMBIENT 2014 technical program committee as well as the numerous reviewers. The creation of such a broad and high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and efforts to contribute to the AMBIENT 2014. We truly believe that thanks to all these efforts, the final conference program consists of top quality contributions.

This event could also not have been a reality without the support of many individuals, organizations and sponsors. We also gratefully thank the members of the AMBIENT 2014 organizing committee for their help in handling the logistics and for their work that is making this professional meeting a success. We gratefully appreciate to the technical program committee co-chairs that contributed to identify the appropriate groups to submit contributions.

We hope the AMBIENT 2014 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in ambient computing research.

We hope Rome provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city.

AMBIENT 2014 Chairs

AMBIENT Advisory Chairs
Maarten Weyn, University of Antwerp, Belgium
Yuh-Jong Hu, National Chengchi University-Taipei, Taiwan (ROC)
Naoki Fukuta, Shizuoka University, Japan
Andrzej Skowron, Warsaw University, Poland
Stephan Hengstler, MeshEye Consulting, USA
Andreas Schrader, University of Lübeck, Germany
Nuno Otero, Linnaeus University, Sweden
Jean Vareille, Université de Bretagne Occidentale, France
Peter Roelofsma, VU - Amsterdam, The Netherlands
Tomasz Rutkowski, University of Tsukuba, Japan
Keith V. Nesbitt, University of Newcastle, Australia
Gunver Majgaard, University of Southern Denmark, Denmark
Thomas Grill, University of Salzburg, Austria
Sven Hartmann, Clausthal University of Technology, Germany
Daniel Berckmans, Katholieke Universiteit Leuven, Belgium
Eugenio Aguirre Molina, Universidad de Granada, Spain
Jean-Paul Sansonnet, LIMSI, University of Paris-Sud 11, France
José Bravo, Castilla-La Mancha University, Spain
Shusaku Nomura, Nagaoka University of Technology, Japan

AMBIENT 2014 Industry/Research Liaison Chairs
Rémi Emonet, Idiap Research Institute, Switzerland
Mieczyslaw A. Klopotek, Polish Academy of Sciences - Warszawa, Poland
Gerhard Nussbaum, Competence Network Information Technology to Support the Integration of People with Disabilities, Austria
Alexander Paar, TWT GmbH Science and Innovation, Germany
Patrick Sayd, CEA SACLAY - NANO INNOV - Gif Sur Yvette, France
Ingo Zinnikus, DFKI GmbH - Saarbrücken, Germany
Jennifer Stein, University of Southern California, USA
Florian Scharf, Institute for Multimedia and Interactive Systems, Germany

AMBIENT Special Area Chairs

Agents
Joseph Giampapa, Carnegie Mellon University, USA
Adina Magda Florea, University "Politehnica" of Bucharest, Romania

Sensing
Sandra Sendra Compte, Polytechnic University of Valencia, Spain

Context-awareness
Sylvain Giroux, Université de Sherbrooke, Canada

AMBIENT Publicity Chairs

Jan Stevens, University of Antwerp, Belgium
Nancy Diniz, Xi'an Jiaotong-Liverpool University, China
Tomas Zizka, technical University of Liberec, Czech Republic

AMBIENT 2014

Committee

AMBIENT Advisory Chairs

Maarten Weyn, University of Antwerp, Belgium
Yuh-Jong Hu, National Chengchi University-Taipei, Taiwan (ROC)
Naoki Fukuta, Shizuoka University, Japan
Andrzej Skowron, Warsaw University, Poland
Stephan Hengstler, MeshEye Consulting, USA
Andreas Schrader, University of Lübeck, Germany
Nuno Otero, Linnaeus University, Sweden
Jean Vareille, Université de Bretagne Occidentale, France
Peter Roelofsma, VU - Amsterdam, The Netherlands
Tomasz Rutkowski, University of Tsukuba, Japan
Keith V. Nesbitt, University of Newcastle, Australia
Gunver Majgaard, University of Southern Denmark, Denmark
Thomas Grill, University of Salzburg, Austria
Sven Hartmann, Clausthal University of Technology, Germany
Daniel Berckmans, Katholieke Universiteit Leuven, Belgium
Eugenio Aguirre Molina, Universidad de Granada, Spain
Jean-Paul Sansonnet, LIMSI, University of Paris-Sud 11, France
José Bravo, Castilla-La Mancha University, Spain
Shusaku Nomura, Nagaoka University of Technology, Japan

AMBIENT 2014 Industry/Research Liaison Chairs

Rémi Emonet, Idiap Research Institute, Switzerland
Mieczyslaw A. Klopotek, Polish Academy of Sciences - Warszawa, Poland
Gerhard Nussbaum, Competence Network Information Technology to Support the Integration of People with Disabilities, Austria
Alexander Paar, TWT GmbH Science and Innovation, Germany
Patrick Sayd, CEA SACLAY - NANO INNOV - Gif Sur Yvette, France
Ingo Zinnikus, DFKI GmbH - Saarbrücken, Germany
Jennifer Stein, University of Southern California, USA
Florian Scharf, Institute for Multimedia and Interactive Systems, Germany

AMBIENT Special Area Chairs

Agents
Joseph Giampapa, Carnegie Mellon University, USA
Adina Magda Florea, University "Politehnica" of Bucharest, Romania

Sensing
Sandra Sendra Compte, Polytechnic University of Valencia, Spain

Context-awareness
Sylvain Giroux, Université de Sherbrooke, Canada

AMBIENT Publicity Chairs

Jan Stevens, University of Antwerp, Belgium
Nancy Diniz, Xi'an Jiaotong-Liverpool University, China
Tomas Zizka, technical University of Liberec, Czech Republic

AMBIENT 2014 Technical Program Committee

Sachin Kumar Agrawal, Samsung - Noida (Delhi NCR), India
Eugenio Aguirre Molina, Universidad de Granada, Spain
Cristina Alcaraz, University of Malaga, Spain
Hani M. Alzaid, King Abdulaziz City for Science and Technology, Saudi Arabia
Cecilio Angulo, Universitat Politècnica de Catalunya, Spain
Eduard Babulak, Sungkyunkwan University, Korea
Costin Badica, University of Craiova, Romania
Youakim Badr, INSA - Lyon, France
Maxim Bakaev, Novosibirsk State Technical University, Russia
Flavien Balbo, University of Paris Dauphine, France
Stephen Barrass, University of Canberra, Australia
Jean-Paul Barthès, Université de Technologie de Compiègne, France
Cecile Belleudy, University Of Nice - Sophia Antipolis LEAT laboratory, France
Orlando Belo, University of Mihno, Portugal
César Benavente, Universidad Politécnica de Madrid, Spain
Daniel Berckmans, Katholieke Universiteit Leuven, Belgium
Daniel Biella, University of Duisburg-Essen, Germany
Stephan Böhm, Hochschule RheinMain, Germany
José Bravo, MAmI Research Lab, Spain
Philip J Breedon, Nottingham Trent University, UK
Ramon F. Brena Pinero, Tecnológico de Monterrey, Mexico
Valerie Camps, Universite Paul Sabatier - Toulouse, France
Juan-Carlos Cano, Universitat Politecnica de Valencia, Spain
Carlos Carrascosa, Universidad Politécnica de Valencia, Spain
John-Jules Charles Meyer, Utrecht University, The Netherlands
Omar Cheikhrouhou, University of Sfax, Tunisia
Albert M. K. Cheng, University of Houston, USA
Cezar Collazos, Universidad del Cauca, Columbia
Carlos Fernando Crispim-Junior, INRIA Sophia Antipolis, France
Manuel Cruz, Immersion Corp, Canada
Giuseppe De Pietro, High Performance Computing and Networking Institute - National Research Council, Italy
Jianguo Ding, University of Skövde, Sweden
Nancy Diniz, Xi'an Jiaotong-Liverpool University, China

Benjamin Weyers, RWTH Aachen University, Germany
Isaac Wiafe, University of Reading, UK
Markus Wiemker, RWTH Aachen University, Germany
Lawrence W.C. Wong, National University of Singapore, Singapore
Chang Wu Yu, Chung Hua University, Taiwan
Zhiyong Yu, Fuzhou University, China
Olivier Zendra, INRIA Nancy, France
Ingo Zinnikus, DFKI GmbH - Saarbrücken, Germany

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Head-Mounted Display in Work Assistance

- An Origami Case Study -

Tadashi Miyosawa, Takayuki Gujo
Department of Business Administration and Information
Tokyo University of Science, Suwa,
Nagano, Japan
miyosawa@rs.suwa.tus.ac.jp, jh110044@ed.tus.ac.jp

***Abstract*—Head-mounted display (HMD) has gained significant attention in recent years. The goal of this study was to verify whether using HMD in work assistance actually improves work efficiency via an evaluation test. We focused on the field of work assistance and reflected on whether people who use HMD become interested in using the technology to help them to work efficiently when they fold origami in their spare time, compared with other methods such as an origami book and cellphone applications. The results of our evaluation test show that HMD technology is effective in the field of work assistance.**

Keywords-head-mounted display; HMD; work assistance; origami.

I. INTRODUCTION

In recent years, head-mounted displays (HMDs) have been sold by a variety of makers and are used in many industries.

For many years, HMDs have garnered attention as the device that realizes the visual aspect of virtual reality (VR), and there has been continued research and development into wearable computers. HMDs are display devices that use a head mount that can be either in the form of goggles or glasses. The HMD displays visuals in front of the eyes and is characterized by having a superior sense of immersion compared with typical displays. HMDs can be separated into four categories, but most of the products on the market today can be split into two main categories: one that shows an image in one eye, and one that shows an image in both. The other two categories are displays where you can see behind the images and see the surrounding landscape (see-through) and displays where you cannot (nonsee-through). HMDs for both eyes are the largest portion of consumer products. Products such as the SONY HMZ-T3 [11] (Fig. 1) represent nonsee-through displays for both eyes. This type of product is suited for the use of some types of audiovisual (AV) content such as when viewing movies or playing games because the user can be immersed into the images without seeing the surrounding background. In addition, see-through displays for both eyes such as the Wrap [13] series by Vuzix and Moverio [12] (Fig. 2) by Epson are also suited for use with AV content. In contrast, single-eye HMDs are unsuitable for the total immersion experience because they do not totally block the user's vision. Instead, they are suitable for use during work and thus these HMDs are designed for businesses.

Figure 1. HMZ-T3 [11].

Figure 2. Moverio [12].

There is also an unprecedented need for improvements in factory floor productivity. For example, it takes time to fix mechanical problems that occur in isolated locations. This then turns into a big problem for the production process. The engineers must be transferred to that location temporarily to address the work that cannot be done by the local staff. Technical succession is becoming difficult as the number of expert engineers declines.

What are the advantages of using HMDs to support daily operations? One is that an employee can work with their hands free. That means an employee can access a user manual or any other materials they need without breaking from their main duties. Therefore, in this paper, we will examine the effectiveness of using HMDs during work. We will consider whether superimposing movies or images onto a real-world view improves work accuracy and efficacy.

We note that fundamental research into HMDs and human anatomy has been explored previously [1] [2] [3] [4].

The aim of this experiment is to validate the work optimization of HMD and propose new usage methods for future HMDs by having students of the Tokyo University of Science, Suwa use and assess the technology.

In present-day origami, it is standard to fold the paper while looking at images in an origami book or on a cellphone application. In this experiment, we considered whether students could shorten the time to comprehend greater amounts of messages, seeking comments like "the directions are easy to understand" and "this is a smoother

 ISBN: 978-1-61208-356-8

process than folding origami while looking at images" by having them fold origami while viewing movies on the HMD, whether they were able to enjoy origami thanks to this new method, and whether they would want to use this method again.

Section I discussed the background that led to this study, and the aim of this paper. Section II introduce the related works. Section III introduces the environment that was prepared for creating the contents used in the HMD, and the developed content for experiment. Section IV introduces the three methods in the experiment that were used for the materials in Section III. Section V discusses the results and observations of each experiment discussed in Section IV. Section VI proposes improvements for the methods in this paper based on the experimental results. Section VII discusses conclusions made based on the experimental results.

II. RELATED WORKS

Currently, research is being conducted on applications for work support and other various applications [6] [7] [8][15], including the ones that we will introduce below.

1) Logistics solutions: In a four-company collaboration between Seiko Epson Corporation [12], Toyo Kanetsu Solutions K.K., Kokusai Kogyo Co., Ltd, and King Jim Co., Ltd, testing of logistics solutions using Epson's Moverio was conducted at Toyo Kanetsu Solutions K.K.'s plant in Chiba, Japan.

As shown in Fig. 3, these brand-new logistics solutions can aid workers in safely and efficiently completing their picking and sorting work while the navigation screen of the picking course is superimposed onto their real-world perception by having workers wear the Moverio in logistics facilities, such as a warehouse or logistics center with using Augmented Reality(AR).

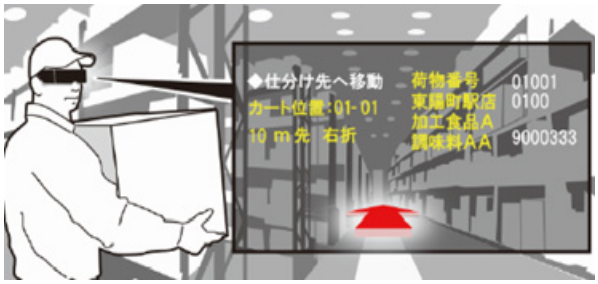


Figure 3. Logistics solutions using AR navigation.

2) Sorting support for physical merchandise distribution: Yamazaki et al. [5] focused on order picking for warehouse storage and proposed an order-picking system that assists workers through mixed reality (MR) technology. They identified problems that order-picking systems have faced in the past, and worked to improve work efficiency and reduce errors by solving those problems using MR technology. The system can intuitively present information to support the current task by superimposing computer-generated images onto the worker's field of vision through an HMD. This research confirmed the behavior of the developed system and carried out evaluative experiments that compared the impact of see-through video HMD and optical HMD on work efficiency.

3) A Highly-portable Markerless User Interface Using Optical See-through HMD and AR Technology : Irie et al. [15] proposed a system "AirTarget", which can point virtual and real object directly with user's finger is proposed for optic see-through HMD devices. The camera attached to the HMD device detects the position of user's fingertip, calibrates the gap between the sight of the camera and eye, and displays the cursor overlapped to the finger on the virtual plain. Finger detection is done in markerless image processing, so that it does not require specific input devices or external computer. This system enables to send control commands by simple gesture, working as a self-contained interface. The user is able to point a virtual object with their finger, and able to cut out an object of the real sight which can be used as a query of the image searching. And the simple gesture operation was achieved highlyaccuracy and recognition rate.

III. CREATING THE HMD CONTENTS

In this section, environment that was prepared for creating the content, and the developed content for experiment are explained.

A. How to Use an HMD

The controller has an Android system. The images and movies installed on the micro memory card are played by the Android system installed in the HMD controller (Fig. 4). The user views the images and movies through the headset.

The screen size changes according to the viewpoint. If the viewpoint is distant, the screen will also appear as if it is far away.

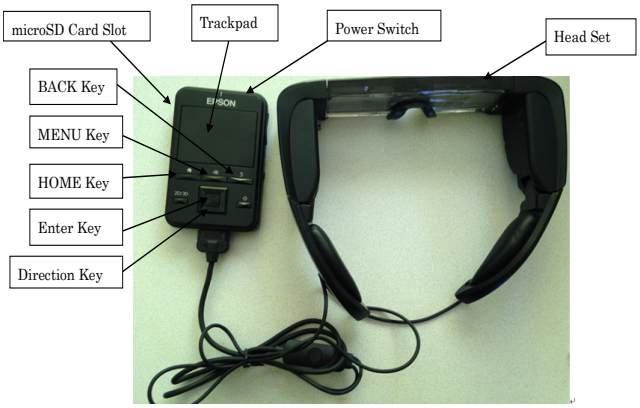


Figure 4. HMD controller and headset.

B. Development/Testing Environment of HMD Images and Movies for this Study

As shown in Table I, As for HMD, MOVERIO was selected for this experiment considering the availability in Japan. A Smart Phone was used to capture the images and movies. And PC was used to edit the images and movies.

TABLE I. DEVELOPMENT/TESTING ENVIRONMENT OF HMD IMAGES AND MOVIES

HMD		Camera		PC	
Product Name	MOVERIO (EPSON)	Product Name	iPhone5	Product Name	Fujitsu FMV-E8290
Movie	MP4(MPEG4+AAC) MPEG2TS (H.264+AAC) SD-Video	Image	8 Million Pxcels (1280×960 dot)	CPU	Intel(R)Core™2Duo CPU P8700(2.53GHz)
Image	JPEG,PNG,BMP,GIF			Memory	2GB
Memory	1 GB microSD(Max 2GB) microSDHC(Max32GB)			OS	Windows7 Professional

The GNU Image Manipulation Program (GIMP) was used to edit the images. A large number of images were prepared in this software using methods such as layering and by cutting the images into shapes.

C. Content Used in the Experiment

1) Origami book: "Present, Decorate, and Enjoy Practical Origami" by Mitsunobu Sonobe was used in this experiment.

2) HMD images: For this experiment, six types of origami with the same estimated completion time were chosen from Sonobe's book. We took one picture per step for each of the six types of origami. The pictures were then transferred to the computer and then saved to the micro memory card. Figure 5 shows one example of an HMD image set of the origami contents used in this experiment on the HMD. This example is of a folding style for a dachshund.

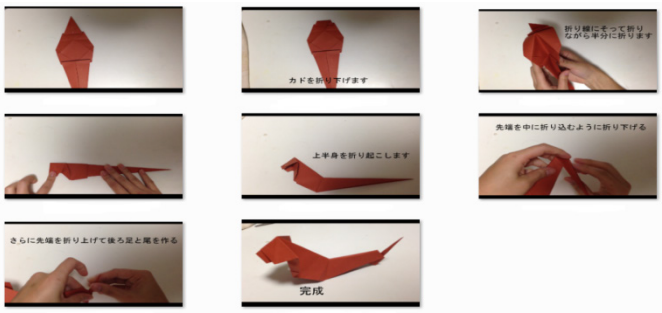
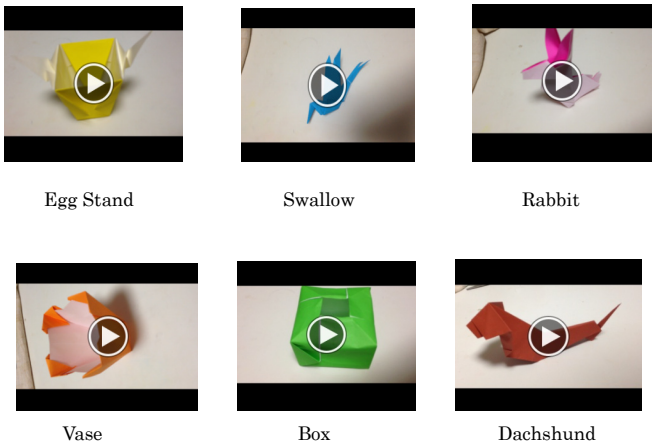


Figure 5. HMD image set.

3) HMD movies: We recorded movies showing the folding of the same six types of origami. Pace was considered while folding the origami so that all six types were completed in the same amount of time.



Egg Stand Swallow Rabbit

Vase Box Dachshund

Figure 6. HMD movies.

The movies were then saved to the micro memory card. Figure 6 shows the six types (movies) of the origami contents used in this experiment on the HMD.

IV. EXPERIMENT

In this section, detailed process of our experiment is explained as follows.

A. Experiment

1) Purpose of this experiment: The purpose of the experiment is to improve the efficiency of folding origami compared with the traditional method of using an origami book through an experiment where participants fold origami using an HMD. Another purpose of this experiment is to listen to the thoughts and opinions of the participants through the user surveys, and verify improvement plans for future materials.

2) Participants: Eighteen students from the Tokyo University of Science, Suwa.

3) Details of the experiment: We implemented three experiments: "origami book," "HMD images," and "HMD movies." Each of the participants performed each experiment in turns using HMD.

In each experiment, the participant created one of the six origami according to the instructions they received. They measured the time it took to complete each origami and answered a user survey after completing all experiments.

4) Materials: Origami book; movies and images created for this experiment; HMD; three sheets of origami paper per person; user survey sheet; timer; and writing tools.

5) Experiment: An origami book and HMDs were used in the experiment. Because most participants were using an

 ISBN: 978-1-61208-356-8

HMD for the first time, we arranged for an explanation and some operational practice before the actual experiment began. For practice, the participants used movies and images that would not be used during the actual experiment. Additionally, to avoid bias in the data because of working order, the order for each participant was based on an order sheet created specifically for that practice. Each participant answered a user survey after completing the experiments with the origami book, HMD movies, and HMD images.

6) The sequence: First, prepare the origami book and turn on the HMD. Prepare the HMD so that it can be used immediately. Give the origami paper to participants so that they can begin folding the origami immediately. Each participant was called on individually and began their practice (see Fig. 7) based on the practice order listed below.

Figure 7. Test scene.

<For HMD images>

- Explanation of experiment
- Distribution of origami
- Prepare HMD images for practice
- Practice
- Prepare HMD images for actual experiment
- Fold origami
- Finish the origami and collect HMD

< For HMD movies>

- Explanation of experiment
- Distribution of origami
- Prepare HMD movies for practice
- Practice
- Prepare HMD movies for actual experiment
- Fold origami
- Finish the origami and collect HMD

<For the origami book>

- Explanation of experiment
- Distribution of origami and book
- Fold origami while viewing instructions in book
- Collect book

After a participant completed all three origami, we distributed a user survey and writing materials, and the participant was encouraged to express their opinions on the experiment freely. The experiment was considered complete when the participant had completed the survey and returned both the survey and writing materials.

There were 18 participants in total. We ensured that all 18 participants were equal and that there was no chance of Bias by having three different folding methods for six different types of origami.

V. RESULTS AND OBSERVATIONS

Based on the experiment described in Section IV, Following result was observed.

A. Survey Results

1) Comparison of survey results: The average of each survey category as shown in Fig. 8 is explained below.

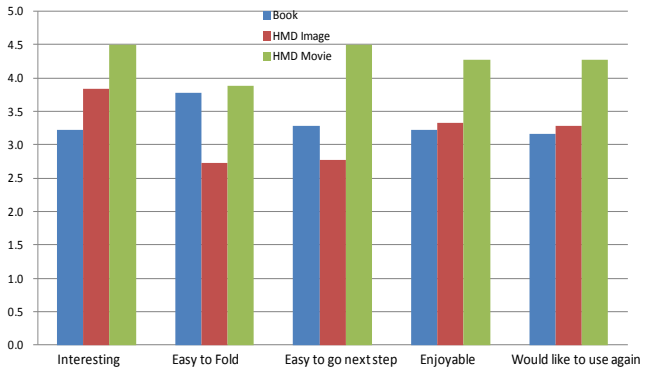


Figure 8. Average of each survey category.

a) The order of the results in the category, "Found it interesting" were HMD movies > HMD images > origami book. These results suggest that the participants developed an interest in HMD technology after the experiment.

b) The order of the results in the category, "Environment in which origami can easily be folded" were HMD movies > origami book > HMD images (there was a slim margin between HMD movies and origami book). These results suggest that folding origami using the HMD was not a hindrance for the participants.

c) The order of the results in the category, "Easy to understand next step" were HMD movie > origami book > HMD images. This shows that movies were easier to understand than images in the HMD environment and that images were inferior to the origami book.

a) The order of the results in the category, "Enjoyed folding the origami" were HMD movies > HMD images > origami book. These results suggest that the participants enjoyed folding origami better with HMD technology.

e) The order of the results in the category, "Would use again" were HMD movies > HMD images > origami book. These results suggest that participants would like to continue using HMD technology.

f) When we look at the average of the sum of all mean values, the average of both the origami book and HMD images are about the same, but the average for HMD movies is much higher than both categories, as shown in Fig. 9.

 ISBN: 978-1-61208-356-8

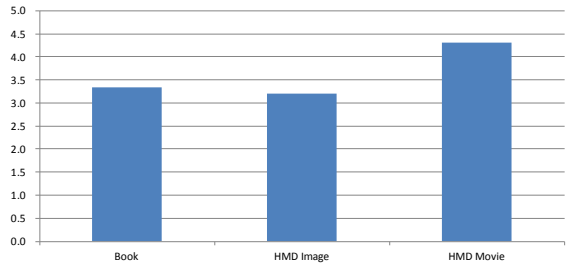


Figure 9. The average of each mean value by category.

2) Analysis of survey results: A one-way analysis of variance was conducted on each category using valuation points to verify whether there is a significant difference in the evaluation points in each category in the survey as shown in Fig. 8.

a) There is a significant difference between the origami book and HMD images in the category, "Found it interesting." There is also a significant difference in the traditional origami book and HMD movies as well as between HMD images and movies.

b) There is a significant difference between the origami book and HMD images in the category, "Environment in which origami can easily be folded." There is also a significant difference between HMD images and movies.

c) There is a significant difference between the origami book and HMD movies in the category, "Easy to understand next step." There is also a significant difference between HMD images and HMD movies.

d) There is a significant difference between the origami book and HMD movies in the category "Enjoyed folding the origami." There is also a significant difference between HMD images and movies.

c) There is a significant difference between the origami book and HMD movies in the category, "Would use again." There is also a significant difference between HMD images and movies.

B. *Measurement of Completion Time*

As shown in Fig. 10, five of the six types of origami were faster to fold using HMD movies. Comparing Fig. 10 and Table II, it can be concluded that the more steps there are to complete the origami, the longer it takes when using the origami book compared with HMD movies and the smoother the process when using HMD movies. However, it is faster to use the traditional origami book rather than the HMD for origami with fewer steps to complete. These results show that the appropriateness of HMD or the original origami book depends on the number of steps needed. These results show that HMD is appropriate for difficult work and inappropriate for simple work.

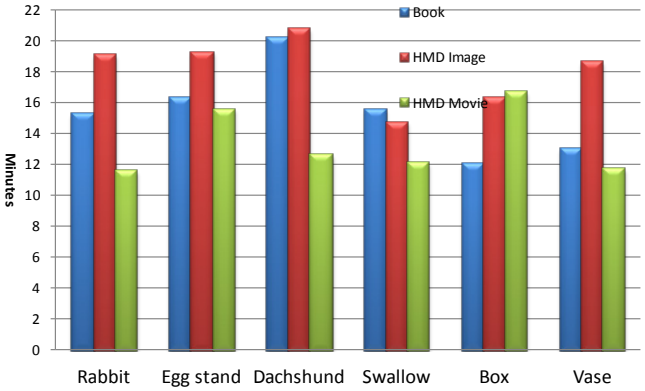


Figure 10. Measurement of completion time.

TABLE II. NUMBER OF STEPS

Rabbit	25 Steps
Egg Stand	22 Steps
Dachshund	25 Steps
Swallow	28 Steps
Box	18 Steps
Vase	21Steps

Figure 11 is the average value of completion time for the method used to complete the origami. This figure shows that the completion time was shorter when using the origami book compared with HMD images, and that the completion time was shortest when using HMD movies. The average work time was fastest with HMD movies.

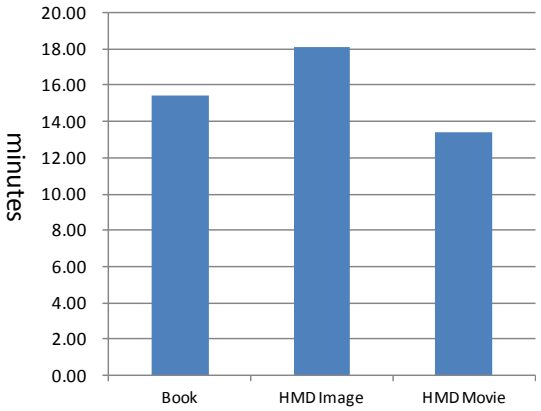


Figure 11 Average completion time by method.

C. *Free Descriptions*

Participants' positive and negative opinions about this experiment are shown in Table III.

TABLE III. OPINIONS

	Good Point	Bad Point
Book	All Process can be seen at the same time. (6 people)	Hard to see complicated process. (7 people)
HMD Image	Can keep his own pace. Can compare his output. (10 people)	Hard to find the folding line. (7 people)
HMD Movie	Because This is see thorough HMD. It was very easy to do same things with movie. (14 people)	Once you get trouble, Operation is bit difficult. (6 people)

The positive responses for HMD movies indicate that close to half of the participants held the opinion that HMD was "Easy to work with because I could copy the folding methods in the movie because the HMD was see-through." Everyone was used to the paper medium of the traditional origami book. As a result, participants felt that it was easy to understand how to fold the origami, easy to complete, and it was also easy to view the entire process.

Other relatively common opinions were that using HMD is "excitingly new" and that they "became interested in the technology." Another opinion that stood out was that the biggest difference from the traditional origami book was that "I was able to fold the origami while viewing the folding method in the see-through glasses, and it was better compared with using the traditional book." Negative opinions were "It's difficult to see and understand where to fold" and as a result, "It was a little confusing and the equipment was bothersome."

The aim was to verify whether the use of HMD improves efficiency over the traditional origami book, and to verify if we can increase interest in HMD. As a result, we can consider that the participants were able to increase their efficiency and work and becoming interested in HMD. It can be considered that there were no obstacles while they were working.

VI. IMPROVEMENT PLAN

The improvement plan for creating future materials for work (this time for folding origami) with HMD was examined.

Many participants held the opinion that "it was difficult to locate the folding crease" with HMD images. Many participants held the opinion that "using the equipment was bothersome" regarding HMD movies.

An improvement plan for HMD images/movies was created based on these two points.

A. Improvement Plan for HMD Images

The instructions for the HMD images used in this experiment were displayed together with the images. We thought that doing work using a different method to the traditional origami book was connected to the improvement of work efficiency; however, it was difficult for participants to locate the folding crease, increasing the completion time.

B. Improvement Plan for HMD Movies

An HMD movie was created for this portion of the experiment. Many participants felt that it was tiresome to have to redo their work if they had trouble for some reason. The participants were able to complete the origami in a shorter amount of time compared with traditional methods. However, it was necessary to coordinate everyone because some people work quickly and some people work slowly.

To address this, we propose a subdivision of the movie into steps by creating an HMD movie that has a pause after each step. If the movies are subdivided by steps, then the participants will be better able to understand the actual situation and they can easily review a step when needed, thus further increasing work efficiency.

VII. CONCLUSION

The aim of this study was to verify whether using HMD in the area of work assistance actually improves work efficiency using an evaluation test. Another aim is to review the improvement plan for future materials based on the experimental results, and to review conclusively the efficacy of HMD in the area of work assistance.

We focused on the field of HMD work assistance and reflected on whether people who use HMD become interested in using the technology to help them to work efficiently when they fold origami in their spare time, compared with other methods such as an origami book and cellphone applications.

In this experiment, 18 participants folded origami while using an origami book, HMD images, and HMD movies, and answered an accompanying user survey after completion of the experiment.

We observed the following three results:

1) To explore work efficiency improvement, participants folded origami using an origami book and HMD movie. Using the HMD movie was overall faster than using the traditional origami book. However, using HMD images is an inferior method to using an origami book. We also discovered that using HMD movies is an inferior method when there are few steps involved in completing the origami.

This may be because a user cannot see the whole picture when using HMD images and it takes time to operate the equipment. These results show that HMD movies increase work efficiency more than do HMD images. However, using HMD takes longer for simple tasks. The average completion time from these procedures show that the order of fastest completion time, in descending order, is HMD images > origami book > HMD movies. This order shows that HMD movies are more effective than an origami book or HMD images.

2) The interest in using HMD technology was examined using a paper user survey. When we examined the average values taken from the survey answers, HMD movies received a higher value than the origami book in three categories: "I became interested in the technology," "Enjoyed folding origami," and "I would like to use HMD again." The average data for each category tells us that

 ISBN: 978-1-61208-356-8

participants preferred, in ascending order, HMD images < origami book < HMD movies. This result shows that using HMD is beneficial to improving participant interest in the technology.

3) In the "free description" section of the survey, six people held the positive opinion that "the origami is easy to understand because I could view the whole thing at once," regarding using the origami book. Seven people held the negative opinion that it was difficult to understand complicated origami instructions when using an origami book. Ten people held the positive opinion that folding origami while looking at HMD images "is comparable to when I fold origami on my own, at my own pace." Seven people held the negative opinion that "it was difficult to locate the folding crease" when using HMD images. Fourteen people held the positive opinion that "it was easy to copy the folding methods because the HMD movie is see through." Six people held the negative opinion that "It was a little confusing and difficult to use the equipment" regarding HMD movies.

From these three results, we can conclude that using HMD technology is effective in the field of work assistance.

In this study, we have not applied AR technology to Origami yet. However, in the next experiment, we are going to compare Origami with AR technology and Movie in HMD environment.

REFERENCES

[1] T. Kawai, T. Iwasaki, T. Inoue and, K. Noro,"Effect of Head Mount Display on Visual Function" Transactions of the Virtual Reality Society of Japan 4(1), 1999, pp. 275-280.

[2] H. Kono, "Visibility and Distance Error on VR Landscape Simulation" Transactions of the Virtual Reality Society of Japan 1(2), 1996, pp. 9-14.

[3] H. Iwase and A. Murata, "Effects of Long Time Task Using Head Mounted Display on Postural Control System" The Transactions of the Institute of Electronics, Information and Communication Engineers. A J85-A(9), 2002, pp.1005-1013.

[4] K. Watanabe, T. Takeuchi, T. Inoue and, K. Okada, "Face-to-Face Collaboration System that Identifies Individual User" The IEICE transactions on information and systems (Japanese edition) J91-D(12), 2008, pp. 2755-2764.

[5] K. Yamazaki, F. Shibata, A. Kimura and, H. Tamura, "Application of Mixed Reality Technology to Order Picking for Warehouse Storage" Proceedings of the Virtual Reality Society of Japan, Annual Conference 18, 2013, pp. 196-199.

[6] Y. Yoshimura "Interactive work support system with movable head-up display" (Nara Institute of Science and Technology) Master's Thesis.

[7] S. Tamiya, M. Kataoka, R. Tenmoku, F. Shibata, A. Kimura and, H. Tamura, "A Network Wiring Support System Using Mixed Reality" Dissertations of Information Processing Society of Japan Kansai Branch, 2007, pp. 13-14.

[8] K. Nitta, "Developing Nuclear Power Plant Maintenance Support Systems Using Augmented Reality", 2002, Kyoto University Master's Thesis.

[9] M. Sonobe, "Present, Decorate, Enjoy Practical Origami" Seibido Shuppan, July 2012.

[10] Nikkei Monozukuri Magazine, October 2012, pp. 158-159.

[11] SONY HMZ-T2 Home page(http://www.sony.jp/hmd/products/HMZ-T2/image.html) [retrieved: June 2014].

[12] Epson Moverio Home Page (http://www.epson.jp/products/moverio/bt100/feature.html) (http://www.epson.jp/osirase/2013/130911.htm) [retrieved: June 2014].

[13] Telescope Magazine (http://www.tel.co.jp/museum/magazine/human/120810_topics_06/04.html) [retrieved: June 2014].

[14] engaget Japanese Edition (http://japanese.engadget.com/2012/08/02/vr-oculous-rift-70-cloud/) [retrieved: June 2014].

[15] H. Irie et al., "AirTarget: A Highly-portable Markerless User Interface Using Optical See-through HMD", Transaction of Information Processing Society of Japan, 55(4), Apr 2014, pp.1415-1427.

 ISBN: 978-1-61208-356-8

A Survey of Rigid 3D Pointcloud Registration Algorithms

Ben Bellekens, Vincent Spruyt, Rafael Berkvens, and Maarten Weyn

CoSys-Lab, Faculty of Applied Engineering
University of Antwerp, Belgium
ben.bellekens@uantwerpen.be

***Abstract*— Geometric alignment of 3D pointclouds, obtained using a depth sensor such as a time-of-flight camera, is a challenging task with important applications in robotics and computer vision. Due to the recent advent of cheap depth sensing devices, many different 3D registration algorithms have been proposed in literature, focussing on different domains such as localization and mapping or image registration. In this survey paper, we review the state-of-the-art registration algorithms and discuss their common mathematical foundation. Starting from simple deterministic methods, such as Principal Component Analysis (PCA) and Singular Value Decomposition (SVD), more recently introduced approaches such as Iterative Closest Point (ICP) and its variants, are analyzed and compared. The main contribution of this paper therefore consists of an overview of registration algorithms that are of interest in the field of computer vision and robotics, for example Simultaneous Localization and Mapping.**

Keywords–3D pointcloud; PCL; 3D registration; rigid transformation; survey paper

I. INTRODUCTION

With the advent of inexpensive depth sensing devices, robotics, computer vision and ambient application technology research has shifted from 2D imaging and Laser Imaging Detection And Ranging (LIDAR) scanning towards real-time reconstruction of the environment based on 3D pointcloud data. On one hand, there are structured light based sensors such as the Microsoft Kinect and Asus Xtion sensor which generate a structured point cloud, sampled on a regular grid, and on the other hand, there are many time-of-flight based sensors such as the Softkinetic Depthsense camera yield an unstructured pointcloud. These pointclouds can either be used directly to detect and recognize objects in the environment where ambient technology is been used, or can be integrated over time to completely reconstruct a 3D map of the camera's surroundings [1], [2], [3]. In the latter case however, point clouds obtained at different time instances need to be aligned, a process which is often referred to as registration. Registration algorithms are able to estimate the ego-motion of the robot by calculating the transformation that optimally maps two pointclouds, each of which is subject to camera noise.

These registration algorithms can be classified coarsely into rigid and non-rigid approaches. Rigid approaches assume a rigid environment such that the transformation can be modeled using only 6 Degrees Of Freedom (DOF). Non-rigid methods on the other hand, are able to cope with articulated objects or soft bodies that change shape over time.

Registration algorithms are used in different fields and applications, such as 3D object scanning, 3D mapping, 3D localization and ego-motion estimation, human body detection. Most of these state-of-the-art applications employ either a simple Singular Value Decomposition (SVD) [4] or Principal Component Analysis (PCA) based registration, or use a more advance iterative scheme based on the Iterative Closest Point (ICP) algorithm [5]. Recently, many variants on the original ICP approach have been proposed, the most important of which are non-linear ICP [6], generalized ICP [7], and non-rigid ICP [8].

The choice for one of these algorithms generally depends on several important characteristics such as accuracy, computational complexity, and convergence rate, each of which depends on the application of interest. Moreover, the characteristics of most registration algorithms heavily depend on the data used, and thus on the environment itself. To our knowledge, a general discussion of each of the above methods is not available in literature. As a result it is difficult to compare these algorithms objectively. Therefore, in this paper we discuss the mathematical foundations that are common to the most widely used 3D registration algorithms, and we compare their strengths and weaknesses in different situations.

This paper is outlined as follows: Section II briefly discusses several important application domains of 3D registration algorithms. In Section III, rigid registration is formulated as a least square optimization problem; Section IV explains the most important rigid registrations algorithms which are PCA, SVD, ICP point-to-point, ICP point-to-surface, ICP non-linear and Generalized ICP; Finally, Section V provides a discussion of the different characteristic of each of these methods in a real world setting; Section VI concludes the paper.

II. APPLICATION DOMAINS

Important application domains of both rigid and non-rigid registration methodologies are robotics, healthcare, augmented reality, and more. In these applications the common goal is to determine the position or pose of an object with respect to a given viewpoint. Whereas rigid transformations are defined by 6 DOF, non-rigid transformations allow a higher number of DOF in order to cope with non-linear or partial stretching or shrinking of the object.

A. Robotics

Since the introduction of inexpensive depth sensors such as the Microsoft Kinect camera, great progress has been made in the robotic domain towards Simultaneous Localization And Mapping (SLAM) [9], [10], [11], [12]. The reconstructed map is represented by a set of pointclouds which are aligned by means of registration and can be used for obstacle avoidance, map exploration, autonomous vehicle control, etc.[3], [13], [14]. Furthermore, depth information is often combined with a

 ISBN: 978-1-61208-356-8

traditional RGB camera [2], [15] in order to greatly facilitate real-world problems such as object detection in cluttered scenes, object tracking and object recognition [16].

B. Healthcare

Typical applications of non-rigid registration algorithms can be found in healthcare, where a soft-body model often needs to be aligned accurately with a set of 3D measurements. Applications are cancer-tissue detections, hole detection, artefact recognition, etc. [8], [17]. Similarly, non-rigid transformations are used to obtain a multi-modal representation of a scene, by combining MRI, CT, and PET volumes into a single 3D model [8].

III. Definitions

Rigid registration can be approached by defining a cost function that represents the current matching error. This cost function is then minimized using common optimization techniques. If the distance between corresponding points in each 3D pointcloud needs to be minimized, this can be simplified to a linear least-squares minimization problem by representing each point using homogeneous coordinates.

In this section, we briefly introduce the least-square optimization problem and discuss the concept of homogeneous transformations since these form the basis of 3D registration algorithms.

A. Least-Square Minimization

A rigid transformation is defined by only 6 DOF, whereas many noisy observations, i.e., point coordinates, are available. Therefore, the number of parameters of any cost function for this problem is much smaller than the number of equations, resulting in an ill-posed problem which does not have an exact solution. A well known technique to obtain an acceptable solution in such case, is to minimize the square of the residual error. This approach is called least-square optimization and is often used for fitting and regression problems.

Whereas a linear least-square problem can be solved analytically, this is often not the case for non-linear least-square optimization problems. In this case, an iterative approach can be used by iteratively exploring the search space of all possible solutions in the direction of the gradient vector of the cost function. This is illustrated by Figure 1, where the cost function $f(d)$ of the ICP registration algorithm is minimized iteratively. The cost function in this case represents the sum of the squared Euclidean distances between corresponding points of two pointcloud datasets.

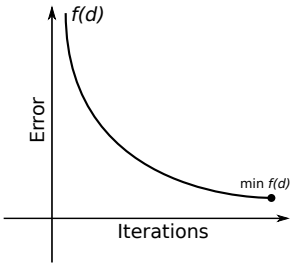


Figure 1. ICP Least square approach.

B. Homogeneous transformations

A homogeneous transformation in three dimensions is specified by a 4×4 affine transformation matrix [18]. This matrix is used to project each point in Cartesian space with respect to a specific viewpoint. In the following, let $\mathbf{v_1} = (x_1, y_1, z_1, 1)^\intercal$ be a point whose base is defined by viewpoint one and let $\mathbf{v_2} = (x_2, y_2, z_2, 1)^\intercal$ be a point whose base is defined by viewpoint two. Then it is possible to express $\mathbf{v_2}$ relative to the base of viewpoint one as $T\mathbf{v_1} = \mathbf{v_2}$, where T is an affine transformation matrix defined by (1). This is illustrated more clearly by Figure 2.

$$T = \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & \mathbf{t}_{1,4} \\ r_{2,1} & r_{2,2} & r_{2,3} & \mathbf{t}_{2,4} \\ r_{3,1} & r_{3,2} & r_{3,3} & \mathbf{t}_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{pmatrix} \tag{1}$$

The transformation matrix shown by (1) represents an affine transformation if $a_{4,1} = a_{4,2} = a_{4,3} = 0$ and $a_{4,4} \neq 0$. Affine transformations are constructed with a 3×3 rotation matrix R and column vector $\mathbf{t}$ representing a translation.

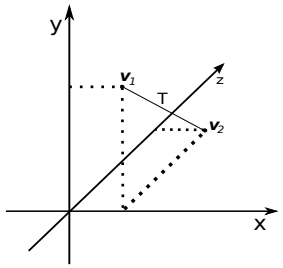


Figure 2. homogeneous transformation.

IV. Registration Algorithms

Both rigid and non-rigid registration algorithms can be further categorized into pairwise registration algorithms and multiview registration methods. Pairwise registration algorithms calculate a rigid transformation between two subsequent point clouds while the multi-view registration process takes multiple point clouds into account to correct for the accumulated drift that is introduced by pairwise registration methods.

In the next sections, we discuss five widely used rigid registration algorithms. Each of these methods tries to estimate the optimal rigid transformation that maps a source point cloud on a target point cloud. Both PCA alignment and singular value decomposition are pairwise registration methods based on the covariance matrices and the cross correlation matrix of the pointclouds, while the ICP algorithm and its variants are based on iteratively minimizing a cost function that is based on an estimate of point correspondences between the pointclouds.

A. Principal Component Analysis

PCA is often used in classification and compression techniques to project data on a new orthonormal basis in the direction of the largest variance [19]. The direction of the largest variance corresponds to the largest eigenvector of the covariance matrix of the data, whereas the magnitude of this variance is defined by the corresponding eigenvalue.

 ISBN: 978-1-61208-356-8

Therefore, if the covariance matrix of two pointclouds differs from the identity matrix, a rough registration can be obtained by simply aligning the eigenvectors of their covariance matrices. This alignment is obtained as follows;

First, the two point clouds are centered such that the origins of their original bases coincide. Pointcloud centering simply corresponds to subtracting the centroid coordinates from each of the point coordinates. The centroid of the pointcloud corresponds to the average coordinate and is thus obtained by dividing the sum of all point-coordinates by the number of points in the pointcloud.

Since registration based on PCA simply aligns the directions in which the pointclouds vary the most, the second step consists of calculating the covariance matrix of each point cloud. The covariance matrix is an orthogonal 3×3 matrix, the diagonal values of which represent the variances while the off-diagonal values represent the covariances.

Third, the eigenvectors of both covariance matrices are calculated. The largest eigenvector is a vector in the direction of the largest variance of the 3D pointcloud, and therefore represents the pointcloud's rotation. In the following, let A be the covariance matrix, let $\mathbf{v}$ be an eigenvector of this matrix, and let λ be the corresponding eigenvalue. The eigenvalues decomposition problem is then defined as:

$$A\mathbf{x} = \lambda\mathbf{x} \tag{2}$$

and further reduces to:

$$\mathbf{x}(A - \lambda I) = 0. \tag{3}$$

It is clear that (3) only has a non-zero solution if $A - \lambda I$ is singular, an consequently if its determinant equals zero:

$$\det(A - \lambda I) = 0 \tag{4}$$

The eigenvalues can simply be obtained by solving (4), whereas the corresponding eigenvectors are obtained by substituting the eigenvalues into (2).

Once the eigenvectors are known for each pointcloud, registration is achieved by aligning these vectors. In the following, let matrix T_t^y represent the transformation that would align the largest eigenvector of the target pointcloud t with the y-axis. Let matrix T_y^s represent the transformation that would align the largest eigenvector of the source pointcloud s with the y-axis. Then the final transformation matrix T_t^s that aligns the source pointcloud with the target pointcloud can be obtained easily, as illustrated by Figure 3.

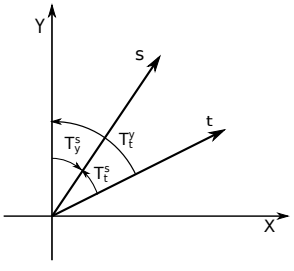


Figure 3. PCA alignment from source to target.

Finally, the centroid of the target data is added to each of the transformed coordinates to translate the aligned pointcloud, such that its center corresponds to the center of the target pointcloud.

B. Singular Value Decomposition

PCA based registration simply aligns the directions of the largest variance of each pointcloud and therefore does not minimize the Euclidean distance between corresponding points of the datasets. Consequently, this approach is very sensitive to outliers and only works well if each pointcloud is approximately normally distributed.

However, if point correspondences between the two pointclouds are available, a more robust approach would be to directly minimize the sum of the Euclidean distances between these points. This corresponds to a linear least-square problem that can be solved robustly using the SVD method [4].

Based on the point correspondences, the cross correlation matrix M between the two centered pointclouds can be calculated, after which the eigenvalue decomposition is obtained as follows:

$$M = USV^{\intercal} \tag{5}$$

The optimal solution to the least-square problem is then defined by rotation matrix R as:

$$R_t^s = UV^{\intercal} \tag{6}$$

and the translation from target pointcloud to source pointcloud is defined by:

$$\mathbf{t} = \mathbf{c_s} - R_t^s\mathbf{c_t} \tag{7}$$

C. Iterative Closest Point

Whereas the SVD algorithm directly solves the least-square problem, thereby assuming perfect data, Besl and Mc. Kay [5] introduced a method that iteratively disregards outliers in order to improve upon the previous estimate of the rotation and translation parameters. Their method is called 'ICP' and is illustrated conceptually by Figure 4.

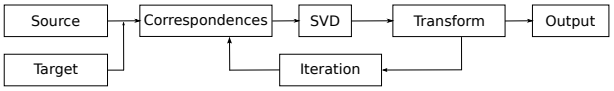


Figure 4. ICP overview scheme.

The input of the ICP algorithm consists of a source pointcloud and a target pointcloud. Point correspondences between these pointclouds are defined based on a nearest neighbor approach or a more elaborate scheme using geometrical features or color information. SVD, as explained in the previous section, is used to obtain an initial estimate of the affine transformation matrix that aligns both pointclouds. After registration, this whole process is repeated by removing outliers and redefining the point correspondences.

Two widely used ICP variants are the ICP point-to-point and the ICP point-to-surface algorithms. These approaches only differ in their definition of point correspondences and are described in more detail in the next sections.

 ISBN: 978-1-61208-356-8

1) ICP point-to-point: The ICP point-to-point algorithm was originally described in [1] and simply obtains point correspondences by searching for the nearest neighbor target point $\mathbf{q}_i$ of a point $\mathbf{p}_j$ in the source pointcloud. The nearest neighbor matching is defined in terms of the Euclidean distance metric:

$$\hat{i} = \arg\min_i \|\mathbf{p}_i - \mathbf{q}_j\|^2, \tag{8}$$

where $i \in [0, 1, ..., N]$, and N represents the number of points in the target pointcloud.

Similar to the SVD approach discussed in section IV-B, the rotation R and translation $\mathbf{t}$ parameters are estimated by minimizing the squared distance between these corresponding pairs:

$$\hat{R}, \hat{\mathbf{t}} = \arg\min_{R,\mathbf{t}} \sum_{i=1}^{N} \|(R\mathbf{p}_i + \mathbf{t}) - \mathbf{q}_i\|^2 \tag{9}$$

ICP then iteratively solves (8) and (9) to improve upon the estimates of the previous iterations. This is illustrated by Figure 5, where surface s is aligned to surface t using n ICP iterations.

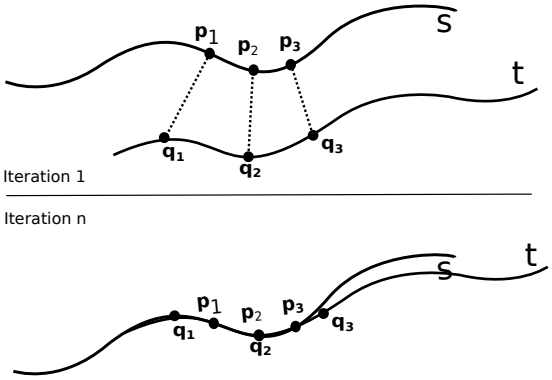


Figure 5. ICP alignment based on a point to point approach.

2) ICP point-to-surface: Due to the simplistic definition of point correspondences, the ICP point-to-point algorithm proposed by [20] is rather sensitive to outliers. Instead of directly finding the nearest neighbor to a source point $\mathbf{p}_j$ in the target pointcloud, one could take the local neighborhood of a correspondence candidate $\mathbf{q}_i$ into account to reduce the algorithm's sensitivity to noise.

The ICP point-to-surface algorithm assumes that the point clouds are locally linear, such that the local neighborhood of a point is co-planar. This local surface can then be defined by its normal vector $\mathbf{n}$, which is obtained as the smallest eigenvector of the covariance matrix of the points that surround correspondence candidate $\mathbf{q}_i$.

Instead of directly minimizing the Euclidean distance between corresponding points, we can then minimize the scalar projection of this distance onto the planar surface defined by the normal vector n:

$$\hat{R}, \hat{\mathbf{t}} = \arg\min_{\hat{R},\hat{\mathbf{t}}} \left(\sum_{i=1}^{N} \|((R\mathbf{p}_i + \mathbf{t}) - \mathbf{q}_i)\mathbf{n_i}\| \right) \tag{10}$$

This is illustrated more clearly by Figure 6.

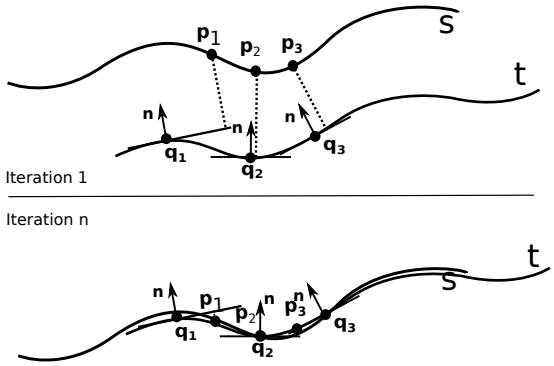


Figure 6. ICP alignment based on a point to surface approach.

3) ICP non-linear: Both the point-to-point and point-to-surface ICP approaches defined a differentiable, convex, squared cost function, resulting in a simple linear least-square optimization problem, known as a L2-optimization, that can be solved numerically using SVD. However, L2-optimization is known to be highly sensitive to outliers because the residuals are squared. An approach that solves this problem is known as L1-optimization where the sum of the absolute value of the residuals is minimized instead of the square. However, the L1 cost function is non-differentiable at the origin which makes it difficult to obtain the optimal solution.

As a compromise between L1 and L2 optimization, the so called Huber loss function can be used as shown by (11). The Huber loss function is quadratic for small values and thus behaves like an L2 problem in these cases. For large values however, the loss function becomes linear and therefore behaves like an L1 cost function. Moreover, the Huber loss function is smooth and differentiable, allowing traditional numerical optimization methods to be used to efficiently traverse the search space.

$$e^2(n) = \begin{cases} n^2/2 & \text{if } |\mathrm{n}| \leq k \\ k|n| - n^2/2 & \text{if } |\mathrm{n}| > k \end{cases} \tag{11}$$

where k is an empirically defined threshold and n is the distance measure.

The ICP non-linear algorithm uses the Huber loss function instead of a naive squared loss function to reduce the influence of outliers:

$$\hat{R}, \hat{\mathbf{t}} = \arg\min_{\hat{R},\hat{\mathbf{t}}} \sum_{i=1}^{N} e^2(n) \tag{12}$$

where

$$n = \|(R\mathbf{p} - \mathbf{t}) - \mathbf{q}\| \tag{13}$$

To obtain the optimal estimates $\hat{R}, \hat{\mathbf{t}}$ in (12), the Levenberg-Marquardt algorithm (LMA) [6] is used. The LMA method is an iterative procedure similar to the well known gradient descent and Gauss-Newton algorithms, that can quickly find a local minimum in non-linear functions.

4) Generalized ICP: A major disadvantage of the traditional point-to-point ICP algorithm, is that it assumes that the source pointcloud is taken from a known geometric surface instead of being obtained through noisy measurements. However,

due to discretization errors it is usually impossible to obtain a perfect point-to-point matching even after full convergence of the algorithm. The point-to-surface ICP algorithm relaxes this constraint by allowing point offsets along the surface, in order to cope with discretization differences. However, this approach still assumes that the source pointcloud represents a discretized sample set of a known geometric surface model since offsets along the surface are only allowed in the target pointcloud.

To solve this, Segal *et al.*[7] proposed the Generalized ICP algorithm which performs plane-to-plane matching. They introduced a probabilistic interpretation of the minimization process such that structural information from both the source pointcloud and the target pointcloud can be incorporated easily in the optimization algorithm. Moreover, they showed that the traditional point-to-point and point-to-surface ICP algorithms are merely special cases of the Generalized ICP framework.

Instead of assuming that the source pointcloud is obtained from a known geometric surface, Segal *et al.* assume that both the source pointcloud $A = \{\mathbf{a}_i\}$ and the target pointcloud $B = \{\mathbf{b}_i\}$ consist of random samples from an underlying unknown pointcloud $\hat{A} = \{\hat{\mathbf{a}}_i\}$ and $\hat{B} = \{\hat{\mathbf{b}}_i\}$. For the underlying and unknown pointclouds $\hat{A}$ and $\hat{B}$, perfect correspondences exist, whereas this is not the case for the observed pointclouds A and B, since each point $\mathbf{a}_i$ and $\mathbf{b}_i$ is assumed to be sampled from a normal distribution such that $\mathbf{a}_i \sim \mathcal{N}(\hat{\mathbf{a}}_i, C_i^A)$ and $\mathbf{b}_i \sim \mathcal{N}(\hat{\mathbf{b}}_i, C_i^B)$. The covariance matrices C_i^A and C_i^B are unknown. If both pointclouds would consist of deterministic samples from known geometric models, then both covariance matrices would be zero such that then $A = \hat{A}$ and $B = \hat{B}$.

In the following, let T be the affine transformation matrix that defines the mapping from $\hat{A}$ to $\hat{B}$ such that $\hat{\mathbf{b}}_i = T\hat{\mathbf{a}}_i$. If T would be known, we could apply this transformation to the observed source pointcloud A, and define the error to be minimized as $d_i^T = \mathbf{b}_i - T\mathbf{a}_i$. Because both $\mathbf{a}_i$ and $\mathbf{b}_i$ are assumed to be drawn from independent normal distributions, d_i^T which is a linear combination of $\mathbf{a}_i$ and $\mathbf{b}_i$, is also drawn from a normal distribution:

$$d_i^T \sim \mathcal{N}(\hat{\mathbf{b}}_i - T\hat{\mathbf{a}}_i, C_i^B + TC_i^A T^\intercal) \quad (14)$$

$$= \mathcal{N}(0, C_i^B + TC_i^A T^\intercal) \quad (15)$$

The optimal transformation matrix $\hat{T}$ is then the transformation that minimizes the negative log-likelihood of the observed errors d_i:

$$\hat{T} = \arg\min_T \sum_i \log\left(p(d_i^T)\right)$$

$$= \arg\min_T \sum_i d_i^{T\intercal}(C_i^B + TC_i^A T^\intercal)^{-1} d_i^T \quad (16)$$

Segal *et al.* showed that both point-to-point and point-to-plane ICP are specific cases of (16), only differing in their choice of covariance matrices C_i^A and C_i^B; If the source point cloud is assumed to be obtained from a known geometric surface, $C_i^A = 0$. Furthermore, if points in the target point cloud are allowed three degrees of freedom, then $C_i^B = I$. In this case, (17) reduces to:

$$\hat{T} = \arg\min_T \sum_i d_i^{T\intercal} d_i^T$$

$$= \arg\min_T \sum_i \|d_i^T\|^2, \quad (17)$$

which indeed is exactly the optimization problem that is solved by the traditional point-to-point ICP algorithm. Similarly, C_i^A and C_i^B can be chosen such that obtaining the maximum likelihood estimator corresponds to minimizing the point-to-plane or the plane-to-plane distances between both point clouds.

V. Results & Discussion

In this section, we illustrate the performance difference between a naive PCA based approach, a correspondences based SVD approach, and the ICP point-to-point registration approach. To allow for a fair comparison, we use the publicly available dataset proposed by Pomerlau *et al.* [21].

Figure 7 shows the matching error plotted against the number of iterations for the ICP point-to-point algorithm (dark-gray) without pre-alignment, and for the ICP point-to-point algorithm (light-gray) where the data has been pre-aligned using the SVD approach. In the latter case, a simple nearest-neighbor matching was used to define point correspondences, after which the SVD algorithm was used to solve the least-squares problem. This result clearly shows the importance of a rough initial alignment before applying the ICP algorithm.

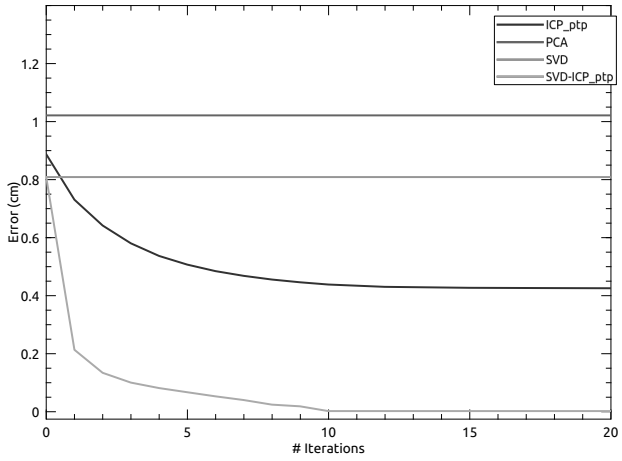


Figure 7. Comparison between PCA, SVD and general point to point ICP

Furthermore, figure 7 shows the results of a single SVD based least-squares iteration, and the results obtained using the PCA based registration approach. It is clear that the PCA based approach yields the largest matching error, due to the fact that it does not incorporate correspondence information, such that this method is highly sensitive to outliers.

On the other hand, a simple PCA or SVD based approach is extremely computational efficient, whereas the iterative ICP scheme is often too computationally expensive for real-time applications. However, Figure 7 shows that convergence can be reached quickly if a rough initial alignment is available.

Finally, it is important to note that result of the variants of ICP such as point-to-plane and plane-to-plane greatly depend on the input data. If the source pointcloud does not contain much noise, while the target pointcloud is mostly smooth and piece-wise planar, the point-to-plane algorithm outperforms

 ISBN: 978-1-61208-356-8

the traditional point-to-point method. On the other hand if the geometric structures in the scene are mostly quadratic or polynomial, the traditional ICP point-to-point algorithm yields better results. Similarly, if a lot of noise is observed in the source pointcloud, ICP plane-to-plane outperforms ICP point-to-plane.

VI. Conclusion

In this paper we provided an overview of six state-of-the-art rigid 3D registration algorithms commonly used in robotics and computer vision. We discussed the mathematical foundation that is common to each of these algorithms and showed that each of them represents different approaches to solve a common least-square optimization problem.

Furthermore, we used a publicly available dataset to compare the results of these algorithms and concluded that the results are extremely data dependent such that the choice for a specific algorithm should mainly depend on the application and input data.

References

[1] R. B. Rusu, "Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments," KI - Künstliche Intelligenz, vol. 24, no. 4, Aug. 2010, pp. 345–348.

[2] R. A. Newcombe, A. J. Davison, S. Izadi, P. Kohli, O. Hilliges, J. Shotton, D. Molyneaux, S. Hodges, D. Kim, and A. Fitzgibbon, "KinectFusion: Real-time dense surface mapping and tracking," in 2011 10th IEEE International Symposium on Mixed and Augmented Reality. IEEE, Oct. 2011, pp. 127–136.

[3] C. Kerl, J. Sturm, and D. Cremers, "Dense visual SLAM for RGB-D cameras," in 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, Nov. 2013, pp. 2100–2106.

[4] S. Marden and J. Guivant, "Improving the Performance of ICP for Real-Time Applications using an Approximate Nearest Neighbour Search," 2012, pp. 3–5.

[5] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," P. S. Schenker, Ed., Apr. 1992, pp. 586–606.

[6] S. Fantoni, U. Castellani, and A. Fusiello, "Accurate and Automatic Alignment of Range Surfaces," in 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission. IEEE, Oct. 2012, pp. 73–80.

[7] A. V. Segal, D. Haehnel, and S. Thrun, "Generalized-ICP," in Proceedings of Robotics: Science and Systems, Seattle, 2009, p. 8.

[8] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast MR images." IEEE transactions on medical imaging, vol. 18, no. 8, Aug. 1999, pp. 712–21. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10534053

[9] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard, "An evaluation of the RGB-D SLAM system," in 2012 IEEE International Conference on Robotics and Automation, vol. 3, no. c. IEEE, May 2012, pp. 1691–1696.

[10] J. Aulinas, Y. Petillot, J. Salvi, and X. Lladó, "The SLAM problem: a survey." CCIA, 2008, pp. 363—-371.

[11] P. F. I. N. D. E. Carrera, "MADRID RGB-D SLAM Author : Jorge García Bueno," no. October, 2011.

[12] K. Berger, S. Meister, R. Nair, and D. Kondermann, "A state of the art report on kinect sensor setups in computer vision," ...and Depth Imaging. Sensors ..., 2013.

[13] J. Sprickerhof and A. Nüchter, "An Explicit Loop Closing Technique for 6D SLAM." ..., 2009, pp. 1–6.

[14] A. Huang and A. Bachrach, "Visual odometry and mapping for autonomous flight using an RGB-D camera," International ..., 2011, pp. 1–16.

[15] M. Ruhnke, L. Bo, D. Fox, and W. Burgard, "Compact RGBD Surface Models Based on Sparse Coding." AAAI, 2013.

[16] S. Savarese, "3D generic object categorization, localization and pose estimation," in 2007 IEEE 11th International Conference on Computer Vision. IEEE, 2007, pp. 1–8.

[17] W. R. Crum, "Non-rigid image registration: theory and practice," British Journal of Radiology, vol. 77, no. suppl_2, Dec. 2004, pp. S140–S153.

[18] J. Kay, "Introduction to Homogeneous Transformations & Robot Kinematics," Rowan University Computer Science Department, no. January, 2005, pp. 1–25.

[19] B. Draper, W. Yambor, and J. Beveridge, "Analyzing pca-based face recognition algorithms: Eigenvector selection and distance measures," Empirical Evaluation Methods in ..., 2002, pp. 1–14.

[20] K. Low, "Linear least-squares optimization for point-to-plane icp surface registration," Tech. Rep. February, 2004.

[21] F. Pomerleau, S. Magnenat, F. Colas, M. Liu, and R. Siegwart, "Tracking a depth camera: Parameter exploration for fast ICP," in 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, Sep. 2011, pp. 3824–3829.

 ISBN: 978-1-61208-356-8

Increasing Stability of Mark Projections on Real World with Angular Velocity Sensor

Kyota Aoki and Naoki Aoyagi
Graduate school of Engineering
Utsunomiya University
Utsunomiya, JAPAN
kyota@is.utsunomiya-u.ac.jp, porutakaya2@gmail.com

Abstract— **There are varieties of methods for realizing the augmented reality. For example, there is a method to display augmented information and video information on a single display using augmented reality markers. However, this method needs the users to look at the display. In order to share same information among multiple persons, everyone should look at the displays that show the same image. Projection type augmented reality solves this problem. Projection type augmented reality is the method projecting augmented information on the surfaces in the real world using projectors. A projection type augmented reality proposes augmented information to the real world. Every person can share the experience at the environment. However, the method needs many preparation works and enormous energy for freely projecting images. It can work on only limited conditions, such as indoors or at night. Our research solves these problems. We propose a method that works in a head-worn type equipment using the method of projection type augmented reality. It recognizes an object from the camera images. It projects a mark that carries a little information onto a recognized object. There is an error of the position where the mark projected when the equipment moves. We decrease this error using an angular velocity sensor. We propose the method, the implementation and the experiments for evaluating the performance.**

Keywords—augmented reality; projection; angular velocity sensor;

I. INTRODUCTION

Augmented reality (AR) is the technology to create a "next generation, reality-based interface" and is moving from laboratories around the world into various industries and consumer markets [1]. In cooperative works, all members must carry devices that show the same information. The device may be a head-worn type or a hand-held type. However, in many cases, it is difficult to carry the devices that communicate among them. In the cases, the projective AR works well. All members may share a single device. This enables to make a very simple system comparing the system based on the personal device.

Many projective ARs work in many environments [2]-[5]. In the environments, all related devices share the proper places and have proper calibrations. This enables to make beautiful projection mappings. In cooperative working environments, we cannot control objects' surfaces. We cannot place the devices at the proper places. The relation between the projector and the surface projected changes time by time. For instance, in connecting works, the connecting terminals must have their surfaces' materials. On pure shining surfaces, we have no way to project proper images on the surfaces. However, almost all surfaces are the mixture of pure shining and pure matt. If there is a matt feature, we can project some kinds of images that our eyes can catch.

Our proposed system will not project beautiful images on a surface. The system will project a mark that catches a human attention. With the system, users may say 'We move that one.' This is the expansion of our pointing ability. This helps many kinds of cooperative work.

In many kinds of cooperative works, there is no proper place to set up the related devices about implementing a projective AR system. They are cameras and projectors. In many cooperative works, workers move in their working environments. In the case, the spatial projective type of AR system has many problems about the occlusions with the workers and related materials. A head-worn type of AR system eases this problem. Our proposed system selects a head-worn type.

Head-worn types of AR systems decrease the problems about the occlusions. However, there are problems about the delay between the image acquisition and the projection. Most of the projective AR systems presume no movement about related devices and objects. At non-projective types of AR systems, there is a little problem about the delay between the image acquisition and the image display. The created images are good with the delay. At projected types of AR systems, the delay between image acquisition and the projection influences the complex both of a projected image and a surface. The complex may be dirty to look. In some cases, the complex carries false information. For instance, in connecting works, the delay may cause the false terminal connection. In the non-projective types of AR systems, there is no chance of these errors. The displayed image may have some direction errors. However, the displayed image shows the proper terminal for connecting.

This paper proposes the stabilization of the marks projected on real objects with a head-worn type projective AR system. First, we discuss about the effects on the projected mark quality with translation and rotation of the proposed head-worn projective AR system. Then, we propose the compensation method about the rotation using the angler velocity sensor. Next, we show our implementation of the

 ISBN: 978-1-61208-356-8

proposed projective AR system. Then, we show our experiments. And last, we conclude our work.

II. TRANSLATIONS AND ROTATIONS

The motion of a solid object is a combination of translations and rotations. Fig. 1 shows the three elements of rotations. Our head-worn projective AR system must include a projector and a camera. The system must include the projector that projects the proper image on an objective surface. It must include the camera that understands the environments around the projective AR system. The camera may be a normal color camera, a depth camera or some other types of sensors that understand the environments. The projector may be a normal image projector or some types of pointers that can show the understandable images on the environments.

If there is no change about the environment and there is a well description of the environment, the projective AR system excludes the camera for understanding the environment. However, our proposed head-worn projective AR system itself the part of the environment. If there is no object that moves or changes, the proposed head-worn projective AR system must move.

The motion of a solid object is a construction of the 3-dimensional translation and 3-dimensional rotation. When the size of an interesting object is much smaller than the size of our human, it makes the projected mark to move onto another object that is a little translation in the plane that is vertical with the line directed to the object.

The proposed head-worn projective AR system works well in the environments where the sizes of interesting objects are similar to the size of a human. In the environments, translation motions of a proposed projective AR system is not larger than the scale of the interesting objects. As a result, the effect about the translation motion is not important about the combination both of the projected image and the surface projected.

Rotation affects the combination of a projected image and a surface projected. The scale of the distance between the surface and the projector is similar to the scale both of an interesting object and a human. The direction of the projector affects the position of the projected image. For easing this problem, we propose the rotation compensation method using an angler velocity sensor.

III. PROPOSED ROTATION COMPENSATION METHOD

A. *Understanding an environment and Making a projection.*

Projective AR systems need to know and understand the environment. For cooperative works, the interesting objects must be known and found. Our application offers no clear image projected. However, the places of the interesting objects are important. With the images captured by a camera, our proposed system knows the relative positions of the interesting objects to the projector.

There is some delay from taking an image to understand the position of the interesting objects. We need some processing time for finding the interesting objects in the captured image. There is a delay from deciding the position of the interesting objects to making the projection of the proper image. We can decrease the delays with the help of expensive very high speed camera, expensive processors and special designed projectors. The total cost of a high speed system easily exceeds the benefit from the usage of the proposed projective AR system.

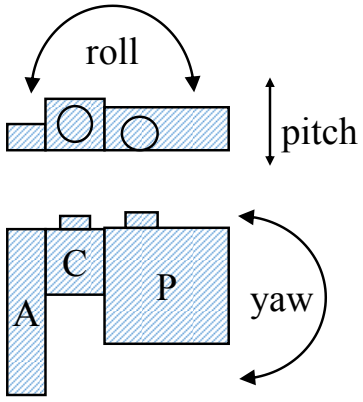


Figure 1. Rotations.

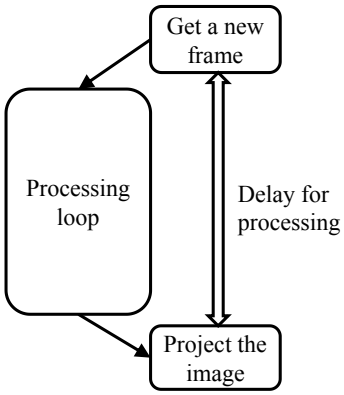


Figure 2. Delay and Processing

B. *Processing loop of our proposed projective AR system*

Without the rotation compensation, a processing loop of a normal AR system has three stages. They are capturing an image, processing the image and displaying a projection image. Using normal cameras, we can have an image at every 33 mS. We can create and display a simple mark at every 33 mS interval. However, for finding an interesting object, we need 100 mS at least in our PC. The PC has Intel Core i3 processor and 8 GB of RAM. Our object finder is based on the object finder included in the OpevCV distribution [2].

The object finder has some stages. They are a pre-processing, a feature description and a feature matching. At each processing stage, we need some processing time. We insert a processing for compensating the rotations. The size of the processing code is small. It is constructed from three

function calls. They are a call for getting an elapsed time, one getting the measurement of an angler velocity sensor and one calculating the compensation amount of the rotations. Fig. 2 shows the overall relation between the processing and the delay about projections.

C. Compensation of rotations

We have the angular velocity using the angular velocity sensor. Fig. 1 shows the experimental device setup. In Fig. 1, 'A' stands for an angular velocity sensor. 'C' does for a camera. 'P' does for a projector. The light axes both of the camera and the projector are parallel. We call the rotation around the light axis as 'roll'. Fig. 1 shows the definitions of 'pitch', 'roll' and 'yaw' in our experimental system.

Our proposed device is a head-worn type. So, the motion of the device depends on the motion of a head. We can shake a head 10 times in 1S at most. In a normal movement, we shake a head one time in 1S. The processing loop needs about 0. 3S. If we have an angular velocity at each processing loop, we have three measurements of angular velocity in a second. In normal environments, these measurements may be enough for observing the head motions. In the processing loop, the projection position error depends on the movements from the image acquisition to the projection. In a single measurement of the pair of angular velocity and time, we cannot estimate the movement between the image capturing and the projection. At least, we need to have two pairs of measurements. With more measurements, the estimation of movements increases its precision. We try to measure the pair of time and angular velocity six times in processing loop. However, the measurements of angular velocity are discrete observation. We must interpolate the observations. There is no information between two successive measurements. We estimate the amount of rotations using the trapezoid interpolation that has no assumption about the measurements.

$$\mathrm{A} = \frac{1}{2}(t_1 - t_0)(a_0 + a_1) \qquad (1)$$

In (1), A is the difference of the direction between the time t_0 and the time t_1. a_0 is the angular velocity obtained at the time t_0. a_1 is the angular velocity obtained at the time t_1.

With the movements of the projector and the surface projected, the complex of the projected image and the surface projected loses consistency. For keeping the consistency of the complex, we must compensate the movements of the projector and the surface projected. However, without the recognition of the surface projected from the captured image, we have no information about the movement of the surface projected. At least, we compensate the movement of the projector. In processing the captured image, we update the image projected with the movement of the projector that is measured with the angular velocity sensor.

With a captured image, we have the position of the object that we interest. We need to compensate the movement of the projector from the time when the image is captured to the time when the image is projected.

When we do not have the position of the object that we interest from the captured image, we must project the image with the recorded position of the object. In the case, the compensation of the movements of the projector is much important. With the measured angular velocity, we update the position of the image projected continuously.

IV. IMPLEMENTATIONS OF THE PROPOSED PROJECTIVE AR SYSTEM

A. Component hardwares and softwares.

The main components are a camera, a projector, an angular velocity sensor and a controlling computer. Fig. 3 shows the devices that construct the head-worn part. In Fig. 3, they are a projector, a camera and a Wii controller with Wii Remote Plus [7] as an angular velocity sensor, from left to right respectively. We use the camera that has a global-shutter. With motions, we cannot avoid motion blur. However, with a global shutter, we avoid the rotate distortions. The camera takes 1024x768 pixels images. In our experimental system, we use the small DLP projector for projecting a mark on the interesting object. The projector's angle of projection is 24 degrees in vertical and 38 degree in horizontal. The projected image is 1280x800 pixels. The weight is 0.8 Kg. For measuring the angular velocity, we use the Wii's controller. The controller has an acceleration sensor, an angular velocity sensor and a Bluetooth transceiver. The controlling computer is Windows PC with a Bluetooth transceiver. The processor is Intel core-i3. The PC has 8 GB memory.

We fix the camera, the projector and the sensor. The distance between the camera and the projector is smaller, the control of the projection is easier. We place the projector's lens and the camera side by side. There is no distortion on the angular velocity sensor with the distance from the projector.

Fig. 1 proposes the proposed experimental system. The experimental system needs the wired connections about the camera. It needs a power cable also. As a result, the experimental hardware is difficult to use in head-worn. The direction of the camera and the direction of the projector are same.

The main software component is an object finder included in OpenCV distribution. The object finder is 'find_obj_fern.' We add the generation of the mark projected, the angular velocity readout and the position calculation.

Figure 3. Devices constructing Head-worn part.

 ISBN: 978-1-61208-356-8

B. Projection Position.

When we have the position of the interesting object on the image captured from the camera, we can calculate the position of the mark projected on a projector's frame. We need to use projective transformation for calculating the precise position on the image projected from the position on the image captured. However, there is a little distortion between two images. We use a simple liner transformation. The (2) is the expression that calculates the horizontal position on the projected frame. The (3) is the expression that calculates the vertical position on the projected frame. In (2) and (3), (x, y) is the position on the image projected. (X, Y) is the position on the image captured. c_x, c_y, $d_x$and d_y are constants.

$$x = c_x X + d_x \quad (2)$$

$$y = c_y Y + d_y \quad (3)$$

In (2), X is the position of the interesting object in the frame captured by the camera. c_x and d_x are constants. In our experimental system, c_x is 1.5 and d_x is -700. In vertical position, c_y is 1.3 and d_y is -64.

C. Position compensation with angular velocity.

We have the position that the mark is projected. With the measured pair of an angular velocity and a time, we compensate the position for projecting the mark. We have the amount of the change of a position with (1). In the find_obj_fern, there are 5 operations in the main loop. At each operation, we measure the pair of the time and the angular velocity. After each operation, we compensate the rotations of the projector and update the projected image. Fig. 4 shows the motion compensation points in a processing loop.

The projector in our experimental system has the 38x24 degree in the projection angle. We compensate the pitch and yaw. We ignore the rotate. In a head-worn system, the rotation around the optical axis does not change the view. We do not slant our head between left and right. As a result, the rotation is small around the light axis both of the camera and the projector.

We compensate the horizontal position with dx in (4) and the vertical position dy in (5).

$$dx = f_x \times A_p/38 \quad (4)$$

$$dy = f_y \times A_y/24 \quad (5)$$

In (3), f_x is the horizontal frame size. A_p is the amount of pitch in degree. In (4), f_y is the vertical frame size. A_y is the amount of yaw in degree. In our experiment, f_x is 1280 and f_y is 800.

When we have the amount of an angular motion and the position of the interesting object on the image captured by a camera, we calculate the position in the image projected with (6) and (7).

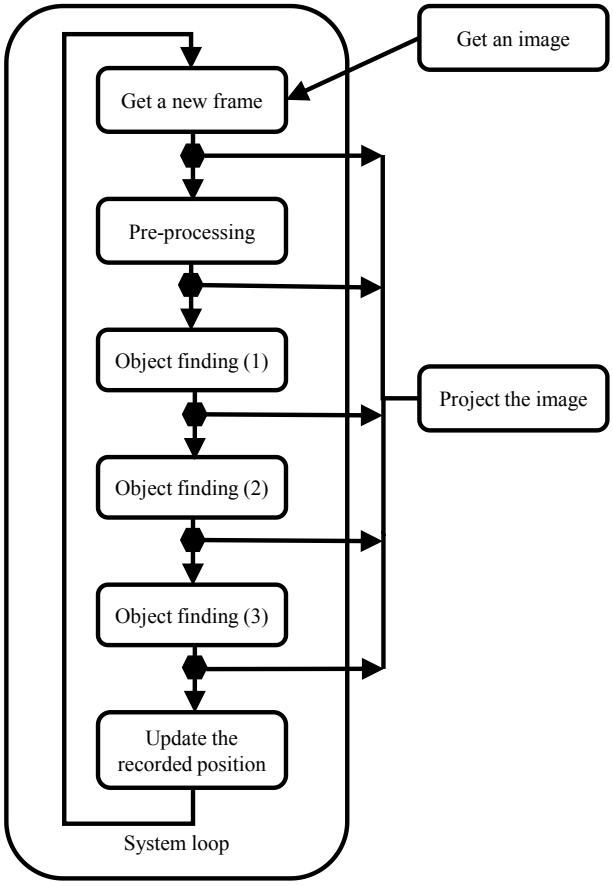


Figure 4. Motion compensation timings.

$$x_{cp} = c_x X + d_x + f_x \times A_p/38 \quad (6)$$

$$y_{cp} = c_y Y + d_y + f_y \times A_y/24 \quad (7)$$

In (6) and (7), (x_{cp}, y_{cp}) is the position on the image projected. (X, Y) is the position of the interesting object on the image captured by the camera.

V. EXPERIMENTS AND DISCUSSIONS

A. Experiment method.

We use a picture of an interesting object for teaching to our proposed projective AR system in experiments. We make three types of experiments based on the projector's motions. They are vertical motion, horizontal motion and circular motion. Vertical motion is derived from the pitch. Horizontal motion is derived from the yaw. And, circular motion is derived from the combination of vertical motion and horizontal motion in the image. The speed of the motion spans from 10 degrees/S to 120 degree/S. We analyze the recorded

 ISBN: 978-1-61208-356-8

TABLE I. EXPERIMENTAL RESULT

Compensation method (the number of measurements)	Type of Motions		
	Vertical motion	*Horizontal motion*	*Circular motion*
No compensation	41	35	36
2 times	57	58	58
6 times	58	63	66

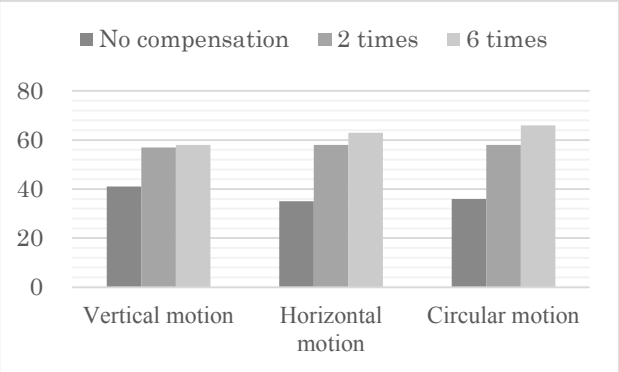


Figure 5. Performance improvements with the number of compensations in a loop.

images that representing the result of object recognitions. The recorded image shows the result of object recognition. In the recorded images, we can see the mark projected. In 100 recorded images, we check the position of the projected mark is good or not. In the case that the mark is on the object, we decide that the system keeps placing the mark. In other cases, we decide that the system cannot keep the mark on the object.

B. *Experimental results.*

Fig. 6 shows the four examples of the captured images in the circular motion experiments. In the images, at the left-upper corner, there is the interesting object. When the object is found, the rectangular is shown around the object found. The small circle is the projected mark that specifies the object. In the images (a) and (b) in Fig.6, the projected marks drop on the interesting object. In the images (c) and (d), the projected marks do not. In the image (d) in Fig. 6, the interesting object is found. In other images, the interesting object is not found. In those images, there is a motion blur.

We show the experimental results in Table I. Table I shows the number of frames that show good results in 100 recorded frames. Without the motion compensation, we have only less than 50% frames that show good results. With the motion compensation, every result shows more performance than the result without the motion compensation.

In the experiment of the vertical motion, there is a little difference between two measurements of angular velocity in a processing loop and six measurements. The horizontal experiments show more performance than the vertical experiment does. The circular motion experiments show the best performance in three types of experiments. At circular motion experiments, the amount of motions is less than the horizontal motion experiments. In 2 measurements, there is no difference between horizontal and circular experiments. In 6 measurements, horizontal and circular experiments show a difference. In circular motion experiments, the angular velocity has four peaks in one cycle motion. In horizontal and vertical motions, there are two peaks in one cycle motion. This difference affects the difference of the performances.

Fig. 5 shows the success rates expressed in percentage. In Fig.5, we easily confirm the differences between the performances of all experiments. With 6 times of angular velocity measurements, in every type of movement, the performance is improved.

VI. CONCLUSIONS

We show the motion compensation method in a projective AR system with an angular velocity sensor. We also confirm the performance of the proposed experimental system. With an angular velocity sensor, the projected mark stays much more periods at the proper place. In all cases of the experiments, there are over 50% of all frames that keep the proper marks' positions. In the experiments, the system moves in all frames. In real usages, the head-worn system has much stable time. If there are 90% stable frames, the total correct mark projections share 97% in all frames in circular movements. This enables to use the proposed stabilized projective AR system in real environments.

The experimental system needs some cables for connecting devices. In the next step, we will make a wireless experimental system.

ACKNOWLEDGMENT

This work is partially supported with the cooperative research project with Soft CDC Corporation.

REFERENCES

[1] D. W. F. van Krevelen and R. Poelman, "A Survey of Augmented Reality Technologies, Applications and Limitations," The International Journal of Virtual Reality, vol. 9, No. 2, pp.1-20, Sep. 2010, DOI: 10.1016/j.suronc.2011.07.002.

[2] M. Mine, D. Rose, B. Yang, J. v. Baar and A Grundhöfer, "Projection-Based Augmented Reality in Disney Theme Parks." IEEE Computer, vol. 45, no. 7, pp. 32-40, July. 2012, doi: 10.1109/MC.2012.154.

[3] Y. Tang, B. Lam, I. Stavness and S. Fels, "Kinect-based augmented reality projection with perspective correction." ACM SIGGRAPH 2011, p.79, Aug. 2011, doi:10.1145/2037715.2037804.

[4] S. Nicolau, L. Soler, D. Mutter and J. Marescaux, "Augmented reality in laparoscopic surgical oncology." Surgical oncology, vol. 20, no. 3, pp. 189-201, Sep. 2011, 10.1016/j.suronc.2011.07.002.

[5] H. Hua, C. Gao, L. D. Brown, N. Ahuja, and J. P. Rolland, "Using a head-mounted projective display in interactive augmented environments," Augmented Reality, 2001. Proceedings. IEEE and ACM International Symposium on, Oct, 2001, pp.217-223, doi:10.1109/ISAR.2001.970540.

[6] OpenCV, "OpenCV", http://opencv.org/, retrieved: June, 2014.

 ISBN: 978-1-61208-356-8

[7] Nintendo, http://wii.nintendo.com/, retrieved: June, 2014.

(a) (b)

(c) (d)

Figure 6. Examples of experiments.

 ISBN: 978-1-61208-356-8

Self-Initiated Intermission: A Better Way of Handling Interrupting Events

Shreejana Prajapati, Koichi Yamada and Muneyuki Unehara

Graduate School of Engineering, Nagaoka University of Technology,
Nagaoka, Niigata, Japan
Email: s105087stn@nagaokaut.ac.jp, yamada@kjs.nagaokaut.ac.jp, unehara@kjs.nagaokaut.ac.jp

***Abstract*—The aspect of handling interruption plays a vital role in human computer interaction. Interruptions become disruptive when they occur at inappropriate times. This is the reason that researchers are paying more attention towards predicting interruptions and their effects over the past few decades. We came up with an idea of finding self-initiated intermission to realize fewer distraction in human-computer interaction. Self-initiated intermission is an initiation to report oneself as being available for interaction. It gives the privilege of choosing appropriate time to handle interruptions, so that it would not hamper the current active tasks. Along with self-initiated intermission, interruptibility levels at the time of application switching and at regular intervals are studied. Then, the interruptibility levels in these three conditions are compared to find the most appropriate moment for handling interrupting events. The study shows that self-initiated intermission is the best moment for interrupting user which is less annoying than at the time of application switching and at regular intervals.**

Keywords–self-initiated intermission; interruption; regular intervals; application switching.

I. INTRODUCTION

An Interruption, as defined by O'Conaill and Frohlich [1], is "a synchronous interaction which is not initiated by the recipient, is unscheduled, and results in discontinuing of recipient's current activity". It is a prominently occurring phenomenon in human computer interaction which drives users' attention away from their regular work. Whenever interruptions occur, they constitute disruptions, i.e., disrupt work flow or make compromises on users' productivity [1][2]. In general, there are two ways of giving information to people working on a computer, i.e., either through pull or push service.

Interruptions might occur as a result of positive or negative feelings [3] of task progress and goal attainment. In most of the cases, interruption forces users to switch their task. On the other hand, users spend very little continuous time on any single task [4][5] before they move to another task. Users keep on interrupting themselves in a certain time interval even if they do not get any external interruptions. A term self-interruption [3][4] has been introduced to discuss this internally motivated interruption. In-depth study of self-interruption [3] shows that self-interruption increases at the time of negative triggers such as frustration, exhaustion, and obstruction than at the time of positive triggers, whereas Jin and Dabbish [4] give the classification on self-interruptions not only as an internally motivated interruptions, but also as a factor of environmental causes.

Regardless of numerous research on interruptions, none of the existing literature has focused on finding the effect of interruption for a short break or intermission. Though they focused on understanding and managing interruptions using external devices, interruption at self-initiated intermission has not been given enough attention. The self-initiated intermission is different from self-interruption mentioned above. Our study seeks to provide interruption management through self-initiated intermission. In this intermission, users can start communicating with a computer without hampering their current work. It gives users the privilege of handling interrupting events at their preferred time. In addition to self-initiated intermission, users' interruptibility levels at application switching [6][7] and at regular intervals [6][8] are studied for comparison.

In human-human communication, people wait for an opportune moment to start communication, but in the case of human computer interaction, such as, in Instant Messenger (IM), communication takes place instantaneously without concerning about the preferable time. In order to manage interrupting events with minimal negative impact, we propose the self-initiated intermission to interrupt users. In this paper, users' interruptibility levels at three different conditions, viz. at the time of application switching, at regular intervals, and at self-initiated intermission are compared to find the best among them. A sequence of experiments are set up where users need to perform two predefined tasks. While users are performing the tasks, they get interrupting messages in the middle of their work, depending on the conditions of interruptions. From the users' response to interruptions, it is found that users' interruptibility level changes depending on the conditions of interruptions. The results of the experiments indicate that users' interruptibility level is higher at the time of self-initiated intermission than at the time of application switching and at the regular intervals.

The rest of the paper is organized as follows. Section II presents the discussion of related work. Section III describes the idea of self-initiated intermission in detail, and shows that self-initiated intermission is different from self-interruption. Section IV, then, proceeds with presentation of experimental methods, the task being assigned, and the procedure carried out for the experiments. Section V demonstrates the results of the experiments and the statistical test. Section VI describe the discussion based on experimental results. Finally, Section VII presents our conclusion and future work.

II. RELATED WORK

Studies on interruptions were started by examining their effects in different aspects. One aspect is based on finding proper moments for the interruption. Czerwinski et al. [9] anticipates the moment of interruption at the beginning, at

 ISBN: 978-1-61208-356-8

the middle, and at the end of the task. Likewise, Miyata and Norman [10] predict that interruption after an important action, or between task execution and evaluation would be less harmful than interruption occurring at other times. But, Gillie and Broadbent [2] discovered that being able to detect one's position in the main task does not protect one from the disruptive effects of an interruption.

Some studies used machine learning techniques [11]–[13] to create predictive statistical models for interruption prediction. CRISP [14], an interruption management algorithm used machine learning algorithm to automatically model users' preference for interruption. It used a rule based algorithm to identify breakpoints, and k-nearest neighbor to find people who behave in a similar way in order to identify the situations that users might encounter while working. Based on the finding of these two algorithms, it decides appropriate time for interruption. Some focused on sensor based interruptibility prediction [8][15][16]. Both the sensor based and machine learning approaches were only concerned with the physically observable interruptible states and not with the intellectually interruptible conditions such as thinking.

Some past work focused on handling interruption that appears in terms of notification [17]–[21]. A haptic notification system [19] is designed for time management during oral presentation, which provides time alert to the presenter by generating different vibration clues. Muller et al. [20] focused on displaying information by employing an ambient timer using LED light along with vibration [21]. It gives gradually increasing cue through light, to inform the subjects that they are running out of the allocated time. Since it uses light as an alternative way for representing notification or pop-up messages, it becomes visible to the co-worker in group work. Sometimes it leads to privacy issues.

Another notification management approach is to schedule notifications at break points [17][18]. It reduces frustration and reaction time as compared to delivering them immediately. Relevance of notification content determines the granularity of breakpoint at which it should be delivered [18]. The core concept of scheduling notifications at breakpoints fitted well with how users prefer notifications to be managed [17]. Breakpoints along with the application specific knowledge are used in sensing appropriate time for notification [22]. It manipulates the running applications based on the granularities and available application specific knowledge. Then, it extracts the information, such as, expected breakpoints, forecasting the incoming breakpoints, and the target application that is most likely to get attention. There are three existing granularities of breakpoints in users' tasks, namely: coarse, medium, and fine [7][23]. And users' interruptibility level increases depending on the granularities of breakpoints [18][24]. However, the model with the breakpoints struggles to differentiate the granularities of breakpoints, and is only useful for detecting them without differentiating the granularities [18]. In spite of this problem, the breakpoints are considered to be more acceptable for delivering notification than during continuous work [6]. Tanaka and Fujita [25] propose a secretary agent to mediation interaction between users and others based on interruptibility estimation at breakpoints [6] along with the concept of avatar. The avatar continuously face towards user's computer screen and appeals for interaction by turning its face to user when breakpoint occurs, and time taken for interaction initiation is calculated [26].

III. Approach of Self-Initiated Intermission

In this section, an idea of self-initiated intermission is proposed to find an appropriate time for interruption. Self-initiated intermission is a way of managing interrupting events at self-initiated time. In this intermission, users deal with interrupting events by taking a short break from their ongoing work. Users are supposed to provide information on their leisure instead of managing the events for that time period. In other words, it is a conscious deviation from the current work. The focus here is to find a proper moment of delivering interrupting events that come in terms of the notification. Especially, we target our study with a long term goal of creating a system that handles notifications in IM and present them at the time of self-initiated intermission. Possible alternative ways of handling notifications or interruptions are: at the time of application switching, at regular intervals, at phase transition, e.g., from execution to evaluation phase and at the timing of the detail task coupling such as copy-paste.

Application switching is users' intentional switching of their working space [6], and in the case of continuous work, users keep interrupting themselves in a certain time interval because of intentionally motivated interruption. Therefore, an interruptibility comparison between the three ways of interruption management, namely: self-initiated intermission, application switching, and regular intervals is made. In our approach of finding the best moment of interruption, unlike detecting switching based on the operating history of the computer, and regardless of task being performed [6], users are provided with pre-defined task and are asked to take an initiative to report their intermission.

Intermission is the time when users' interruptibility becomes higher, which means interruptions are acceptable without any distraction at that time. A hypothesis is generated to study the conditions in which users' interruptibility becomes higher among the three ways of interruption management.

Hypothesis: *Users' interruptibility becomes higher at the time of self-initiated intermission than at the time of regular intervals and at application switching.*

IV. Experimental Method

Three experiments are designed to study users' availability for interruption. They are conducted to examine users' interruptibility at the time of 1) Application Switching (hereafter referred to as AS interruption), 2) Regular Intervals (RI interruption), and Self-Initiated Intermission (SI intermission). Users are interrupted at the time of every application switching, at every regular intervals, or on their intermission time in each experiment and are asked to report on their interruptibility levels.

A. Subjects

Twelve subjects participated in all three experiments. The subjects are familiar working with a computer and, are similar in working environment. Among them, three are females (25%) and nine are males. Their age ranges from 20-35 years. The

 ISBN: 978-1-61208-356-8

average age is 28.1 with a standard deviation of 3.3. All of the subjects are from technical field whose academic qualification ranges from undergraduate to post-graduate. One subject is from undergraduate program, 6 are from graduate program and 5 are from post-graduate program. Each subject needs to take all three experiments.

B. Task

The subjects are provided with two pre-defined tasks in each experiment. The first task is summarizing a document and the other task is solving a crossword puzzle. In document summarization, the subjects are provided with 5-6 pages of documents with approximately 2600 words. The documents is extracted from novels 'Luminance', 'One Night @ Call Center' and a book 'An Essay Concerning Human Understanding'. They are asked to summarize the given document within a page. Along with the summarization, they need to answer two questions related to the document. One of the question is simple while another is a tricky one for which the subjects need to go through the whole document. In the case of solving the crossword puzzle, some hints are given such as an initial letter, final letter and synonyms. They are asked to solve the crossword puzzle on their own. If they need help they can go through on-line dictionaries to search for words or information related to the given hints. In the experiments, document summarization force the subjects to concentrate on their task as they need to answer the questions and summarize the document. In the case of solving crossword puzzle, they need to think possible words in order to solve the puzzle, which can be thought of as a thinking task in the real environment.

Three sets of document summarization and crossword puzzle task are prepared with different content and assign one of them to the subjects randomly so that each subject would use a different set for the three experiments.

C. Procedure

Before starting, the subjects are briefed about the tasks they need to perform during the experiments. They are asked to perform the given tasks on their own pace. There is no time limitation for completing them because if they are given a time limit, they will race towards completing the tasks within the allocated time instead of concentrating on them.

The subjects' activities, both desktop and physical, are monitored constantly through a program and via video data throughout the experiment. They are informed about this prior to the experiment. The information monitored are key press events, mouse events, the current active application, and the switches between applications. When the subjects start the experiments, this program keeps running in the background and records the above mentioned information. The program generates interruption messages at the time of application switching, or at regular time intervals. The interruption messages ask the subjects to give their interruptibility levels. The interruptibility levels here indicate degrees of being available for interruption. In other words, this gives information about how much they are distracted when those messages appear. In our study, there are four levels of interruptibility, they are: Level 1 (Not at all), Level 2 (Interruptible), Level 3 (Un-interruptible), and Level 4 (Highly Un-interruptible).

During the experiments, in both AS interruption and RI interruption, if the subjects do not give their interruptibility level for thirty seconds then the interruption message disappears. The program considers the subjects' interruptibility level as "Highly Un-Interruptible". It assumes that they were busy with their work and could not provide their interruptibility level. If they "Cancel" or close the message instead of giving interruptibility level, it again considers the interruptibility level as "Highly Un-Interruptible". In the experiment for RI interruption, 12 minutes is considered as the time interval for the interruption, since subjects spend only about 12 minutes on a particular task before switching to the next task [4][5].

In the experiment for SI intermission, when the subjects start with an experiment, a notification icon will appear in the task bar. At the initial stage, the notification icon will appear in red color indicating "Un-interruptible" state, i.e., unavailable for interruption. It is made to represent "Un-interruptible" state at the beginning, so that, once the subjects turn into "interruptible state" they can change the notification icon into green by clicking it. It remains green for two minutes, and the program asks the subjects whether they are still in "Intermission". If the answer is yes, the icon remains green for another two minutes. Otherwise, the color goes back to red.

Similar to AS interruption and RI interruption, in SI intermission, if the subjects do not respond to the message for thirty seconds then it disappears. It considers the subjects' interruptibility level as "Un-Interruptible" state. In the case of SI intermission, the subjects' interruptibility levels are binary, that is, either in "Interruptible" or in "Un-interruptible" state.

Order of the three experiments as well as the tasks to be performed in the experiments are chosen randomly for every subject.

V. RESULTS

The three sequential experiments are performed to study subjects' responses on interruptions in three different conditions. Based on the obtained data, subjects' interruptibility levels are analyzed in terms of being available and not being available for interruption.

A. User's Response on Interruptibility

The results of AS interruption, RI interruption, and SI intermission are shown in Tables I-II, III-IV, and V-VI, respectively. Tables I, III, and V show the minimum, maximum, and average time spent by the subjects for completing the assigned tasks in AS, RI, and SI interruptions/intermission along with the standard deviation. Tables II and IV show the minimum, maximum, and average frequencies of interruptions and interruptibility levels, which are caused by application switching and regular intervals, respectively. Table VI shows the minimum, maximum, and average frequencies of intermission in SI intermission. The interruptibility levels in SI intermission are binary as noted before.

The frequency of application switching is higher during crossword puzzle than during document summarization because of it's complexity. Most of the subjects find difficult to solve it on their own. So, in order to find the information

 ISBN: 978-1-61208-356-8

TABLE I: Time taken to complete the tasks in the experiment of AS interruption (in minutes)

Task	Min	Max	Average	Standard Deviation
Total Task	85	176	117.9	26.9
Document Summarization	32	94	61.3	16.9
Crossword Puzzle	22	118	56.4	26.1

TABLE II: Frequency of Interruption and Interruptibility Levels in the experiment of AS interruption

Frequency	Min	Max	Average	Standard Deviation
Highly Un-interruptible	21	252	97.0	67.5
Un-interruptible	0	38	11.6	13.3
Interruptible	0	8	3.9	2.8
Not At All	0	19	4.6	5.6
Interruption	32	283	117.1	77.8

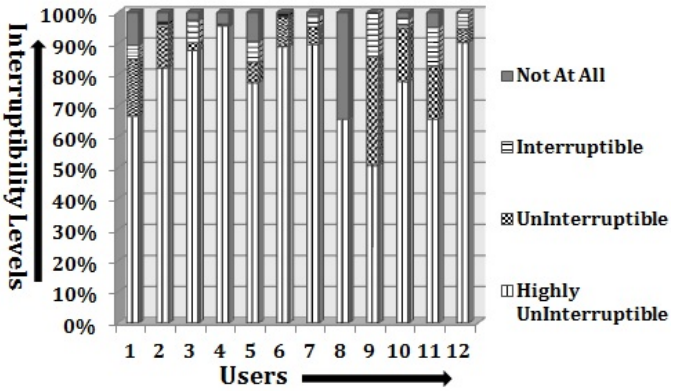


Figure 1. Subjects' responses on their Interruptibility levels in the experiment of AS interruption

related to given hints, they search the Internet for which they need to switch the applications. In the case of document summarization, they switch the applications when they look at the original document to check for the keywords while writing the summary.

In percentage wise interruptibility level, as shown in Figure 1, the subjects considered themselves being in "Highly Un-interruptible" state for almost 80% of the time, when they get interruption messages during AS interruption. Around 5 to 15% of time they considered themselves to be in "Un-interruptible" state. They considered themselves either in "Interruptible" or "Not at all" state for remaining 5% of the time. This shows that there are very few chances for the subjects to be available for interaction at the time of application switching. Every application switching does not always indicate subjects are free for interaction or available for interruption.

In Figure 2, most of the subjects find themselves either in "Highly Un-interruptible" or "Un-interruptible" state whenever they get interruption messages. In this experiment, out

TABLE III: Time taken to complete the tasks in the experiment of RI interruption (in minutes)

Task	Min	Max	Average	Standard Deviation
Total Task	52	144	97.9	29.7
Document Summarization	28	85	56.2	17.4
Crossword Puzzle	23	74	41.7	17.2

TABLE IV: Frequency Of interruption and interruptibility levels in the experiment of RI interruption

Frequency	Min	Max	Average	Standard Deviation
Highly Un-interruptible	2	10	6.5	3.0
Un-interruptible	0	4	1.8	1.1
Interruptible	0	6	2.0	2.2
Not At All	0	3	1.3	1.5
Interruption	4	12	8.4	2.8

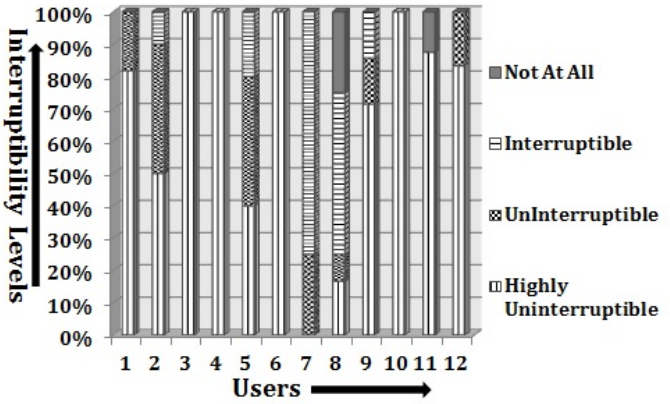


Figure 2. Subjects' responses on their Interruptibility levels in the experiment of RI interruption.

of twelve subjects, four report themselves as "Highly Un-interruptible" all of the time throughout the experiment, six report either "Highly Un-interruptible" or "Un-interruptible" for almost 90% of the time. The rest of two report either "Interruptible" or "Not at all" for almost 80% of the time.

In the case of SI intermission, the time taken for completing the assigned task and the frequency of self-initiated intermission are shown in Tables V and VI, respectively. The subjects report on their intermission maximum 12 times and the minimum 4 times (Figure 3). In self-initiated intermission, whenever subjects take intermission, their interruptibility level is "Interruptible". In other words, reporting on intermission is nothing but reporting about their interruptible state.

Results of these three experiments demonstrate that interruptibility levels change based on the conditions of interruptions as shown by the graphical data in Figures 1, 2, and 3. Also, it is found that subjects take initiations to report on their intermission at least 4 times within the task completion period.

 ISBN: 978-1-61208-356-8

TABLE V: Time taken to complete the tasks in the experiment of SI Intermission (in minutes)

Task	Min	Max	Average	Standard Deviation
Total Task	74	174	115.7	26.9
Document Summarization	58	91	66.7	9.8
Crossword Puzzle	11	83	49.1	20.4

TABLE VI: Frequency Of intermission in the experiment of SI Intermission

Frequency	Min	Max	Average	Standard Deviation
Interruptible	4	12	7.3	2.5
Un-Interruptible	0	0	0.0	0.0
Intermission	4	12	7.3	2.5

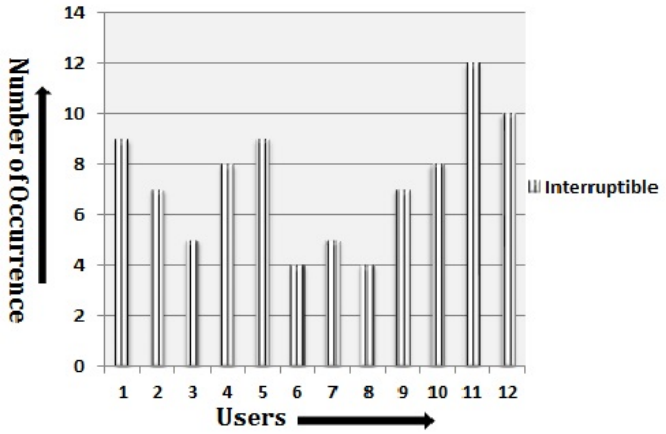


Figure 3. Subjects' responses on their Interruptibility levels in SI Intermission and also shows the frequency of intermission.

B. Evaluation based on Interruptible state

In the experiment of SI intermission, there are two levels of interruptibility, i.e., either "Interruptible" or "Un-interruptible". When the subjects report on their intermission, they are always in "Interruptible" state. On the other hand, when they are not in intermission they always seem to be in "Un-interruptible" state. Here in this section, the subjects' interruptibility level is analyzed statistically.

In order to compare the three ways of interruptions, viz. AS interruption, RI interruption, and SI intermission, repeated paired wise t-test is performed. The available four levels of interruptibility in the experiments of AS and RI interruption is transformed into binary to compare with the results of SI intermission experiment. The two interruptibility levels, level 4 (Highly Un-interruptible) and 3 (Un-interruptible) are transformed to Un-interruptible. The next two interruptibility levels, level 2 (Interruptible) and 1 (Not at all) are transformed into Interruptible. The value '0' is assigned for Interruptible and '1' for Un-interruptible, respectively. Then, the data of 12 subjects is mixed into a data set. Three data sets of the 12 subjects are obtained with 1405, 101, and 88 data, resulted from AS interruption, RI interruption, and SI intermission, respectively (shown in Table VII).

TABLE VII: Data sets from three different experiments

Experiments	Data	Average	Standard Deviation
AS Interruption	1405	0.93	0.26
RI Interruption	101	0.84	0.37
SI Intermission	88	0.00	0.00

The average values suggest that SI intermission is the most interruptible, whereas RI interruption is the 2nd most interruptible and AS is the 3rd. Since there is unequal variance, Welch's t-test is performed for every pair of the data sets. The results of one-sided t-tests also show that SI Intermission is more interruptible than AS and RI interruptions with significance level of 1%. The p-value is almost zero between the results of SI intermission and AS interruption, and 4.23×10^{-42} between the results of SI intermission and RI interruption. This supports the Hypothesis in Section III. As for the comparison between AS interruption and RI interruption, significant difference at 5% can be seen with p-value of the one-sided test being 0.0114, which concludes RI interruption is more interruptible than AS interruption.

VI. Discussion

In this study, the subjects' responses on their interruptibility levels are explored to find the best condition of interruption. From the experimental data and the statistical test, it is verified that SI intermission is the best moment for interruption management. In the interruptibility comparison, it is found that degree of availability for interruption is higher at AS interruption, at RI interruption, and at SI intermission in increasing order. This result contradicts with the suggestion from past study [6], which says application switching is a good approach of getting subjects' attention. In their study, regardless of the nature of the task they only considered computer operation records to check subjects' interruptibility during application switching. This might be one reason for getting application switching acceptable for the interruption. Whereas, in our case, the task being assigned require more concentration. They made an interruptibility comparison at the time of application switching and at the middle of continuous work. The interval of interruption at continuous work was very short (5 minutes), that might be another reason for getting application switching more acceptable for interruption than continuous work (regular intervals). In our study, the time of interruption for RI interruption is increased, which was 12 minutes. As studies [4][5] suggested that even if people do not get external interruptions, they spend only about 12 minutes of time before moving to another task because of the self-interruption. This was the reason to increase the time of interruption in RI to 12 minutes. Tanaka and Fujita [6] also generate interruptibility estimation rules based on the operating history of the computer. They made the estimation that physical activity such as keystroke and mouse operation reflect 'Interruptible state' and close coupling of tasks such as copy and paste reflect 'Un-interruptible state' but the accuracy was only around 50%.

 ISBN: 978-1-61208-356-8

There was a post questionnaire session to examine subjects' reactions towards the appearance of interruption messages at different conditions of interruptions. After the experiments, the subjects described that it was annoying to get messages at every application switching and at regular intervals. Similar to the fact mentioned in past studies [27]–[29], out of twelve subjects, ten reported interruption messages as annoying and distracting while two expressed reporting on their intermission was annoying. This demonstrates that when subjects are fully concentrated on their assigned task, they prefer pull service and when they are doing their work just for the sake of completing, they prefer push service, i.e., interruption.

Subjects found it more convenient when they could handle the incoming interruption messages at their preferable time. Among the four methods of coordinating interruptions [30]: Immediate, Mediated, Scheduled, and Negotiated, the approach of handling interrupting events at SI intermission supports the negotiated method of coordinating interruptions in human computer interaction.

VII. Conclusion and Future Work

In this paper, a new approach of self-initiated intermission was presented in order to handle the interruption related events or notification at the appropriate time without distracting users from their current work. Self-initiated intermission is a conscious diversion from current work. Unlike other interruptions, it leaves very less negative impact. Through the experimental data, it is proved that self-initiated intermission is the best way to handle interrupting events without creating annoyance and anxiety to users. This study is limited to the verification of our anticipation on how well self-initiated intermission works. We plan our future work on implementing self-initiated intermission for minimizing interruptions by presenting notification in an appropriate moment in an instant messaging system. We also plan to develop a system with self-initiated intermission to get users' attentions for interaction.

References

[1] B. O' Conaill and D. Frohlich, "Timespace in the Workplace: Dealing with Interruptions," in Proc. CHI 95 Conference Companion on Human Factors in Computing Systems, 1995, pp. 262–263, ISBN: 0-89791-755-3.

[2] T. Gillie and D. Broadbent, "What Makes Interruptions Disruptive? A Study of Length, Similarity, and Complexity," Psychological Research, 1989, pp. 243–250, ISSN: 0340-0727, 1430-2772.

[3] R. F. Adler and R. Benbunan-Fich, "Self-interruption in Discretionary Multitasking," Computer in Human Behavior, vol. 29, 2013, pp. 1441–1449.

[4] J. Jin and L. A. Dabbish, "Self-interruption on the Computer: A Typology of Discretionary Task Interleaving," in Proc. of the SIGCHI Conference on Human Factors in Computing Systems, 2009, pp. 1799–1808, ISBN: 978-1-60558-246-7.

[5] V. M. Gonzlez and G. Mark, "Constant, Constant, Multitasking Craziness: Managing Multiple Working Spheres," in Proc. of the SIGCHI Conference on Human Factors in Computing Systems, 2004, pp. 113–120, ISBN: 1-58113-702-8.

[6] T. Tanaka and K. Fujita, "Study of User Interruptibility Estimation Based on Focused Application Switching," in Proc. of the ACM conference on computer, 2011, pp. 721–724, ISBN: 978-1-4503-0556-3.

[7] P. D. Adamczyk and B. P. Bailey, "If Not Now When? The Effect of Interruptions of Different Moments Within Task Execution," in Proc. of the SIGCHI conference on Human factors in computing systems, 2004, pp. 271–278, ISBN: 1-58113-702-8.

[8] M. Nilsson, M. Drugge, U. Liljedahl, K. Synnes, and P. Parnes, "A Study on Users'Preference on Interruption When Using Wearable Computers and Head Mounted Displays," in In Communications IEEE PerCom'05, 2005, pp. 149–158, ISBN:0-7695-2299-8, doi:10.1109/PERCOM.2005.5.

[9] M. Czerwinski, E. Cutrell, and E. Horvitz, "Instant Messaging and Interruption:Effects of Relevance and Timing," in People and Computers XIV: Proceedings of HCI 2000, British Computer Society, 2, 2000, pp. 71–76.

[10] Y. Miyata and D. A. Norman, "Psychological Issues in Support of Multiple Activities," In: D. A. Norman and S. W. Draper (Eds.), User Centered Systems Design: New Perspectives on Human-Computer Interaction, Hillsdale: Lawrence Erlbaum Associatesn, 1986, pp. 265–284.

[11] J. Fogarty, S. E. Hudson, and J. Lai, "Examining the Robustness of Sensor-Based Statistical Models of Human Interruptibility," in Proc. of the SIGCHI Conference on Human Factors in Computing Systems, 2004, pp. 207–214, ISBN:1-58113-702-8.

[12] S. E. Hudson, J. Fogarty, C. G. Atkeson, D. Avrahami, J. Forlizzi, S. Kiesler, J. C. Lee, and J. Yang, "Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study," in Proc. of the SIGCHI Conference on Human Factors in Computing Systems, 2003, pp. 257–264, ISBN:1-58113-630-7.

[13] E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelset, "The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users," in Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, Madison, WI, 1998, pp. 256–265, ISBN:1-55860-555-X.

[14] T. Shrot, A. Rosenfeld, and J. Golbeck, "CRISP: An Interruption Management Algorithm Based on Collaborative Filtering," in Proc. of the SIGCHI Conference on Human Factors in Computing Systems, 2014, pp. 3035–3044, ISBN: 978-1-4503-2473-1, doi:10.1145/2556288.2557109.

[15] P. Vorburger, A. Bernstein, and A. Zurfluh, "Interruptability Prediction Using Motion Detection," 2011, international Workshop on Managing Context Information in Mobile and Pervasive Environments MCMP-05, doi:10.1.1.87.380.

[16] M. Haller, C. Richter, P. Brandl, S. Gross, G. Schossleitner, A. Schrempf, H. Nii, M. Sugimoto, and M. Inami, "Finding the Right Way for Interrupting People Improving their Sitting Posture," in Human-Computer Interaction-INTERACT, 2011, pp. 1–17.

[17] S. Iqbql and B. P. Bailey, "Oasis: A Framework for Linking Notification Delivery to the Perceptual Structure of Goal Directed Tasks," ACM Transactions on Computer-Human Interaction, vol. 17, 2010, doi:10.1145/1879831.1879833.

[18] S. Iqbal and B. P. Bailey, "Effects of intelligent notification management on users and their tasks," in Proceedings of the ACM Conference on Human Factors in Computing Systems, 2008, pp. 93–102, ISBN: 978-1-60558-011-1.

[19] D. Tam, K. MacLean, J. McGrenere, and K. Kuchenbecker, "The Design and Field Observation of a Haptic Notification System for Timing Awareness During Oral Presentation," in Proc. CHI'13 ACM, 2013, pp. 1689–1698, ISBN: 978-1-4503-1899-0.

[20] H. Muller, M. Kazakova, A. Pielot, W. Heuten, and S. Boll, "Ambient Timer Unobtrusively Reminding Users of Upcoming Tasks with Ambient Light," in Pro. of Human Computer Interaction- INTERACT, 2013, pp. 2111–228.

[21] H. Muller, M. Pielot, and R. Oliveira, "Towards Ambient Notification," 2013, in Pheripheral Interaction: Embedding HCI into Everyday Life-Workshop at INTERACT.

[22] T. Okoshi, H. Tokuda, and J. Nakazawa, "Application as a sensor (AaaS) Approach for User Attention Sensing," in Proc. of The 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2013, URL: http://www.ubicomp.org/ubicomp2013/dc/okoshi-jr-ds-crc.pdf.

[23] S. Iqbql and B. P. Bailey, "Understanding and Developing Models for Detecting and Differentiating Breakpoints During Interactive Tasks," in Proc. CHI 2007, 2007, pp. 697–706, ISBN:978-1-59593-593-9.

[24] S. T. Iqbal and B. P. Bailey, "Leveraging Characteristics of Task Structure to Predict the Cost of Interruption," in Proceedings of the

 ISBN: 978-1-61208-356-8

SIGCHI Conference on Human Factors in Computing Systems, 2006, pp. 741–750, ISBN:1-59593-372-7.

[25] T. Tanaka and K. Fujita, "Interaction Mediate Agent Based on User Interruptibility Estimation," Human Interface and the Management of Information. Interacting with Information, vol. 6771, 2011, pp. 152–160.

[26] K. Fujita and T. Tanaka, "Secretary Agent for Mediating Interaction Initiation," in Proc. of Human Agent Interaction, 2013, URL: http://hai-conference.net/ihai2013/proceedings/pdf/II-2-p5.pdf.

[27] D. C. McFarlane and K. A. Latorella, "The Scope and Importance of Human Interruption in Human Computer Interaction Design," Human Computer Interaction, vol. 17, 2002, pp. 1–61.

[28] D. C. McFarlane, "Interruption of People in Human-Computer Interaction: A General Unifying Definition of Human Interruption and Taxonomy," 1997, naval Research Laboratory (NRL/FR/5510979870),Washington,DC.

[29] G. Mark, V. M. Gonzalez, and J. Harris, "No Task Left Behind, Examining the Nature of Fragmented Work," in Proc. of the SIGCHI Conference on Human Factors in Computing Systems, 2005, pp. 321–330, ISBN:1-58113-998-5.

[30] D. C. McFarlane, "Comparison of Four Primary Methods for Coordinating the Interruption of People in Human-computer Interaction," Human Computer Interaction, vol. 17, 2002, pp. 63–139.

 ISBN: 978-1-61208-356-8

Enhancing Collaborative Learning and Management Tasks through Pervasive Computing Technologies

Christophoros Karachristos
Educational Content Methodology and Technology Laboratory, Hellenic Open University, Patras, Greece
karachrist@eap.gr

Christos Goumopoulos
Educational Content Methodology and Technology Laboratory, Hellenic Open University & Information and Communication Systems Engineering Department, Aegean University, Hellas
goumop@{tutors.eap.gr; aegean.gr}

Achilles Kameas
Educational Content Methodology and Technology Laboratory, Hellenic Open University, Patras, Greece
kameas@eap.gr

***Abstract*— We are examining in this paper how ubiquitous technology enhanced classrooms can foster opportunities for enhancing teaching and learning. Our concept of Smart Classroom is shaped upon an ambient intelligent environment which supports three major objectives of the educational process: assisting course creation and presentation, classroom management and student assessment and collaboration. Distance learning poses additional requirements on smart classrooms since both local and remote students should have an equal educational experience. This paper describes a Smart Classroom prototype, which combines a number of pervasive computing technologies such as RFId, Microsoft Kinect, magnetic cards and Android applications. The main contribution of our work is combining such different technologies to support classroom attendance management, lecture presentation handling through physical interaction and student collaboration through android applications. Finally, this work includes a survey on related systems and their contribution on strengthening the educational procedure.**

Keywords- Smart Classroom; collaborative learning; pervasive computing; physical interaction.

I. INTRODUCTION

Traditional classrooms have a simple structure as the instructor is the only responsible person for planning, management and guidance of all activities in the classroom as well as for the presentation of the educational content. On the other hand, smart classrooms (SC) provide a complex structure that allows innovative ways of learning. A SC is a physical classroom in which information and communications technologies have been integrated. The technological infrastructure of a typical SC includes equipment for multimedia content rendering, projectors, large screens for video projection, computers, smart boards and desks, data centers, etc. This infrastructure facilitates students to watch a lecture without concerning on keeping detailed notes, since after the lecture they can have access to the teaching material including notes and annotations. In such an environment students can interact with the material, the instructor and the other students in new ways.

The deployment of a SC has many advantages such as enabling active learning (students become active inventors of knowledge rather than passive recipients), enhancing content and presentation, harnessing the wealth of the Internet, bringing in experts with videoconferencing, allowing lecture and notes capturing and enhancing distance learning (instructing remote and local students). As an example, students can use SC resources (e.g., interactive desks that are connected to the smart board) and under the supervision of their instructor can work educational activities in teams as part of a course. Afterwards, each team can present their solutions, compare them, discuss under the guidance of the instructor, and as a consequence create joint knowledge (complementing one another).

Hellenic Open University (HOU) has a mission to offer university level education using distance learning methodology and to develop the appropriate material and teaching methods. Currently, HOU offers 31 undergraduate and postgraduate Study Programmes with a total of approximately 30,000 students, coached by 1,700 tutors in 1,550 groups (20 students per group on average). Students of the HOU usually live in disparate locations all over the country. Besides being students they usually have families and working obligations so they have pressing time constraints for studying. Once per month students of a particular group meet face-to-face with their tutor.

In this paper, we describe a SC prototype developed in the HOU. The prototype enhances a typical SC with pervasive computing technology components in order to accelerate management services and to support an enhanced learning experience by providing services that enhance student-tutor interaction. A critical objective of the HOUs SC is to decrease the gap between a traditional classroom and a virtual classroom regarding the educational experience acquired by students. A basic approach to achieve this is to move the user interface from the static computer environment of a typical virtual class software system to the 3D space of the SC so that the instructor can interact with the remote students with multiple natural ways, in equal terms as with the local students.

The remaining of the paper is organized as follows. Section II gives an overview of the SC prototype and the enabling platforms used for developing a number of services. Section III describes how user authentication and status update is performed in the context of the SC using a combination of pervasive computing components. The management of the educational content presentation through gesturing movements using the Kinect technology

 ISBN: 978-1-61208-356-8

is presented in Section IV. The importance of enhancing collaborative learning through suitable collaboration tools is discussed in Section V. The results of a preliminary evaluation of the SC prototype are discussed in Section VI. Related work is presented in Section VII followed by our conclusions in the final section.

II. SYSTEM OVERVIEW

The SC model that was developed incorporates computational intelligence in such a way to be invisible to the audience and its final goal is the improvement of learning experience both for the local students and the remote students. It constitutes a mixture of a classroom with traditional characteristics but as well as a web-based learning system which provides both synchronous and asynchronous communications services. The classroom that is used has a total coverage of 66 m2. It includes bearings for the students, a Student Board and a Course Board connected with a projector. The SC incorporates audiovisual content recording and processing system (server, sound processing and video processing software), digital content distribution / streaming, a sensor package to measure environmental parameters (sound, temperature, light, etc.), a location tracking system based on the commercial Ubisense system, Radio Frequency Identification-based (RFId) applications, biometric devices, a voice recognition and text rendering system and an eyetracking system for building affect sensitive intelligent tutoring systems. A basic setup of the SC is shown in Fig. 1.

When possible, open source technologies are used for the enhancement of the learning process considering that the system must be inexpensive so that it can be easily adopted by the educational institutions. On the other hand, commercially available software is necessary to cover the needs that the open source software cannot. A characteristic example is the Saba Centra Teleconference software, which the HOU employs for synchronous communication in the context of virtual classrooms.

The features of the SC that were developed apply to different stages of a lecture. The first stage refers to the authentication of the user either the local one or the remote one. For the authentication task there are various available technologies that can be used such as the RFId technology in combination with single board computers and magnetic cards technology which are used by students attending a face-to-face meeting with the instructor in the SC. Remote students authentication is accomplished through JSP pages. Regarding the educational process stage, the aim is to be enhanced with various technological means. Thus, the students will have in their disposal tablets with android operating as an enabling platform. Alternatively, students can bring their personal laptops. There are a lot of reasons why tablets are used in the educational process. Some of them are concerning their low cost, their storage capacity and their ease of use. Finally, a primary factor is that these devices have peer-to-peer connection capabilities but also broad network capabilities through various communication protocols such as Wi-Fi, Bluetooth or NFC. Thereby, in the context of the SC the learners are given the chance to have access to the educational content as well as to operate cooperatively in groups. In the educational process stage, another parameter that is taken into account is the enhancement of the students' participation in the process. Students are no longer passive viewers but they are able to be active, to ask questions and to handle the educational content. Via the Android applications they can intervene to the discussion and handle the content, which is projected to the Course Board remotely. Hence, the student from her place can be connected to a central computer and can use the input/output capabilities of her tablet to interact with material of an educational task collaboratively with other students. Finally, in this prototype the instructor can move around in the SC and interact with educational material in a natural manner without being static in front of a computer. For example, she is able to manage her lecture and navigate the presentation slides projected on the board via gesturing movements. This can be accomplished by using Microsoft Kinect technology which can recognize the tutors' movements through infrared rays.

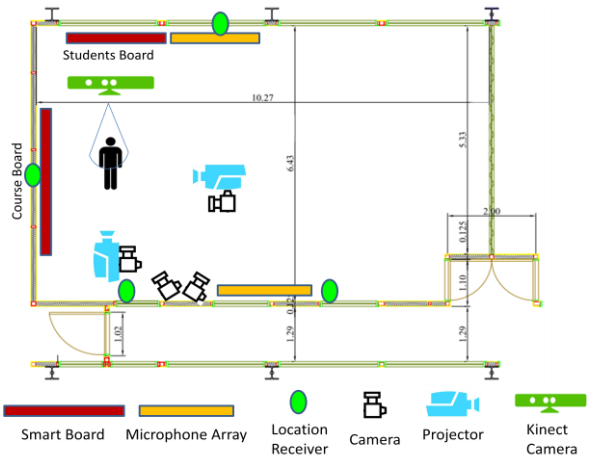


Fig. 1. Smart Classroom setup.

In this prototype, we aim to use devices and objects which have embedded computational capabilities that can be accessible by the student in a non-intrusive manner. This prototype continues to develop with new applications and services which enhance even more the concept of ubiquitous computing and improve the educational process both for the local and remote students. Each one service that was developed is analyzed below.

III. UBIQUITOUS USER AUTHENTICATION AND STATUS UPDATE

One of the offered services in the SC prototype is the user authentication module serving both the local and remote students and the recording of their status (e.g., present, absent) for every lecture/meeting taking place. In this particular implementation a combination of technologies are used, such as microcontrollers (RFId), Web development and magnetic cards. This combination came out through the analysis of the users' requirements as different needs have emerged by the different user categories of the system. Thus, the developed service constitutes a combination of three modules. Two modules are responsible for the physical authentication of the user and the recording of her status. In this case, both RFId sensors and magnetic card technology are used (Fig. 2, 3).

The RFId represents a major change in how applications handle object identification and tracking. RFId

 ISBN: 978-1-61208-356-8

uses radio waves and allows the automatic identification of the data which are transferred into smart RFId "tags". The RFId tags can be automatically monitored by stationary or mobile devices without the need of being serially scanned.

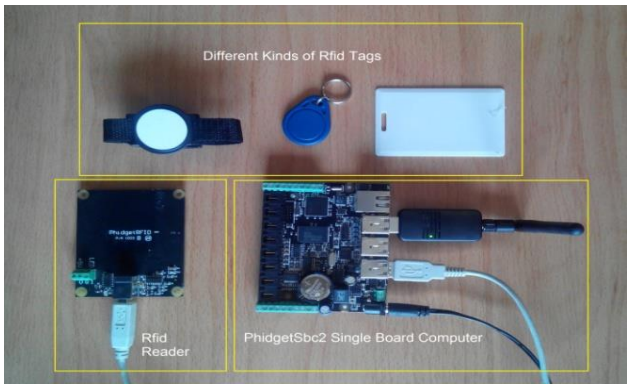


Fig. 2. Phidget SBC – RFId Technology.

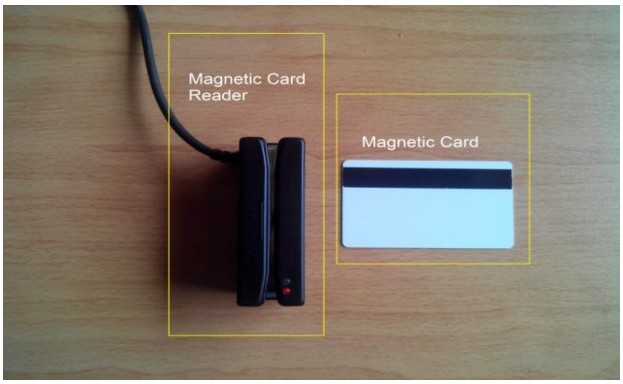


Fig. 3. Magnetic Card Technology.

Using this way of gathering information the university can dramatically reduce the time and the cost of management tasks such as the management of the status of its human resources. For instance, the instructor is not anymore required to inform the status of the students who attended a face-to-face meeting (this is a mandatory task according to HOUs management procedures) as this is achieved in an automatic manner.

The user needs to pass the RFId tag through the RFId reader which is placed in the classroom's entrance. Alternatively, the user can carry a magnetic card. A magnetic card is provided to every student of the HOU integrating in its magnetic tracks information that uniquely identifies each one of them. This card is also used in a number of other activities such as using library services or joining university's social network. If the user is a member of the group that has assigned a meeting at that time in the classroom she will be granted entrance in the classroom. Simultaneously, the user's attendance status will be updated. If the user is not a member of that group he/she will be informed with an appropriate message on the screen.

Additionally, in the context of the authentication process a web interface was developed. This interface gives the user the capability to be connected and authenticated remotely. The remote student provides her personal details (e.g., email and id number) via a web interface and is authenticated by the system. When the user is connected she can watch information regarding the lecture that is taking place (e.g., location, opening time, number of participants) as well as information concerning her status both to this particular lecture and those preceding. The authentication does not mean the automatic participation in the meeting in which she has been connected. This procedure is accomplished through the Centra platform, a teleconference service (implementing virtual classes) which provides both the instructors and students the capability of attaining meetings and seminars regardless the place of their physical presence.

Status update refers to the recording of the users' status with respect to the authentication time and mode. The potential user states are the following: *present, late, in distance* present, *in distance late*, *absent.* User's status recording allows continuous statistical analysis with respect to the HOU's students' attendance profile with an ultimate aim to provide improvements and/or adjustments of the offered services towards them.

A. RFId-based Authentication

For the development of this module PhidgetSBC2 Single Board Computers are used serving the execution of independent and autonomous applications without the need of a host computer on which they must be connected. A PhidgetSBC2 node is running a light version of a Debian GNU/ Linux operating system. Development services like the creation of a new application running on the board or its network management features are managed through a Web interface. Its advanced features can be accessed through an embedded SSH server. The identification process is performed through PhidgetSBC2 wireless connection with the Central Server on which the system database is running.

Those boards are placed in the entrance hall of every classroom. The PhidgetRFId Reader reads the RFId tags which are attached to the reader instantly and returns the unique authentication number that identifies each tag. This requires that each participant in the meeting has her own tag which is provided to her during the registration process. The tag's unique authentication number is recorded in the specific field in the corresponding database table. The software, which runs on PhidgetSBC2 board, is responsible for identifying the tag, establishing a connection to the database if the user is authenticated successfully and updating the status of the tag holder with the proper indication. Finally, it is responsible for the system's output to the user giving a short beep after successful identification from the user or a longer beep with a different pitch after unsuccessful identification.

B. Web-based Authentication

For the development of the user interface for the remote students' authentication the JSP technology (JavaServer Pages) is used. This technology helps software developers to create dynamically generated web pages based on HTML, XML, or other document types. The core of the developed application uses also the Java programming language. Specifically two JavaBeans components have been created (Fig. 4). The first is responsible for the storage of data variables entered by the student (username, email) through the interface and the second is responsible for establishing a connection with the database and performing SQL queries. After the successful

 ISBN: 978-1-61208-356-8

user validation the same procedure as in the RFId module is applied and the status update of the user is performed. Then, the student who successfully entered the credentials can retrieve information about the lecture attending and other relevant information. Finally, the application enables the student either to disconnect or to connect remotely to the lecture which is carried through the teleconference system.

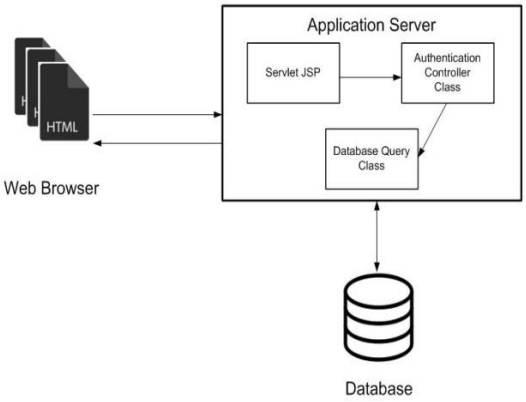


Fig. 4. Web Authentication Architecture

C. Magnetic Card-Based Authentication

In this module, the JSP technology is used in conjunction with the magnetic cards technology. The Web interface developed in the previous module is modified so that instead of username and email it receives a string which is a unique identifier for each student. Specifically this identifier consists of three parts. The first part is a 66 characters length secret key that uniquely identifies each user. The second part is the user's password. Finally, the third part is the user's name in ASCII encoding. Each part has a starting character and an ending character. In the first part the starting character is "%", in the second part is the character "#" and in the third part is the "+" character. In all three parts the terminating character is ";". The string is stored in the three tracks of the magnetic card. So when the magnetic card is passed through the card reader, authentication is performed by a series of SQL queries to the database including the students' status update. Simultaneously, on a screen, which is positioned at the entrance of the SC, the user can also retrieve information about the lecture to attend. After a few seconds the display returns to the original home screen for next student authentication. The software architectural design follows the pattern of the previous module making use of Java beans in the same way. Here the major difference is the credentials that must be managed by the application. Another difference with the previous module is the number of users that it handles. This module must give the ability to a number of users to enter their credentials in a short period of time. So it is designed in a way that every single user can be validated and disconnected after a while so the next user can be authenticated.

IV. GESTURE-BASED PHYSICAL INTERACTION

A second service that was developed in the context of the SC prototype is the management of the educational content presentation through gesturing movements using the Kinect technology. Microsoft Kinect [10] is a motion sensing device that enables physical interaction with the computer. Gesturing movements replace the standard human-computer interaction which is achieved with the mouse and the keyboard. The innovative technology behind Kinect is a combination of hardware and software contained within the Kinect sensor accessory that can be added to any computer device running Windows, Linux or Mac OS. Kinect cameras use three hardware innovations working together. The first one is a Color VGA video camera, which is called RGB camera, referring to the color components it detects. The second one is a depth sensor. It's a combination of an infrared projector and a monochrome complimentary metal-oxide semiconductor (CMOS) sensor working together to "see" the room in 3-D regardless of the lighting conditions. Finally, there is a multi-array microphone which is an array of four microphones that can isolate the voices of the users from the noise in the room. This allows the users to be a few feet away from the microphone and still use voice controls. This camera was positioned to the area where the instructor moves between the smartboard and the bearings of the students and within the distance of 5 feet from him (Fig. 1). The reason for this distance is the optimal performance of the camera, which ranges from 4 feet to 6 feet.

The tutor by using gesturing movements can change slides which are projected to the Course Board. Specifically pointing the Kinect camera at her direction and standing at least five feet away she can extend her right arm to activate the "right" or "forward" gesture. If she extends her left arm she activates the "left" or "back" gesture. These gestures will send a right or left arrow key signal to the foreground application, respectively. The up and down movement can be added if necessary. These two movements can be used to create a gesturing handled document reader with which the user will be able to move smoothly to the next page (not instantaneously as in the slides presentation). Additionally, one can mark on the document or 'printscreen' the document.

It should be noted that this application is independent of any slideshow software architecture (such as PowerPoint or Adobe Reader) which means that basic movements like *Next_Slide()*, *Previous_Slide()* can be easily applied to any of them. The way this is achieved is that the application uses *Window.Topmost* property of Windows Presentation Foundation (WPF) windows, which is a graphical subsystem for rendering user, interfaces in Windows-based applications by Microsoft. This property defines how in a group of windows, the window that has set the *Topmost* property to true is currently activated as the topmost window. Likewise for the group of windows that has set *Topmost* property to false are currently deactivated. The only thing the instructor has to do is to run the slides presentation software with the lecture's slides loaded to it. Since the slides presentation software has the *this.Topmost* value set to true this means that the window is not overlapped by any other window and the instructor can use the gesturing movements *Next_Slide()*, *Previous_Slide()* and change slides back and forth in a natural interaction manner using a pervasive computing infrastructure.

The HOU teleconferencing platform for virtual classrooms provides tutors and students with the ability to

 ISBN: 978-1-61208-356-8

participate in meetings and seminars regardless of their physical location by organizing and monitoring "electronic events". A typical teleconference session (event) consists of a central leader - presenter who is usually the tutor, and the other participants/guests who are usually the students. HOU uses the Saba Centra system which enables learners to connect remotely and watch lectures in real time. One of its main functions, inter alia, is the lectures slides presentations. This enables remote students to see on their screen the lecture's slides and listen to the sound from the classroom in real time. The developed Kinect-based application has been adapted to the needs of the Centra system (Fig. 5) providing the teacher with the ability to handle slides of lectures by using gesturing movements.

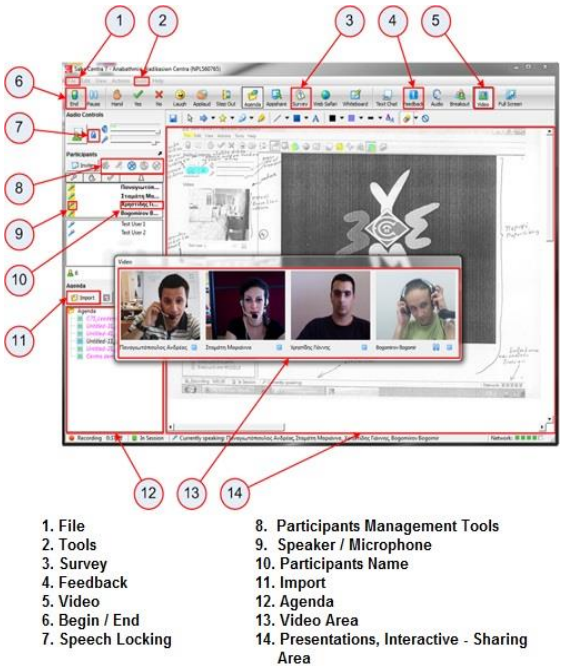


Fig. 5. The HOU's teleconference system.

The modification needed to be done in the existing software consisted of the operation of the movements that the tutor could accomplish. The loaded slides at Centra are converted to images independent of each other. To move to the previous/next image the user has to move with the up/down arrow keys and therefore two new gesturing movements which send a down or up arrow key signal to the foreground application were developed.

V. ENHANCING COLLABORATIVE LEARNING

The need for an active involvement of the students in the educational process led us to the design of an application which gives the student the capability to perform and share their thinking and understanding using annotation techniques. The aim of this application is particular classroom resources, such as the Course Board, to become available to the students in order to unfold their thoughts or to annotate comments to the content that is under discussion. In the context of a traditional lecture this seems difficult due to the fact that the student in order to be able to have access to the Smart Board must move from her seat and stand in front of the computer in which the operation software of the smart board is installed. Then she can handle smart board operations and eventually put forward her views. With this particular application the student is given the chance to interact with and handle this specific resource remotely, without moving from her place.

The rationale is that the student, given the permission of the instructor, is able to connect remotely with the smart board and handle its content either through the provided tablet or through her own Android device (Fig. 6). This specific application makes use of an android application which gives the ability to the android device to operate like a wireless mouse and keyboard for a specific computer. Thus, when the student is given the permission by the instructor to interact, she is connected remotely and gains access to the resource making use of the mouse and keypad in distance from her place. The application gives the instructor the ability to determine who is connected remotely each time as well to terminate the connection.

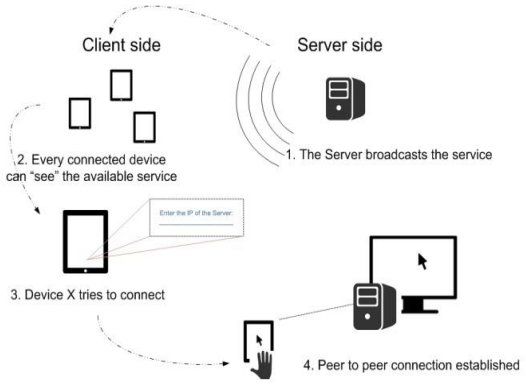


Fig. 6. Collaborative Tools

The architecture of this particular application is based on the client-server model in which the role of the client is played by the android devices and the role of server is played by the computer on which are connected and in our case is the same to that which runs the smart board software. On the server side special software is installed as a windows remote service which gives the capability to set the correct IP as well the TCP port in which the service will start. Once the windows remote service is started, every android device which belongs to the same network can search for its tracking. When the android device tracks the remote service it can be connected through the connect command with the server and have access to the handle of the mouse and the keyboard and therefore to the content of the lecture which is projected on the smart board. On the server side the tutor can monitor the interactions in the log section, to see which device is connected each time and terminate the connection when this is necessary.

VI. EVALUATION

The proposed authentication model was evaluated in a pilot setting including five trial meetings allocated to three groups of 20 students. Both the students and the tutors were provided with magnetic cards and RFId tags for their authentication as well as codes for their remote connection. From the data collected, 53% of the students on average participated physically to the meetings, 31% participated remotely and 16% were not attending. Regarding the

 ISBN: 978-1-61208-356-8

technology preference, there was a clear preference to the magnetic cards (87%) versus to RFId (13%), which is easily explained, as the former is a technology familiar to most of the students due to other services provided to them through it. Another reason for this preference was the additional feedback that the magnetic card based module provides to the students as it informs them not only about their successful authentication but also for their participations to the preceding sessions. Regarding the use of RFId technology we noted an increase on its usage meeting after meeting due to the fact that students were more familiarized to the new technology. The 100% of the tutors agreed that the automated way of recording the status of the students no matter the technology employed was time saving in relation to the traditional status recording.

On the other hand, it was noticed that alongside with the ubiquitous mode of authentication the traditional way of recording should be maintained as there were cases in which the students had forgotten to bring either their RFId tag or their magnetic card. In technological level, after the use of the system it was found out that the specific system was operating robustly within the failure margins given by the RFId/magnetic tag/card manufacturer (2%).

The gesture based physical interaction service was also evaluated. At the end of each session both tutors and students who participated in the meeting were asked to answer a questionnaire stating their view on the use of the particular application during the educational process. Regarding tutors who are the ones that are directly involved with the new service as well as being the main operators of the system, at first they found the application very difficult to handle as they were not familiar with this way of interaction with the system. Alongside they found it difficult to memorize the various combinations of gesturing movements that the application requires for handling slides. They reported that their concentration was focused on combining the movements with their hands to cause an action (e.g., next slide) which made them lose their sequence of thoughts. However, after the first two sessions the tutors got familiar with the handling of the application which led them to the conclusion that the use of gesturing movements increased their performance giving them the ability to interact in a more physical manner with the material and the students. The Kinect technology gives them the capability to navigate through the slides of the lecture along with their verbal communication without being forced to stop the lecture, approach the computer, change the slide from the keyboard and then continue their discussion. It should also be pointed out that one of the objectives of the SC prototype is the recording of the training material and the online streaming as a service to students. This application therefore helps achieving a better quality of the produced material since the video recording of lectures contains material which is more natural.

On the other hand, students, mainly those who are remotely connected to the SC accepted very well the use of this technology as their access to the classroom is achieved through the Centra teleconference system so using such an application gives them the illusion of even more physical presence in the classroom. Regarding the technical part there is one point in which the application should be improved in the future. That is the gesturing movement combinations which should be simplified so that the system can be used with greater ease.

Finally, students who participated in the meetings were provided with network capability to their laptops. The remote desktop application, which was mentioned in section V, was distributed to students via the network. Alongside all students were informed about the use of the application in the teaching process. In conclusion the use of the application strengthened the participation of students in the educational process something which is particularly difficult in traditional classrooms. Students were asked to rate the use of the application at a five point likert scale questionnaire. The result revealed the positive attitude shown by the students in the adoption of the application as a tool that enhances the involvement in the learning process.

VII. Related Work

In this section, we present research efforts that attempt to convert simple classrooms to smart ones by using pervasive computing technologies.

In 2003, researchers at Tsinghua University of China proposed a smart classroom standard for enhancing the traditional classroom with pervasive computing technologies and supporting remotely connected students [5][6]. The classroom is equipped with two wall-size projector screens, one in the front wall called Mediaboard and another called Studentboard. In the first board, which is touch sensitive, the teacher can present his slides and notes, while in the Student screen, images of the remote students and information about them are displayed. The smart classroom supports user authentication services via face and voice recognition also. A multiagent software platform coordinates multiple processes which are hosted in any computer of the classroom.

The George Maison University presented in 2005 Network EducationWare (NEW) an open source and distance learning education system [4]. Students via the NEW have the ability to hear instructor's voice and view lecture's presentation material, on their screen. They are also given the opportunity to intervene in the process and make questions either asynchronously or synchronously via instant messaging (chat) during the lecture. Finally, the system supports recording and playing back recorded lectures whenever the student wants in a later time. An important advantage over other commercially available systems in this domain, such as Saba Centra and Microsoft LiveMeeting, is the low cost of purchase and maintenance for educational institutions.

The CyLab laboratory of the Carnegie Mellon University in Japan proposed its own version regarding the Smart Classroom topic [9]. The main aim of this project is to link two or more classrooms through the internet. For this purpose a set of technologies are used such as the BlackBoard system, which provides online learning opportunities and recording lectures, the Lecture On Demand system (LOD), which enables students to have access to all previous lectures and material, and finally the VTE system (Virtual Training Environment), which is a library of educational tools and learning objects such as demonstrations, lectures slides, videos, hands-on labs.

In 2009, the University of Tsinghua in collaboration with the University of Kyoto in Japan developed the Open

 ISBN: 978-1-61208-356-8

Smart Classroom [7] model based on Smart Classroom project [5]. The developing model proposes the interaction among several Classrooms which are geographically dispersed. Two Smart Classrooms have been used, one at Tsinghua University and another one at the University of Kyoto. All processes that (are used), run on a software infrastructure called Open Smart Platform, which is based on the software platform of Smart Classroom that was developed at the university Tsinghua and consists of Agents, Containers and Directory Services. To optimize performance in the interconnection of the two classrooms the platform adopted services such as instant messaging, live data streaming and transferring large amounts of data (bulk data transfer).

The LOCAL project (Location and Context Aware Learning) is based on the principles of pervasive computing [1]. This project utilizes location and contextual information of learners aiming to promote the learning process. The model uses description learner's standard called PAPI to describe the user profiles [2]. The services that have been developed and are being supported are: sending messages to specific trainee who has been authenticated by the system; sending messages to a specific group of learners; and massive messaging. The messages addressed to students represent learning objects and are correlated to the goals set by the students themselves.

In 2010 the University of Cairo developed its own Smart Classroom standard which is named iClass [8]. This project uses a classroom equipped with various types of sensors such as light, temperature and humidity sensors and a weather station. All of those sensors are being interconnected via Lonworks, a building automation protocol, commercially available which provides HTML interfaces to users and which interfaces with the IP network infrastructure. RFId technology is also used, for user authentication. Specifically when the teacher enters the room a Smart Agent recognizes him and automatically opens the smart board. Then students can be authenticated by the Smart Agent too. Finally, using voice recognition in the classroom the instructor is allowed to interact with the devices of the classroom through voice commands.

Researchers at the Taipei University have developed a Smart Classroom standard [3] in which the teaching table agent has several modules for classroom management. One module takes care of the power supply and lighting system and another module is responsible for managing entrance issues since the classroom has digital locks and smart user authentication system. A third module manages the smart blackboard system and is responsible for controlling the projector. This model beyond the management capabilities focuses on providing a smart timetable system which enables the instructor to interact with it. The application has several levels of interaction which are defined according to the distance of the user from the screen and the gesturing movements he uses.

VIII. CONCLUSIONS

In this paper, we introduced services deployed in a SC for enhancing collaborative learning and management tasks through pervasive computing technologies. The specification and the development of such services require an understanding of the needs and characteristics of all the involved educational stakeholders. The proposed system is a first prototype which makes use of tracking technologies such as RFId and natural interaction modalities with the system to support a collaborative learning methodology. The objective is the proposed model to be enriched with new technologies and services that will enable the students to participate in the educational process through contemporary and interactive distance learning environments.

ACKNOWLEDGEMENT

The research described in this paper has been co-funded by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) (Funding Program: "HOU").

REFERENCES

[1] J. Barbosa, R. Hahn, S. Rabello, and D. Barbosa, "LOCAL: A Model Geared Towards Ubiquitous Learning," ACM SIGCSE Bull.,2008 vol. 40, no. 1, pp. 432-436.

[2] PAPI - Draft Standard for Learning Technology. Public and private information (PAPI) for learners (PAPI learner). Last Accessed [retrieved: May, 2014]. Available at ltsc.ieee.org/wg1/files/ltsa_06.doc.

[3] Y. Yu, S. D. You, and D. Tsai, "Smart Timetable Plate for Classroom," 2010, icalt 10th IEEE International Conference on Advanced Learning Technologies, pp.552-554.

[4] C. Snow, J. M. Pullen, and P. McAndrews, "Network EducationWare: an open-source web-based system for synchronous distance education," Education, IEEE Transactions on, 2005, vol. 48, pp. 705-712.

[5] Y. C. Shi, et al.: 'The smart classroom: merging technologies for seamless tele-education', IEEE Pervasive Computing, IEEE CS Press, 2003, 2, (2), pp. 47-55.

[6] X. Li, L. Feng, L. Zhou and Y. Shi, "Learning in an Ambient Intelligent world: Enabling Technologies and Practices." IEEE Trans. Knowl. & Data Eng., June 2009, vol. 21, no. 6, pp. 910-924.

[7] Y. Suo, N. Miyata, H. Morikawa, T. Ishida, and S. Yuanchun, "Open Smart Classroom: Extensible and Scalable Learning System in Smart Space Using Web Service Technology," IEEE Transactions on Knowledge and Data Engineering,2009, pp. 814-828.

[8] R. Ramadan, H. Hagras, M. Nawito, A. Faham, and B. Eldesouky, (2010). The Intelligent Classroom: Towards an Educational Ambient Intelligence Testbed. 2010 Sixth International Conference on Intelligent Environments, 2007 pp. 344-349. Cairo.

[9] D. Pishva, "Smart Classrooms Bring Top-Quality Education around the Globe," International Symposium on Applications and the Internet Workshops (SAINTW07), pp.40.

[10] Kinect- Xbox.com, http://support.xbox.com/en-US/browse/xbox-on-other-devices/kinect-for-windows [retrieved: May, 2014]

 ISBN: 978-1-61208-356-8

Mobile Augmented Reality for Distributed Healthcare

Point-of-View Sharing During Surgery

Lauren Aquino Shluzas, Gabriel Aldaz,
Joel Sadler, Shantanu Joshi, Larry Leifer
Center for Design Research, Stanford University
Stanford, California, USA
{larenino, zamfir, jsadler, joshi4, larry.leifer}@stanford.edu

David Pickham
Office of Transdisciplinary Research
Stanford Hospital & Clinics
Stanford, California, USA
dpickham@stanfordmed.org

***Abstract*—This research examines the capabilities and boundaries of a hands-free mobile augmented reality (AR) system for distributed healthcare. We use a developer version of the Google Glass head-mounted display to develop software applications to enable remote connectivity in the healthcare field; characterize system usage, data integration, and data visualization capabilities; and conduct a series of pilot studies involving medical scenarios. This paper presents the software development and experimental design for a pilot study that uses Glass for augmented point-of-view sharing during surgery. The intended impact of this research is to: (i) examine the use of technology for complex problem solving and clinical decision making within interdisciplinary healthcare teams; (ii) study the impact of enhanced visualization and auditory capabilities on team performance; and (iii) explore an AR system's ability to influence behavior change in situations requiring acute decision-making through interaction between centralized experts and point-of-impact delivery personnel.**

Keywords—ambient; systems; distributed; healthcare; team-based collaboration; head-mounted display, surgery; wearable computing.

I. INTRODUCTION

In the healthcare field, the need for improved tools to enhance collaboration among patients and providers has become increasingly urgent – due, in large part, to a global rise in aging populations and chronic disease prevalence, coupled with increasing health care costs and physician shortages worldwide [1][2]. To address these challenges, this research examines the use of wearable mobile computing to mediate interdisciplinary communication and collaboration in healthcare. We first present an overview of collaborative technologies and interaction modalities for home-healthcare and hospital use. These advances in mobile computing provide remote patient monitoring, automate simple knowledge-base procedures, and facilitate the delivery of interventions. We then present the experimental design for a pilot study that uses augmented point-of-view (POV) sharing during surgery through leveraging the camera and integrated sensors of Google Glass' head mounted display (Figure 1).

A. Remote Patient Monitoring

For home healthcare applications, existing systems enable medical staff to remotely monitor patients suffering from advanced chronic disease and provide prompt support regarding health education and treatment compliance [3]. The *Vsee* video collaboration system (Sunnyvale, CA), for instance, simplifies patient-doctor interactions through web-based video calling, coupled with medical device integration.

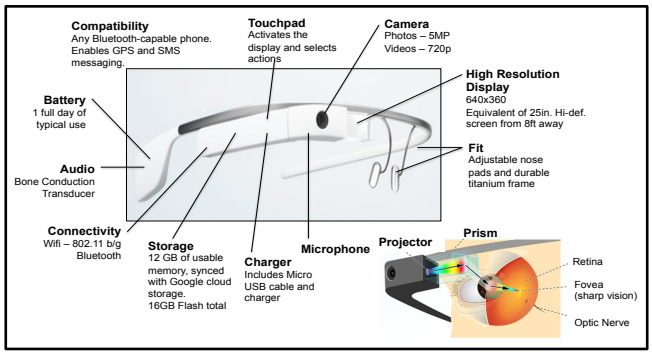


Figure 1. *Google Glass* (Mountain View, CA) wearable computer and head-mounted display features

In the consumer health and wellbeing space, a variety of products provide self-patient monitoring to encourage behavior modifications aimed at promoting healthier lifestyles. These include activity monitors such as the *Nike Fuelband* and *Fitbit Flex*, and the *LUMOback* real-time posture feedback system. Yet, despite current advances, home monitoring technologies are often limited by the dependency on appropriate bandwidth, customized networks and high-cost equipment, as well as a lack of integration into electronic medical record (EMR) systems.

In the hospital setting, remote monitoring systems, such as the *tele-ICU*, aid clinicians in the delivery of care to ICU patients. By collaborating with the bedside team, the *tele-ICU* support care without distraction, while assisting in the delivery of timely interventions [4]. Yet, financial barriers associated with installing and operating such systems has limited widespread adoption.

B. Automating Knowledge-based Procedures

For patient-virtual agent interaction, the animated virtual nurse (VN) is an automated system that teaches patients their post-discharge self-care regimen directly from their hospital beds [5]. This system incorporates a VN who embodies best practices in health communication for patients with inadequate health literacy, and illustrates a growing field in mobile computing aimed at increasing universal healthcare access.

 ISBN: 978-1-61208-356-8

Cognitive aids, such as dynamic checklists, present another example of tools to facilitate collaboration in the clinical setting, through automating knowledge-based procedures. A recent study involving dynamic checklists found that medical crisis care situations reveal "a physically complex information space, and relatively high-tempo time scales of gaze, action, and team-based coordination and communication" [6].

C. *Facilitating the delivery of an intervention*

Collaborative technologies that facilitate the delivery of an intervention may include those in which (i) a clinician (expert) aids a non-expert in delivering an intervention, and (ii) a clinician delivers an intervention remotely through a robotic interface. With the commercialization of high-speed data networks, the implementation of these scenarios may be realized through the use of augmented reality (AR) systems. Many AR applications provide the benefit of visualizing three-dimensional data captured from non-invasive sensors, and range from remote 3D image analysis to advanced telesurgery [7][8].

Despite existing technologies, however, there is a growing need for new tools capable of augmenting a clinician's knowledge base and his/her complex problem solving ability, while performing an intervention. Such augmentation can be accomplished if a system simultaneously connects the clinician (expert) to relevant medical databases, other experts, and a live telemetry of patients' relevant vital statistics. The first order challenge required to accomplish these connections is the ability to manage the resulting high-bandwidth information flow between human and computer agents, and to enable agents to collectively work as a design team.

In this paper, Section II describes the methods used to conduct this research study. In Section III, we provide an overview of point-of-view sharing during surgery, including software and hardware development, a usability assessment, and a proposed pilot study. Section IV provides a conclusion and discussion of future work.

II. METHODS

To address the challenges discussed above, this ongoing research effort examines the use of hands-free mobile AR for distributed healthcare collaboration. We hypothesize that a mobile AR system (head-mounted display) will shorten communication cycles and reduce errors associated with point-of-care decision-making and distributed collaboration in healthcare scenarios.

To test our hypothesis, we obtained four pairs of Google Glass as an initial platform for research. Our methodology includes: (i) clinical needs finding to ground the study in the context of high-impact clinical problems; (ii) software development to create customized applications for the head-mounted display that are specific to two or more clinical areas; and (iii) a series of pilot tests to characterize the AR system's usage, data integration, and data visualization capabilities in medical scenarios.

A. *Clinical Needs Finding*

We conducted needs finding at Stanford Hospital and Clinics from October to December 2013. Two members of the research team interviewed ten nurses and two physicians, and shadowed four additional nurses. Each interview lasted approximately one hour, and nurses were shadowed from one to three hours at a time. We recorded interviews using a digital recording device, and took hand-written notes during each interview and clinical observation.

From this process, we captured 135 clinical needs and grouped needs into 15 areas. To further narrow our research focus, we filtered needs based on degree of importance to the hospital, alignment with research interests, fit for Glass, and feasibility. We ordered our top three needs in order of increasing clinical risk. These included: (i) wound and skin care photography, (ii) point-of-view sharing during surgery, and (iii) vital sign communication when patients are suffering from cardiac arrest.

B. *Initial Pilot Study*

We developed a Google Glass application to capture and document images for chronic wound photography, using the Android 4.0.4 (API 15) SDK and Glass Development Kit (GDK) add-on. The application leverages Glass' camera, inertial measurement unit (IMU), infrared sensor, and microphone. The initial pilot study focused on the use of voice and gestural-based interaction commands for photo capture and documentation, and the transfer of annotated images to a patient's EMR. The software development and pilot study for wound care management have been separately documented [9]. In this paper, we focus on the software application development and experimental design for a second pilot study to conduct point-of-view sharing in the operating room, using Google Glass.

III. POINT-OF-VIEW SHARING DURING SURGERY

A range of studies has demonstrated the use of head-mounted cameras for clinical use and education. Bizzotto et al. [10], for instance, used the GoPro HERO3 in percutaneous, minimally invasive, and open surgeries with high image quality and resolution. Several reports document the live two-way broadcasts of patient surgeries by doctors wearing Glass in the operating room [11][12].

Figure 2. Center for Immersive and Simulation-based Learning

 ISBN: 978-1-61208-356-8

To build on existing work in this field, we plan to conduct a pilot study within Stanford's Center for Immersive and Simulation-based Learning (CISL), as shown in Figure 2. For this study, we will simulate a medical scenario using augmented POV sharing for the treatment of an acute condition, in which relevant patient data is superimposed on the Glass display, within the surgeon's field of view. The simulation will focus on complex problem solving with multiple potential solution paths (e.g., situations in which there is no optimal, context-independent solution).

A. *Software Development and Hardware Enhancement*

The development of a software application for POV sharing during surgery focuses on three key areas [13]: voice commands, bi-directional communication, and EMR data transfer. We intend to use voice commands to control the application in a hands-free manner, in order for clinicians to maintain heightened sterility while performing surgical procedures. Bi-directional communication is required to enable collaboration between surgeons wearing Glass and remote colleagues, and to establish connectivity with remote sensors in the healthcare environment. We are in the process of developing a communication interface between the wearable computer and remote sensors using WiFi and Bluetooth. Using an Arduino microprocessor as a proxy for medical sensors, we have demonstrated that wireless connectivity via WiFi may be achieved by connecting Glass to remote sensors through Android's WiFi Peer-to-Peer (P2P) API. For Glass, this may also be achieved through the Google Mirror API, as an intermediary. For Bluetooth connectivity, we demonstrated the ability to connect Glass with a remote sensor (Arduino microprocessor), pair the two devices, and transfer data between them. With Glass, the connection may be launched manually using the device's touch pad, physical gestures, or voice commands. We aim to apply the connectivity protocols, demonstrated with an Arduino, towards capturing the actual vital signs of patients from monitoring systems in the hospital (e.g., the Phillips Intellivue Solution System). We also intend to build on existing work in wound care photography to wirelessly transfer surgical image and video data to a patient's EMR [9].

Based on feedback from consulting surgeons, in addition to software development, we aim to physically enhance the Glass hardware for surgery in three ways. We intend to add a transparent splash shield that surgeons may adhere to the front frame of the Glass display for protection from infectious disease, attach an optical loupe to the frame (in front of a surgeon's left eye) in order to increase magnification for surgical procedures, and encase elements of the head-mounted display in a protective cover for improved cleaning and robustness during routine clinical use.

B. *Initial Feasibility and Usage Assessment*

Following software development and hardware enhancements, we will conduct an initial assessment of the AR system's ability to (i) connect with remote study participants, (ii) capture a patient's sensory data, and (iii) transmit information to/from a hypothetical EMR. We intend to evaluate the system's usage characteristics with 10-15 individuals as they wear the Glass technology while performing a series of manual tasks. We will evaluate each user's ability to: (i) use and navigate the system through voice and tactile commands, (ii) integrate multiple audio and visual data inputs, and (iii) transfer data. We will assess the time required for each user to complete a specified task, and the associated error rate.

C. *Pilot Study*

1) Experimental Set-up: In a pilot study at Stanford's CISL facility, an attending surgeon will wear the Glass head-mounted display, while conducting a hypothetical procedure in collaboration with one assisting surgeon and one nurse. The local medical team will be in direct communication with a remote team of collaborators (in Germany and the Netherlands) via the Glass interface. The team will be presented with a hypothetical medical scenario and asked to develop a solution in 20 minutes using standard operating room supplies. A 'patient', e.g., computer driven mannequin that replicates physiologic parameters (pulse, heart and breath sounds, blood pressure, etc.) – will be used in the simulation. The patient's sensory data will transmit to the remote team and to a hypothetical EMR, until the patient's desired health state is achieved during the simulation. Using voice commands, the attending surgeon will be able to switch views in the Glass display between EMR data and wound images with information overlays. The images within the Glass display will be projected on an adjacent wall, so that the assisting surgeons and nurse may see the procedure from the attending surgeon's point-of-view. A schematic illustration of the experimental set-up is shown in Figure 3.

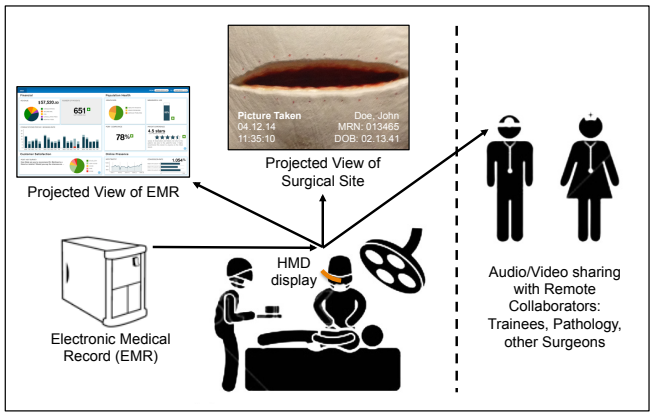


Figure 3. Point-of-View Sharing with Remote Collaborators

We will repeat the simulation three times with different medical teams. Local study participants will include registered nurses and medical residents from Stanford Hospital & Clinics.

2) Performance Measures: We plan to assess team performance using the AR system based on a combination of qualitative and quantitative measures. The independent

variables, based on a Likert 5-point scale, include: X_1 – degree of team collaboration; X_2 – degree of task coordination (e.g., the ability to co-locate activity with visual instruction, and switch between data inputs and verbal communication modes); and X_3 – the degree of visual and auditory human perforance augmentation. The dependent variables include: Y_1 – the time for task completion and Y_2 the error rate in completing each task. Secondary measures include (i) the most/least effective mode to enhance feedback capability; (ii) the most/least effective data visualization display method; and (iii) each participant's percevied degree of user behavior change, based on the test conditions. Scores will be ascertained through a survey administered to study participants, field observations conducted by the research team, and post-experiment interviews with study participants.

3) Data Analysis Methods: We will calculate the mean time and survey scores for the three simulations. Multivariate statistical methods (e.g., ANOVA) will be used to determine if correlational relationships exist between the independent and dependent variables. Qualitative data, including field observations and interviews, will be transcribed, coded, and analyzed using NVivo qualitative analysis software.

IV. CONCLUSION AND FUTURE WORK

Augmented point-of-view sharing during operating room procedures advances the state-of-the-art in surgery through enhancing a clinician's knowledge base and decision-making capabilities, while performing surgical interventions. Specifically, since several individuals are needed to perform a surgical procedure (e.g., the attending surgeon, assisting surgeons, an anesthesiologist, and multiple nurses), the incision site is frequently crowded and obstructed from one's field-of-view. With augmented POV sharing, images projected from an attending surgeon's vantage point onto a remote screen provide visual clarity to individuals directly involved in the surgery, as well as remote participants and expert advisors. It also helps trainees and surgical fellows to learn procedures more accurately, by viewing an intervention from the same perspective as an attending surgeon, rather than from the reverse perspective (e.g., standing on the opposite side of the table) - which is often the case today. Finally, the use of POV sharing with bi-directional communication capabilities and remote sensor connectivity enables real-time collaboration with a pathology lab or expert consultants while performing a surgical intervention. Through visual overlays, POV sharing can co-locate one's visual field with information critical to performing a procedure – such as vital sign information, procedural descriptions, or MRI scans.

In the future, we aim to further enhance the robustness and reliability of a mobile AR system for acute care scenarios in the operating room and emergency room settings. This involves developing applications with data security and privacy features that are in compliance with strict hospital security protocols.

In the broader context of distributed collaboration for improved healthcare delivery, this research aims to examine the use of technology for complex distributed problem solving through interdisciplinary collaboration; gain an improved understanding of the benefits of human augmentation through enhanced visualization and auditory capabilities, on healthcare team performance; and explore an AR system's ability to influence behavior change in situations requiring acute decision-making through interaction between centralized experts and point-of-impact delivery personnel.

ACKNOWLEDGMENT

We gratefully acknowledge the Hasso Plattner Design Thinking Research Program (HPDTRP) for their support of this work. We thank the clinical staff at Stanford Hospital & Clinics for their time and participation in this research study.

REFERENCES

[1] S. Mattke et al., Health and Well-Being in the Home: A Global Analysis of Needs, Expectations, and Priorities for Home Health Care Technology. Santa Monica, CA: RAND Corporation, Dec. 2010, pp. 1-39.

[2] L. Magnusson, E. Hanson, and M. Borg, "A literature review study of information and communication technology as a support for frail older people living at home and their family carers," Technology and Disability 16, Dec. 2004, pp. 223–235.

[3] L. C. Baker, S. J. Johnson, D. Macaulay, and H. Birnbaum, "Integrated telehealth and care management program for Medicare beneficiaries with chronic disease linked to savings," Health Affairs 30, 9, Sept. 2011, pp.1689-1697.

[4] S. F. Goran, "A second set of eyes: An introduction to Tele-ICU," *Critical Care Nurse* 30, 4, Aug. 2010, pp. 46-55.

[5] T. W. Bickmore, L. M. Pfiefer, and B. W. Jack, "Taking time to care: empowering low health literacy hospital patients with virtual nurse agents," CHI 2009, ACM Press, Apr. 2009, pp. 1265-74, ISBN: 978-1-60558-246-7.

[6] L. Wu et al., "Interactive cognitive aids: maintaining shared mental models in anesthesia crisis care with nurse tablet input and large-screen displays," UIST'11, ACM press, Oct. 2011, pp. 71-72, ISBN: 978-1-4503-1014-7.

[7] D. W. F. van Krevelan and R. Poelman, "A survey of augmented reality technologies, applications, and limitations," The International Journal of Virtual reality 9,2, Jan. 2010, pp. 1-20.

[8] F. Zhou, H. B. L. Duh, and M. Billinghurst, "Trends in augmented reality tracking, interaction and display: a review of ten years of ISMAR," Proc. 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, (ISMR 2008), Sept. 2008, pp. 193-202. doi: 10.1109/ISMAR.2008.4637362.

[9] G. Aldaz et al., "SnapCap: Improved hands-free image capture, tagging, and transfer using Google Glass," unpublished.

[10] N. Bizzotto, A. Sandri, F. Lavini, C. Dall'Oca, and D. Regis, "Video in operating room: GoPro HERO3 camera on surgeon's head to film operations—a test," Surgical Innovation, Dec. 2013, Published online before print, doi: 10.1177/1553350613513514. [retrieved: June, 2014]

[11] B. J. Lutz, and N. Kwan, "Chicago surgeon to use Google Glass in operating room," Dec. 2013. [Online]. Available

 ISBN: 978-1-61208-356-8

from: http://www.nbcchicago.com/news/tech/google-glass-surgery-237305531.html. [retrieved: June, 2014]

[12] Medical Xpress. First US surgery transmitted live via Google Glass (with video). Oct. 2013. [Online]. Available from: http://medicalxpress.com/news/2013-08-surgery-transmitted-google-glass-video.html. [retrieved: June, 2014]

[13] L. Shluzas et al. A wearable computer with a head mounted display for hands-free image capture and bi-directional communication with external sensors. US Provisional Patent No. 61968857, Mar. 2014.

 ISBN: 978-1-61208-356-8

High Speed Wireless Access Based on Visible Light Communication Utilizing Maximum Ratio Combination of Multi-Detectors

Jiehui Li, Xingxing Huang, Nan Chi
Department of Communication Science and Engineering
Fudan University
Shanghai, China
e-mail: jiehuili13@fudan.edu.cn, xinghuang90@gmail.com, nanchi@fudan.edu.cn

***Abstract*-In this paper, we propose and experimentally demonstrate a novel space-reception diversity model based on Maximal Ratio Combining (MRC) algorithm in Single Input Multiple Output (SIMO) Visible Light Communication (VLC) system. Different from the previous schemes, the data was received by two detectors located in different space positions at the same time, and then, combined using MRC algorithm. As a result, the Bit Error Ratio (BER) of receivers can be reduced to one to two orders of magnitude. The performance of VLC system is significantly improved.**

Keywords-LED; multi-detector; maximun ratio combining; visible light communication; spatial positions.

I. INTRODUCTION

Visible Light Communication (VLC) based on Light Emitting Diodes (LEDs) offers an entirely new paradigm in wireless technologies in terms of communication speed, flexibility, usability and security [1]-[3]. The VLC is being touted as one of the next generation wireless communication systems, because it can simultaneously support illumination and communication [4]. With the further advancement of the project in future, VLC can be a replacement to almost all indoor and short range communication purpose.

Multipath propagation is a key feature of wireless communication. It has been conceived of as a bottleneck that hampers the creation of efficient wireless communication systems [7]. In VLC system, the multipath interference effect is mainly caused by the distance difference between the light and the receiver and the light reflected the walls. In a multipath environment, the spatial diversity can average out a substantial part of defecting effects of sporadic fades between different detectors. In the simplest form of it, there is just a single transmitter (Tx), in which case the problem reduces to combining the signal from multiple receivers (Rxs) in the most efficient way. The optimum solution to it is well established. We called the algorithm Maximal Ratio Combining" (MRC). MRC weighted after the merger, significantly lower Bit Error Rate (BER) for receiving signals, which can effectively enhance the received signal [8]. Since it is optimum in Single Input Multiple Output (SIMO) transmission systems, it can be considered as a benchmark in comparison with more recently proposed Multiple Input Multiple Output (MIMO) systems.

This paper is organized as follows. Section Ⅱ introduces the related work information, Section Ⅲ introduces the proposed system model and its mathematical presentation. The simulation and experiment results are presented and discussed in Section Ⅳ, before we conclude this paper in Section Ⅴ.

II. STATE OF THE ART

The performance variation at the receiver is one of important issues for VLC. Due to the sensitivity of the detector the performance is not same, even though the same transmitter and receiver are being used. Many researchers proposed some methods for improving receiver performance in VLC system, but no one has considered the performance can be improve according to automatically weight two different receivers. *Lee et al.* [5] have proposed a receiver structure to improve the VLC system where separate receivers with specific spectral response are used for the detection of different colors of channels. *Muhammad et al.* [6] proposed a receiver considering the performance variation on the different color of channels according to the VLC band plan, and proved it functionality mathematically as well as by simulation, but they did not consider the performance variations of the receivers located in different space positions.

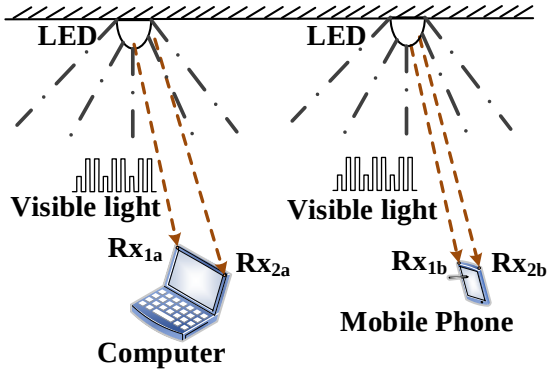


Figure 1. The schematic of VLC 1×2 SIMO system.

In this paper, we proposed a 1×2 SIMO transmission model considering two receivers located in different space positions. Figure 1 shows the sketches of VLC 1×2 SIMO system. This dual function of LED, for lighting and communication. MRC was used to combine the data with diversity from two detectors located in different space

 ISBN: 978-1-61208-356-8

positions in the computer and mobile phone. Then, the data was outputted after combined. With the scheme, the transmission distortion will be improved efficiently in visible light communication system. As a matter of fact, our simulation and experiment results also proved it. Our proposition can be extended to more detectors to enhance the performance in the future.

III. SIMULATION AND RESULTS

LEDs are used to transmit desired optical signal in VLC system. The desired optical signals then travel through air before reaching optical detector. The receiver collects some undesirable optical signals which cause severe degradation to the overall system performance. In a VLC system, the receiver performance variation is occurred due to the angle and space position. We proposed a visible light SIMO communication system, the combining algorithm of MRC is used to improve the performance of the receiver.

A. Sprinciple of MRC

As well as most of the receive diversity algorithm, MRC combiner linearly combines the individually received branch signals so as to maximize the instantaneous output Signal-to-Noise Ratio (SNR) [9]; that is to say, the output result is the weighted sum of different fading channels. MRC algorithm is applicable to any modulation way, arbitrary diversity route and branch decline.

A SNR can express the quality of a communication system. We assume that the transmitter transmits the signal using On-Off Keying (OOK) modulation technique. Among all modulation techniques for visible light communication link, OOK is the simplest one and very easy to implement. In a single receiver, the average SNR is defined as the ratio of the received signal to the aggregated noise, and the SNR is proportional to the detector area when the shot noise is the dominant noise source [10]. The signal component of the SNR is measured by *Uddin et al*. [6]:

$$S = \eta^2 \xi^2 (n_{Rx}) P_{rsignal}^2$$

The signal received power $P_{rsignal}$ is:

$$P_{rsignal} = \int_0^T \left(\sum h_i(t) \otimes P_i(t)\right) dt$$

where η represents the photo sensitivity of the photo-detector (in A/W), $P_i(t)$ is the instantaneous input power, the symbol " $\otimes$ " denotes convolution, $h(t)$ resembles the impulse response, $\xi(n_{Rx})$ is the performance variation balancing factor. The time average transmitted optical power is given by *Uddin et al*. [6]:

$$P_t = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} P_i(t) dt$$

where $P_i(t) \geq 0$ since the instantaneous input power must be nonnegative.

We assume OOK with rectangular transmitted pulses of duration equal to the bit period. Gaussian noise having a total variance N_0 that is the sum of contributions from shot noise, thermal noise and Inter-Symbol Interference (ISI) by an optical path difference:

$$N_0 = \sigma_{shot}^2 + \sigma_{thermal}^2 + \eta^2 P_{rISI}^2$$

Therefore, the SNR is given by:

$$SNR = \frac{\eta^2 \xi^2 (n_{Rx}) H^2(0) P_t^2}{\sigma_{shot}^2 + \sigma_{thermal}^2 + \eta^2 P_{rISI}^2}$$

And BER is given by:

$$BER = \frac{1}{2\pi} \int_{\sqrt{SNR}}^{\infty} e^{-y^2/2} dy$$

where $H(0) = \int_{-\infty}^{\infty} h(t) dt$ is the channel DC gain.

The received power by inter-symbol interference P_{rISI} is given by:

$$P_{rISI} = \int_T^{\infty} \left(\sum h_i(t) \otimes P_i(t)\right) dt$$

Short noise variance is given by:

$$\sigma_{shot}^2 = 2q\eta \left(P_{rsigal} + P_{rISI}\right) B_{en} + 2q I_{bg} I_2 B_{en}$$

where q is the electronic charge, B_{en} is equivalent noise bandwidth, I_{bg} is background current and I_2 is noise bandwidth factor.

The thermal noise variance is given by *Uddin et al*. [6]:

$$\sigma_{thermal}^2 = \frac{8\pi k T_k}{G} \delta A I_2 B_{en}^2 + \frac{16\pi^2 k T_k E}{g_m} \delta^2 A^2 I_3 B_{en}^3$$

where k is Boltzmann's constant, $T_k$is absolute temperature, G is the open-loop voltage gain, δ is the fixed capacitance of photo-detector per unit area, E is the Field-Effect Transistor (FET) channel noise factor, g_m is the FET transconductance.

B. Simulation of VLC system

The schematic of visible light SIMO system is shown in Figure 2. For high speed telecommunications, Quadrature Amplitude Modulation (QAM) technique that is used in the communication system would be helpful in which the desired waveform is modulated onto the instantaneous power of the carrier. On the other hand, QAM demodulation algorithm is the requirement proportional to the received data in receivers. In our channel model, the information carrier is a light wave. Moreover, detector dimensions are in the order of thousands of wavelengths, leading to efficient spatial diversity, which prevents multipath interference. Therefore, multipath interference can be greatly reduced in optical wireless transmitting.

 ISBN: 978-1-61208-356-8

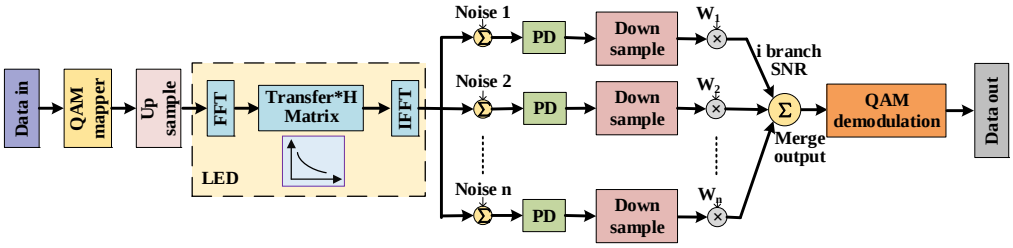


Figure 2. The simulation model of visible light SIMO system.

C. The results

When studying the performance with QAM modulation, the abscissa of the curves is the QAM order and delta frequency of the two detectors. The simulation results are depicted in Figure 3 and Figure 4. These curves are essential to understand the diversity gain that is attainable by concatenation of the SIMO. A threshold line for BER is set as 3.8×10^{-3}.

Figure 3 shows the receive BER variation for different detectors and proposed MRC combining according to the different QAM orders. It can be intuitively explained by noting that MRC weighted can significantly improve the performance of the separate detectors even though the BER of the MRC algorithm and separate channels looks almost the same when 128 QAM or higher. The BER curve has a lowest point with the QAM order increasing, due to the high frequency fading of low-order signal is faster [11]. The 4-QAM order (16QAM) shift is worth noting between the SIMO schemes. When the QAM order is 4 or below, the BERs of the detector change little, but above 4-QAM order, the BERs increase exponentially. This further illustrates that the MRC algorithm is suitable for low-order QAM modulation rather than high-order.

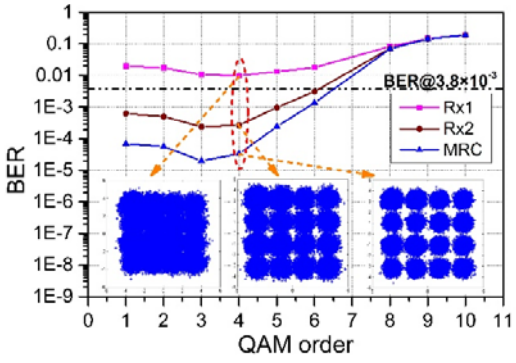


Figure 3. Simulation BER versus QAM order.

The simulation results show that it is hard to discern any differences in performance between the optimum and sub-optimum 16QAM demodulations. Considering the transmitting speed, we determine the optimal modulation format for 16QAM. From Figure 4, the receiver BER increases with the increase of transmission frequency difference. It illustrates that the higher the delta frequency, the greater the noise in the receiver, the easier to produce error. The BER of MRC weighted combining significantly lower than the separate detector. When the delta frequency bellow 7Hz, the receiver has a better performance, and the BER of MRC weighted combining is below the Forward Error Correction (FEC) limit of 3.8×10^{-3}.

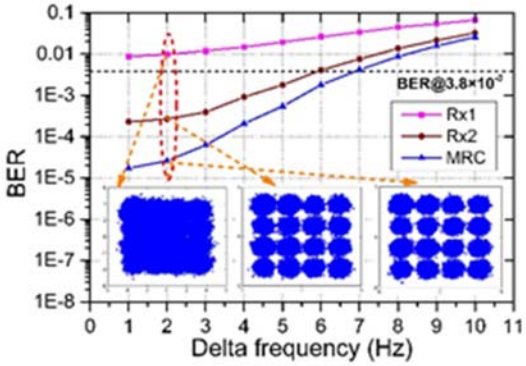


Figure 4. Simulation BER versus delta frequency.

IV. EXPERIMENTAL SETUP

In a VLC system, every receiver expects the same performance, even though the two receivers have different SNR. Therefore, our scheme focuses on reducing the performance variation at the receiver side. In order to verify the simulation results, we tried to send an image with our proposed VLC system. At the receiving end, we placed two receivers at in different spatial positions to receive signal at the same time. Due to the limited experiment conditions, two receives are used. Then the received data was analyzed with MRC algorithm. The experiment system is shown as Figure 5. To increase the separation distance between a light transmitter and a receiver, lenses are often used. A light receiver may use a lens to collect the weak light from the transmitter and focus it onto the optical detector for processing. However, the lens will always collect extra light from the environment that is not wanted. Hence, all of the received data from detectors in different spatial positions are combined by MRC algorithm, then the output data requires further signal processing to restore the desired image. As a result, the performance of the VLC system will be significantly improved.

 ISBN: 978-1-61208-356-8

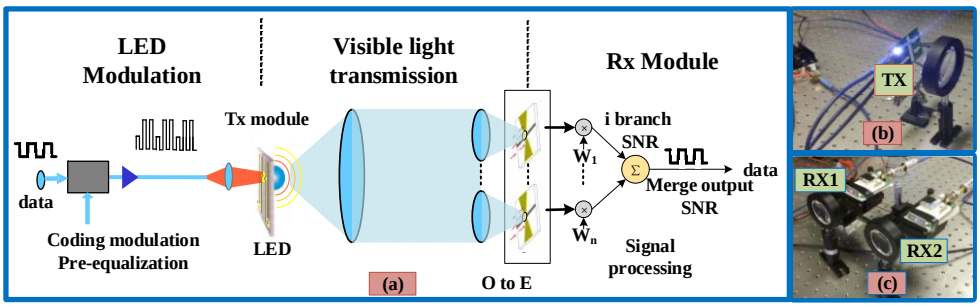


Figure 5. The experiment system: (a) the schematic of VLC system; (b) the transmitter (Tx); (c) the receives (Rx).

We measured the BER performances versus bias voltages of red LED. In this demonstration, the distance between transmitter and receivers is 0.6m, and the electrical input power is fixed at 12dbm, which is relatively small to avoid reaching the saturation area of LED. The modulation format is 64QAM-OFDM, considering the actual data transfer rate. The BER versus bias voltage results are depicted in Figure 6, and the constellations are shown for the bias voltage of 2.2V. We can find the BER of MRC algorithm can improve one to two orders of magnitude compared with any separate detector. Therefore, it is very effective that space diversity reception with MRC algorithm adopts multiple receivers in order to improve the performance of the VLC system.

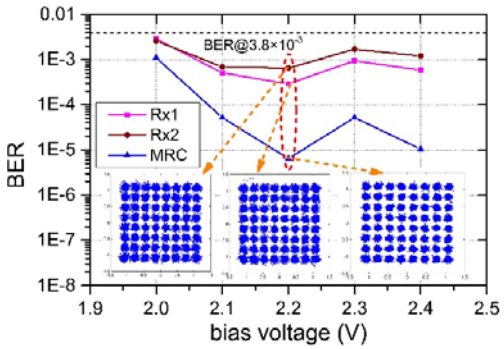


Figure 6. Measure BER versus bias voltage.

V. CONCLUSIONS

Diversity combining, which skillfully combines multiple replicas of received signals has long been as one of the most efficient techniques to overcome the destructive effects of multipath fading in wireless communication systems [5]. In this paper, we have proposed a novel 1×2 SIMO transmission model in VLC system based on two detectors located in different space positions. With the scheme, the transmission distortion of images can be improved efficiently. The BER of receivers can be reduced to one to two orders of magnitude. The BERs of signal combining algorithm of MRC are always below the FEC limit of 3.8×10^{-3}. Therefore, space-reception diversity based on MRC algorithm could be an effective technique to achieve high data rate and high fidelity transmitting images in visible light communication system.

ACKNOWLEDGMENT

This work is partially supported by NHTRDP (973 Program) of China (Grant No. 2010CB328300), NHTRDP (863 Program) of China (No. 2013AA013603), NNSF of China (No. 61177071), and Key Program of Shanghai Science and Technology Association (No. 12dz1143000).

REFERENCES

[1] Punith P. Salian, Sachidananda Prabhu, Preetham Amin, Sumanth K. Naik, and M. K. Parashuram, “Visible Light Communication,” Texas Instruments India Educators’ Conference. India, 2013, pp. 379-383.

[2] Yuanquan Wang, Yiguang Wang, Nan Chi, Jianjun Yu, and Huiliang Shang, “Demonstration of 575-Mb/s downlink and 225-Mb/s uplink bi-directional SCM-WDM visible light communication using RGB LED and phosphor-based LED,” Opt. Express, vol. 21, 2013, pp. 1203-1208.

[3] Yiguang Wang, Minglun Zhang, Yuanquan Wang, Wuliang Fang, Li Tao, and Nan Chi. “Experimental demonstration of visible light communication based on sub-carrier multiplexing of multiple-input-single-output OFDM,” OECC 2012, 2012, pp. 745-746.

[4] Toshihiko Komine, “Fundamental Analysis for Visible-Light Communication System using LED Lights,” IEEE Transactions on Consumer Electronics, vol. 50, 2004, pp. 100-107.

[5] It Ee Lee, Moh Lim Sim, and Fabian Wai-Lee Kung, “Performance Enhancement of Outdoor Visible-Light Communication System Using Selective Combining Receiver,”. IET Optoelectron, vol. 3, 2009, pp. 30-39.

[6] Muhammad Shahin Uddin, Jae Sang Cha, Jin Young Kim, and Yeong Min Jang, “Mitigation Technique for Receiver Performance Variation of Multi-Color Channels in Visible Light Communication,” Sensor, 2011, pp. 6231-6144.

[7] Risto Nordman, “Soft decision decoding of the orthogonal complex MIMO codes for three and four transmit antennas,” Physical Communication, vol. 5, 2012, pp. 33-46.

[8] Fawaz S. Al-Qahtani, Salam A. Zummo, Arun K. Gurung, and Zahir M. Hussain, “Spectral efficiency of maximum ratio combining (MRC) over slow fading with estimation errors,” Digital Signal Processing, vol. 20, 2010, pp. 85-96.

[9] Andrea Goldsmith, Wireless Communications, Cambridge University, 2005, ISBN: 0521837162.

[10] Joscph M. Kahn, and John R. Barry, “Wireless Infrared Communications,” Proc. IEEE, vol. 85, 1997, pp. 265-298.

[11] Changxin Fan, and Lina Cao, “Principles of Communications the Seventh Edition,” Beijing: National Defence Industrial Press, 2012, pp. 191-198.

 ISBN: 978-1-61208-356-8

Driver Body Information Analysis with Near-miss Events

Momoyo Ito
Institute of Technology and Science,
The University of Tokushima,
Tokushima, Japan
momoito@is.tokushima-u.ac.jp

Kazuhito Sato
Faculty of Systems Science and Technology,
Akita Prefectural University,
Yuri-Honjyo, Akita, Japan
ksato@akita-pu.ac.jp

Minoru Fukumi
Institute of Technology and Science,
The University of Tokushima,
Tokushima, Japan
fukumi@is.tokushima-u.ac.jp

***Abstract*—This study examines safety verification behaviors associated with near-miss events at nonregulated intersections with poor visibility. From an assessment of a driver's eye-gaze movements and facial orientation associated with the sudden appearance of bicyclists encountered while approaching a nonregulated intersection, we attempt to analyze the distinctive motion of the safety verification behaviors before and after near-miss events. Finally, the experimental results suggest that the sudden appearance of a bicyclist in the vehicle path has an increased chance of becoming a near-miss event. Specifically, the trajectory characterized by the eye-gaze movements and facial orientations of the driver before and after near-miss events show close correlation with the interview results of daytime on a sunny day.**

Keywords-driving behavior analysis; facial orientation; eye-gaze movements; near-miss event.

I. INTRODUCTION

In Japan, the proportion of fatal bicycle accidents is low. However, there are about 120,000 bicycle-related nonfatal casualties, which is about 20% of the total vehicle-related casualties [1]. Although, from a recent survey [2], the proportions of casualties appear to be decreasing, a serious situation exists.

In this study, we focus on near-miss events and safety verification behaviors at nonregulated intersections. Specifically, by devoting attention to the facial orientation and eye-gaze movements of a driver before and after near-miss events involving the sudden appearance of a bicyclist at an intersection, we attempt to analyze the causal relations between the distinctive safety verification behaviors, and interview .

According to statistics, about 3/4 of the human errors in crossing collisions in nonregulated intersections where many crossing collisions have occurred involve cognitive distraction states. However, cognitive distraction depends on changes in a driver's internal state, and external observation of the distraction is difficult. A method of estimating distraction has not yet been established.

Several methods of detecting distraction have been proposed by Dong et al. [3] and classified into four types based on the modalities of measurement [4]. The application of what Dong et al. defined as "subjective evaluation" and "physiological information" for actual driving situations is not realistic because of the burden these methods place on the driver. Furthermore, the "operation information" that the authors defined as steering or braking lead to accidents directly, and when the distracted state appears in the operation information, there is a possibility to miss the accidents avoiding.

In related studies of the estimation of driver distraction, Abe et al. [5] examined what they referred to as the driver's "thinking state", and Honma et al. [6] discussed what they referred to as the "blank states". These studies not only have confirmed the basic characteristic scene wherein discovery delay occurs but also overlooked changes in ambient conditions. In this study, we focus on the following three types of distraction states, i.e., thinking states and blank states reflecting how attention resources are distributed, and "impatience and frustration states" where the driving task occurs under time constraints.

In Section II, experimental procedure is described. In Section III, we show the experimental results focusing on facial orientation and eye-gaze movements. Section IV concludes this study and indicates the future plans.

II. EXPERIMENTAL PROCEDURE

A. *Experimental system*

This study employs a driving simulator (DS), which is able to freely design load environments and traffic conditions and to measure driving behaviors under different environments.

Fig. 1 illustrates the experimental system architecture. The DS used for these experiments is composed of a control device, the size of which is equal to a standard-sized vehicle,

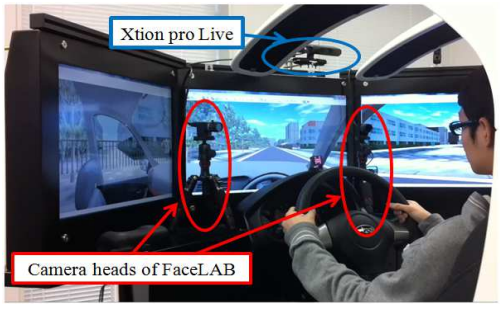


Figure 1. Experimental system for measuring driver behaviors.

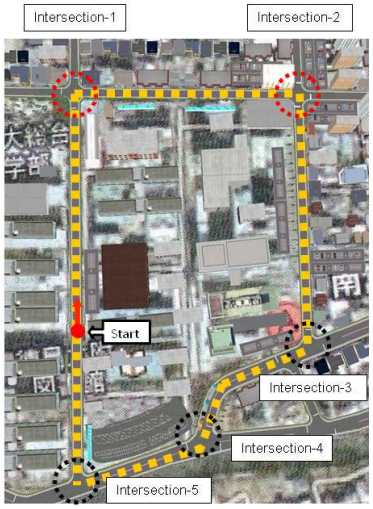
 ISBN: 978-1-61208-356-8

and a compact 6-axis motion platform (SUBARU design). The system includes three color liquid crystal displays (LCD) at the front of the cabin, and these are capable of freely displaying the horizontal viewing angle and driving environments. To achieve unconstrained monitoring and measuring of the driver's head motions, facial orientations, and eye-gaze movements, we set two camera heads on the right and left side of the center LCD, and an infrared pod at the upper center front of the cabin, as shown in Fig. 1.

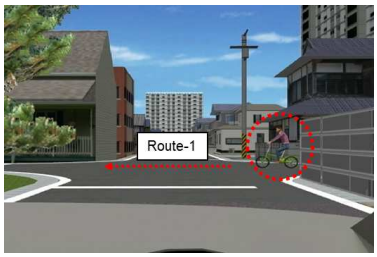
B. Near-miss events and running scenarios

Fig. 2 depicts the driving route and the two definitions of near-miss events due to the sudden appearance of a bicyclist at an intersection. At intersection-1, a bicyclist suddenly appears from the right front of the vehicle and passes through sthe intersection (route-1). At intersection-2, a bicyclist appears suddenly from the left front side of the vehicle and interferes directly in the vehicle's path (route-2).

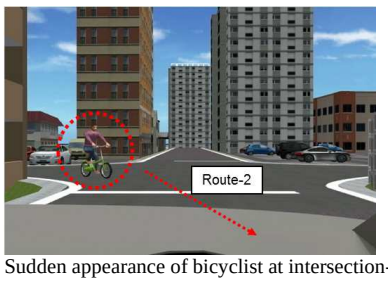
Subsequently, an overview of the various running scenarios is as follows. The fundamental scenario is to run three laps of the driving route described in Fig. 2-(a). We performed a control of the near-miss events as follows. The first lap is made without near-miss events at intersections. During the second lap, the driver encounters the crossing bicyclist indicative of route-1 at intersection-1, in addition to the bicyclist defined by route-2 at intersection-2. Specifically, we constructed four types of running scenarios at varying



(a) Running route for simulation

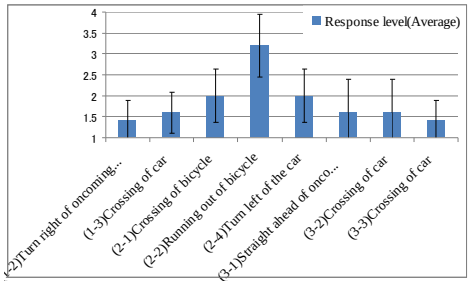


(b) Sudden appearance of bicyclist at intersection-1

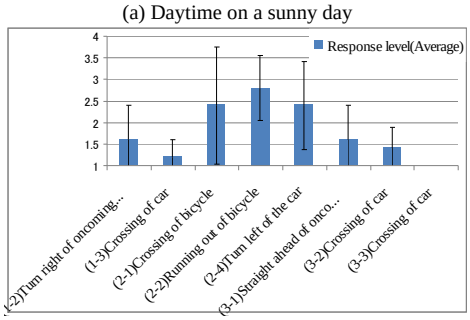


(c) Sudden appearance of bicyclist at intersection-2

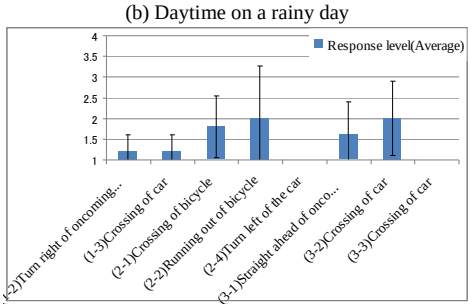
Figure 2. Simulation course with two types of near-miss events (Route-1/Route-2).



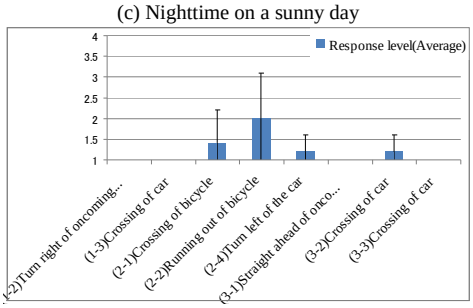
(a) Daytime on a sunny day



(b) Daytime on a rainy day



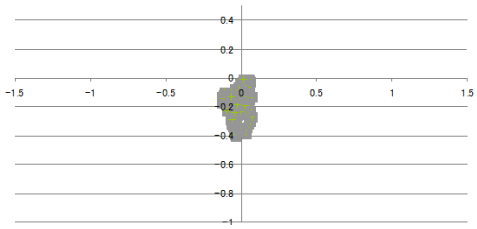
(c) Nighttime on a sunny day



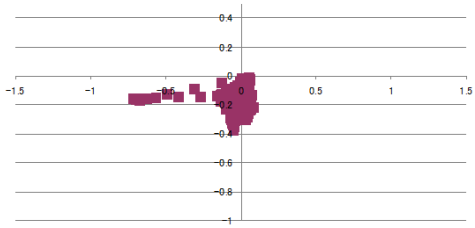
(d) Nighttime on a rainy day

Figure 3. Interview results of traffic events at intersections (1, 2, and 3).

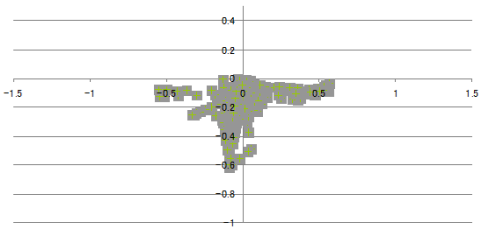
 ISBN: 978-1-61208-356-8



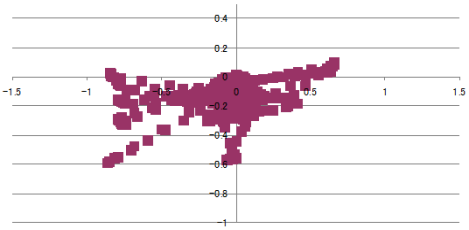
(a) Eye-gaze movements from approach to intersection-1 to pause



(b) Facial orientations from approach to intersection-1 to pause

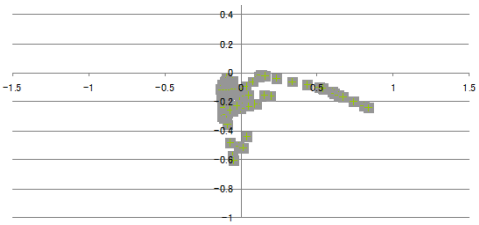


(c) Eye-gaze movements from safety check to turn right

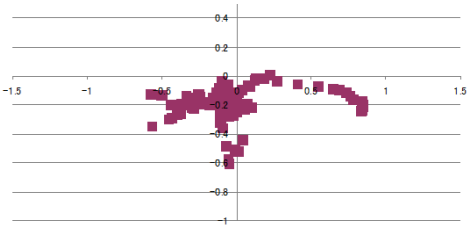


(d) Facial orientations from safety check to turn right

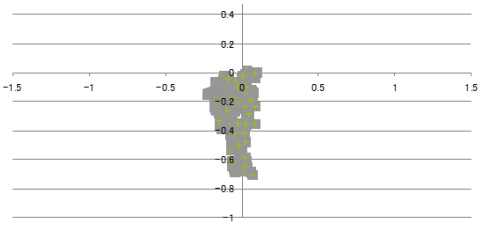
Figure 4. Sudden appearance of bicyclist crossing right to left at intersection-1 (subject C: daytime on a sunny day).



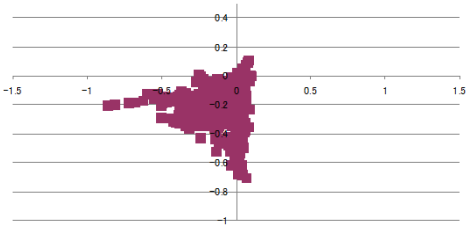
(a) Eye-gaze movements from approach to intersection-2 to pause



(b) Facial orientations from approach to intersection-2 to pause



(c) Eye-gaze movements from safety check to turn right



(d) Facial orientations from safety check to turn right

Figure 5. Sudden appearance of bicyclist interfering with the vehicle path at intersection-2 (subject C: daytime on a sunny day).

times of the day (i.e., day or night) and under various weather conditions (i.e., sunny or rainy) by designing near-miss events involving the two intersection types, which control the timing and routes for suddenly appearing bicyclists.

III. EXPERIMENTAL RESULTS AND DISCUSSION

The subjects were 5 men (Subject D and Subject E were 21 years old, whereas subjects A, B, and C were 22) and one woman (Subject F was 23 years old), all of whom were university students and had driver's licenses for ordinary vehicles. During test runs, we instructed all subjects to maintain standard speeds and traffic rules, such as pause, etc.

A. Interview results of near-miss events

We interviewed all subjects to assess the response level of surprising near-miss events occurring in nonregulated intersections (i.e., intersection-1, intersection-2, and intersection-3) using a method including four scenarios. The interview results in sunny-daytime, rainy-daytime, sunny-nighttime, and rainy-nighttime are shown in Fig. 3-(a), Fig. 3-(b), Fig. 3-(c), and Fig. 3-(d), respectively. The results indicate that the route-2 near-miss event in which a bicyclist

 ISBN: 978-1-61208-356-8

suddenly appears in the path of the vehicle, as shown in Fig. 2-(c), was confirmed as the most effective for all subjects.

B. Analysis of facial orientation and eye-gaze movements

We targeted near-miss events at nonregulated intersection-1 and intersection-2 with poor visibility, and classified driver behavior according to two segments. The first and the second segments are defined as the period from the approach to the intersection to pausing for the oncoming bicyclist and the period from the safety verification to turning right to avoid collision, respectively. Focusing on the gaze movements and facial orientation measured by the head-gaze tracking device (FaceLAB), we portray the results of intersection-1 and intersection-2 for Subject C driving on a sunny day as scatter diagrams, as shown in Fig. 4 and Fig. 5, respectively.

First, we focus on the eye-gaze movements and facial orientations in the case of encountering the sudden appearance of the bicyclist defined by route-1, as shown in Fig. 4. In the first segment, the driver is aware of the stop position of intersection-1, gazing toward the front of vehicle path. The scatter diagrams of gaze movements and facial orientation are concentrated in the center of the screen, as shown in Fig. 4-(a) and Fig. 4-(b). In the second segment, the driver visually confirmed the crossing bicyclist by safety verification after a pause, and then the driver was tracking the bicycle route. The scatter points of the facial orientations are distributed to the left and right of the screen, as shown in Fig. 4-(c) and Fig. 4-(d). We believe that these results quantitatively support the safety verification behavior against the near-miss event defined by route-1.

Next, Fig. 5 shows the scatter diagrams of eye-gaze movements and facial orientations in the case of encountering the sudden appearance of the bicyclist defined by route-2. During the first segment, the driver visually confirmed that a bicyclist suddenly appeared from the front left of the vehicle path, and then the driver was tracking the bicycle route. The scatter diagrams of gaze movements and facial orientations form trajectories that are continuous from the center of the screen to the right side, as presented in Fig. 5-(a) and Fig. 5-(b). In the second segment, safety verification behavior after a pause is concentrated on the left side of the intersection where the sudden appearance of the bicyclist has occurred. The scatter points of the facial orientations formed a trajectory that is continuous to the left side while concentrating on the center of the screen, as shown in Fig. 5-(d). This result quantitatively supports the safety verification behavior against the near-miss event involving the sudden appearance of a bicycle along route-2.

IV. CONCLUSION

In this study, we concentrated on the facial orientation and eye-gaze movements of a driver before and after near-miss events involving the sudden appearance of a bicyclist at a nonregulated intersection. Our results provide the following points.

1) The sudden appearance of bicyclists in a vehicle's path indicates that the probability of a near-miss event is likely to be high.

2) The trajectory characterized by the eye-gaze movements and facial orientations of the driver before and after near-miss events show close correlation with the interview results of daytime on a sunny day.

In future work, we will add the number of subjects and experiments, and a quantitative analysis of relation between the near-miss event and the safety verification behavior will be carried out. Moreover, we are planning analysis of influence of the driving style and the workload sensitivity for driving behaviors.

REFERENCES

[1] The Tokyo Metropolitan Police Department, "The situation of bicycle accidents," http://www.keishicho.metro.tokyo.jp/toukei/bicycle/bicycle.htm (in Japanese) (April, 2014).

[2] The National Police Agency, "Statistical information of traffic accidents in Japan," https://www.npa.go.jp/toukei/index.htm (in Japanese) (May, 2014).

[3] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," IEEE Trans. on Intelligent Transportation Systems, vol. 12, no. 2, pp. 596–614 (2011).

[4] T. Hirayama, K. Mase, and K. Takeda, "Timing analysis of driver gaze under cognitive distraction toward peripheral vehicle behavior," The 26th Annual Conference of the Japanese Society for Artificial Intelligence, pp. 1–4 (2012).

[5] G. Abe, K. Kikuchi, R. Iwaki, and T. Fujii, "Effects of cognitive distraction on driver's visual attention," Transactions of the Japan Society of Machanical Engineers. C, vol. 76, no. 767, pp. 14–20 (2010) (in Japanese).

[6] R. Honma, G. Abe, K. Kikuchi, R. Iwaki, and T. Fujii, "Characteristics of visual attention while driving under the state of drowsiness," Journal of Society of Automotive Engineeris of Japan, vol. 42, no. 5, pp. 1217–1222 (2011) (in Japanese).

[7] M. Ishibashi, M. Okuwa, and M. Akamatsu, "Development of Metrics for Driver's Individual Characteristics", Matda Technical Review, no. 22, pp. 155-160 (2004) (in Japanese).

 ISBN: 978-1-61208-356-8

Ambient Sensor System for In-home Health Monitoring

Toshifumi Tsukiyama

School of Information Environment
Tokyo Denki University
Inzai, Chiba, Japan
Email: t-tsuki@mail.dendai.ac.jp

***Abstract*—In this paper, we deal with a sensor-based monitoring system, which evaluates the health status of the elderly based on daily living activities and provides the forecast of an emergency situation to a local nursing center without explicit user interaction. Here, the main focus is on reasonably priced and noncontact sensors which assure direct recognition of quotidian activities. For meeting these criteria, water flow sensors that are attached to faucets in the kitchen, washroom, and to the toilet are a promising solution. An advantage of this solution is that the system can be readily installed in any type of housing. In addition, this system does not require personal data to be saved or transmitted outside. We will present initial results from some experiments.**

Keywords–Ambient monitering system; Health care; Water flow sensor; Active RFID tag; Vibration sensor; Recognition of quotidian activity.

I. INTRODUCTION

We are confronted with an increasing population of solitary elderly many of whom live in their own housing and for whom dangerous situations that may require medical attention are insidious. However, the number of caregivers available for frequent home visits is limited. Thus, new care services, such as the use of monitoring systems, are needed to cut costs in health care while still providing security and adequate medical treatment for people who live alone and are restricted physically.

There are a number of compact wearable sensors used for the detection of emergencies, such as sensors for the observations of vital signs [1]. This kind of emergency sensor has one disadvantage: it has to be worn permanently and operated actively, making it highly limited in regards to functionality and comfort. Recently, research for this aspect of health care has focused on capturing the activities of daily living by using ambient monitoring systems. Such monitoring systems can be divided into two categories: to identify short-term emergencies and long-term variations in health status. In this paper, we focus on long-term variations in the health status, in other words, a sensor-based monitoring system which evaluates the health status of an elderly person based on his/her daily living activities and provides the forecast of an emergency situation to a local nursing center without explicit user interaction.

One method for the early detection of emergency situation is monitoring systems using position sensors. The best-known representatives of position sensors are infrared-ray position sensors, which are installed in the living room, a bedroom, corridors and so on [2]-[4]. Positional information for the person is acquired from the detection place of body heat as he/she moves through the house. A model of the normal day-to-day behavior patterns is created based on the individual's at-home actions, such as their movement patterns, living room use frequency, and living room use time, which are derived from the positional data [5][6]. Problems with the individual's physical condition can be detected when there is a major deviation between the model behavior and actual behavior patterns. The development of the day-to-day behavior models with machine learning methods requires a solid database with a substantially large number of cases in order to achieve reasonable results. Therefore, the installation of such technologies takes a great deal of time.

These position sensors provide only indirect information on daily living activities. From the viewpoint of reliability, it is desirable to directly specify normal daily activities and then to detect variations which may be signs of a dangerous situation. TV cameras can observe daily activities and detect dangerous situations, but from the viewpoint of privacy, their introduction into private homes is limited. A smart meter, which is used for the billing of electricity, can also be used for activity recognition [7]. Daily variations in power consumption are recorded in the smart meter and household appliance use can be ascertained from the variations, allowing the inference of daily living activities. However, it is very difficult to analyze the variations in the current.

In this paper, we propose a monitoring system for the early detection of an emergency situation by using water flow sensors which are attached to faucets in the kitchen and washroom sinks, and in the toilet in the bathroom. These sensors assure direct recognition of quotidian activities, such as urination, kitchen work, and activities related to keeping neatness. For example, the activity of washing one's hands can be derived from the signal of a water flow sensor at the washroom sink. An advantage of this method is that rule-based methods [8][9] can be used for analysis and interpretation of the sensor data. In addition, these sensors are available at reasonable price. The system can be installed easily in any type of housing, and no interaction by the user is required. No personal data, such as photographs or video recording have to be saved in the system or transmitted outside.

This paper is organized as follows. In Section II we describe the brief outline of the proposed method and monitoring system with water flow sensors. Section III describes the technical aspect of the water flow sensors. Section IV contains the implementation of the monitoring system and experimental results. Concluding remarks and future works are given in Section V.

 ISBN: 978-1-61208-356-8

II. Overview of the Proposed Monitoring System

Many quotidian activities that are involved in maintaining a healthy life are accompanied by the usage of tap water, and there is a strong connection between such activities and the time and duration of tap water use. For example, a large volume of tap water is used for cooking and washing-up before and after a meal. Therefore, based on the duration of tap water use at each faucet at particular times, day-to-day activities are directly recognizable. There are some conventions about the usage of tap water in daily living. The proposed monitoring system uses such conventions to understand the daily living activities and to deduce signs of ill health in the elderly.

The water flow sensors of the monitoring system are fixed on faucets in the kitchen and washroom sinks, and on the toilet in the bathroom in the user's house, and the monitoring system records the duration of tap water use at one-hour intervals. The knowledge and conventions used in the monitoring system are as follows. According to medical knowledge, humans urinate an average of five or six times a day. If there is much less or much more frequent use of the toilet, then some diseases are suspected [10]. It is possible to infer this by checking the frequency of water usage in the bathroom. If the user is a healthy person, he/she washes his/her face and rinses his/her mouth when he/she gets up in the morning and before going to bed at night. He/she also performs activities, such as hand washing during the day. Such activities can be inferred based on the data from the water flow sensor at the sink. In addition, at meal times, people use a large volume of water in the kitchen for cooking and washing dishes. Some peaks of water usage appear in the distribution of the use duration in the kitchen. Based on these conventions and this knowledge, the health status of the user can be inferred.

The model for the normal water usage is created based on such conventions. When there are major deviations between the model and actual usage patterns, problems with the user's physical condition are detected. The reasoning can be based on a decision tree which is organized using the conventions and knowledge of water usage. At midnight, the reasoning program checks the time and duration of water use at every faucet. If there is a major deviation between the model and the activities of the day it reports the forecast of an emergency situation to a local nursing center. Then, a caregiver will visit the user to verify his/her condition.

III. Equipment for the Monitoring System

This section describes the technical implementation of the proposed monitoring system. We considered the following factors for the system design. The two most important factors are the selection of suitable sensors and software technical realization. The sensors must be reasonably priced and must assure direct recognition of quotidian activities. The sensors must also be easy to install in any type of housing without high-cost remodeling. A wireless connection is also an important factor. In addition, the monitoring system must not be operated or configured by the user. No personal data should be saved in the system or transmitted outside.

Fig.1 illustrates the functional block diagram of the monitoring system. The monitoring system consists of two main elements: water flow sensors attached to a water pipe near a faucet and a notebook computer that is placed in the house of the user. The sensors are linked with the computer wirelessly. The computer collects data from the sensors, processes it, and sends the report to the nursing center via the Internet at the fixed time.

The water flow sensor consists of a vibration sensor and an active Radio Frequency IDentification (RFID) tag. Fig.2 illustrates the prototype of the water flow sensor. Mechanical vibration in the range of 1000 to 1500Hz occurs at the water pipe near the faucet while tap water is running. The vibration sensor is designed to be attached to the water pipe near a faucet to pick up the mechanical vibrations. A ready-made vibration microphone used for tuning musical instruments can be applied to the vibration sensor, which is available for about 10 euros in the market. Fig.3 shows an example of the vibration microphone clipped to a water pipe near a faucet.

The active RFID tag has a unique ID code, which provides information about its location in the house. When the microphone detects mechanical vibrations from the faucet, the active RFID tag is switched on. Consequently, the tag sends a radio frequency signal (315MHz) with its ID code. The signal is transmitted at one-second intervals while the water continues to flow through the faucet. The tags can send signals to a range of up to about 10m, which is sufficient for collecting the ID codes in a normal house. The proposed monitoring system measures the time of running water from each faucet to obtain the use frequency and duration of tap water use in everyday life.

Fig.4 illustrates an RFID reader with a notebook computer. The RFID reader receives the RF signal from the tags, obtains the ID code and reports it to the computer through an RS-232c serial port. The number of ID codes received indicates the amount of time that there is running tap water and is proportional to the amount of water use because ID codes are transmitted steadily at one-second intervals. The computer accumulates the received ID codes at one hour-intervals and then the distribution of the duration of tap water use can be obtained for each faucet. The computer is responsible for recording the ID codes and evaluating the health status based on the distributions of the duration of tap water use. It also makes the report of the day at the fixed time, which includes warning messages if there is a forecast of an emergency situation.

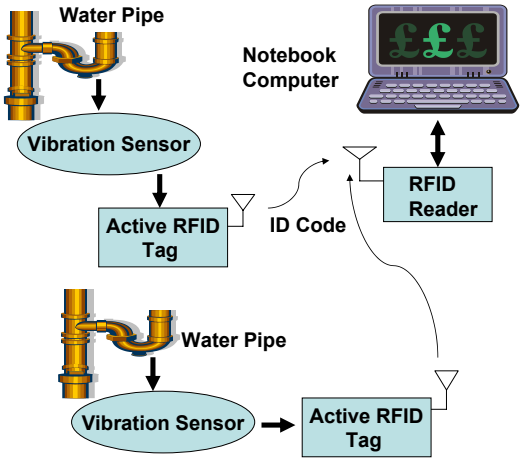


Figure 1. Functional block diagram of the monitoring system, which consists of vibration sensors, active RDIDs and a notebook compute with a RFID reader.

 ISBN: 978-1-61208-356-8

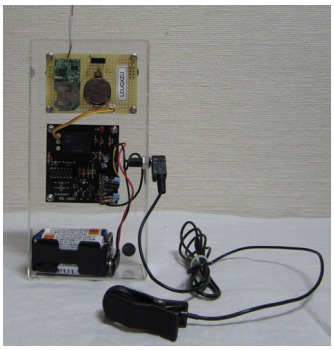


Figure 2. Prototype of the water flow sensor which consists of a vibration microphone, an audio amp., an active RFID tag with an antenna, and a battery.

Figure 3. The water flow sensor is installed at the sink. The vibration microphone is clipped to a water pipe near the faucet.

IV. Implementation and Experimental Results

The monitoring system was installed in a real housing environment and the experiment was conducted to see whether there is a strong connection between quotidian activities and the time and frequency of tap water use. The house had a living area of 108 square meters (roughly 15m by 8m) and the resident was a 65-year-old man in good health. The water flow sensors were set at the kitchen sink, the washroom sink and the toilet's flush tank. The maximum distance between the RFID reader and the water flow sensors was 7m. First, the sensitivity and reliability of the water flow sensors was checked. The sensors could detect gently running water such as would be poured into a glass.

Fig.5 illustrates an example of the distribution of the duration of water use at the toilet. The water flow sensor was set on the water pipe connected to the toilet's flush tank in the bathroom. The horizontal axis of the graph denotes time, which starts at 2 a.m., at one-hour intervals. The flow sensor transmits its ID code at one-second intervals while water is flowing into the flush tank. It continues for about 60 seconds to fill up the tank. The vertical axis shows the duration of water use for the received ID code. From the graph you can see that the resident urinated seven times on this day, at an average interval of 2.5 hours.

Fig.6 shows the graph of the duration of tap water use in the kitchen. The water flow sensor was set on the water pipe connected to the sink's faucet. The time period when the frequency of water use becomes high was observed twice. The first was during the lunch hour and the second one was at mealtime in the evening. The graph shows that the resident prepared meals twice on this day. The total time of water

Figure 4. RFID reader and a notebook computer for data acquisition and processing.

flow is relatively short because the house is equipped with a dishwasher. The peak of the distribution in the middle of the night appears to be caused by drinking tea or something.

Fig.7 shows the graph of the duration of tap water use at the washroom sink. The water flow sensor was set on the water pipe connected to the sink's faucet. This sensor is intended to detect quotidian activities, such as face washing and brushing of teeth. In other words, this is a sensor to detect activities for keeping neatness. The distribution of the duration had a peak in the morning. This means that the resident used tap water for brushing his teeth and washing his face. He also rinsed his mouth around midnight and performed activities, such as washing his hands during the day.

The model of the normal water usage was created based on the following medical knowledge and conventions. According to medical knowledge, humans urinate an average of six to eight times a day. If there is much less or much more frequent use of the toilet, then, some diseases are suspected. If you awake two or more times at midnight to urinate it is also the sign of a poor state of health. At meal times, people use a large volume of water in the kitchen for cooking and washing dishes. The time period when the use frequency of tap water is high appears in the distribution graph, and it is very likely to be observed in the morning, at around lunch time, or in the evening. You wash your face and rinses your mouth when you get up in the morning and before going to bed at night. You also perform activities, such as hand washing during the day. A large volume of water is not used at the washroom sink but the use frequency of tap water is apt to become high.

At every midnight (2 a.m.), the reasoning program checked the use frequency of tap water at the faucets. The program counted the duration of water usage at each interval as one quotidian activity if it was above a fixed threshold time. When there were major deviations between the time and number of activities and the model it was considered that some problems occurred in the user's physical condition. The reasoning for detecting the problems was done on a decision tree which was organized based on the model, and the report of the day was sent out by e-mail. In this experiment, the mail was sent to the author's office at midnight. Fig.8 shows an example of the e-mail message, which includes three items: "Urination", "Kitchen Work", and "Activities for Neatness". Each item consists of a judgment and the list of the time of the major water usage, which is used for the judgment. If there is a deviation between the normal model and the actual activity pattern the judgment of each item is denoted as "Something Wrong". If not, the message is "Normal".

 ISBN: 978-1-61208-356-8

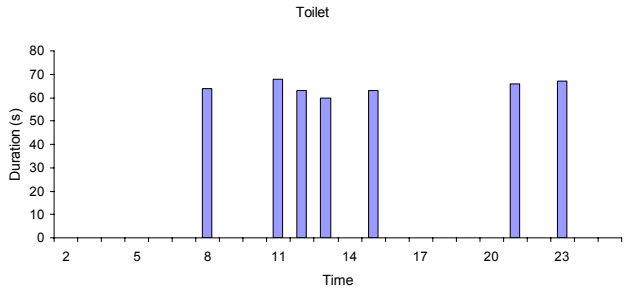


Figure 5. Distribution of the duration of water flow at the toilet's flush tank.

Figure 6. Distribution of the duration of tap water use at the kitchen sink.

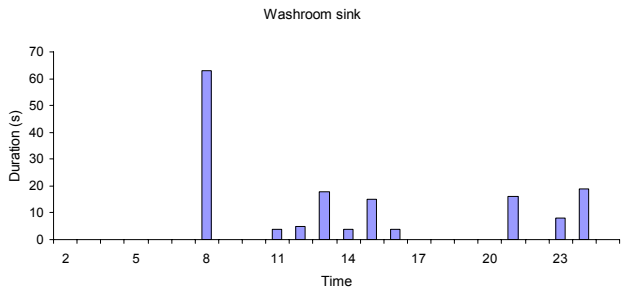


Figure 7. Distribution of the duration of tap water use at the washroom sink.

Report of Daily Living Activities
Client ID: 235476 Date: 10/14/2013

- Urination: Normal
  (Time Log of Water Usage: 9 10 13 15 16 19 22 1)
- Kitchen Work: Normal
  (Time Log of Water Usage: 12 15 19 20 22 23)
- Activities for Neatness: Normal
  (Time Log of Water Usage: 9 10 16 19 23)

Figure 8. Example of the e-mail massage which is sent to a caregiver at regular intervals.

V. Concluding Remarks

In this paper, we dealt with a sensor-based monitoring system which evaluates the health status of the elderly based on daily living activities. Here, the main focus was on reasonably priced and contactless sensors which assure direct quotidian activity recognition. To meet these demands we proposed the water flow sensors attached to faucets in the kitchen and washroom and to the toilet. The advantage of this solution is that the monitoring system can be built affordably into any kind of housing. We made a prototype of the monitoring system from electric parts which are all available in the market. The prototype of the monitoring system was checked in a real house and the experiments showed that the expected results from the water flow sensors were obtained. The initial results will be followed by more practical trials where the system will be installed in various home environments and criteria for generating automatic alert messages will be derived. This is currently in preparation.

References

[1] A. Pantelopoulos, and N. G. Bourbakis, "A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis," IEEE Transactions on Systems, Man, and Cybernetics - Part C: Application and Reviews, Vol. 40, No. 1, pp.1-12, 2010.

[2] X. H. B. Le, M. D. Mascolo, A. Gouin, and N. Noury "Health Smart Home for elders -A tool for automatic recognition of activities of daily living," 30th Annual International IEEE EMBS Conference, pp. 3316-3319, IEEE Press, Vancouver 2008.

[3] M. Skubic, R. D. Guevara, and M. Rantz "Testing Classifiers for Embedded Health Assessment," M. Donnelly, et al. Eds. ICOST 2012, LNCS vol. 7251, pp. 198-205, 2012.

[4] S. Ohta, H. Nakamoto, Y. Shinagawa, and T. Tanikawa, "A Health monitoring system for elderly people living alone" J Telemed Telecare 8, pp. 151-156, 2002.

[5] S. Aoiki, M. Onishi, A. Kojima, and K. Fukunaga "Learning and Recognizing Behavioral Patterns Using Position and Posture of Human" The 2004 IEEE Conference on Cybernetics and Intelligent Systems, pp. 1299-1302, IEEE Press, Singapore 2004.

[6] P. Chahuara, A. Fleury, F. Portet, and M. Vacher "Using Markov Logic Network for On-Line Activity Recognition from Non-Visual Home Automation Sensors" F. Paterno, et al. Eds. AmI 2012, LNCS vol. 7683, pp. 177-192, 2012.

[7] S. Chiriac, and B. Rosales "An Ambient Assisted Living Monitoring System for Activity Recognition -Results from the First Evaluation Stages" R. Wichert, and B. Eberhardt Eds. AAL-Kongress 2012, pp. 15-27, Berlin, Springer 2012.

[8] C. Marzahl, P. Penndorf, I. Bruder, and M. Staemmler "Unobtrusive Fall Detection Using 3D Images of a Gaming Console -Concept and First Results" R. Wichert, and B. Eberhardt Eds. AAL-Kongress 2012, pp. 135-146, Berlin, Springer 2012.

[9] C. A. Siebra, M. D. C. Silva, F. Q. B. Silva, A. L. M. Santos, and R. Miranda, "A Knowledge Representation for Cardiovascular Problem Applied to Mobile Monitoring of Elderly People," The Fifth International Conference on eHealth, Telemedicine, and Social Medicine, pp. 314-319, 2013.

[10] http://www.nlm.nih.gov/medlineplus/ency/article/003140.htm

 ISBN: 978-1-61208-356-8

Finding Probability Distributions of Human Speeds

Ha Yoon Song
Department of Computer Engineering,
Hongik University, Seoul, Korea
Email: hayoon@hongik.ac.kr

Jun Seok Lee
Department of Computer Engineering,
Hongik University, Seoul, Korea
Email: rolunoa@gmail.com

***Abstract*—For various research and industry purposes, knowledge of the characteristics of human mobility is required. In this paper, we will estimate the speed distribution of everyday human mobility using the unit of 0.01 m/s. From various smart mobile devices, a huge number of positioning data were collected, from which human mobile speed values are calculated. In a range of speed up to 108 Km/hour and 180Km/hour, we fit the speed data into probability distribution functions in order to establish a base for human mobility research, to which we believe this paper can make a significant contribution.**

Keywords–Speed distributions; Human speed; Human Mobility; Distribution Fitting.

I. INTRODUCTION

Monitoring human mobility is necessary for various purposes, including mobile computing and transportation engineering. Such work often requires natural science knowledge to ascertain the probability distributions of molecular mobility, wind velocity, and so on. The major purpose of this type of research is to model irregular motions into models with predictability at a certain level. Using the probability distribution of human speed, it is possible to predict human velocity, to detect abnormal motion, or to detect positioning system errors. If a monitoring system detects abrupt human mobility, this may imply an emergency or surveillance situation. There are a great number of methods for detecting abrupt human mobility. In addition, it is possible for an end-user to carry a mobile device that can report his or her position by use of embedded positioning functionalities and positioning systems. Such devices include dedicated GPS receivers and commercial smartphones, the latter using various combinations of positioning systems [1]. It is also possible to calculate distance between two points from positioning data, generally using the Haversine method [2]. From the distance and time, speed values may also be delivered from two consecutive positioning data.

The speed values impose distance information as well as the length of time interval between two positions, i.e., they are normalized values. In order to detect abrupt mobility, the speed value is key for calculating an abrupt change of position for a given time interval. However, there is at present inadequate knowledge of the probability distribution of speeds as related to everyday human life. Specifically, the question is, how can we scientifically define abrupt human mobility?

In this paper, we are going to analyze the probability distribution of speeds and will present several outstanding well-fits designed for practical use across a reasonable speed range found in everyday human life. Once we have found this distribution of speed, which forms the basis of human mobility, it will also be possible to calculate abrupt human mobility. Our aim is to provide the well-fit speed distribution of everyday human life.

In Section II, we will discuss the results of past research and will indicate the positioning data collection procedure used herein. Section III will show several important distributions that result from our fitting. We will conclude this paper in Section IV.

II. STATISTICAL BACKGROUND

There are about three previous research to be discussed as prerequisites for this research. The first one is previous research on the probability distribution for human mobile speed, the second one is positioning system technologies and positioning data collection, and the final one is data fitting between real world data and probability distribution.

A. Previous Research

There is very little research regarding human mobile speed and human mobile distance. In the very first research, cellular network-based location system used to collect mobile phone user's location data and analyzed mobile distance [3]. Cellular network-based location system can identify the location of mobile phone user based on cellular station and consecutive mobile user location data used to identify distance. It is found that the probability distribution function of human mobile distance is truncated power-law distribution. In addition Access Point (AP) of Wireless LAN (WLAN) can be used to analyze human mobility [4]. Pre-identified location of AP can be used and then the MAC address of each device in combination with the time analyzed to identify the mobile distance of each device. The result of this research is that human mobile distance follows log-normal distribution. Another research utilized GPS data of taxi [5]. This research reveals human mobile speed follows exponential distribution.

In our previous research [6], we tried to figure out proper probability distributions of human speeds in various categories. We tried to calculate the proper probability distributions of human speeds across various categories. The units of speeds were 0.1 m/s, 0.5 m/s and 1.0 m/s. However, this is somewhat inaccurate considering human micro-mobility. Once we have bigger unit then 1.0 m/s, it is hard to fit for continuous probability distributions. Unlike previous methodologies, we now collect more positioning data, using the more precise unit of 0.01 m/s. Our analysis is also carried out using the Kolmogorov-Smirnov test [7]. The fit between raw data and each candidate probability distribution can be found by K-S test.

 ISBN: 978-1-61208-356-8

B. Data Collection

Positioning data sets were collected for this research. The longest collection period was from March 2013 to Feb 2014. Several individuals carried their mobile devices whilst in a resting or moving state outside of their own home. The area of collection was mostly the metropolitan area of Seoul, Korea, however it included other part of Korea, as well as countries such as the USA, Canada, France, Italy, the Netherlands, Austria. A total of 2,218,020 speed values were collected.

The devices used for collection were as follows: iPhone 3Gs, iPhone 4 [8], iPhone4S [9], Galaxy S3 [10], Galaxy Note2 [11], Garmin Edge 800 [12], Garmin EDGE 810 [13], Garmin 62s [14]. For dedicated devices, such as Garmin, no app is requrired; however, apps for collecting positioning data from smartphones are required. Such smartphones use hybrid positioning system rather than just GPS in order to figure out the position of devices. We thus developed positioning data collecting apps for iPhone and Android phones and some of the positioning data were also collected using commercial apps. Using a variety of collection methods was intentional since we needed to cope with every possible situation of positioning data collection to guarantee the generality of the data collection.

TABLE I. Number of data in unit of 0.01 m/s.

Speed Region (m/s)	Count
Total	2,218,020
0	355,832
0.01 - 50	1,853,045
0.01 - 30	1,836,608
2.78 - 50	860,672
2.78 - 30	844,235

As the actual speed data calculated had 12 digits below the point, we needed to prune out the data to achieve reasonable values. We chose units of 0.01 m/s, which correspond to 36 meter/hour. These values are likely to be precise enough for use as everyday human speed units. One of the clear phenomena contained in speed values is that the data set will contain a lot of zero speed values, i.e., the distribution is zero inflated. In such situations, even though speed value is not zero, but it is rather less than 0.01 m/s, such value will be treated as zero. Therefore, we decided to exclude zero values. And considering that a person can move anytime in a certain distance, for example 2.77 meter in a second, we also investigated the distribution so as to excluded speed values less than 2.77 m/s. In other words, we intentionally exclude speed values less than 2.78 m/s since we are considering only the meaningful mobile situation. We also needed to decide the upper speed bound. From various possible upper bounds, we chose the values of 30 m/s and 50 m/s, which correspond to 108 Km/hour and 180 Km/hour, respectively. We considered these values to represent reasonable limits for speed in everyday human life. We also conducted distribution fitting for the four categories of speed data. The frequency of each speed value was calculated and the Maximum Likelihood Estimation (MLE) [15] was applied in order to find the parameters of probability distribution, and then K-S statistics were obtained [7].

Table I is an overall summary of the number of data classified into four categories.

III. RESULTS

Table II - V present the results of fitting for each of the positioning data categories. Table II summarizes 20 distinguished distributions for the speed range of 0.01 m/s to 30 m/s. It contains the rank of probability distribution by K-S statistic, the name of the distributions, the K-S statistic values, and the parameters for the corresponding distributions. For example, lognormal distribution shows the best fit with statistic 0.03859 for the speed data ranging from 0.01 m/s to 30 m/s. Lognormal distribution requires three parameters and the parameter values are listed accordingly. Figure 1 shows the CDF of raw data and that of Lognormal distribution. Since these 20 distributions have statistic values lower than 0.1, they are considered a relatively good fit for practical purposes. The graphs show similar results and the similarity is proven by their statistic values.

TABLE II. Parameters for Distributions of Speeds from 0.01 m/s to 30.00 m/s.

0.01 - 30.00			
Rank	**Distribution**	**Statistic**	**Parameter**
1	Lognormal	0.03859	σ = 1.4072, μ = 0.89733, γ = 0
2	Fatigue Life(3P)	0.04272	α = 1.3456, β = 2.8829, γ = -0.18893
3	Lognormal(3P)	0.04303	σ = 1.2761, β = 0.98033, γ = -0.07749
4	Log-Logistic(3P)	0.04461	α = 1.2533, β = 2.5724, γ = 6.7623E-4
5	Phased Bi-Weibull	0.04772	α1 = 1.06, β1 = 3.619, γ1 = 0, α2 = 0.72755, β2 = 4.2933, γ2 = 2.49
6	Frechet(3P)	0.05133	α = 1.1741, β = 2.213, γ = -0.59554
7	Dagum	0.05244	κ = 0.69969, α = 1.441, β = 3.8884, γ = 0
8	Johnson SB	0.05525	γ = 1.2294, δ = 0.50855, λ = 28.515, ξ = 0.21335
9	Dagum(4P)	0.05645	κ = 0.63721, α = 1.4529, β = 4.3813, γ = 0.01
10	Log-Pearson 3	0.05663	α = 11.523, β = -0.41454, γ = 5.6741
11	Pearson 6(4P)	0.06449	α1 = 0.87855, α2 = 3.8949, β = 17.718, γ = 0.01
12	Burr(4P)	0.06806	κ = 6.9702, α = 0.89148, β = 36.702, γ = 0.01
13	Gen. Gamma(4P)	0.06837	κ = 0.75574, α = 1.1357, β = 3.8121, γ = 0.01
14	Gen. Pareto	0.06908	κ = 0.33377, σ = 3.6824, μ = 0.01
15	Pareto 2	0.06970	α = 3.0493, β = 11.325
16	Pearson 6	0.07269	α1 = 0.93678, α2 = 3.6704, β = 15.511, γ = 0
17	Weibull(3P)	0.07308	α = 0.817585, β = 4.7174, γ = 0.01
18	Burr	0.07552	κ = 13.116, α = 0.86664, β = 87.161, γ = 0
19	Gamma	0.07870	α = 0.71417, β = 7.4253, γ = 0
20	Weibull	0.07929	α = 0.83261, β = 4.7886, γ = 0

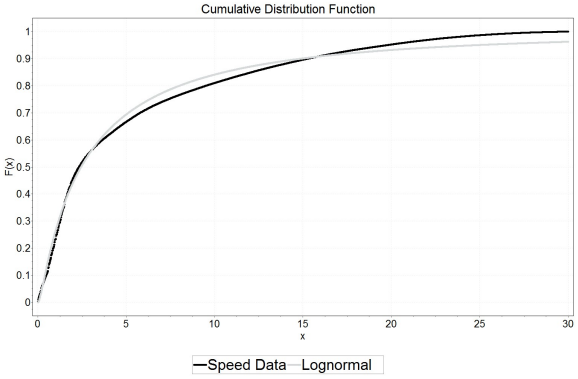


Figure 1. Lognormal Distribution from 0.01 m/s to 30.00 m/s.

TABLE III. Parameters for Distribution of Speeds from 0.01 m/s to 50.00 m/s.

0.01 - 50.00			
Rank	**Distribution**	**Statistic**	**Parameter**
1	Lognormal	0.03669	σ = 1.4235, μ = 0.92117, γ = 0
2	Log-Logistic(3P)	0.04015	α = 1.2355, β = 2.6226, γ = 0.00216
3	Lognormal(3P)	0.04172	σ = 1.3013, μ = 0.99766, γ = -0.07077
4	Fatigue Life(3P)	0.04462	α = 1.3809, β = 2.9546, γ = -0.18301
5	Phased Bi-Weibull	0.04706	α1 = 1.0588, β1 = 3.6587, γ1 = 0, α2 = 0.74503, β2 = 4.3098, γ2 = 2.48
6	Pearson 6(4P)	0.04929	α1 = 0.93411, α2 = 2.7027, β = 10.366, γ = 0.01
7	Frechet(3P)	0.05082	α = 1.1418, β = 2.2015, γ = -0.56667
8	Dagum(4P)	0.05108	κ = 0.62862, α = 1.4854, β = 4.3338, γ = 0.01
9	Dagum	0.05157	κ = 0.72888, α = 1.3983, β = 3.8113, γ = 0
10	Log-Pearson 3	0.05609	α = 13.024, β = -0.39444, γ = 6.0584
11	Burr(4P)	0.05750	κ = 5.6601, α = 0.8748, β = 29.514, γ = 0.01
12	Gen. Pareto	0.06641	κ = 0.37942, σ = 3.6578, μ = 0.01
13	Pareto 2	0.06701	α = 2.6758, β = 9.8706
14	Pearson 6	0.06824	α1 = 0.96916, α2 = 2.8621, β = 11.213, γ = 0
15	Burr	0.07062	κ = 5.3507, α = 0.89947, β = 27.138, γ = 0
16	Weibull(3P)	0.07084	α = 0.7923, β = 4.8503, γ = 0.01
17	Gamma	0.07284	α = 0.65148, β = 8.5636, γ = 0
18	Weibull	0.07926	α = 0.81451, β = 4.9539, γ = 0
19	Gen. Gamma(4P)	0.08034	κ = 0.89068, α = 0.82445, β = 6.6063, γ = 0.01
20	Gamma(3P)	0.07084	α = 0.70851, β = 7.8447, γ = 0.01

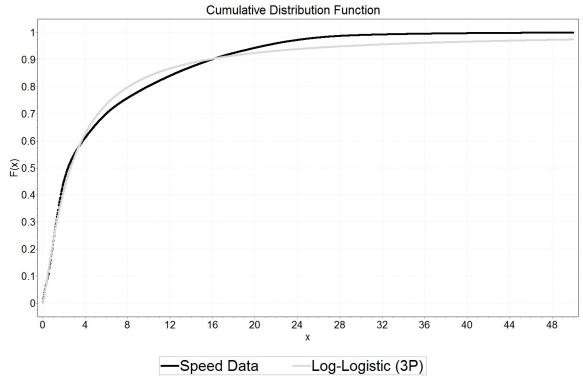


Figure 2. Log-Logistic(3P) Distribution from 0.01 m/s to 50.00 m/s.

Table III summarizes 20 distinguished distributions for the speed range of 0.01 m/s to 50 m/s. Again, lognormal shows the highest rank along with log-logistic(3P) and lognormal (3P) distribution with even smaller statistic values.

It is clear that these distributions must have different parameters than in the case for the speed range of 0.01 m/s to 30 m/s. That is, there are slight shifts in the parameters of the distributions. However we successfully deduced the common probability distribution for human mobility speeds. Figure 2 shows CDF of raw data versus CDF of Log-Logistic (3P) in the speed range of 0.01 m/s - 50.00 m/s. Phased Bi-Weibull distribution is composed of two independent Weibull Distribution over two disjoint domain.

Table IV summarizes 20 distinguished distributions for the speed range of 2.78m/s to 30m/s. Distributions such as Kumaraswamy as shown in Figure 3 is the most distinguished one.

TABLE IV. Parameters for Distributions of Speeds from 2.78 m/s to 30.00 m/s.

2.78 - 30.00			
Rank	**Distribution**	**Statistic**	**Parameter**
1	Kumaraswamy	0.01883	α1 = 0.77339, α2 = 2.0933, a = 2.78, b = 31.475
2	Gen. Gamma(4P)	0.02436	κ = 2.0551, α = 0.34749, β = 16.827, γ = 2.78
3	Beta	0.02918	α1 = 0.67033, α2 = 1.8405, a = 2.78, b = 30.349
4	johnson SB	0.03251	γ = 0.87714, δ = 0.67248, λ = 27.818, ξ = 2.3027
5	Gamma(3P)	0.03949	α = 0.93084, β = 8.0734, γ = 2.78
6	Weibull(3P)	0.04194	α = 0.99716, β = 7.4181, γ = 2.78
7	Burr(4P)	0.04217	κ = 9.3570E+5, α = 1.0387, β = 4.1967E+6, γ = 2.7778
8	Exponential(2P)	0.04502	λ = 0.13589, γ = 2.78
9	Erlang(3P)	0.04509	m = 1, β = 7.3576, γ = 2.78
10	Pearson 6(4P)	0.05484	α1 = 0.98461, α2 = 403.79, β = 2933.8, γ = 2.78
11	Fatigue Life	0.05600	α = 0.66622, β = 8.3005, γ = 0
12	Log-Pearson 3	0.05799	α = 918.16, β = 0.02129, γ = -17.436
13	Lognormal	0.05964	σ = 1.3013, μ = 0.99766, γ = -0.07077
14	Gen. Pareto	0.06220	κ = -0.25534, σ = 9.3381, μ = 2.78
15	Lognormal(3P)	0.06696	σ = 0.95749, μ = 1.6937, γ = 2.1059
16	Pearson 6	0.06816	α1 = 43.735, α2 = 2.8405, β = 0.45116, γ = 0
17	Pearson 5(3P)	0.06930	α = 2.7254, β = 18.727, γ = -0.05896
18	Pearson 5	0.06977	α = 2.6836, β = 18.223, γ = 0
19	Reciprocal	0.06991	a = 2.78, b = 30.0
20	Frechet(3P)	0.07047	α = 2.2951, β = 8.3746, γ = -2.1033

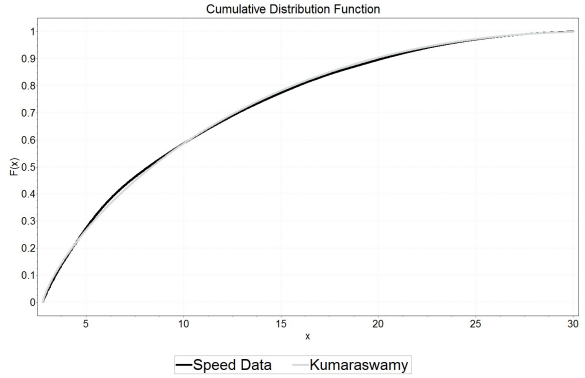


Figure 3. Kumaraswamy Distribution from 2.78 m/s to 30.00 m/s.

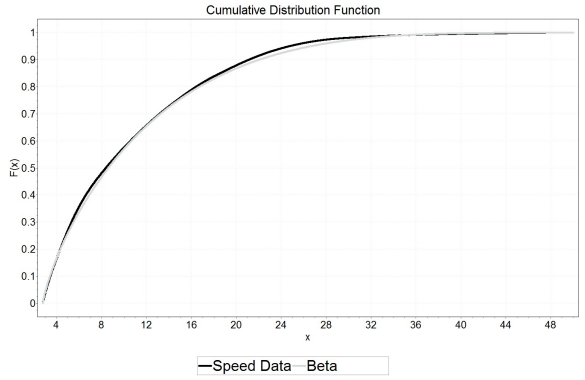


Figure 4. Beta Distribution from 2.78 m/s to 50.00 m/s.

Table V summarizes 20 distinguished distributions for the speed range of 2.78 m/s to 50 m/s. Distributions such as Beta as shown in Figure 4, Weibull(3P) as shown in Figure 5 and

TABLE V. Parameters for Distributions of Speeds from 2.78 m/s to 50.00 m/s

2.78 - 50.00			
Rank	**Distribution**	**Statistic**	**Parameter**
1	Beta	0.02177	$\alpha 1$ = 0.77961, $\alpha 2$ = 3.9036, a = 2.78, b = 51.667
2	Weibull(3P)	0.03077	α = 0.96766, β = 7.8915, γ = 2.78
3	Burr(4P)	0.03243	κ = 7.5068E+8, α = 1.0113, β = 4.7291E+9, γ = 2.7796
4	Exponential(2P)	0.03411	λ = 0.12722, γ = 2.78
5	Erlang(3P)	0.03424	m = 1, β = 7.8579, γ = 2.78
6	Gen. Pareto	0.03703	κ = -0.07541, σ = 8.3914, μ = 2.78
7	Gen. Gamma(4P)	0.03925	κ = 0.99856, α = 0.9238, β = 8.4445, γ = 2.78
8	Pearson 6(4P)	0.04931	$\alpha 1$ = 0.91756, $\alpha 2$ = 20.15, β = 161.89, γ = 2.78
9	Gamma (3P)	0.05284	α = 0.91908, β = 8.1668, γ = 2.78
10	Log-Pearson 3	0.05538	α = 205.5, β = 0.04676, γ = -7.4669
11	Fatigue Life	0.05690	α = 0.69581, β = 8.5747, γ = 0
12	Lognormal	0.05941	σ = 0.67026, μ = 2.1414, γ = 0
13	Lognormal(3P)	0.06369	σ = 1.0099, μ = 1.6995, γ = 2.2037
14	Log-Logistic(3P)	0.06389	α = 1.4469, β = 5.1199, γ = 2.6294
15	Pearson 6	0.06416	$\alpha 1$ = 86.972, $\alpha 2$ = 2.5986, β = 0.20853, γ = 0
16	Burr	0.06526	κ = 1.2717, α = 2.2942, β = 9.8812, γ = 0
17	Pearson 5	0.06531	α = 2.5322, β = 17.465, γ = 0
18	Fatigue Life(3P)	0.06538	α = 1.0624, β = 5.4173, γ = 2.0879
19	Log-Gamma	0.06617	α = 10.207, β = 0.20979
20	Pearson 5(3P)	0.06718	α = 2.4103, β = 15.982, γ = 0.18981

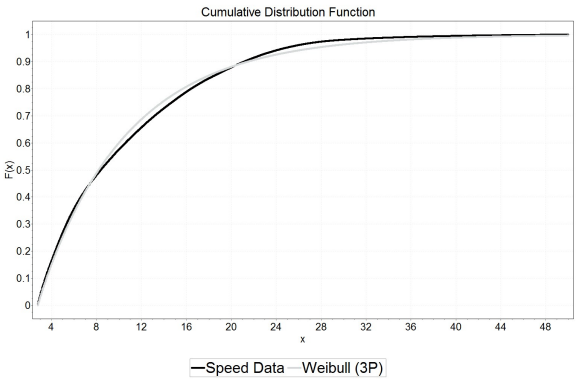


Figure 5. Weibull(3P) Distribution from 2.78 m/s to 50.00 m/s.

Burr(4P) were the best tree distributions in this speed category.

IV. Conclusions and Future Work

In this research, we showed several possible probability distributions of speed. Here, speed is used to mean the possible speed found within everyday human life. Using the positioning data sets, speed values were calculated. These positioning data sets were collected by the use of mobile positioning devices such as GPS receivers or smartphones. Volunteers carried such devices in order to collect positioning data.

It is normal that people stay at a certain place for a while; thus, our data included many speed values of zero and there was a zero inflated probability distribution. Using the unit of 0.01 m/s, we divided the range of speeds in four groups. For each category, we executed the Kolmogorov-Smirnov test to find an acceptable approximation of probability distribution. As a result, we provided several well-fit probability distributions for speed. Different from our previous research, we found better fits through using this more precise unit and with more data. One of the notable distributions is exponential distribution, which is ranked 4th in Table V in the speed range of (7.2 Km/h, 180.0 Km/h). Since it has pretty nice statistic, the exponential distribution could be used as an alternative for less strict applications for the devices with lower computational power than other complicated distributions.

We expect that this basic research will help other researchers develop or assess location based services, mobile computing, positioning devices, and others. For example, detection of positioning error is a likely use of our findings. It is well known that positioning data develops errors, which are in fact mostly due to systematic and environmental errors. More precisely, errors in positioning data in the form of (latitude, longitude) follow bivariate normal distribution. Therefore, the speed values derived from positioning data also show propagated errors. It is very hard for mobile devices to detect positioning errors and the derived errors since a user can rarely touch the underlying positioning system or change the operating environment.

One possible scenario can be described: once we have a sequence of speed values of 5.1, 5.2, 5.2, 5.3, 11.8 and 5.3 we can easily identity that 11.8 is the abrupt change of speed value and it may imply an error on the positioning tuple related to the speed value. In such a case, we need to develop sophisticated method to determine the *abruptness* of speed change and we guess that the method may be based on a statistical method. The results shown in this paper may thus be a basis for developing positioning error detection as shown in [16] where it was assumed that human mobile speed is up to normal distribution. However, more precise distribution such as lognormal distribution and beta distribution can be used as basis of erroneous positioning data detection and moreover exponential distribution can be used as a real-time application of this approach since exponential distribution is simple enough to be used on mobile devices.

V. Acknowledgments

This research was supported by the Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (NRF-2012R1A2A2A03046473).

References

[1] "iOS 7: Understanding location services," http://support.apple.com/kb/HT5594 [retrived: Jul, 2014].

[2] C. Robusto, "The cosine-haversine formula," American Mathematical Monthly, 1957, pp. 38–40.

[3] M. C. Gonzalez, C. A. Hidalgo, and A-L Barabasi, "Understanding individual human mobility patterns," Nature, vol. 453, no. 7196, 2008, pp. 779–782.

[4] M. Kim, K. David and S. Kim, "Extracting a mobility model from real user traces." in INFOCOM, vol. 6, 2006, pp. 1–13.

[5] L. Xiao, Z. Xudong, L. Weifeng, Z. Tongyu, and X. Ke, "The scaling of human mobility by taxis is exponential," Physica A: Statistical Mechanics and its Applications, vol. 391, no. 5, 2012, pp. 2135–2144.

[6] K.W. On, and H.Y. Song, "On the probability distribution of speeds derived from positioning data obtained by portable positioning devices," in Proceedings of International Conference on Advances in Mobile Computing and Multimedia. New York, NY, USA: ACM, 2013, pp. 316–320.

[7] F. J. Massey Jr, "The kolmogorov-smirnov test for goodness of fit," Journal of the American statistical Association, vol. 46, no. 253, 1951, pp. 68–78.

 ISBN: 978-1-61208-356-8

[8] “iPhone 4,” http://support.apple.com/kb/sp587 [retrived: Jul, 2014].

[9] “iPhone 4s,” http://www.apple.com/iphone-4s/specs/ [retrived: Jul, 2014].

[10] “Galaxy s3,” http://www.samsung.com/global/galaxys3/ [retrived: Jul, 2014].

[11] “Galaxy note2,” http://www.samsung.com/global/microsite/galaxynote/note2/index.html?type=find [retrived: Jul, 2014].

[12] “Garmin gpsmap800,” https://buy.garmin.com/en-US/US/into-sports/cycling/edge-800/prod69043.html [retrived: Jul, 2014].

[13] “Garmin gpsmap810,” https://buy.garmin.com/en-US/US/into-sports/cycling/edge-810/prod112912.html [retrived: Jul, 2014].

[14] “Garmin gpsmap62s,” https://buy.garmin.com/shop/shop.do?pID=63801 [retrived: Jul, 2014].

[15] F. Scholz, “Maximum likelihood estimation,” Encyclopedia of Statistical Sciences, 1985.

[16] K.W. On, and H.Y. Song, “Detection and estimation of erroneous positioning data,” Procedia Computer Science, vol. 18, 2013, pp. 2533–2536.

 ISBN: 978-1-61208-356-8

Ambient Assisted Living: Benefits and Barriers From a User-Centered Perspective

Christina Jaschinski[1,2] and Somaya Ben Allouch[1]
Research Group Technology, Health & Care
Saxion University of Applied Sciences[1] , University of Twente[2]
Enschede, The Netherlands
{c.jaschinski, s.benallouch}@saxion.nl

***Abstract*—Older adults have a growing desire to remain independent and age in their own home environment. Policy makers support this wish, as the quality and quantity of institutional care cannot be guaranteed with the present health care budget. Ambient Assisted Living (AAL) technologies can meet the interests of both parties, by facilitating healthy and active aging in the home environment. However, those technologies are still in their infancy and the likelihood of their adoption remains uncertain. By conducting a literature study and a user-requirement study for a conceptual AAL application called SONOPA (Social Networks for Older Adults to Promote an Active Life), benefits and barriers to the adoption of AAL technologies, as perceived by the elderly user, were identified. The user-requirement study consisted of focus groups and interviews with older adults and elder care professionals, conducted in the UK, France and Belgium. Together, the literature study and the user-requirement study led to several design guidelines which direct the future development process of SONOPA and related AAL applications.**

Keywords-AAL; elderly; benefits; barriers; design.

I. INTRODUCTION

Worldwide the proportion of elderly people is increasing. With 18.2% of the population being 65 years or older in 2013, Europe has one of the highest shares of elderly people in the world [1]. It is expected that this proportion will rise to almost 30% by 2050 [2]. This goes along with a sharp increase in the old-age dependency ratio, meaning that the number of potential recipients of health and pension funds rises (65 years and older), while the number of potential providers of funds belonging to the working age population (15-64 years), continues to decline [2][3]. While global aging can be considered as a great accomplishment of today's socially and technologically advanced culture, it creates immense challenges for governments in terms of healthcare regulations, pension schemes and state budgets [2].

To meet these challenges the concept of AAL was introduced. AAL is an umbrella term for innovative Information and Communication Technology (ICT) based products, services and systems which support healthy and active aging at home, the community and at work [4]. By promoting a healthy and autonomous lifestyle, AAL technologies meet both the elderly people's desire to remain independent and age in place and the demand for controlling healthcare cost [5][6].

Despite the fact that AAL technologies offer a promising perspective on successful aging, the likelihood of adoption of those technologies remains uncertain [7]. Usability problems [6][8], the lack of perceived benefits [5][7][9] and technology self-efficacy [7][10] can form, among other factors, severe barriers to technology adoption among elderly people. In our view, this heightens the need for a user-centered approach when designing AAL technologies, to access the wishes and needs of the intended user and identify potential benefits and barriers at an early stage of development.

The presented work is part of the SONOPA project [11] which is carried out in the framework of the AAL Joint Programme. The aim of the SONOPA project is to employ a set of available ICTs for stimulating and supporting activities at home. Various sensors and behavior modelling techniques will be used to generate a personal activity profile of the elderly user and track variations in the daily activities over time. When the user's activity level is low, the system will provide a recommendation, suggesting individual activities or social interactions with peers. SONOPA seeks to empower elderly people to stay active, independent and socially involved in their home environment, so to maintain their psychological and physical well-being. The work described in this paper was carried out in the first year of the project to access the perceptions and needs of the user at an early stage of development. Results are used for the further development of the SONOPA technology over the remaining two years of the project life span.

In this paper, several design guidelines for AAL technologies are identified. Findings are based on both insights from a literature study and a user-requirement study conducted as part of the SONOPA project. Section II and Section III provide an overview of the literature study with regard to the perceived benefits and perceived barriers of assisted living technologies. The user-requirement study points out to what degree benefits and barriers could be supported by our user data. Results are described in Section IV. Together, these insights led to several design guidelines which are directive for the future development process of SONOPA and related AAL applications. These guidelines are described in Section V. Section VI provides a general conclusion and implications for the future development process of the SONOPA system and related AAL technologies.

 ISBN: 978-1-61208-356-8

II. PERCEIVED BENEFITS

To get an insight in the perceived benefits and barriers of AAL technologies, relevant literature about AAL and related technology applications designed for the purpose of healthy and active aging in place was reviewed. We searched several scientific databases (Scopus, Web of Science, Google Scholar) with keywords such as 'older adult', 'assistive technology', '(ambient) assisted living', 'robots' 'monitoring', 'independent', 'adoption', 'use' or synonyms of these words. After initial screening of titles and abstracts, we included sixteen articles applying the following criteria :

- Peer –reviewed
- English language
- Published between 1999 - 2014
- Systematic review, qualitative study, quantitative study or mixed method approach
- Research focusing on the adoption and use of technology applications designed for healthy and active aging

The selected papers studied different AAL applications such as in-home sensor technologies, social network applications, domestic robots or the more general concept of technologies for aging in place (see Table 1).

TABLE I. OVERVIEW SELECTED PAPERS

Ref.	Name of Journal	Technology Application
5	International Journal of Medical Informatics	in-home sensor/monitoring technology
6	Informatics for Health and Social Care	technologies for aging in place
7	International Journal of Medical Informatics	consumer health ICT
12	Gerontechnology	in-home sensor/monitoring technology
13	Journal of Housing for the Elderly	technologies for aging in place
14	International Journal of Medical Informatics	technologies for aging in place
15	International Journal of Medical Informatics	in-home sensor/monitoring technology
16	Journal of Applied Gerontology	in-home sensor/monitoring technology
17	Proc. of the SIGCHI conference on Human factors in computing systems	in-home sensor/monitoring technology
18	International Journal of Social Robotics	domestic robots
19	Social Science & Medicine	technologies for aging in place
23	HCI and Usability for e-Inclusion	social network application
29	Journal of the American Medical Informatics Association	consumer health ICT
30	International Journal of Social Robotics	assistive social robot
31	Disability and Rehabilitation	emergency alert, door monitor, stove controle
32	Journal of Telemedicine and Telecare Telecare	in-home sensor/monitoring technology

Several studies tested the assistive device (mostly prototypes) in the field. We extracted the perceived benefits and barriers of those technologies from the selected papers and grouped them into categories. Consequently, six benefit and eight barrier categories could be identified. The benefits are discussed below. The barriers are discussed in Section III.

A. Independent Living and Aging in Place

Independent living and aging in place are perceived as essential benefits of assisted living technologies [5][12]-[16]. Steel, Lo, Secombe, and Wong [5] reported, that independence is of utmost importance to elderly people and technology which can facilitate autonomous living, is perceived as useful. This is contributed by the fact that many elderly people have a negative view on nursing homes and regard institutionalization as a last resort [5][15]. The desire for independence is so strong, that it often supersedes other concerns, such as privacy and intrusiveness [16][17].

B. Health and Safety

Health and safety are prerequisites for aging in place [15] and perceived as important benefits of assisted living technologies. Hence, responding to emergencies [5][6][12] [13][15][16][18][19]; detecting and preventing falls [5][6][12][13][15][16][19]; and monitoring medical parameters [5][6][12][16][18] are regarded as key features of those technologies. Other valued features include property security [6][18] and detecting safety hazards, e.g., fire or unlocked doors [6][15][18]. Automatic and around-the-clock monitoring is viewed as a major advantage of sensor-based assistive living technologies in comparison with existing solutions, such as an emergency button or a human caregiver [5][15].

C. Social Involvement

Another benefit of assisted living technologies concerns the improvement of the user's social involvement. Social connectedness has been described as a key element of a good quality of life [20][21] and successful aging [22]. In the 'Building Bridges' project [23], elderly people met fellow seniors via online calls and chat to discuss a broadcast they had commonly listened to. Participants stated that they were very keen to arrange real-life meetings and get to know their conversations partners. The field trial of the 'Digital Family Portrait' project, revealed that the female participant felt less lonely, knowing a family member was watching over her with the help of technology [17].

D. Support with the Activities of Daily Living (ADLs)

With older age physical, cognitive and sensory impairments such as muscle stiffness, memory decline and poor vision increase [24]-[27]. Assisted living technologies can help elderly people to compensate for these deficits and help them with their ADLs. Indeed, Smarr et al. [18] found that elderly people would value the assistance of domestic robots in helping them with chores such as cleaning, fetching objects or reminders, e.g., taking their medicine. With those

 ISBN: 978-1-61208-356-8

tasks robotic assistance is even preferred over human assistance. Similarly, Demiris et al. [6] found that older adults identify assistance with impairments and a reminder function as potential advantages of assisted living technologies.

E. Support Care Network

Both, informal caregivers and the elderly people themselves perceive assisted living technologies as good tools to support the care network because they provide some piece of mind and reduce the overall burden of family caregivers [14]-[17]. With the help of in-home monitoring, caregivers can gain a better understanding of the elderly person's well-being, and it allows them to detect functional and cognitive decline at an early stage [16][17].

F. Education and Leisure

Opposed to common stereotypes, a good proportion of elderly people are still capable of learning new things and is still fairly active and productive [28]. In the 'Building Bridges' Project [23], participants were positive about the educational element of the tested device. Several of the participants acknowledged that modern technologies could help them to develop and share their personal interest with others.

III. PERCEIVED BARRIERS

Besides benefits, eight perceived barriers which could interfere with the successful adoption of AAL technologies were extracted from the literature study. The insights on those barriers are discussed below.

A. Perceived Need and Perceived Usefulness

The subjective need and the perceived usefulness of a new technology are essential for elderly people to adopt it [7][9][14][29]-[31]. Consequently, the lack of subjective need and perceived benefits forms a major barrier to accepting assisted living technologies [5][14][16][19]. The subjective need for assisted living seems to be influenced by the elderly person's perceived well-being in terms of health, activity and social involvement. Steele et al. [5] found that elderly persons with good social ties were less likely to feel the need for such a technology. Greenhalgh et al. [19] discovered that their participants saw no value in assistive technologies if they had never needed to use it before. Zimmer and Chappell [31] indicated that the subjective health status stimulated the felt need for a technology which can improve the independence. However, many elderly people struggle to imagine future deterioration where they might benefit from features such as monitoring [16]. This is confirmed by Peek et al. [14] who concluded that many elderly people talk about a hypothetical older person who could benefit from assisted living technology rather than themselves. The use of existing technologies, such as an emergency button and the help of family members or a spouse can also reduce the perceived need for assisted living technologies [14]. This is contributed by the fact that many elderly people do not fully understand the additional benefits assisted living technologies can provide [5][19]. While the perceived benefits are more abstract, the concerns related to those technologies are very specific [14].

B. Privacy, Obstrusiveness and Controle

Concerns about privacy, security and possible intrusion are perceived as important barriers to the adoption of assisted living technologies [6][12][14][19][23]. Elderly people are worried that their personal information can get in the wrong hands and be misused. Some are reluctant to the monitoring aspect of assisted living technologies, as it feels like surveillance to them. Especially the use of cameras, is strongly rejected [6][12]. In contrast, some studies find that privacy is just a minor concern to their elderly participants [5][15][16]. They regard some loss of their privacy as a valid trade-off for their safety, independence and health. Another reason could be the lack of awareness of potential security risks.

The perceived obtrusiveness of assisted living technologies is another concern which is voiced by the elderly target group [5][14][15][19][23]. They are worried that technologies are too visible in their home environment, and could interfere with their normal routine. Indeed, some participants in the study by Van Hoof, Kort, Rutten and Duijnstee [15] complained about visible cables, annoying sounds and interference with other devices, such as the TV.

Finally, the level of user control is a matter of concern to the elderly user. Most elderly people want to have some level of control about the technology, e.g., turn it off manually. Consequently, the lack of user control is perceived as a barrier [5][14][16]. On the other hand, some elderly people argue that a monitoring system cannot assure safety, unless it is switched on all the time. Emergencies could happen when the system is switched off or when users forget to switch it back on [5]. A low level of user-control would also be more suitable for people who are not very confident in interacting with technologies [6].

C. Lack of experience, technology anxiety and self efficacy

Several elderly people are apprehensive towards technology and worry about their abilities concerning technology use [5]-[7][14][15][23][29]. They perceive technology to be very complex and inaccessible for elderly people who miss the necessary skills and experience. Ease of use and making mistakes when interacting with the technology, are major concerns. However, some of them are willing to undertake training and believe that this knowledge could make the interaction with the technology easier [5].

D. Social Stigma

A social stigma is also identified as a potential barrier to the acceptance of assisted living technologies [5][6][14][19]. Many elderly people are hesitant to use technologies which could stigmatize them as frail or needing assistance. Some admit to be ashamed of wearing existing solutions, such as a panic button and wanted sensor systems to be as discreet and unobtrusive as possible. The concern about the social stigmatization seemed to be especially prevalent for female seniors [5].

 ISBN: 978-1-61208-356-8

E. Reliability

Many elderly people worry about the reliability of assisted living technologies and question the accuracy and ability of those technologies in ensuring the health and safety of the user, compared to a human caregiver [5][6][14][15]. Indeed, several participants in an earlier study reported false emergency alarms when using an assisted living technology application [15].

F. Lack of Human Interaction

The lack of human interaction is also a matter of concern to the elderly target group. They think that assisted living technologies cannot and should not replace human assistance and human interaction [5][6][32]. Indeed, Smarr et al. [18] revealed that while robot assistance is accepted for certain tasks, human assistance is preferred for personal care tasks (e.g., wash hair), leisure activities (e.g., playing games) and most health related tasks (decide which medication to take). Van Hoof et al. [15] found that the video-call feature of their assisted living application was hardly used and did not help to improve the user's social connectedness or loneliness. Steele et al. [5] found that the elderly participants rejected the suggestion to incorporate social aspects in an in-home monitoring application as they did not believe this could impact their social life.

G. Cost

Another barrier concerns the cost of assisted living technologies [5][6][14][32]. Several elderly people have stated that, due to their limited income, such systems would either not be affordable to them, or they would not to be willing to spend a lot of money on such technologies. Elderly people also mentioned that cost should be subsidized by the government.

H. Health Concerns

Finally, the last barrier regards health concerns. Several elderly people worry that electromagnetic radiation caused by wireless sensors could cause health problems [5][14].

IV. USER-REQUIREMENT STUDY

To evaluate the perceived benefits and barriers identified from the literature study in the context of the SONOPA technology, a user-requirement study with older adults and elder care professionals was conducted.

A. Method

Three focus groups (UK: n = 8; FR: n = 5; BE: n = 9) and semi-structured interviews (n = 21) were conducted in the UK, France and Belgium. In total, 28 older adults aged between 55 and 86 (M = 71.36, SD = 9.45) participated in the study. Six older adults participated in both focus-groups and in-depth interviews. Of all participants, twelve were male and sixteen were female. Nine participants lived on their own, while the other participants lived with a partner, family members or a friend. The older adults lived independently and without the regular help of a formal or an informal caregiver. A few seniors depended on their family members or external help for certain chores such as cleaning, transport, grocery shopping or gardening. The physical well-being ranged from "perfectly alright" to "I don't feel myself at all at the moment". However, the majority felt fairly healthy. Overall, participants also felt fairly active, ranking their own activity level at an average of 7.06 (SD = 2.07) on a 10-point scale. Moreover, the majority of the older adults felt socially involved, ranking their own level of social involvement at an average of 7.32 (SD = 1,59) on a 10-point scale.

The Belgium focus group was conducted with four male and five female elder care professionals. The professionals were aged between 36 and 61 years (M = 46.50, SD = 9.89) and had an average of M = 14.44 years of work experience in the care sector (SD = 6.32).

A video was used to visualize two potential user-scenarios of the future SONOPA technology [11]. Subsequent questions targeted the following topics:

- Problems related to ADLs and the level of social involvement
- Opinion about the SONOPA solution
- General level of technical skills and design requirements for technology for elderly

The recorded material was then coded according to the benefits and barriers perceived by the participants.

B. Perceived Benefits

Almost all of the benefits found in the literature study were supported in the user-requirement study with regard to the future SONOPA technology, with the exception of the benefit 'independent living and aging in place'.

1) Health and Safety: Safety was an attribute which was highly valued with regard to the future SONOPA technology. Older adults and elder care professionals both felt that embedded sensors could provide added safety and security by detecting abnormal behavior such as falls or other emergencies, and automatically contact help. Thus, like in previous studies, 'fall-detection' and 'emergency response' were identified as key features. Automation was regarded as the main advantage in comparison with existing solutions: "I have a panic button on my mobile […]. But as far as I'm concerned it is practically useless. Because if something serious happens it is either going to be on the other side of the room, or in your hand bag, or you're not capable to press the button. So really what you are talking about, is a lot more helpful". Again, this confirms previous findings [5][15]. Another feature which was suggested to be incorporated to the SONOPA system was a reminder for turning off the stove.

2) Social Involvement: Social involvement was perceived as an important advantage of the future SONOPA technology. Participants from both groups liked that the technology would allow elderly people to make new friends and strengthen the neighbourhood network: "It's like a social club." They also valued that one could stay in touch with family and other existing contacts. Participants appreciated that contact would be one-on-one and could

 ISBN: 978-1-61208-356-8

lead to real-life interaction. They concluded that SONOPA could prevent social isolation by getting people outside the house, motivate them to participate in social life and therefore give them back a sense in life. By aiding social involvement, SONOPA could simultaneously stimulate the elderly people's activity level: "If you meet someone, you get ready, you clean the house and you get busy with other daily chores. And in this way this kind of technology could contribute to staying active". While this is in line with finding from some researchers [17][23], it contradicts findings from Steele et al. [5] who found that their elderly participants strongly rejected the suggestion to incorporate social aspects in an assisted living technologies. However, one elder care professional argued that particularly these social aspects could be the reason that the more healthy and active elderly people would be interested in SONOPA: "For some people safety would not be such a big problem at first, and if that is all there is, they probably would not get [the technology] installed. But it also includes some social elements which could maybe convince people to get it installed anyway. This way they get familiar with [the technology] […] and by the time it is needed for safety purposes than there is already a good [activity] profile of this people and that I consider a strength". Mynatt and Rogers [33] also implicated that the more technologies can be incorporated in the homes of fit elderly, the more likely they will be to adopt more advanced assistive technologies when their health declines.

3) Support with the ADLs: In line with previous studies [6][18], assistance with chores and reminders (e.g., medicine, important appointements) was much appreciated among older adults and elder care professionals in the context of the future SONOPA technology. A few older adults especially liked that there would be a possibility to get personal advice from peers or family members via video-chat. One of the elderly UK participants even suggested to use SONOPA to recruit help for chores through the network feature: "But imagine if you want to decorate your kitchen and you put it on there, you could have five people come around and you could go shopping and come back and it would all be done". Additionaly, elder care professionals and older adults found the automatic door openers which could be incorporated in SONOPA quite helpful in aiding people with mobility problems.

4) Support Care Network: The older adults stated that SONOPA could be very valuable to support the care network and provide peace of mind for the relatives. One participant regretted that a similar technology was not available when she was an informal caregiver: "When mom was older I looked after her to be sure she is well. And I think this kind of solution would have been very valuable in that situation". Again, this is line with previous findings [14]-[17].

5) Education and Leisure: Some older adults also saw the potential SONOPA social network feature as an opportunity to share common interests and educate themselves. As one elderly participant stated: "I do watercolour painting, I might find somebody who wants to come in with me once a week and sit." Another participant suggested to incorporate online classes or educational videos in the SONOPA system. Wherton and Prendergast [23] had similar findings.

A possible explanation why 'independent living and aging in place' was not explicitly mentioned with regard to SONOPA, is that SONOPA was already presented as a conceptual technology for healthy and independent aging at home. Therefore participants might have felt that this was an obvious advantage and therefore unnecessary to recall. However, various statements made clear that independence is very important to the participants. This, and the fact that it was a major advantage in previous studies lead to the conclusion that 'independent living and aging in place' indeed should be emphasized as a benefit of AAL technologies.

C. Perceived Barriers

Besides 'health concerns', all barriers identified in the literature study were supported in the user-requirement study with regard to the future SONOPA technology.

1) Perceived Need and Perceived Usefulness: Although the majority of the older adults liked the general idea of SONOPA, many felt no need for it in their current situation. They found the concept of SONOPA more beneficial for people who are less independent, active and healthy; and who are more isolated: "I mean we're not in the position at the moment to need any of those things. But thinking of other people, I think it is marvellous". In line with previous findings [16], some older adults found it hard to imagine that they might feel less healthy in the near future and would need more assistance. Like Peek et al. [14], it was observed that many older adults talked about a hypothetical older person who could benefit from SONOPA, rather than themselves. However, eleven older adults indicated that they have no need for it at the moment, but could imagine to use it in the future, when they felt less healthy and active, or in case they would lose their partner. Some older adults found that the future SONOPA technology would not offer a lot of added benefits. Several older adults indicated to already use a paper diary for overlooking their appointments, or a pill-box to remember to take their medications. However, it also became clear that the concept of the technology was still quite abstract and therefore some of the participants did not fully understand all benefits the SONOPA technology could offer to them.

2) Privacy, Obstrusiveness and Controle: In line with previous studies [12][14][19][23], participants from both groups considered the loss of privacy as a negative aspect of the future SONOPA technology. Some of SONOPA's

 ISBN: 978-1-61208-356-8

potential functionalities were also regarded as intrusive. Several elderly participants felt that the SONOPA technology would invade their personal space, and that they would feel observed: "I think it is big brother, being watched all the time". The older adults worried that they would feel restricted in their freedom and loose spontaneity: "But I don't know whether you would creep around the house, thinking oh dear they can see me […] That would be horrible, sort of spy on the wall". Some of the older adults were concerned that the data could get in the wrong hands. However, the majority of the older adults found the idea of sensors acceptable because they perceived them to benefit their personal well-being and safety at home: "When I know that the sensors are installed in my home for my well-being, I don't have any problems with them being in my home". Earlier studies found that the loss of some privacy is an acceptable trade-off for safety and health [5][15][16]. Most older adults wanted to be able to switch the future SONOPA system on and off, be aware of which data are shared and decide with whom the data are shared. On the contrary, other participants thought that the system would only work to its full potential, when it could not be switched off. Furthermore, most of the participants who were comfortable with sensors, were comfortable to have them in every room of the house as "you can fall anywhere in the house". However, a few older adults would not like to have sensors in the toilet, bathroom and bedroom.

3) Lack of experience, technology anxiety and self efficacy: The older adults were worried about the complexity of the future SONOPA technology. It was repeatedly emphasized that they did not grew up with technology and therefore, might lack the necessary skills, experience and confidence: "I think a lot of our generation are computer shy". They were worried about the potential complexity of the interface, and how much user participation is needed to operate the system: "But if you got to go to an iPod thing and should do tututututu [push buttons] before you find out what you are supposed to do, that is not helpful". Again, this confirms earlier findings [5]-[7][14][15][23][29].

4) Social Stigma: While assistance with chores was well perceived by a few older adults, others felt no need for assistance and almost felt insulted by the idea: "I don't need anybody to tell me how to make a stew". We observed that some older adults were very proud of their independence and therefore, rejected anything which would imply otherwise. Indeed, one older adult pointed out that seniors might be resistant to accepting that they need assistance and therefore, would not want to use technology that stigmatizes them as frail and dependent. This was also found by other researchers [5][6][14][19].

5) Reliability: Confirming earlier findings [5][6][14][15], older adults were concerned about the reliability of the future technology, especially the sensors. They worried that SONOPA could give false alarms: "It might just go off with your natural things". Two seniors regarded the activity recommendations as ineffective: "I am not convinced that a single technology application and especially a screen can motivate people to do stuff". One older adult stated that the technology could even work the opposite way, by providing too much assistance and making people less active because then they do not have to go outside the house to have social contact: "It could be that you shackle them behind the computer". Seniors also wondered if all parts of the system could be installed in different domestic environments: "I can't honestly visualize it to be a possibility. Not in an old house".

6) Lack of Human Interaction: Participants from both groups stated that SONOPA could not and should not replace human care and human interaction: "For me human contact is still most important […} Thus, I prefer no computer". Another participant said: "The negative point is that this person's family and the environment cannot fully rely on this application. Because the application cannot replace the human". This concern was also found in earlier studies [5][6][32].

7) Cost: Although cost came not up as a top-of-the-mind concern among the older adults, when asked about what they would be willing to pay for the SONOPA technology, it became clear that the technology has to be affordable for a person living on a pension. Several French and Belgium seniors demanded that the government would have to cover parts of the costs. Again, cost was identified as a critical issue in earlier studies [5][6][14][32].

V. DESIGN GUIDELINES

Based on the findings from the literature study and the user-requirement study conducted within the SONOPA project, several design guidelines are formulated and discussed below.

A. Clear, Specific and Flexible Benefits

To motivate people to use AAL technologies like SONOPA, it must not just offer added benefits, but at the same time those benefits have to be clear, specific and profound. Benefits which should be emphasized include: independence, safety, social involvement, support with ADLs, support of the care network and education and leisure. Keeping in mind that the intended target group is partially still very active and social, and therefore, might not feel an immediate need for an assistive technology, social, leisure and educational benefits should be further developed to target this segment. Because the concept of AAL technologies is often perceived as abstract, elderly should be able to try out or experience SONOPA without immediately being obliged to buy it.

 ISBN: 978-1-61208-356-8

B. Ensuring Privacy, Security and Unobtrusiveness

AAL technologies like SONOPA contain sensitive data such as the personal activity patterns. Measures must be taken to ensure the security of this sensitive information. Privacy concerns can be reduced by giving the user control over whether the system is active, where the sensors are placed and which data are shared and with whom. However, user control has to be weighed against the proper functionality and reliability of the system. To avoid that people forget to switch the system back on, a time limit for deactivation could be applied. Furthermore, caregivers could be informed that the system has been switched off. To counter obtrusiveness, the technology and the sensors should be embedded in the elderly people's home environment and blend with the surroundings. The system should be able to communicate wirelessly and without noise, and not interfere with other devices in the home environment. For the social network element, the use of a closed network is recommended. Finally, it should be emphasized that the monitoring feature is for the sole purpose of the elderly person's health, safety and well-being.

C. Simplicity and Familarity

The interaction with the system should be simple, consistent and easy to use and learn. The SONOPA interface has to be intuitive and clearly structured. Technical slang should be avoided and textual elements should fit the elderly's frame of reference. The challenge is to create a simple design but not limit the functionality [23].

D. Training and Low Level of Active Interaction

To simplify the interaction with AAL technologies like SONOPA, it is suggested to automate most processes and to opt for a minimal level of active user interaction, if desired by the user. Special training programs should be designed to teach the elderly how to use SONOPA and thereby improve the perceived ease of use and the confidence in their skills.

E. Emphasizing Abilities rather than Disabilities

When designing and marketing AAL technologies like SONOPA, emphasis should be put on the abilities rather than the disabilities of the target group. This can be achieved by further developing and embedding social, leisure and educational features. SONOPA's functionalities should be helpful but not patronizing and be flexible to the wishes of the still healthy and active user.

F. Reliability and Technial Support

Given that the average experience with technology in the elderly target group is rather low, robustness to mistakes is another important demand for designers to keep in mind. Furthermore, sensors should be accurate and reliable to avoid false alarms. Technical support in form of a helpline or a well-written manual should be available to all users to minimize technology anxiety and promote a successful interaction with the technology.

G. Flexibility and Adaptiveness

AAL technologies like SONOPA should be adaptive to differences in physical constraints, personal preferences, technological skills, context and environment. By offering high flexibility in content, functionalities and level of control, SONOPA can appeal to the different needs of this highly divers target group.

H. Promoting not Replacing Social Interaction

AAL technologies like SONOPA should promote and not replace social interaction. For instance, it is recommended to use a local social network so that face-to-face interaction is a possibility.

I. Low Cost and Spread Payments

Keeping in mind that the average income in parts of the intended target group is rather low, costs should fit into the available resources of the users. Also, a monthly payment scheme is recommended. Furthermore, one should keep in mind that users might expect that costs are partially covered by social security means.

VI. Conclusion and Future Work

AAL technologies can offer a promising solution to help elderly people to age independently in their own home environment, and at the same time control healthcare cost [5][6]. However, it is still uncertain if elderly people who are generally technology shy and have not grown up with technology will be ready to adopt these technologies [7][14]. This paper identified six benefits and eight barriers which are perceived by the elderly user with regard to assisted living technologies. Those benefits and barriers were found as a result of an extensive literature study and then supported and further specified by the findings of focus groups and interviews conducted within a user-requirement study. Together, findings led to the following design guidelines: (1) clear, specific and flexible benefits, (2) ensuring privacy, security and unobtrusiveness, (3) simplicity and familiarity, (4) training and low level of level of active interaction, (5) emphasizing abilities rather than disabilities, (6) reliability and technical support, (7) flexibility and adaptiveness, (8) promoting not replacing social interaction, (9) low cost and spread payments.

Our approach is not without limitations. The literature study did not follow a strictly systematic approach and therefore relevant articles might have been missed. However, a very recent systematic review on technology for aging in place was included [14]. Secondly, at this stage benefits, barriers and consequent design guidelines are based on qualitative data which were collected at an early project stage and with the use of only video scenarios. Therefore, these guidelines should be considered as an initial blueprint which will be further evaluated and specified as the SONOPA project matures.

Future work will focus on gathering quantitative data to further verify benefits, barriers and other factors relevant for the adoption of AAL technologies. Furthermore, instead of using scenarios a SONOPA prototype will be developed and evaluated in the field.

 ISBN: 978-1-61208-356-8

Although design guidelines need further specification, they form a valuable directive for the developers of SONOPA and other AAL technologies.

ACKNOWLEDGMENT

Part of this research is supported by the AAL Joint Program under contract number AAL-2012-5-187. We would like to thank our partners for their effort, insights and continuous support.

REFERENCES

[1] Central Intelligence Association. *The World Factbook-People and Society*: *European Union*. [Online]. Available from: https://www.cia.gov/library/publications/the-world-factbook/geos/ee.html, 2013.04.01.

[2] United Nations, DESA, World Population Ageing: 1950-2050, New York: United Nations Publications, 2002.

[3] European Union, Eurostat., The EU in the world 2013 – A statistical portrait, Luxembourg: Publications Office of the European Union, 2012.

[4] Ambient Assited Living Association. *About*. [Online]. Available from: http://www.aal-europe.eu/about/, 2013.04.01.

[5] R. Steele, A. Lo, C. Secombe, and Y. K. Wong, "Elderly persons' perception and acceptance of using wireless sensor networks to assist healthcare," International Journal of Medical Informatics, vol. 78, no. 12, pp. 788-801, 2009, doi:10.1016/j.ijmedinf.2009.08.001.

[6] G. Demiris et al., "Older adults' attitudes towards and perceptions of 'smart home' technologies: a pilot study," Informatics for Health and Social Care, vol. 29, no. 2, pp. 87-94, 2004, doi: 10.1080/14639230410001684387.

[7] T. Heart and E. Kalderon, "Older adults: Are they ready to adopt health-related ICT?," International Journal of Medical Informatics, vol. 82, no. 11, pp. e209-e231, 2011, doi:10.1016/j.ijmedinf.2011.03.002.

[8] A. D Fisk., W. A. Rogers, N. Charness, S. J. Czaja, and J. Sharit, Designing for older adults: Principles and creative human factors approaches. Boca Raton, FL: CRC press, 2004.

[9] A. S. Melenhorst, W. A. Rogers and E. C. Caylor, "The use of communication technologies by older adults: exploring the benefits from the user's perspective," Proc. of the Human Factors and Ergonomics Society (HFES) Annual Meeting, HFES press, Oct. 2001, pp. 221-225.

[10] A. Morris, J. Goodman, and H. Brading, "Internet use and non-use: views of older users," Universal Access in the Information Society, vol. 6, no. 1, pp. 43-57, 2007, doi:10.1007/s10209-006-0057-5.

[11] SONOPA Consortium. *Welcome to the SONOPA projec*t. [Online]. Available from: www.sonopa.eu, 20140601

[12] L. Boise et al., "Willingness of older adults to share data and privacy concerns after exposure to unobtrusive in-home monitoring," Gerontechnology: international journal on the fundamental aspects of technology to serve the ageing society, vol. 11, no. 3, pp. 428–435, 2013, doi:10.4017/gt.2013.11.3.001.00.

[13] A. Mahmood, T. Yamamoto, M. Lee, and C. Steggell. "Perceptions and use of gerotechnology: Implications for aging in place," Journal of Housing for the Elderly, vol. 22, no. 1-2, pp. 104-126, 2008, doi:10.1080/02763890802097144.

[14] S. Peek et al., "Factors influencing acceptance of technology for aging in place: a systematic review," International Journal of Medical Informatics, vol. 83, pp. 235-248, 2014, doi:10.1016/j.ijmedinf.2014.01.004.

[15] J. Van Hoof, H. S. M. Kort, P. G. S. Rutten, and M. S. H. Duijnstee, "Ageing-in-place with the use of ambient intelligence technology: Perspectives of older users," International journal of medical informatics, vol. 80, no. 5, pp. 310-331, 2011, doi:10.1016/j.ijmedinf.2011.02.010.

[16] K. Wild, L. Boise, J. Lundell, and A. Foucek, "Unobtrusive in-home monitoring of cognitive and physical health: Reactions and perceptions of older adults," Journal of Applied Gerontology, vol. 27, no. 2, pp. 181-200, 2008, doi: 10.1177/0733464807311435.

[17] J. Rowan and E. D. Mynatt. "Digital family portrait field trial: Support for aging in place," Proc. of the SIGCHI conference on Human factors in computing systems, ACM Press, Apr. 2005, pp. 521-530.

[18] C. Smarr et al., "Domestic Robots for Older Adults: Attitudes, Preferences, and Potential," International Journal of Social Robotics, vol. 6, no. 2, pp. 229–247, 2014, doi:10.1007/s12369-013-0220-0.

[19] T. Greenhalgh et al., "What matters to older people with assisted living needs? A phenomenological analysis of the use and non-use of telehealth and telecare," Social Science & Medicine, vol. 93, pp.86-94, 2013, doi:10.1016/j.socscimed.2013.05.036.

[20] C. Victor, S. Scambler, J. Bond, and A. Bowling, "Being alone in later life: loneliness, social isolation and living," Reviews in Clinical Gerontology, vol. 10, no. 4,pp. 407-417, 2000, ISSN (print) 0959-2598.

[21] Z. Gabriel and A. Bowling, "Quality of life from the perspectives of older people," Ageing and Society, vol. 24, no. 5, pp. 675-691, 2004, doi: 10.1017/S0144686X03001582.

[22] H. Bouma, J. L. Fozard, D. G. Bouwhuis, and V. Taipale, "Gerontechnology in perspective," Gerontechnology, vol. 6, pp. 190-216, 2007, doi: 10.4017/gt.2007.06.04.003.00.

[23] J. Wherton and D. Prendergast, "The building bridges project: involving older adults in the design of a communication technology to support peer-to-peer social engagement," in HCI and Usability for e-Inclusion, A. Holzinger and K. Miesenberger, Eds. Berlin, Germany: Springer, pp. 111-134, 2009.

[24] F. Schieber. "Human factors and aging: Identifying and compensating for age-related deficits in sensory and cognitive function," in Impact of technology on successful aging, K. W. Schaie and N. Charness, Eds. New York: Springer, pp. 42-84, 2003.

[25] F. I. Craik, "Memory changes in normal aging. Current directions in psychological science, ", vol. 3, no. 5, pp. 155-158, 1994, doi: 10.1111/1467-8721.ep10770653.

[26] J. Massion, "Postural control system," Current opinion in neurobiology, vol. 4, no. 6, pp. 877-887, 1994, doi:10.1016/0959-4388(94)90137-6.

[27] B.E. Maki and W.E. McIlroy, "Control of rapid limb movements for balance recovery: age-related changes and implications for fall prevention," Age and ageing, vol. 35, suppl. 2, pp. ii12-ii18, 2006, doi:10.1093/ageing/afl078.

[28] M. Ory, M. K. Hoffman, M.Hawkins, B. Sanner, and R. Mockenhaupt. "Challenging aging stereotypes: strategies for creating a more active society," American Journal of Preventive Medicine, vol. 25, no. 3, pp. 164-171, 2003, doi:10.1016/S0749-3797(03)00181-8.

[29] C. K. L. Or and B. T. Karsh. "A systematic review of patient acceptance of consumer health information technology." Journal of the American Medical Informatics Association, vol. 16, no. 4, pp. 550-560, 2009,doi: 10.1197/jamia.M2888.

[30] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, "Assessing acceptance of assistive social agent technology by older adults: the Almere model," Int J Soc Robot, vol. 2, pp. 361–375, 2010, doi:10.1007/s12369-010-0068-5.

 ISBN: 978-1-61208-356-8

[31] Z. Zimmer and N. L. Chappell, "Receptivity to new technology among older adults," Disability and Rehabilitation, vol. 21, no. 5/6, pp. 222–230, 1999, doi: 10.1080/096382899297648.

[32] A. J. Sixsmith, "An evaluation of an intelligent home monitoring system," Journal of Telemedicine and Telecare,vol. 6, pp. 63–72, 2000, doi:10.1258/1357633001935059.

[33] E. D. Mynatt and W. A. Rogers, "Developing technology to support the functional independence of older adults," Ageing International, vol. 27, no. 1, pp. 24-41, 2001, doi: 10.1007/s12126-001-1014-5.

 ISBN: 978-1-61208-356-8

Using an Autonomous Service Robot for Detection of Floor Level Obstacles and its Influence to the Gait

A Futher Step to an Automated Housing Enabling Assessment

Nils Volkening, Andreas Hein
Department of Health Services Research
Carl von Ossietzky University Oldenburg
Oldenburg (Oldb.), Germany
{Nils.Volkening, Andreas.Hein} @uni-oldenburg.de

***Abstract*—Depending on the aging society, new care concepts for older people are needed, especially the preservation of the personal mobility should be in focus. A solution could be the use of Information and Communication Technology (ICT) based solutions. A key role to preserve the autonomy and social interaction of older persons is their mobility. The prevention of fall events is a goal for the Housing Enabling (HE) Assessments by adaption of room, e.g., by detecting and removing tripping hazards. It was pointed out, that an automated HE Assessment executed by an autonomous service robot could reach a better quality and a higher acceptance. In this paper, we present the first try to detect relevant unevenness of the floor in home environments with an autonomous service robot and the resulting problems. For the gait analysis, we used a Microsoft® Kinect and for measurement of the unevenness of the floor we used the Primesense Carmine 1.08 depth sensor. First results explain which kind of influence the environment to gait parameters has (gait speed, step / stride length and the variation) and that it is mandatory to factor the conditions of the floor into an in-home gait analysis.**

Keywords-Mobile robot; gait analysis; floor level; Housing Enabling

I. INTRODUCTION

Industrial countries have to cope with different problems caused by the demographic change [1]. A possible way to cope with these upcoming problems is the use of ICT in the Ambient Assisted Living (AAL) area. There are two approaches to bring the technology to the homes of elderly people. The first and older solutions are smart homes [8], where the whole technology is integrated in the flat. The second is the field of autonomous service robots. In this case the sensors, actuators and the computational unit are mounted on a mobile base. The simplest representatives are household robots like autonomous vacuum cleaners, which raise the acceptance of users. Advanced Systems could support the caretakers and assist elderly in an independent lifestyle and preserve their indoor mobility up to a high age [2][4]. One advantage of service robots is reduced costs compared to smart homes. They need only few sensors to generate a good coverage which depends on their mobility, so they can bring them in the area of interest [3]. We will use the mobility of these platforms to realize a new approach of the HE assessment. A first step is the evaluation of the flat, for example the examination of the floor to detect stumbling risks. The rest of this paper is organized as follows. Section II gives a short motivation about the topic, followed by the state of the art and the current limitation of it (Section III). In Section IV, we present our first approach to measure the unevenness of the floor and the results in Section V. The conclusions and further steps close the article (Section VI).

II. MEDICAL MOTIVATION

Fall-related costs are one of the major factors influencing the proportionally higher costs to the health care system caused by elderly people. From a clinical perspective long-term monitoring of changes in mobility has a high potential for early diagnosis of various diseases and for assessment of fall risk [4]. As important as the age and potential diseases of the patient [5][6] are the condition of the floor for the self-selected gait velocity and in general for the risk of stumbling. Especially in an unsupervised environment the additional information about the quality of the floor could increase the precision of the gait analysis [9][10] and a precise gait analysis could be very helpful for the HE assessment to estimate the personal factor. In our approach, we try to realize both, good results for the HE assessment and also additional information for a gait analysis to increase their precision.

III. STATE OF THE ART

A. Trend Analysis of mobility in Domestic Environments

There are different approaches for gait analysis, so is it possible to upgrade a home with various sensors, especially from the home automation or security domain to a (health) smart home [13]. Most systems are used for a trend analysis and only some approaches use ambient sensors for detailed gait analysis. Various groups use home automation technologies like motion sensors, light barriers or reed contacts placed in door frames or on the ceiling other than Cameron et al. [14], which use optical and ultrasonic sensors. These were placed on both sides above the door frames to determine the walking speed and direction of a person passing. Kaye et al. [15] presented a study based on sensors covering different rooms of a flat. Also, laser range scanners are used for Time Up and Go (TUG) assessment [16] or for detailed gait analysis [31]. Poland et al. [17] used a camera

 ISBN: 978-1-61208-356-8

attached to the ceiling, recording a marked floor evenly divided into rectangles (virtual sensor). For persons within these, the approach 'activates' the virtual sensor in which they are currently located. Stone and Skubic [18] used the Kinect to analyse the gait in a home environment. Especially, the variation of different gait parameters like step length and self-selected gait speed over time was measured and identified as independent factors for the personal stumbling risk. Also, Gabel et al. [19] used the MS Kinect for a full body gait analysis which is capable for a precise in home gait analysis. A similar approach for a long time in house gait analysis by using the Kinect is published by Stone and Skubic [30].

B. Mobility Assessments Using Service Robotics

Service robots combine ideas of different fields of robotic research into one system to target at a specific application. Most available platforms are still in (advanced) research states. There are different fields of interest, e.g., acting autonomously in home environments [20], learning of environmental factors and user behaviour [21][29] and as well as robot designs itself [24]. Within our own work [23] we have recently presented a new approach to enhance mobile robot navigation in domestic environments by the use of mobility assessment data. The advantage of a mobile robot is that it can bring the needed sensor technology to the Optimal Observation sLots (OOL) for monitoring, as introduced in [24]. In the observation phase the robot stands at a safe place in the initial room of the flat and observes the human behaviour and environment. These data are used to compute new OOL, which fulfil different safety and quality criteria. After that phase the robot will travel to that OOL and measure different gait parameters by using the laser range scanner and the Kinect, which can be used in HE.

C. Housing Enabling

A quite popular assessment in the Scandinavian countries is the housing enabling assessment. It reduces the risk of fall in home environments and the near surrounding. The flats will correlate relating to the personal health status of the inhabitants [25] and the structure of the flat itself. This rating gives advice how to change the flat with its furniture etc. so that it is suitable for the resident. The housing enabling assessment is split into three parts. The first part is the descriptive part to collect some general information about the flat and the condition of the user. The second part is the evaluation of functional limitations and dependence on mobility aids. Detailed information about medical condition of the user is collected, e.g., severe loss of sight or limitation of stamina. The last part is based on different questionnaires, which are related to the flat and the surroundings. After completion of all questions, the score of the flat in relation to the actual health status of the user [27] could be computed [26]. The adaption of the flat is related to the rating [28] in order to reduce the risk of falling is also possible.

D. Determine the unevenness of the floor

There are several different building regulations [12], which identified different levels, which should not be exceed. These regulations are only obligatory for public buildings, but unevenness also influences the gait velocity [9]. To raise the validity of domestic gait analysis it is important to have detailed information about the floor. Udsatid et al. [29] used a mobile robot and a Kinect sensor to measure the ground and calculate a virtual ground plane. But, only for a background subtraction for a foot tracking algorithm, which was used by a side by side navigation algorithm. Currently, there are no mobile service robots to determine the unevenness of the floor.

E. Limitation of the State of the Art

As shown in section III-A, most of the systems use ambient sensors and do not observe the user continuously. This means, that only presence at specific known points is measured. The problem of this kind of monitoring is, that it can only be used for trend analysis instead of a detailed assessment to determine different mobility parameters of a person. For precise assessments of the mobility, laboratory equipment and a well-known surrounding are needed. On the one hand, the installation affords and costs are too high to install it in domestic homes, on the other hand homes are "floating", this means that, e.g., the furniture changes over time. All of the automated gait analyses don't respect the influence of the floor cover. Within the domain of health care and rehabilitation service robotics there are quite few systems commercially available. Further, there is no robotic system that is capable of doing HE assessments and tries to present advice to reduce the risk of falling. The current HE tests suffer from some drawbacks, e.g., the estimation of the personal disorders, the investigation and also the following adaption of the flat depends highly on the skill of the person executing the test. This could lead to different or insufficient results. Furthermore, this assessment is mostly not done as a continuously assessment, but rather as an event triggered assessment after accident. In summary, there is currently no system or approach available, that is capable of doing precise and continuous housing enabling assessments in domestic environments and using this additional information from to raise the precision of gait assessment results.

IV. APPROACH

A. Detection of Unevenness

Our own approach provided an automated and continuous detection of relevant unevenness/texture of the floor assembly, which will be used to rate the flat during the HE Assessment and to raise the quality of the gait analysis. To implement a stable algorithm in an unsupervised environment, we include at the start a self-calibration to calculate the ground level and the sensor orientation for a better error correction. This step is necessary, because it could happen that the orientation of the sensor changes a little between runs or the sensor underlies a drift over time. In this case a pre-calculated ground plain would lead to a wrong detection of relevant unevenness of the floor. In a first step we estimate the quality of the current depth image of the sensor, by calculating the Root-Mean-Square (RMS) deviation of each pixel.

 ISBN: 978-1-61208-356-8

Figure 1. Left side: Depth values from the Sensor in grayscale (White near, dark grey far away) with a 10mm doorstep in a distance of 80cm, right side: Visualisation after ground subtraction and convert to a binary image of depth values with the RMS as threshold

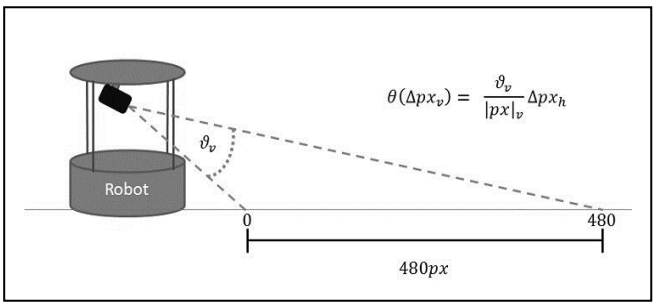


Figure 2. Schematic draw of the mobile service robot with the Primesense Sensor and the calculation of the vertical aperture angle between two points.

The median value of these results is used as a quality factor for the selection of depth points with a low noise. To calculate the virtual ground two points of the middle row and two of the middle column of the depth frame are selected which satisfy three criteria. The first is that both points have the lowest possible RMS (minimum below the quality factor otherwise use other column or row), the second is to maximize the distance between these points and the third criterion is that they don't belong to a known obstacle like walls. This information came from the navigation map of the mobile platform. In the following section we only look on the estimation of a vertical ground line, in fact the calculation of the horizontal ground line and therefore the ground plane is straight forward. After the selection of two vertical points we're able to calculate the first ground line and the vertical orientation of the sensor. Only five parameters are known, the two distance values of the two selected points and the pixel distance between both points. The vertical aperture angle of the Primesense Sensor [11] and the resolution of current depth frame are known. Figure 2 shows the aperture angle calculation of each pixel. Together with the pixel distance between the selected points we get the angle between it. For all Examples, we used a resolution of 640px times 480px, which is the highest possible depth resolution of the Primsense Carmine Sensor. Using the law of cosines, it's possible to estimate the missing parameters like the height of the sensor or the vertical angle. After the complete calculation all relevant values are known to estimate the vertical ground line. The next step is similar to the background subtraction. We use the ground line as a kind of background and calculate the difference to the current depth image. Figure 1 shows the normal depth image and a binary picture, which is generated by a root-mean-square deviation approach. If the difference is higher than the RMS, the pixel is set to 1, otherwise to 0. Now, it is easier to cluster this picture and find relevant trip hazards. Therefore many approaches are published, e.g., edge detection and many more. After we found interesting blobs (e.g. size or shape), we calculated the height of these obstacles from the depth picture and saved this information into the navigation map of the robot. After that we can use it during the scoring of the flat and to raise the precision of the gait analysis and the balance analysis on the different areas.

B. Calculate Balance Parameter

In our first approach, we use the Microsoft Kinect [31] to track the person because of the low price and the existing openNI skeleton tracking algorithm from ROS. The mobile platform does not move during the measurements, because of the specification from the openNI Algorithm. During the observation phase the timestamp, x-, y- and z- coordinates of the following skeleton joint point from the openNI tracker node will be saved:

- Foot and hip (each: left, right)
- Torso and Neck

In respect to the low processor capacity of the Turtlebot 2 netbook we used an offline approach. After the observation phase, different balance and gait analysis parameters are calculated. As a first validation the distances of the joint points are checked, whether they are between ranges of 0.80 – 3.00m, which is the effective distance of the Primesense sensor. After that, we calculate the gait speed, step and stride length and related to those values the stance and swing phase for each foot. First, we estimate the different phases from each foot during a measurement by using formula 1.

$$\lfloor x_i - x_{i+1} \rfloor_{i=0}^{n} = \begin{cases} \leq 0.02\,m, & stance\ phase \\ > 0.02m, & swing\ phase \end{cases} \quad (1)$$

This means that a foot needs a minimum acceleration of approx. 0,6m/s to mark as moving. This value reflected a compromise of literature values and a kind of error correction of the drift from skeleton tracking. After that the middle index of each phase for each foot was calculated, these are used to estimate the stride and step length. Also, the calculation of the gait speed used these indexes, by choosing the first and the last stand phase of each measurement and calculates the distance between these points. Now the corresponding timestamps are used to determine the elapsed time and by dividing the distance through the time we get the gait speed for each measurement. We used two facts to get a better reliable between measurements, the first is that the mobile robot stands on a defined OOL, so the global coordinates and the direction are nearly equal between the measurements; the second helpful point is that humans used more or less the same path between two points in the home environment. These points help to get a bigger and comparable data base from same OOL's

 ISBN: 978-1-61208-356-8

V. RESULTS

A. Detection of Unevenness

To test and verify our approach, we used the OFFIS IDE-AAL Lab, which provides a complete demo flat for first measurements in a realistic environment. As mobile platform a Turtlebot 2 (Kobuki) is used with Primesense Carmine 1.08 Sensor, which is mounted upside down below the third level of the platform and looks down to the ground with an angle of approx. 35 degrees at a height of approx. 34 cm. The resolution of the depth sensor is set to 640px times 480px and a frame rate of 30 Hz. The platform, the sensor and the fixing of both have not been changed during the measurements. To get comprehensive measuring results, we used the IDEAAL Lab and the normal office space to test our approach on different floor types. So we got results from two different carpets, laminate and PVC coating. The measurement in between two floors represents the change between coatings (laminate / carpet). To measure normalized height difference, 5 wooden footsteps layers are used. Each piece has a height of 5 mm, so that we're able to measure between uneven doorways (0mm) up to 25 mm. For each test set-up 30 single frames are saved and the mean values and the standard deviation for each pixel, to verify the precision of the sensor, are calculated. According to different buildings regulations [12], the requirement is to detect differences of a minimum of 4 mm between two surfaces. The measured minimal standard deviation is approx. 3.94 mm and the median value is 6.29 mm. This means that the precision of the Primesense Carmine 1.08 sensor is near to the required precision of 4mm. After this result we performed further tests to verify our first results. Therefore we made different measurements in the IDEAAL Lab and at the office with the wooden doorsteps. The proceeding for each measurement was the same, first we took 30 frames of the even surface, 30 frames with a 5 mm doorstep in a distance of 80 cm followed by 30 frames with 10 mm doorstep and so on until we reached the maximum of 25 mm. After that we reduced the distance to 40 cm and started over without any obstacles and then raised the doorsteps in 5 mm steps. After the measurement, we calculated the virtual ground plan and subtracted it from the different test images.

The result was unexpected; in the first approach, we had only two small areas around the selected points for calculation the ground plane, which provided good results, even for a floor without any unevenness. After a small modification (also considered in the description of the approach) in the algorithm, which selects the point for estimate the ground plane, we had a vertical ground line which only matched in the lower third of the depth picture. Figure 3 shows that in the upper two thirds the difference between the calculated ground and the real ground was too big to detect any relevant barriers. After that failure we tried to solve this problem in our approach or setup. The First step was to verify the measurements, therefore we subtracted the mean value of the even ground from the mean values of the modified ground. After these results showed acceptable results for the detection of barriers from 5mm up to 25mm, we searched for further reasons. The next test was the linearity of the sensor over the distance. If it has a linear characteristic for the depth sensor, then our approach should work in general. Therefore, we made different measurements from an even surface, a 5 mm and 10 mm barrier in a distance of 40 cm. The result in Figure 3 shows the ground and the calculated virtual ground and a 5mm doorstep obstacle. This shows that the sensor has not a perfect linear characteristic; so, it is difficult to calculate a virtual ground which is represented by a plain or line and use it for a simply background subtraction. The difference between the calculated ground and the real ground is bigger than the standard deviation, which means that we would detect false positive barriers. Also, the difference to the 5 mm footstep is only few mm above the standard deviation and in comparison to the error between real ground to calculated ground, it seems to be difficult to detected obstacles below 10 mm, but for the HE Assessment we need a resolution up to 3 – 4 mm. On the other hand, Figure 3 shows a good difference level between ground and 5 mm barrier, which points out that the choice of the points to calculate the virtual ground plane has a big influence on the further results. So, it is difficult to find a selection algorithm such that the correct points are chosen to get optimal result by minimal calculation cost.

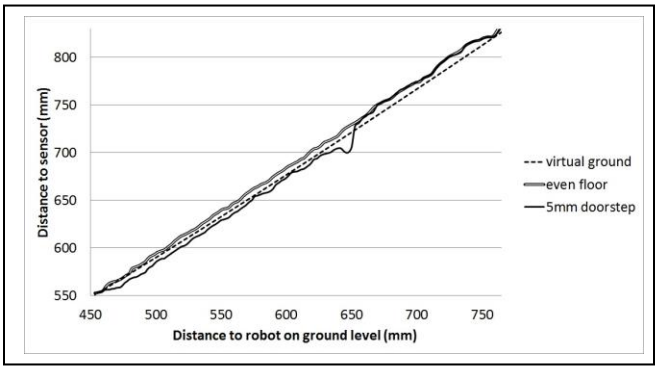


Figure 3. Visualisation of the calculated ground (dotted line) and the measurement from the ground (double line) and the 5 mm doorstep (single line).

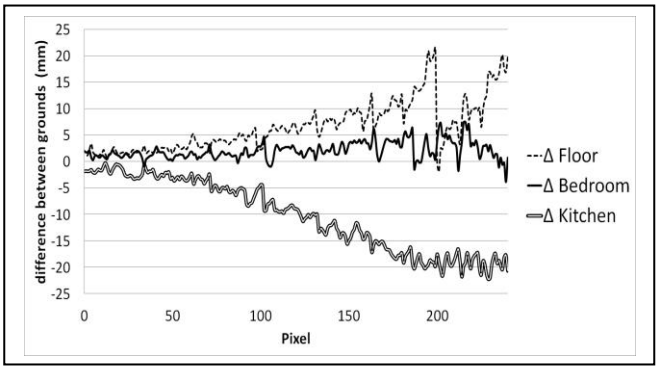


Figure 4. Visualisation of dependency of different floorings in comparison to the general mean ground value. One vertical row is plotted as reference

B. Sensor independence related to the surface

Also, the independence of the sensor compared to the flooring was tested, by measuring 4 different floor types. Two different kinds of carpet, PVC coating and laminate and the transition from laminate to carpet were tested. For each surface we made 30 single measurements and computed the mean value over all 30 single frames on pixel base. Then, we used these mean values to calculate the overall mean value for the ground. From each measurement we selected the mean values of middle column and subtracted it from the corresponding value of the overall mean depth picture. The results are shown in Figure 4 and lead us to the fact, that the different floorings have an influence on the distance values and the reliability of the sensor. As you can see on Figure 4, only the differences in the first 50 pixel, which are equivalent to a distance of 40cm to 55 cm in front of the sensor, are between the first standard deviation (about 6mm). This measurement represents a distance to the sensor between approx. 20cm to 84 cm. This result pointed out that it is advisable to calibrate the sensor for each subsurface and every day to reduce the errors during the measurement or use another model of this sensor type, e.g., the Primesense Carmine 1.09 with higher depth resolution or a complete other type of sensor to detect the unevenness of the floor.

C. Gait parameters vs. floorings

Parallel to the test for the detection of unevenness of the floor we made first measurements in a domestic environment with 5 users (two females/ three males) in the age between 42 – 76 years for a first validation of our approach to calculate gait speed, stride and step length and, when possible, to see differences between different floorings by using the Microsoft Kinect and the openNI tracker. For all measurements the Turtlebot 2 stands at a predefined position, similar to final setup when the mobile robot drives to different OOL's for measurement. Each subject has to walk 5 times in direction to the mobile robot for the same conditions. Each test person has to fulfil a test with 10 different conditions. Two different coatings (carpet / parquet), three different doorsteps (5mm, 10mm and 25 mm height) and each condition under dark and normal lighted condition. So we get a data base of 250 measurements over all conditions and subjects. The first results for the step-, stride length and self-selected gait speed (SGS) on parquet, high pile carpet and different doorsteps are presented. As you can see in Figure 5 and Figure 6 a difference between the stride length and the SGS could not be only detected for elderly persons, also for mid-aged persons, depending on the floorings. Also, it seems like as if the variation of the step- and stride length depends on the coatings. But further tests with more measurements, longer walking distances and time periods must be evaluated to verify our first results. Nevertheless evidence that the floorings have an impact on the gait analysis in the domestic environment was shown. Without the knowledge of the characteristic of the flooring, e.g., like the most classical automated approaches it could lead to false decisions related to the decreasing of the SGS on some coatings. These give first evidence that the quality of balance and gait analysis depended also on the floorings. Further tests must be made to get reliable facts, what kind of obstacles has an influence and how big is the impact.

VI. CONCLUSION

A new approach for the detection of fall relevant unevenness and a first idea of an advanced gait analysis which used this information for better results in the context of an automated Housing Enabling assessment was presented. Therefore, we used a mobile robot platform the Turtelbot 2. As depth sensor a Primesense Carmine 1.08 is used for the detection of unevenness with the original OpenNI driver v.2.1.0 and a Microsoft Kinect with the ROS openNI tracker Node for the balance and gait analysis. The Carmine sensor was mounted up-side down below the third level of the Turtelbot platform in a height of approx. 34 cm. The Kinect was mounted on the highest level (height approx. 55cm). We were able to determine the position and orientation of the sensor, only from a small knowledgebase. Our approach is aimed at calculate a virtual ground, which is the reference for barriers, because in a normal scenario it is unrealistic to have the chance to make a clean depth picture from each part of the room without any carpets on the subsurface or other stumping blocks. But, our measurements have shown that the combination of our approach with this sensor, the mounting and the needed resolution does not work in a proper way. This depends on tree facts.

- First point is the depth resolution of the sensor. The noise of the sensor is near to the values that we want to detect.
- Second point is the instability of the sensor, depending on different factors, is too big. As we can show, the floorings and the gloss of it have a big influence on the depth values. The difference is sometimes even more than the third standard deviation.
- Third point is the quality of our algorithm to select the points for the calculation of the virtual ground. We should add a validation step if the virtual ground matched with the most points. Otherwise we should select new points or to change to a spline based approach.

Finally, we could say that the Primesense Carmine 1.08 Sensor has some advantages, like the price and the relatively good resolution and a low noise in fact of the price and range. But, the quality is not high enough for this application in the frame of the housing enabling Assessment or to determine relevant unevenness of the floor.

Our second approach to use the additional information about the floorings to raise the quality of gait analysis in the domestic environments seems to be essential to generate reliable data. For the first results we could show that an influence of the flooring exists, but for final statements we have to evaluate this approach with more users and with more flooring and other important facts. The first results allow the statement that all automated gait analysis in unsupervised environments should consider the texture and unevenness of the flooring.

 ISBN: 978-1-61208-356-8

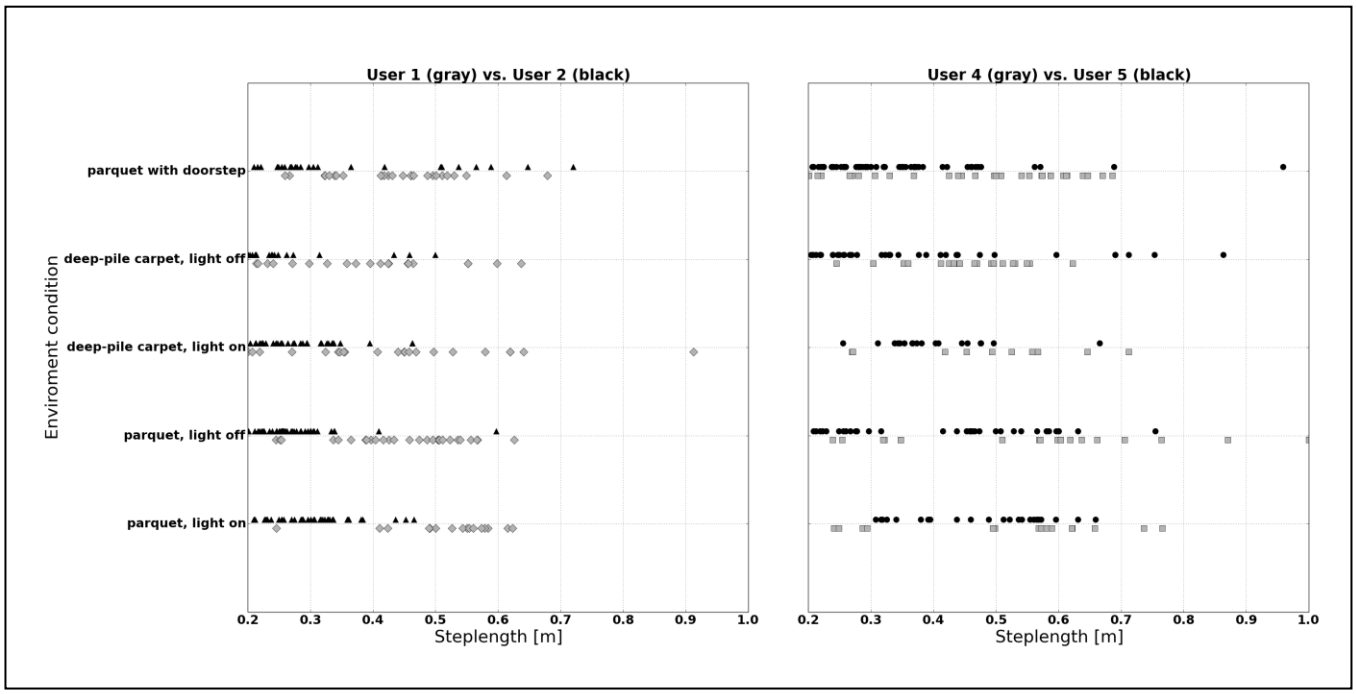


Figure 5. Influence of floor conditions to the step-length of different subjects (grey: mid-age, black: elderly). Left side: two female subjects and on the right site two male subjects.

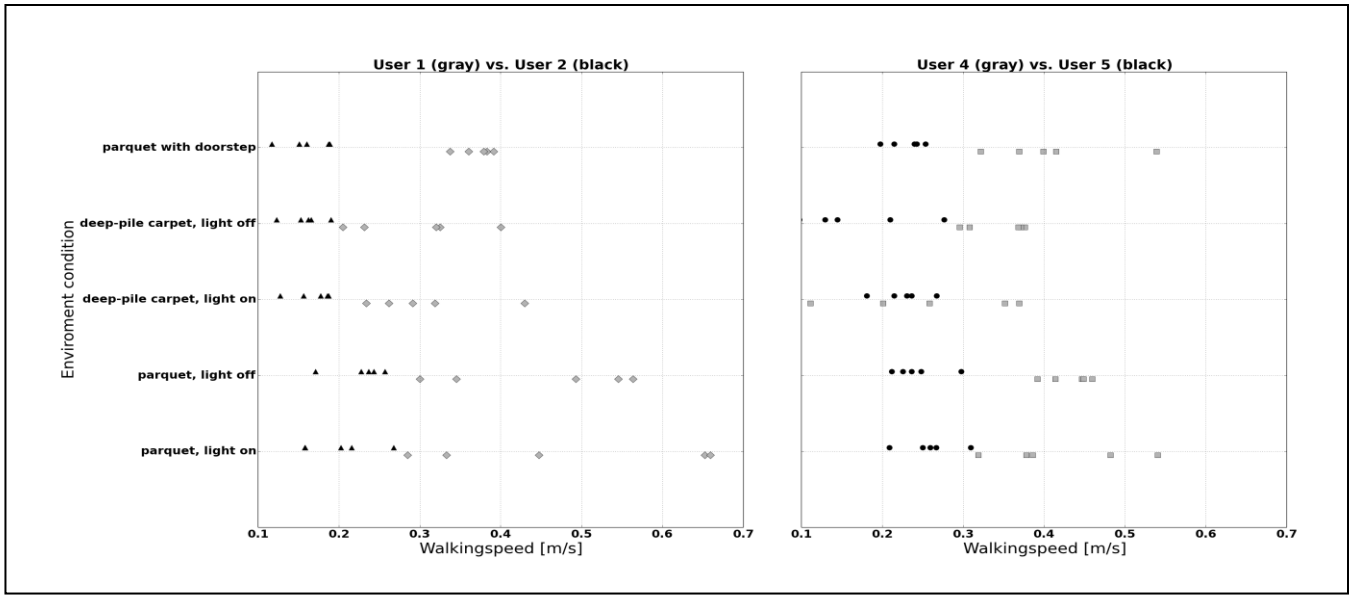


Figure 6. Influence of floor conditions to the gait speed of different subjects (grey: mid-age, black: elderly). Left side: two female subjects; Right site: two male subjects.

REFERENCES

[1] K. Böhle, K. Bopp, and M. Dietrich, “The "Artificial Companion" - a useful guiding principle for development and implementation of technical assistance systems in care arrangements?,” In Proceedings of: 6. Deutscher AAL-Kongress: "Lebensqualität im Wandel von Demografie und Technik", Berlin, VDE 2013.

[2] J. Meyer, M. Brell, A. Hein, and S. Gessler; “Personal Assistive Robots for AAL Services at Home - The Florence Point of View,” 3rd. IoPTS workshop, Brussels, 2009.

[3] T. Frenken, M. Isken, N. Volkening, M. Brell, and A, Hein, “Criteria for Quality and Safety while Performing Unobtrusive Domestic Mobility Assessments Using Mobile Service Robots,” Ambient Assisted Living, Advanced Technologies and Societal Change 2012, VDE, 2012, pp. 61-76.

[4] T. Rehrl et al., “The Ambient Adaptable Living Assistant is Meeting its Users,” AAL Forum 2012.

[5] F. J. Imms and O. G. Edholm, “Studies of gait and mobility in the elderly”, Age Ageing, vol. 10, no. 3, Aug. 1981, pp. 147–156.

[6] M. Montero-Odasso et al., “Gait velocity as a single predictor of adverse events in healthy seniors aged 75 years and older,” J Gerontol A Biol Sci Med Sci, vol. 60, no. 10, Oct. 2005, pp. 1304–1309.

[7] N. Volkening, A. Hein, M. Isken, T. Frenken and M. Brell, “Housing Enabling – Detection of imminent risk areas in domestic environments using mobile service robots,” 6. Deutscher Kongress Ambient Assisted Living at Berlin, Germany, VDE Verlag 2013, pp. 479-485.

[8] D. J. Cook and S. K. Das, “How smart are our environments? An updated look at the state of the art,” Pervasive and Mobile Computing, vol. 3, no. 2, 2007, pp. 53 – 73.

[9] S. B. Thies, J. K. Richardson, and J. A. Ashton-Miller, “Effects of surface irregularity and lighting on step variability during gait: A study in healthy young and older women,” Gait & Posture, vol. 22, iss. Aug. 1, 2005, pp. 26-31, ISSN 0966-6362.

[10] D. S. Marigold and A. E. Patla, “Age-related changes in gait for multi-surface terrain”, Gait&Posture, vol. 27, iss. 4, May, 2008, pp. 689-696.

[11] Primesense – 3D Carmine 1.09 Sensor, Product Information, Available Online: http://i3du.gr/pdf/primesense.pdf, last access: June 27, 2014.

[12] Professional association rules for safety and health at work, BGR 110, 04-2007, Federation of Trade Associations, Online: http://publikationen.dguv.de/dguv/pdf/10002/bgr-110.pdf, last access: June 26, 2014.

[13] C. N. Scanaill et al., “A review of approaches to mobility telemonitoring of the elderly in their living environment,” Ann Biomed Eng, vol. 34, no. 4, Apr., 2006, pp. 547–563.

[14] K. Cameron, K. Hughes, and K. Doughty, “Reducing fall incidence in community elders by telecare using predictive systems,” in Proc. 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 3, 1997, pp. 1036–1039.

[15] J. A. Kaye et al., “Intelligent Systems For Assessing Aging Changes: home-based, unobtrusive, and continuous assessment of aging,” The journals of gerontology. Series B, Psychological sciences and social sciences, vol. 66, iss. 1, July 1., 2011, pp. i180–i190, doi: 10.1093/geronb/gbq095.

[16] T. Frenken et al., “A novel ICT approach to the assessment of mobility in diverse healht care enviroment,” CEWIT-TZI-acatech Workshop "ICT meets Medicine and Health" (ICTMH 2013), April, 2013.

[17] M. P. Poland, D. Gueldenring, C. D. Nugent, H. Wang, and L. Chen, “Spatiotemporal Data Acquisition Modalities for Smart Home Inhabitant Movement Behavioural Analysis,” ICOST '09, Proceedings of the 7th International Conference on Smart Homes and Health Telematics, Springer, 2009, pp. 294-298.

[18] E. E. Stone and M. Skubic, “Passive In-Home Measurement of Stride-to-Stride Gait Variability Comparing Vision and Kinect Sensing,” 33rd Annual International Conference of the IEEE EMBS, Boston, Massachusetts, USA, 2011, pp. 6491-4.

[19] M. Gabel, R. Gilad-Bachrach, E. Renshaw, and A. Schuster, “Full Body Gait Analysis with Kinect,” 34th Annual International Conference of the IEEE EMBS, San Diego, USA, 2012.

[20] A. Petrovskaya and A. Y. Ng, “Probabilistic mobile manipulation in dynamic environments, with application to opening doors,” in International Joint Conference on Artificial Intelligence (IJCAI), 2007, pp. 2178-2184.

[21] C. L. Breazeal, “Sociable machines: Expressive social exchange between humans and robots,” Ph.D. dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2000.

[22] C. Ray, F. Mondada, and R. Siegwart, “What do people expect from robots?,” in IEEE/RSJ International Conference on Intelligent Robots and Systems, 2008, pp. 3816–3821.

[23] M. Isken et al., “Enhancing Mobile Robots’ Navigation through Mobility Assessments in Domestic Environments,” in Proceedings 4. Deutscher Kongress, Ambient Assisted Living, VDE Verlag, 2011, pp. 223-238.

[24] M. Brell, J. Meyer, T. Frenken, and A. Hein, “A Mobile Robot for Self-selected Gait Velocity Assessments in Assistive Environments,” in The 3rd International Conference on Pervasive Technologies Related to Assistive Environments (PETRA’10), Samos, Greece, June 2010, ISBN 978-1-4503-0071-1.

[25] G. Carlsson, B. Slaug, A. Johannisson, A. Fänge, and S. Iwarsson,. “The Housing Enabler - Integration of a computerised tool in occupational therapy undergraduate teaching,” CAL Laborate, June, 2004, pp. 5 – 9,

[26] T. Helle et al., “The Nordic Housing Enabler: Interrater reliability in cross-Nordic occupational therapy practice,” Scandinavian Journal of Occupational Therapy, Dec. 17, 2010, pp. 258-66.

[27] A. Fänge, “Strategies for evaluation of housing adaptations – Accessibility, usability and ADL dependence”, ISBN91-974281-5-9. Doktorsavhandling. Institutionen för klinisk neurovetenskap, Lunds Universitet. Lund, Sverige, 2004.

[28] M. Cesari et al, “Prognostic Value of Usual Gait Speed in Well-Functioning Older People—Results from the Health, Aging and Body Composition Study”, Journal of the American Geriatrics Society, vol. 53, 2005, pp. 1675–1680.

[29] P. Udsatid, N. Niparnan, and A. Sudsang, “Human Position Tracking for Side By Side Walking Mobile Robot using Foot Positions,” Proceedings of the 2012 IEEE International Conference on Robotics and Biomimetics, Dec. 11-14, 2012, pp. 1374 - 1378 Guangzhou, China.

[30] E. E. Stone and M. Skubic, “Unobtrusive, Continuous, In-Home Gait Measurement Using the Microsoft Kinect,” IEEE Transactions on biomedical engineering, vol. 60, no. 10, Oct. 2013, pp. 2925-32.

[31] T. Pallej, M. Teixid, M. Tresanchez, and J. Palacn, “Measuring Gait Using a Ground Laser Range Sensor,” Sensors, vol. 9, no. 11, 2009, pp. 9133–9146.

[32] Zhengyou Zhang, "Microsoft Kinect Sensor and Its Effect," IEEE Multimedia, vol. 19, no. 2, pp. 4-10, April-June 2012, doi:10.1109/MMUL.2012.24

 ISBN: 978-1-61208-356-8

Transient Stress Stimulus Effects on Intentional Facial Expressions

- Estimation of Psychological States based on Expressive Tempos -

Kazuhito Sato
Department of Machine Intelligence and Systems Engineering,
Faculty of Systems Science and Technology, Akita Prefectural University
Yurihonjo, Japan
ksato@akita-pu.ac.jp

Takashi Suto
IAI Corp.
Mechanical Design Section 2
Shizuoka, Japan
takashi-suto@iai-robot.co.jp

Hirokazu Madokoro
Department of Machine Intelligence and Systems Engineering,
Faculty of Systems Science and Technology, Akita Prefectural University
Yurihonjo, Japan
madokoro@ akita-pu.ac.jp

Sakura Kadowaki
Smart Design Corp.

Akita, Japan
sakura@smart-d.jp

***Abstract*-This paper presents a framework of tempos and rhythms to clarify the relevance between psychological states and facial expressions, particularly addressing repetitive operations of intentional facial expressions after giving a stress stimulus. By acquiring image datasets of facial expressions under the states of pleasant–unpleasant stimulus for 20 subjects, we extracted expressive tempos for respective subjects. Consequently, averages of extraction rates show that the pleasant state was 81.1%. The unpleasant state was 77.8%. Regarding effects of pleasant–unpleasant stimulus on the expressive tempos, particularly addressing the variation of the number of frames constituting one tempo, the variation in unpleasant stimulus became greater than that in the pleasant stimulus. The results show that the analysis using expressive tempos and rhythms is valid as an indicator for estimating the psychological state.**

Keywords-Psychological measures, stress; Intentional facial expression; Machine learning approaches; Behavior modeling.

I. INTRODUCTION

Humans can feel rhythms from all of their personal surroundings that are moving, especially any emitting sound. Additionally, they feel rhythms from engaging in daily life, such as rhythms related to conversation and rhythms of human life [1][2]. Among these, biological rhythms are based on personal tempos. In other words, personal tempos are individual-specific, not derived from physiological functions. It has been reported that personal tempos also vary depending on environments and moods [1]. In daily life, for behaviors such as walking or talking, personal tempos represent individual-specific speeds, which are expressed naturally in a free-action situation without constraint. For a facial expression as a daily life behavior, we infer that individual-specific rhythms can exist also.

To clarify the relevance between psychological states and facial expressions, we propose a framework of rhythms and tempos that specifically examines actions to repeat intentional facial expressions after giving a stress stimulus. We define one rhythm as one tempo repeated several times. In addition, we regard one tempo as the period during which facial expressions transform from a neutral face (i.e., expressionless) to the next neutral face through the maximum number of facial expressions, in a time-series variation of Expression Levels (ELs), i.e., labels quantifying exposed levels from the neutral face [3]. We use Hidden Markov Models (HMMs) [4] of left-to-right type to extract expressive tempos. As a method to classify categories extracting the occurrence part of patterns from the time series data, HMMs are widely used in fields of signal processing and speech recognition. They can extract expressive tempos, which represent the occurrence pattern of exposed intensities. Stress reactions appear in relation to biological phases (e.g., changes in heart rate, changes in blood pressure), psychological phases (e.g., depression, irritability), and behavioral phases (e.g., increase of drinking and smoking, fidgety state) [5]. Here, the facial expression is classified as a behavioral phase among the stress reactions. For this reason, by analyzing the rhythm and tempos that appear after exposure to different stress conditions, we infer that the inference of psychological states, such as comfort

 ISBN: 978-1-61208-356-8

and discomfort, from the changes of individual-specific facial expressions will become possible in the future.

In this study, as the basis for objectively expressing the ambiguity and complexity of facial expressions attributable to the psychological stress states of human, we propose a framework of exposed rhythms and tempos on intentional facial expressions. This study might derive the following advantages in applications. One familiar case of those is to develop a training tool to create an attractive smile that hospitality mind is easily transmitted to the customer. Foreseeable future, we could be sure that this study is valid as new indices for detecting the distraction state of driver by time-series changes of eye-gaze and facial expressions.

This paper is presented as the following. We review related work to clarify the position of this study in Section II. In Section III, we define a new framework of exposed rhythms and tempos for analyzing relations of psychological stress and facial expressions. In Section IV, we describe the method to capture facial expression images, preprocessing, classification of facial expression patterns with self-organizing maps, integration of facial expression categories with fuzzy adaptive theory, extraction of expressive tempos using HMMs. We explain our originally developed facial expression datasets including stress measurements in Section V. In Section VI, we optimize the number of states of HMMs by extracting expressive tempos from facial expressions and .analyze the transient stress stimulus of pleasant–unpleasant effects on the expressive rhythm of facial expressions. Finally, we present conclusions and intentions for future work in Section VII.

II. RELATED WORKS

Open datasets [6][7][8]of facial expression images are released from some universities and research institutes to be used generally in many studies for performance comparisons of facial expression recognition or automatic analysis of facial expressions. These datasets contain a sufficient number of subjects as a horizontal dataset. However, images are taken only once for each person. As one of recent researches using these datasets, there is the study by Das and Yamada [9]. They used the Cohn–Kanade [6] and the Extended Cohn–Kanade (CK+) datasets [7] to obtain emotional mixture or percentage composition of emotion data, because cross-sectional datasets are valid rather than time-series datasets in evaluating stress. The CK+ datasets contain Action Units (AUs) coded facial image data with lead emotion label for each peak expression. Therefore, they considered the peak and few intermediate states of each facial expression taking care that the difference in intensities is not large enough to represent another emotion altogether. Das and Yamada conducted two moderate sized surveys to correlate individual emotions to stress and to find relationship between predicted emotional mixtures of facial expressions and stress levels [9]. After predicting emotional composition, they selected facial expression images for two surveys. However, the respondents were just only instructed to look at the static facial image and label the stress levels from 0 to 9 according to each individual perception. Consequently, Das and Yamada did not carry out analysis that focused on the expressive process of individual-specific facial expressions, in spite of lurking clue in there.

In a study particularly addressing the dynamic aspects of facial expressions, Hirayama et al. [10] found the kinetic period of face parts. They have proposed an expressive notation as a representative format that describes the timing structure on facial expressions. They were seeking linear systems (i.e., modes) to the bottom-up from feature vector sequences. The modes represent various motional states or stationary states of face parts. For example, in the case of the mouth, there are open, remain open, close, and keeping closed as elements of mode sets. The method explained by Hirayama et al. tracked feature points, i.e., a total of 58 points is assessed from the outline of the lower half face including each eyebrow, each eye, nose, and lips. Then using Active Appearance Models (AAMs) [11] for time-series facial expressions at the beginning, a feature vector sequence was obtained for each part of the face. Then, they acquired expressive notation of involuntary and spontaneous facial expressions based on providing an automatic phrase of the mode from the obtained feature vectors. The experimentally obtained results show that by particularly addressing timing structures of the two expressive notations that were obtained, Hirayama et al. analyzed how two facial expressions differ. In the analytical results, a difference was found in the timing of movement of the muscles between lifting the cheek and moving the mouth for two facial expressions. Consequently, for describing the timing structure of facial expressions, the time resolution of the model and the image sequence are set high using expressive notation. However, because the spatial resolution of the model representing facial expressions is low, analysis of differences of the expressive intensities representing the intermediate facial expression have not yielded satisfactory results.

Otsuka et al. [12] proposed a method to extract six individual basic expressions described by Ekman et al. [13]. Modeling the movement of the facial expressions by HMMs, which carry out the state transition corresponding to the motion of different facial muscles, i.e., relaxation, contraction, stationary, and elongation, Otsuka et al. sought to recognize the facial expressions by analyzing the motion vectors of their surroundings, noticing that AUs of Facial Action Coding System (FACS) are distributed around the eyes and mouth. In their method, they first obtained the motion vectors around the eyes and mouth using the gradient method [14] from the facial expression image sequences, e.g., facial expressions of two kinds for 20 subjects. Next, by performing two-dimensional Fourier transform in a matrix component of the image, they

 ISBN: 978-1-61208-356-8

acquired a time series of 15-dimensional feature vectors. As the input time series of the feature vectors, Otsuka et al. extracted individual facial expressions by application of HMMs of left-to-right type. In this case, the experimenters confirmed the determination of true or false facial expressions. In a section of actual facial expressions, they treated the corresponding facial expression that had been extracted as a correct answer. In contrast, they treated the following two cases as incorrect answers: when no facial expression was extracted; when different facial expressions were extracted at once. An extraction rate of 90% was achieved in their experimentally obtained results: 40 facial expressions were extracted in the 20 subjects. Then, 36 facial expressions were accurately extracted in them. However, it is not always the precise period because being extracted represents the start and end points of facial expressions. The correctness checker is treated as a correct answer when the corresponding facial expression is expressed within the period.

According to the most recent study of the emotion–expression relationship based on evidence from laboratory experiments [15], high coherence has been found in several studies between amusement and smiling; low to moderate coherence between other positive emotions and smiling. Additionally, insufficient emotion intensity and inhibition of facial expressions could not account for the observed dissociations between emotion and facial expression. Furthermore, as a statistical indice of the coherence between emotion and facial expression, R. Reisenzein et al. reported that the most informative indice was "the average intra-individual correlation between emotion and expression". In this study, we actively do challenge to elucidate the correlations between the expressive process of individual-specific facial expressions and psychological states, particularly focusing on the correlations between pleasant-unpleasant stimulus and smiling process of intentional facial expressions.

III. FRAMEWORK OF EXPOSED RHYTHMS AND TEMPOS

A. *Facial Expression Levels*

As an index for quantifying the individual facial expression spaces, we proposed the framework of expression levels (ELs) [3]. The ELs include both features of the pleasure and arousal dimensions based on the arrangement of facial expressions on Russell's circumplex model [16]. Specifically, we extract the dynamics of topological changes of facial expressions of facial components such as the eyes, eyebrows, and mouth. Here, topological changes show the structure defining the connection form of the elements in the set [2]. The ELs obtained in this study are sorted categories according to their topological changes in intensity from expressions that are regarded as neutral facial expressions. As discussed above, the ELs in this study include both features of the pleasure and arousal dimensions. In Russell's circumplex model, all emotions are constellated on a two-dimensional space: the pleasure dimension of pleasure–displeasure and arousal dimension of arousal–sleepiness. In the intentional facial expressions covered in this study, directly handling the facial expressions for the influence of pleasure dimension is difficult. Therefore, as a method of measuring transitory stress response, we conduct an evaluation using the salivary amylase test. Therefore, as a method of measuring transitory stress response, we conduct an evaluation using the salivary amylase test through the task of watching emotion-evoking videos caused a pleasant-unpleasant state. Focusing on the values of salivary amylase activity between before and after watching videos, we can effectively perform stress measurements by the salivary amylase test to assess the stress state transiently. Consequently, we target the intentional facial expressions under stimulating states of pleasant and unpleasant.

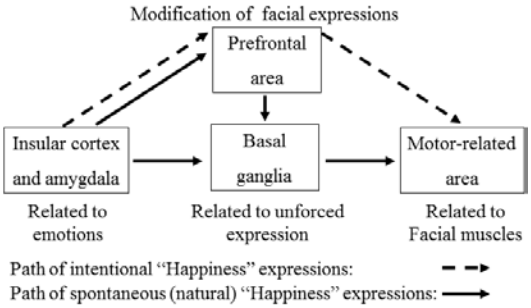


Figure 1. Expression paths based intentional and spontaneous facial expressions.

B. *Definition of Exposed Rhythms and Tempos*

Blair [17] has reported that, for facial expressions, four brain domains are mutually related: (1) parts producing feelings (insular cortex and amygdala), (2) parts forming facial expressions involuntarily (basal ganglia), (3) parts embellishing facial expressions according to the surrounding circumstances (prefrontal area), and (4) motor-related areas actually moving mimic muscles. Yamaguchi [18] reported that the brain memorizes experiences in a rhythm: according to specific brain waves, nerve cells work cooperatively, and experiences are memorized. In perceptual recognition, it is explained that nerve cells function simultaneously according to the gamma waves, which are brain waves having quick rhythms. From the results of these studies, we infer that the rhythms of nerve cells participate in the expressional process of facial expressions. As presented in Figure 1, in cases where facial expressions are embellished intentionally or spontaneously, time-sequential differences exist based on the route through

 ISBN: 978-1-61208-356-8

which facial expressions are revealed. The basis of our hypothesis is as follows. According to specific brain waves of four brain area, nerve cells of each brain area are used to work cooperatively, in the case of the repetition process of facial expressions under a pleasant-unpleasant stimulus particularly. Mimic muscles is activated by coordination of nerve cells with different speed, a unique expression is exposed through the individual path of each facial expression.

In this study, using temporal variation of ELs, we intend to visualize rhythms and tempos of facial expressions that humans create. We defined one rhythm as a tempo that is repeated several times. One tempo indicates the period during which facial expressions are transformed from a neutral state to the next neutral state. Facial expressions exposed intentionally by humans form an individual space based on dynamic diversity and static diversity of the human face [19]. Facial expression dynamics can be regarded as "topological changes in time-sequential facial expression patterns that facial muscles create." Static diversity is individual diversity that is configured by the facial componential position, size, and location, consisting of eyes, nose, mouth, and ears. In contrast, dynamic diversity represents that human can move facial muscles to express internal emotions unconsciously and sequentially or to express emotions as a message. After organizing and visualizing topological changes of face patterns by ELs, we attempt to use the framework of rhythms and tempos with expressions to express ambiguities and complexities of facial expressions attributable to a psychological state.

IV. PROPOSED METHOD

Facial expression processes differ among individuals. Therefore, Akamatsu [19] described the adaptive learning mechanisms necessary for modification according to individual characteristic features of facial expressions. In this study, our target is intentional facial expressions. We use Self-Organizing Maps (SOMs) [20] to extract topological changes of facial expressions and for normalization with compression in the direction of the temporal axis. After classification by SOMs, facial images are integrated using Fuzzy ART [21], which is an adaptive learning algorithm with stability and plasticity. In fact, SOMs perform unsupervised classification input data into a mapping space that is defined preliminarily. In contrast, Fuzzy ART performs unsupervised classification at a constant granularity that is controlled by the vigilance parameter. Therefore, using SOMs and Fuzzy ART, time-series datasets showing changes over a long term are classified with a certain standard. Figure 2 presents an overview of the procedures used for our proposed method. In the following, we describe extraction of time-sequential changes of ELs, and also explain detection of expressive

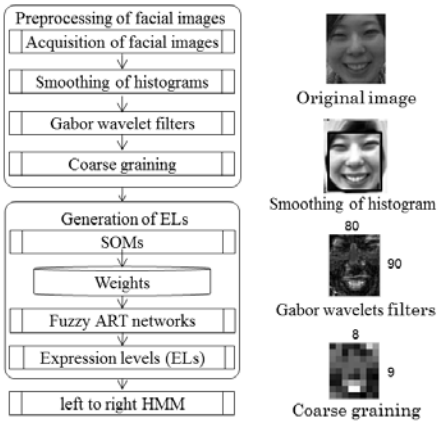


Figure 2. Overview of the procedures used for our proposed method.

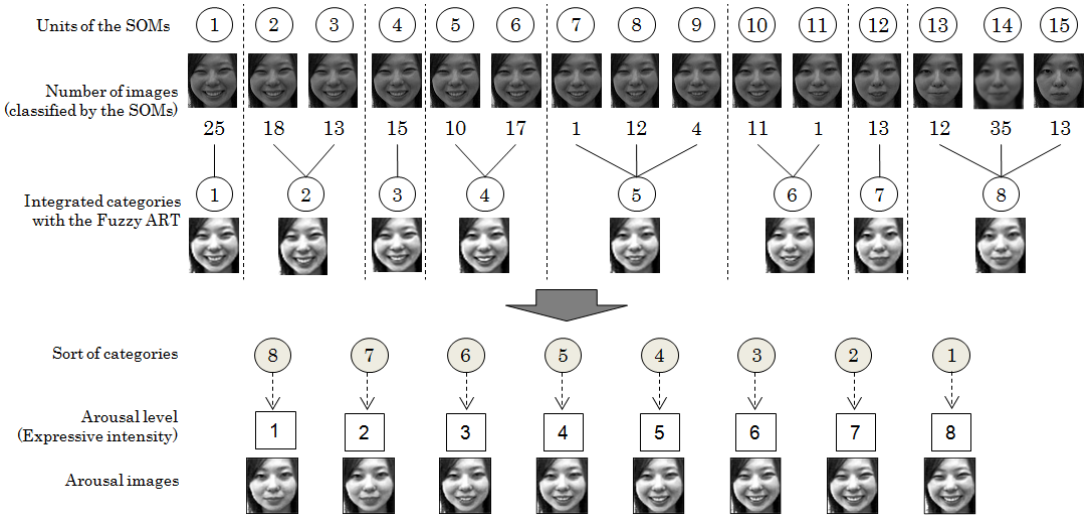


Figure 3. Procedure details for acquiring a time-series variation of ELs.

tempos by HMMs.

A. Acquisition of Time-series Variation of ELs

We set the Region of Interest (ROI) to 90 × 80 pixels, including the eyebrows, which all contribute to the impression of a whole face as facial feature components. With preprocessing, brightness values are normalized for time-series images of facial expressions. The influence of brightness values attributable to illumination conditions is thereby reduced. Moreover, smoothing the histogram is useful to adjust contrast and clarify the images. In addition, using the orientation selectivity of Gabor Wavelets filtering as a feature representation method, the facial parts characterizing the dynamics of facial expressions are emphasized, such as eyes, eyebrows, mouth, and nose. By down-sampling (i.e., 10 × 10 pixels) time-series facial expressions converted with Gabor Wavelets filtering [22], the effects of a slight positional deviation when taking facial images are minimized. Then data size compression is conducted.

Figure 3 presents details of procedures for acquiring a time-series variation of ELs. First, we use SOMs to learn the time-series images of facial expressions with down-sampling. The face images that show topological changes of facial expressions that are similar are classified into 15 mapping units of SOMs. Next, similar units (i.e., Euclidean distances of the weight vectors are close) among 15 mapping units of SOMs are integrated into the same category by Fuzzy ART. By sorting the facial expression categories integrated by Fuzzy ART from neutral facial expression to the maximum of facial expression, we obtain ELs labeled as expressive intensities of facial expressions quantitatively. The sorting procedure of integrated categories is based on the two-dimensional correlation coefficient of the average image of the facial expression images classified into each category. Finally, we conduct corresponding ELs with each frame of the facial images to produce time-series variations of ELs.

B. Extraction of Expressive Tempos by HMMs

As a method of recognizing words by estimating phonemes from acoustic signals, HMMs were first used in the speech recognition field. Takeda et al. [23] performed an automatic accompaniment and score tracking of MIDI music using HMMs. Actually, HMMs have been established as a technique for extracting an occurrence pattern from time-series datasets and classifying it as a category. Datasets used for this study are directed to time-series facial images, an expressive tempo consists of occurrence pattern of ELs. Therefore, we use HMMs to extract expressive tempos. HMMs are simple Markov models with multiple nodes, defined by transition probabilities between mutual nodes and output probabilities of multiple symbols from each node. By preparing HMMs to extract a target, each HMM is trained in the symbol sequence of each training dataset for

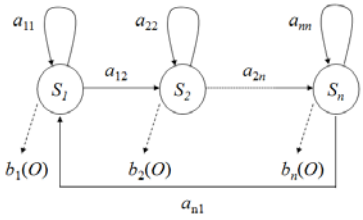


Figure 4. Configuration of HMMs used for this study (Type of Left to Right).

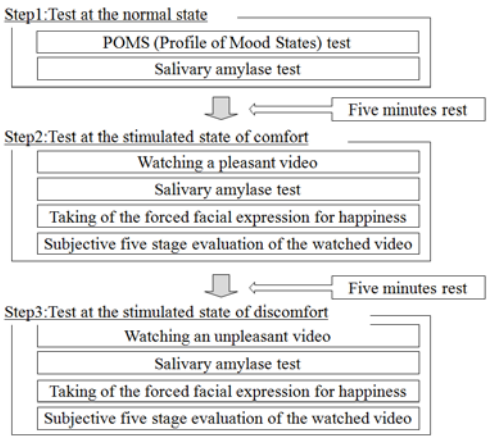


Figure 5. Details of experimental protocols.

targets. Training of HMMs is useful to estimate two parameters of symbol output probabilities and state transition probabilities that generate a high probability of training symbol sequence. Additionally, using Baum–Welch algorithm [24], training is repeated until the parameters converge i.e., the change in the output likelihood is sufficiently small. The configuration of HMMs used for this study is a type of Left to Right, as shown in Figure 4, we set the internal state of nodes to S_1, S_2, · · ·, S_n from left to right. Here, S_1 is the initial state of facial expressions (neutral facial expression), S_2 · · ·S_{n-1} are the intermediate states, and S_n is designated as the final state (maximum value of ELs). To obtain the updated values of state probability of S_i (i = 1, · · ·, n), we define the probability of following equations. State transition probabilities (a_{ij}) mean the transition probability from state S_i to state S_j, only the self-transition and transition to the right state in Left to Right HMMs are permitted. Therefore, the following constraints are satisfied.

$$a_{ij} = 0 \qquad (j < i) \tag{1}$$

$$0 <= a_{ij} <= 1 \qquad (j >= i) \tag{2}$$

$$\sum a_{ij} = 1 \tag{3}$$

Symbol output probabilities bi (O) denote the probability density distribution for outputting a symbol sequence O in state S_i, we use a discrete distribution of allocating

probabilities to discrete symbols that are commonly used in the field of speech recognition.

V. DATASETS

In this study, we constructed an original and long-term dataset for the specific facial expressions of one subject. Figure 5 presents details of experimental protocols. One experiment comprises three steps, i.e., step 1 is under a normal state, step 2 is in watching pleasant video, and step 3 is in watching unpleasant video. As shown in Figure 6, we gave subjects the task of watching emotion-evoking videos caused a pleasant–unpleasant state, and performed stress measurements by salivary amylase tests to assess the stress state transiently. In addition, the watching time is about 3 min for each emotion-evoking video, we prepared unpleasant videos (i.e., implant surgery and cruel videos) and the pleasant videos (i.e., comic videos of three types). The subjective assessment of five stages was also conducted at watching videos. For all subjects, we fully explained the experiment contents in advance based on the research ethics policy of our university, and also obtained the consent of experiment participants in voluntary writing of subjects. Moreover, from all subjects, we received agreement to publish face images as part of their experimental participation.

A. Facial Expression Images

Open datasets of facial expression images are open to the public through the internet from universities and research institutes. However, the specifications vary among datasets because of imaging with various conditions. As static facial images, the dataset presented by Ekman and Friesen [13] is a popular dataset comprising collected various facial expressions used for visual stimulation in psychological examinations of facial expression cognition. As dynamic facial images, the Cohn–Kanade dataset [6] and Ekman–Hager dataset [25] are widely used, especially in experimental applications. In recent years, the MMI Facial Expression Database presented by Pantic et al. [8] and the CK+ dataset [7] have become a widely used open dataset containing both static and dynamic facial images. These datasets contain a sufficient number of subjects as horizontal datasets. However, facial images are taken only once for each subject. No dataset exists in which the same subject has been traced over a long term. Therefore, we created original and longitudinal datasets that include collections of the specific facial expression of the same subject during a long term.

Six basic facial expressions proposed by Ekman et al. [13] are "happiness", "anger", "sadness", "disgust", "fear", and "surprise". Among the six basic facial expressions, we specifically examined the facial expression of "happiness", which is believed to be most likely exposed spontaneously. As the target facial expression of "happiness" under stimulating states of pleasant and unpleasant, we acquired the facial expressions of 20 subjects. As a method of stimulation, we pre-selected emotion-evoking videos that elicit emotions that are pleasant or unpleasant, with all subjects expressing the facial expression of "happiness" immediately after watching them. Subjects were 10 men (Subject J was 20 years old; Subjects B, G, H, and I were 21; Subjects A, E, and F were 22; Subjects C and D were 23) and 10 women (Subjects K, M, O, and P were 20 years old; Subjects L, Q, R, S, and T were 21; Subject N was 23), all of whom were university students. The imaging period was three weeks at one-week intervals for all subjects. The imaging environment for facial expressions was an imaging space partitioned by a curtain in the corner of the room. We took frontal facial images with conditions including the head of the subject in each image. In advance, we instructed each subject to expose the facial expression without any head movement. Consequently, imaging the face region to fit within the scope has been possible. However, with respect to extremely small changes caused by body motion, we used template-matching methods to trace the face region by setting the initial template to include facial parts. By consideration of the application deployment and ease of imaging in future studies, we used commercially available USB cameras (QcamOrbit; Logicool Inc. [26]). When taking images of each facial expression, the same expression was repeated three times based on the neutral facial expression during the image-taking time of 20 s. We previously instructed all subjects to express an emotion three times at their own timing according to a guideline for 20 s. One dataset consisted of 200 frames with the sampling rate of 10 frames per second.

B. Stress Measurement Method

Because types of psychological stress are regarded as affecting facial expressions, we assessed transient stress and chronic stress. Chronic stress is that which humans have on a daily basis, whereas transient stress is that caused by a temporary stimulus. To assess transient stress stimulus to the subjects in this study, we applied the salivary amylase test, which is one method of measuring transient stress reactions. As a biological reaction, salivary amylase activity is detected as a low value if one is in a pleasant state. In contrast, the value is high if one is in an unpleasant state. As stress reactions when subjected to external transient stimulus, Yamaguchi et al. [27] confirmed that salivary amylase activity is an effective means of stress evaluation. For this study, using the emotion-evoking videos as an external transient stimulus, we used the salivary amylase test method to measure stress reactions immediately after participants watched the videos.

 ISBN: 978-1-61208-356-8

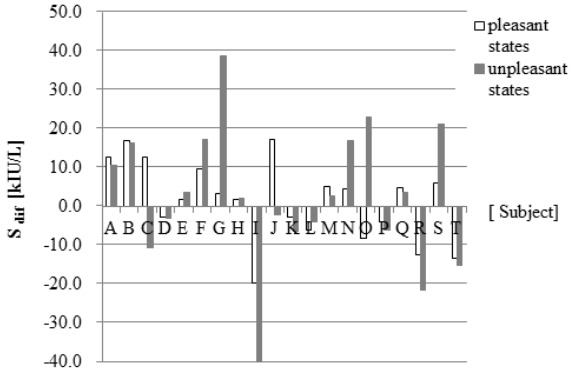


Figure 6. Results of S_{dif} obtained for target to the 20 subjects of A-T.

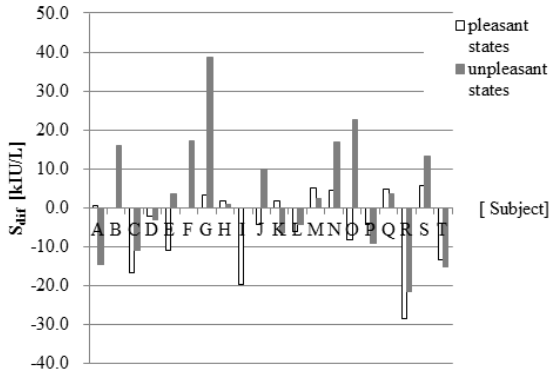


Figure 7. Results of S_{dif} addressed only the score of 4 and 5 with subjective evaluations.

VI. EXPERIMENT

We verified the validity of emotion-evoking videos, which give a pleasant–unpleasant stimulus. Next, we optimized the number of states of HMMs by extracting expressive tempos from facial expressions. Subsequently, using the HMMs with an optimized number of states, we verified the accuracy of the extracted expressive tempo obtained from a time-series change of ELs. Finally, we analyzed the transient stress stimulus of pleasant–unpleasant effects on the expressive rhythm of facial expressions.

A. Effectiveness of Pleasant–unpleasant Stimulus

Using the salivary amylase test, we examined the validity of emotion-evoking factor in watching the video used as a pleasant–unpleasant stimulus. The following were shown for salivary amylase activity. The value of salivary amylase activity is reduced if in a pleasant state. In contrast, its value is increased if one is in unpleasant circumstances [27]. Accordingly, letting S_{normal} be the value of salivary amylase activity at normal state, and letting S_{stimu} be the value of salivary amylase activity after watching the video, then the difference of salivary amylase activity between the normal state and after watching video (S_{dif}) is defined by the following equation.

$$S_{dif} = S_{stimu} - S_{normal} \quad (4)$$

$S_{dif} < 0$ (i.e., after watching pleasant videos)

$S_{dif} > 0$ (i.e., after watching unpleasant videos)

Figure 6 presents results of S_{dif} obtained for target to the 20 subjects of A–T. In this case, the perception for the pleasant–unpleasant videos differs slightly among subjects, so this fact might cause the results of salivary amylase activity of C and B differ with previous studies [27]. Therefore, we decided to calculate the salivary amylase activity only for data for which subjective evaluation of the subject is high. The subjective evaluation receives a score of 1–5, score 1 (i.e., not at all), score 5 (i.e., strong) at watching each emotional video. Figure 7 presents results of salivary amylase activity in the case of particularly addressing only the score of 4 and 5 because we consider that the emotional video is effectively working as a pleasant–unpleasant stimulus. Based on this result, the average of all S_{dif} indicates -2 [kIU/l] at a pleasant state, 5 [kIU/l] at an unpleasant state. Therefore, results show that the emotion-evoking video functioned as a pleasant–unpleasant stimulus.

B. Examination of HMM Parameters

Otsuka et al. [12] pointed out that the process of facial expressions was made up with state transitions such as "neutral state" → "expression state" →"neutral state". In this case, the operation of facial muscles was to be the acts of "relaxation" → "contraction" → "rest" → "extension" → "relaxation". In the method explained by Otsuka et al., under conditions in which the state of facial muscles and the state of HMMs are associated with initial values, they modeled the state transitions of facial muscles by setting the number of states of HMMs to five [12]. However, by varying the initial state transition probability and number of states of the HMMs, our experiments were conducted to ascertain the optimum value of the highest extraction rate shown in equation (5). Therefore, it is possible to obtain parameters (i.e., the initial state transition probability and number of states of the HMMs) that represent the best movement of facial muscles under conditions of transient stress stimulus.

As the accuracy judgment of extraction with HMMs, we set as Ground Truth (GT) the average value of the frames for which three evaluators judged that the transition state had returned to a neutral state by their visual observation of the videos showing facial expressions. The extraction rate of accuracy judgment is defined by equation (5).

$$x_1, x_2, x_3 = \begin{cases} 1, \mathrm{E} \in \mathrm{R} \pm 5 \\ 0, \neg(\mathrm{E} \in \mathrm{R} \pm 5) \end{cases}$$

$$A = \frac{x_1, x_2, x_3}{C} \times 100[\%] \quad (5)$$

In that equation, *A* represents the extraction rate, *C* denotes the number of facial expressions, *E* represents the final frame of the facial expressions extracted with HMMs, and *R* denotes the frame indicating the end of the facial expressions obtained as the GT.

For this study, we performed experiments by obtaining the number of states to represent the movement of facial muscles optimally in a stress stimulus. In the pleasant–unpleasant state, we compared the extraction rate by varying the transition probability *b* to the next state, the self-transition probability *a*, and a number of states of the HMMs. Figure 8 presents the results. In the experimentally obtained result, the average extraction rate is the largest with setting the number of states to three. The average extraction rate is reduced later peaked at the state number of 3. Furthermore, the average extraction rate becomes a maximum under conditions of self-transition probability *a* of 0.70, and state transition probability *b* of 0.30. Based on consideration of the results described above, the parameters of the HMMs in this study were determined as follows. The number of states is 3, the self-transition probability *a* is 0.70, and the state transition probability *b* is 0.30.

C. Extraction Results of Expressive Tempos

As the extracted results of expressive tempos by application of HMMs, Figure 9 depicts the expressive tempos of six cases of subjects A, C, J, K, Q, and S. The top of each figure shows the time-series change of ELs. The bottom of each figure shows the transitional state of HMMs. Additionally, we marked the dashed vertical lines as GT. The GT indicates the average value of the frames, in which three evaluators judged that the facial expression had been completed by their visual observation for the original image. In consideration of variation among three evaluators, we

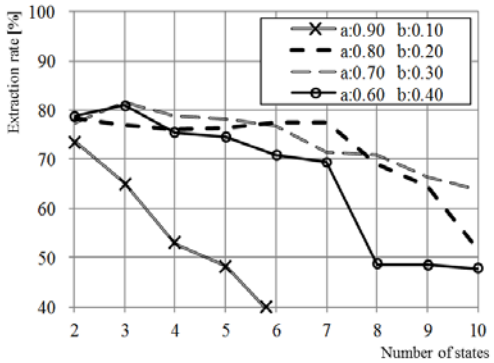
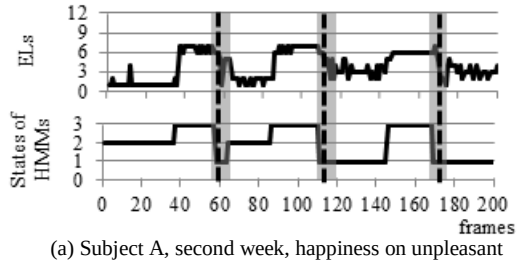
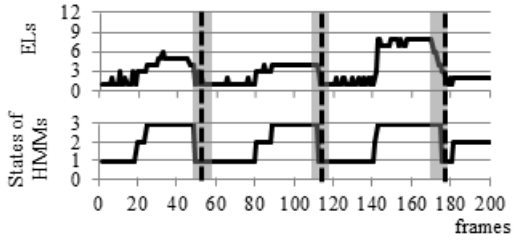


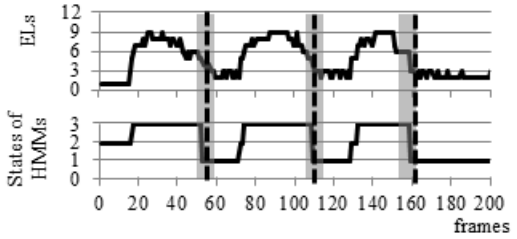
Figure 8. Extraction rates by varying a number of states of HMMs.



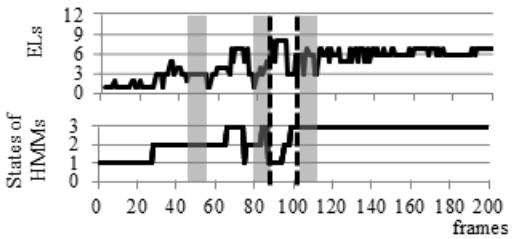
(a) Subject A, second week, happiness on unpleasant



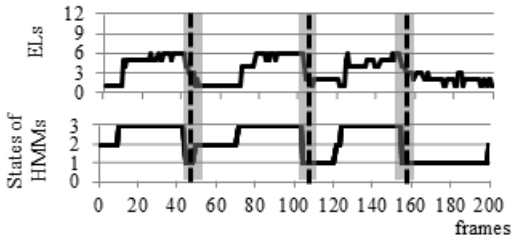
(b) Subject C, third week, happiness on unpleasant



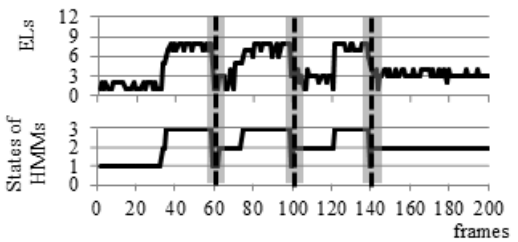
(c) Subject K, second week, happiness on pleasant



(d) Subject J, first week, happiness on pleasant



(e) Subject Q, first week, happiness on pleasant



(f) Subject S, first week, happiness on unpleasant

Figure 9. Extracted results of expressive tempos for six subjects.

 ISBN: 978-1-61208-356-8

presented a shaded gray pattern as the period of extraction, indicating a range of ± 5 frames with respect to each GT frame.

For subjects A, C, K, Q, and S, the extraction rates are 100% because all frames extracted by HMMs were included in the extraction range. In subject J, the extracted frames by HMMs are 60, 76, and 88, whereas the frames of GT are 40, 71, and 98. Therefore, in this example, only the second tempo was extracted successfully. Turning to the time-series change of ELs in the top of figure, extraction results of HMMs do not correspond to the timings of facial expressions. A major cause of that lack of correspondence is that evaluators have difficulty dividing the periods of facial expressions by visual observation because expressive levels of facial expressions appearing on the original image are small. For this study, we used view-based feature representation of facial expression datasets. Given difficulty in identifying the periods of facial expressions by human visual observation, we believe that automatic extractions of expressive tempos generally become difficult. Therefore, when acquiring facial expression datasets, we must ensure an instruction for each subject to expose the maximum ELs possible.

Subsequently, targeting the facial expression datasets of three weeks for subjects A–T (i.e., 20 cases), Figure 10 presents extraction rates of expressive tempos for each subject. Taking the average of the extraction rates in three weeks, the pleasant state was 81.1%. The unpleasant state was 77.8%. Even including a difficult case of identification of the facial expression period by visual inspection, such as Figure 10(d), the average extraction rate of 79.5% was obtained for all subjects.

D. Effects of Pleasant–unpleasant State on Expressive Rhythms

For subject G, Figure 11 presents the extraction result of expressive tempos and the time-series variation of ELs with "happiness" after watching pleasant videos. The three extracted tempos are as follows. The first tempo comprises 60 frames, the second tempo comprises 57, and the third tempo comprises 36. As described above, there are variations in the three expressive tempos which constitute one rhythm. Therefore, by calculating the average and standard deviation of number of frames constituting one tempo for all subjects A to T, we discuss the relation of expressive rhythms with a pleasant–unpleasant state.

Table I presents the standard deviation of tempos and average number of frames constituting one expressive tempo for all subjects of three weeks. Considering the average frames constituting one tempo in the pleasant–unpleasant state, the pleasant state is 49.1 [frames], the unpleasant state is 49.2 [frames]. Therefore, we conclude that the pleasant–unpleasant state does not affect the average number of frames that constitute one tempo. In contrast, particularly addressing the standard deviation of the number of frames constituting one tempo, the pleasant state is 8.4 [frames]; the unpleasant state is 6.1 [frames]. Comparison of the pleasant state and unpleasant state showed variation in the unpleasant state in the frames constituting one tempo. As a tendency among all subjects by transient stress stimulus watching unpleasant videos, we demonstrated quantitatively that fluctuations occurred in expressive tempos that were components of the expressive

TABLE I. STANDARD DEVIATION OF TEMPOS AND AVERAGE NUMBER OF FRAMES CONSTITUTING ONE TEMPO FOR ALL SUBJECTS

	Pleasant states	Unpleasant states
Average of frames	49.1	49.2
Standard deviation	6.1	8.4

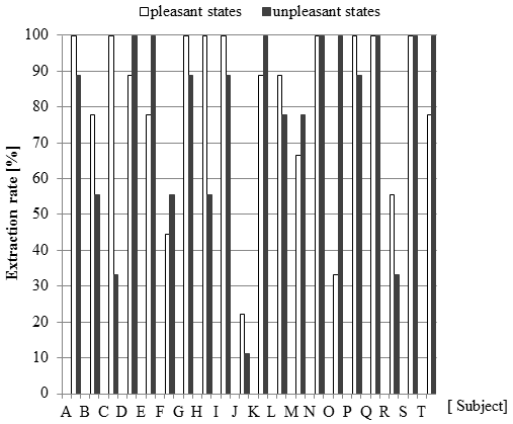


Figure 10. Extraction rates of expressive tempos for each subject.

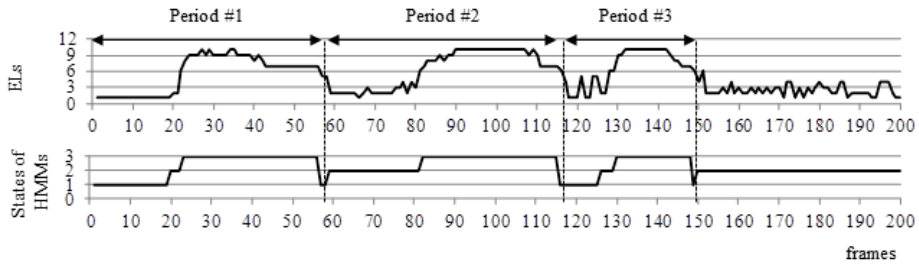


Figure 11. Expressive tempos and the time-series variation of the ELs with “happiness” after watching pleasant videos.

rhythm. The results described above reveal one indicator estimating the psychological state of humans. We conclude that the analysis of expressive tempos and rhythms is valid, with emphasis on repeated operations of intentional facial expression with "happiness".

VII. CONCLUSION AND FUTURE WORK

In this study, using the framework of expressive tempos and rhythms in facial expressions, we examined the relation between the psychological state (i.e., pleasant or unpleasant) and the time-series variation of ELs with exposure of intentional facial expressions. Acquiring image datasets of facial expressions under the states of pleasant–unpleasant stimulus for 20 subjects, we extracted expressive tempos of each subject. Consequently, taking the average of the extraction rates, the pleasant state was 81.1%, and the unpleasant state was 77.8%. By taking the effects of pleasant–unpleasant stimulus on the expressive tempos, particularly addressing the variation of number of frames constituting one tempo, the variation in unpleasant stimulus became greater than that in pleasant stimulus. The results presented above demonstrate that analysis using expressive tempos and rhythms is valid to indicate the psychological state. Moreover, by quantifying fluctuations of expressive tempos and rhythms, we can ascertain differences of the expressive path between intentional and spontaneous facial expressions.

ACKNOWLEDGMENTS

The authors thank the 20 students at our university who participated as subjects by letting us take facial images over such a long period. This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 25330325 and the Cosmetology Research Foundation.

REFERENCES

[1] N. Nobutani and Y. Nakatani, "Talk support system by rhythmical sound based on personal tempo," Information Processing Society of Japan, The 71st National Convention, pp. 4.227-4.228, Mar. 2009.

[2] S. Ohishi and M. Oda, "The personal tempo's effect on dialogue smoothness – from an index of switching pause –," The Institute of Electronics, Information, and Communication Engineers, Technical Report, pp. 31-36, 2005.

[3] H. Madokoro, K. Sato, and S. Kadowaki, "Facial expression spatial charts for representing time-series changes of facial expressions," Japan Society for Fuzzy Theory, vol. 23, no. 2, pp. 157-169, 2011.

[4] T. Sakaguchi, J. Ohya, and F. Kishino, "Facial Expression Recognition from Image Sequence Using Hidden Markov Model," The Institute of Image Information and Television Engineers, vol. 49, no. 8, pp. 1060-1067, 1995.

[5] H. Kumano, "Stress evaluation," http://hikumano.umin.ac.jp/StressAssess.pdf [retrieved: July, 2014]

[6] T. Kanade, J. F. Cohn, and Y. L. Tian, "Comprehensive database for facial expression analysis," Proc. of the 4th IEEE Int. Conf. on Automatic Face and Gesture Recognition, pp. 46-53, 2000.

[7] P. Lucey et al., "The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression," Proc. of the 3rd Int. Workshop on CVPR for Human Communicative Behavior Analysis, pp. 94-101, 2010.

[8] M. Pantic, M.F. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," Proc. IEEE Int'l. Conf. Multimedia and Expo, Amsterdam, The Netherlands, Jul. 2005. doi: 10.1109/ICME.2005.15214.

[9] S. Das and K. Yamada, "Evaluating instantaneous psychological stress from emotional composition of a facial expression, " Journal of Advanced Computational Intelligence and Intelligent Informatics, vol. 17, no. 4, pp. 480-492, 2013.

[10] T. Hirayama, H. Kawashima, M. Nishiyama, and T. Matsuyama, "Facial expression representation based on timing structures in faces," Human interface: the transaction of Human Interface Society, pp. 271-281, May 2007.

[11] T. F. Coots, G. J. Edwards, and C. J. Taylor, "Active Appearance Model: Proceedings of European Conference on Computer Vision," vol. 2, pp. 484-498, 1998.

[12] T. Otsuka and J. Ohya, "A study of spotting segments displaying facial expression from image sequences using HMM," The Institute of Electronics, Information, and Communication Engineers, Technical Report, pp. 17-24, Nov. 1997.

[13] P. Ekman and W. V. Friesen, "Unmasking the face: a guide to recognizing emotions from facial clues," Malor Books, 2003.

[14] B. K. P. Horn and B. B. Schunck, "Determining optical flow," Artificial Intelligence, vol. 17, pp. 185-203, 1981.

[15] R. Reisenzein, M. Studtmann, and G. Horstmann, "Coherence between emotion and facial expression: evidence from laboratory experiments, " Emotion Review, vol. 5,no. 1, pp. 16-23, 2013.

[16] J.A. Russell and M. Bullock, "Multidimensional scaling of emotional facial expressions: similarity from preschoolers to adults," Journal of Personality and Social Psychology, vol. 48, pp. 1290-1298, 1985.

[17] R.J.R. Blair, "Facial expressions, their communicatory functions and euro-cognitive substates," Philos. Trans. R. Soc. Lond., B358, pp. 561-572, 2003.

[18] Y. Yamaguchi, "Contextual information rhythmically processed in the brain," IEEJ Transactions on Electronics, Information and Systems C, vol. 128, no. 8, pp. 1068-1071, Aug. 2000.

[19] S. Akamatsu, "Recognition of facial expressions by human and computer [I]: facial expressions in communications and their automatic analysis by computer, " The Journal of the Institute of Electronics, Information, and Communication Engineers, vol. 85, no. 9, pp. 680-685, Sep. 2002.

[20] T. Kohonen, Self-organizing maps, Springer Series in Information Sciences, 1995.

[21] G. A. Carpenter, S. Grossberg, and D.B. Rosen, "Fuzzy ART: fast stable learning and categorization of analog patterns by an adaptive resonance system," Neural Networks, vol. 4, pp. 759-771, 1991.

[22] M. Haghighat, S. Zonouz, M. Abdel-Mottaleb, "Identification Using Encrypted Biometrics, " Computer Analysis of Images and Patterns, Springer Berlin Heidelberg, pp. 440-448, 2013.

[23] H. Takeda, T. Nishimoto, and S. Sagayama, "Automatic accompaniment system of MIDI performance using HMM-based score following," Information Processing Society of Japan, SIG Technical Report, pp. 109-116, Aug. 2006.

[24] The Institute of Electronics, Information and Communication Engineers Edition, "Speech recognition by probabilistic model," Corona Publishing Co. Ltd., ISBN**:** 978-4-88552-072-3, pp. 29-66, 1988.

[25] M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, "Measuring facial expressions by computer image analysis," Psychophysiology, vol. 36. pp. 253-264, 1999.

[26] QcamOrbit; Logicool Inc., http://www.logicool.co.jp/ja-jp/webcam-communications/webcams [retrieved: July, 2014]

[27] M. Yamaguchi, T. Kanamori, M. Kanemaru, Y. Mizuno, and H. Yoshida, "Correlation of stress and salivary amylase activity," Japanese Journal of Medical Electronics and Biological Engineering: JJME vol. 39, no. 3, pp. 46-51, Sep. 2001.

 ISBN: 978-1-61208-356-8

SCUID^{Sim}: A Platform for Smart Card User Interface Research, Development and Testing

Markus Ullmann

Federal Office for Information Security
D-53133 Bonn, Germany
www.bsi.bund.de
and
University of Applied Sciences Bonn-Rhine-Sieg
www.h-brs.de
Email: markus.ullmann@bsi.bund.de

Ralph Breithaupt

Federal Office for Information Security
D-53133 Bonn, Germany
www.bsi.bund.de
Email: ralph.breithaupt@bsi.bund.de

***Abstract*—The latest advances in the field of smart card technologies allow modern cards to be more than just simple security tokens. Recent developments facilitate the use of interactive components like buttons, displays or even touch-sensors within the cards body thus conquering whole new areas of application. With interactive functionalities the usability aspect becomes the most important one for designing secure and popularly accepted products. Unfortunately the usability can only be tested fully with completely integrated hence expensive smart card prototypes. This restricts application specific research, case studies of new smart card user interfaces, concerning applications and the performance of useability tests in smart card development. Rapid development and simulation of smart card interfaces and applications can help to avoid this restriction. This paper presents SCUID^{Sim} a tool for rapid user-centric development of new smart card interfaces and applications based on common smartphone technology.**

Keywords–Smart Card; User Interface Design, Interactive Smart Card Applications; Rapid Prototyping.

I. INTRODUCTION

Recently developed interactive components allow the integration of input devices, like buttons, keypads or gesture interfaces as well as output devices like displays and LEDs directly into a smart card. This offers especially new security services like "on-card" user authentication and trusted displays and avoids the use of external terminals which are potentially vulnerable to active and passive attacks.

With the interactive functionalities the usability aspect becomes the most important one for designing a usable smart card and adds many new demands to the development process. Now aspects like the adequate size of a button, the visibility of a touch interface, the resolution, contrast and speed of a display and the overall design of the card have to be addressed as well as an appropriate hardware/software-codesign to ensure clear user guidance and high overall usability. This can only be achieved by conducting extensive field tests with as many different people as possible. Creating the necessary card prototypes with the complete design and full hardware and software functionality can be very expensive and time consuming which makes usability centered security research difficult. In this paper we present an alternative approach to allow the necessary testing in order to determine the requirements for design, hardware components and the software without the need to build costly prototypes. By using common smartphones almost all user related aspects can be investigated by simulating the "look & feel" of a new smart card design before any real hardware integration is needed.

SCUID^{Sim} is an android application and therefore usable on a wide range of smartphones which combine in a single compact device all the necessary hardware input/output components as well as communication links, cryptographic services, the processor power and memory needed for simulating a large variety of current and future smart card interfaces and applications. With SCUID^{Sim} the visible aspects of a multi-component smart card can be designed on the smartphone. Based on a simple SCUID^{Sim}-API, user defined card applications can be executed while SCUID^{Sim} simulates the behavioural properties of all interactive components. New requests and requirements can be implemented, simulated and evaluated instantly. This way SCUID^{Sim} supports detailed requirement engineering for software as well as hardware and the development of new user interface concepts hand in hand. This is especially useful for the design and integration of new usable user centric security algorithms in smart cards.

The rest of the paper is organized as follows: Section II starts with a description of related work. Section III provides a brief overview of the software architecture of SCUID^{Sim} and its functionality. Next, Section IV describes a first case study of a contactless smart card with a low cost user interface. The user interface consists of a touch slider component for user input and a display component implemented as 3×5 LED matrix which can only illustrate one character with very limited details. The use case "on-card" user authentication shows how concepts for user guidance and visual user feedback can be tested and evaluated in SCUID^{Sim}. Finally, Section V summarizes the findings of this paper and gives an outline of open issues regarding SCUID^{Sim}.

II. RELATED WORK

The first research and development projects investigating the idea to integrate input and output elements in smart cards go back as far as the late 1990s, see [1]. With the advances in low power and low profile embedded technologies

 ISBN: 978-1-61208-356-8

many different component technologies have been successfully developed and integrated in ID1-compatible smart cards during the last decate. Primarily a variety of display types and buttons even fingerprint scanners are discussed for integration, see [2] and [3]. Moreover, in [4] smart cards with an integrated display as security enforcing component are introduced. A first approach to integrate a 2D on-card gesture input sensor implemented as capacitive touch matrix is introduced first in [5]. It has also been an important topic for public funding in many countries (e.g. the INSITO-project of the german federal office for information security and the SECUDIS-project of the german ministry of education and research, see [6], and [7]). Despite all the effort and the growing number of available components, interactive smart cards have not yet been used in many real applications. Amongst other reasons this is due to high production costs and the much higher complexity of such smart cards. With the recent advances in printed electronics capacitive sensors have become standard technology and even printed displays are available today, see [8], [9], and [10]. But the complexity issue is still a serious obstacle on the way to the final product. At least regarding the system integration issues of combining several hardware components there have been approaches for rapid prototyping tools. One of the first was the FlexCOS system suggested by Beilke at al. [11], which uses FPGAs for a very flexible and rearrangable interface to connect separate component prototypes into one complete system. Although this approach became standard procedure for many manufacturers and researchers it only covers the technological aspects. Such functional prototypes are much too bulky and fragile to conduct real world tests with many people in real application scenarios outside the lab. The usability aspects that first and foremost define how the smart card should interact and therefore what the requirements for the hardware and software components really are can not be tested without fully integrated and designed card prototypes. Unfortunately each version of real prototypes to test for user acceptance require huge expenses of time and money. This lack of end-user centered rapid prototyping tools was the starting point for the development of the SCUID^{Sim} tool. Simulation of user interfaces was very popular in the beginning of ubiquitous computing. One approach was the iStuff toolkit to support the development of user interfaces for the post-desktop age for multiple displays, multiple input devices, multiple systems, multiple applications, and multiple concurrent users, see [12]. Alternative technologies were developed by the Stanford Interactive Workspaces project for multi-person and multi-device collaborative work settings, see [13]. To the best of our knowlege SCUID^{Sim} is the first approach to model, simulate and analysis user interfaces for (contactless) smart cards.

III. SCUID^{Sim} Architecture

SCUID^{Sim} consists of two modules: a card designer which enables a flexible but simple arrangement of smart card layouts based on preconfigured components and a card simulator. In the card simulator such a card layout can be paired with a smart card application in a real time simulation. It was a design decision to separate the card design process and the card simulation process in two independant software modules. Figure 1 illustrates the SCUID^{Sim} architecture.

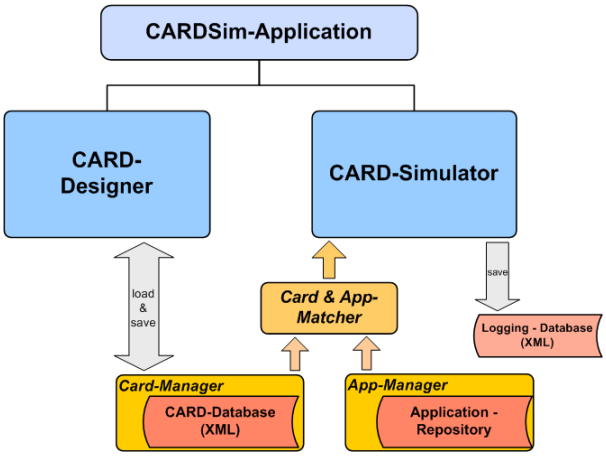


Figure 1. Overview of the SCUID^{Sim} software architecture.

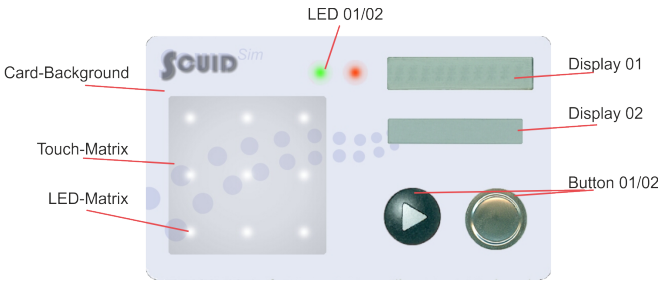


Figure 2. Available card components (in the card designer)

A. Card Designer

The card designer is a simple tool to engineer smart card layouts. Figure 2 gives an overview of the available components in the current version of SCUID^{Sim}. Currently the following predefined components: push buttons, segmented displays (7- and 14-segments), matrix displays (RGB, greyscale and black & white), LEDs, n × m LED-matrixes, 2D-touch sensors, image boxes and the overlay image of the smart card are supported. There are also non-visible components like e.g. acceleration sensors that are automatically available to all cards if the used android smartphone is supporting it. So based on this predefined components, SCUID^{Sim} can simulate a huge variety of smart card layouts. Figure 3 depicts a real card prototype opposite to a replicated design of this card within SCUID^{Sim}. This Figure illustrates the very realistic replication capabilities of our tool.

Within the card designer the properties of each compo-

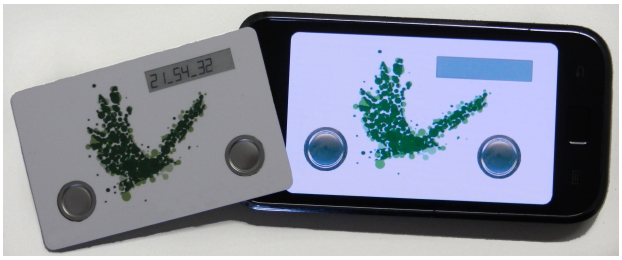


Figure 3. Confrontation real - and simulated card layout within SCUID^{Sim}

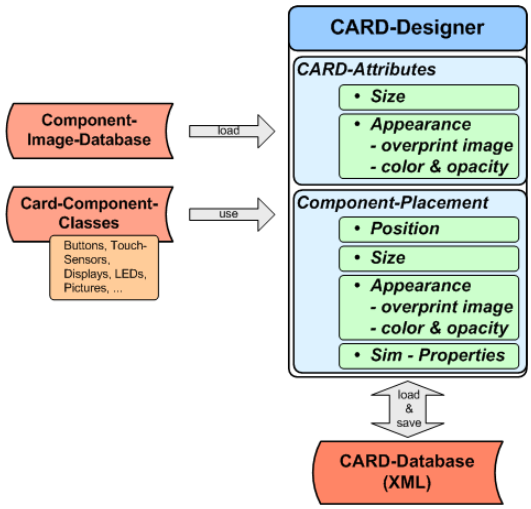


Figure 4. Software architecture of the card designer module

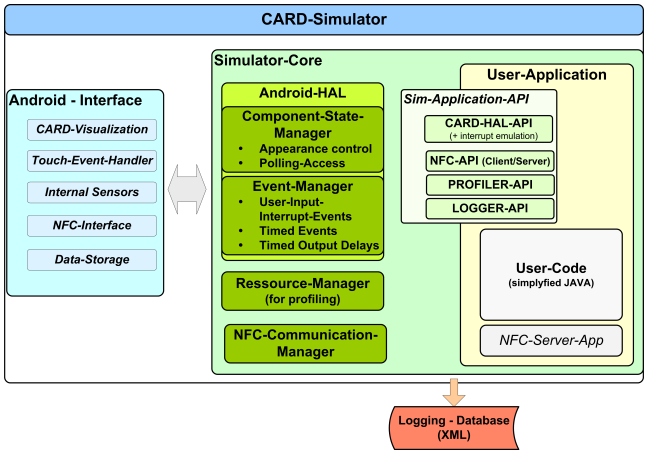


Figure 5. Software architecture of the card simulator module

nent like selection, position, size, and deletion can easily be controlled via simple finger gestures commonly known from many other mobile applications. Additional properties like overlay image (appearance of the component), color modifier (to the overlay image, in RGB and alpha for transparency) or component specific properties like X/Y-resolution of a matrix display, or update delay time for a display component can be set in a component property page that is dynamically generated based on all the properties of a selected card component. Each card component, its properties and its specific simulated behaviour (e.g. delay of the visual update) is defined in the respective class within the component library of SCUID^{Sim}. To add new components or behavioural functionality to this library the developer simply inherits and modifies the provided component base class. All administrative support like the list of available component types and the components property page are generated "on the fly". The complete card design can be loaded from and saved to a card library in an XML-format that can be read and edited outside SCUID^{Sim} with all existing standard XML-viewers/editors. Figure 4 depicts the software architecture of the card designer.

B. Card Simulator

The two main objectives of the card simulator are to provide a flexible framework for the development and evaluation of card applications and to simulate an user interaction as close to a real smart card as possible. For creating card applications the card simulator offers a simple API in order to access the interactive components of the simulated card. In order to keep as close to a real card program as possible the API allows input components to be polled as well as providing a simulated interrupt event handling. The concept of the API is based on the intention to shield the application developer from Android Java specific constructs in order to facilitate application code that can easily be transferred to real smart cards. In addition the card simulator consists of a ressource manager module for simple profiling purposes as well as a flexible logging system. Since most applications for contactless smart cards implicate a communication to a reader/server via NFC (ISO 14443) the card simulator offers an interface to the real NFC component of a smartphone. This way the simulated card can also be used in the targeted environment. If the real NFC component cannot be used, the framework allows the execution of NFC server applications in order to also simulate the reader functionality. A manual describing the usage and programming of applications can be found in [14]. Figure 5 depicts the software architecture of the card simulator module.

IV. Case Study: Smart Card with 3 × 5 LED Matrix Display

In this Section a contactless smart card with a very restricted user interface is described and analyzed. Typically contactless cards follow the ISO 14443 specification, see [15]. This means that contactless smart cards have no battery. They are powered by a magnetic field of a terminal device. So the available energy on real contactless smart cards for powering additional components is very limited. Due to the very limited available power the user interface consists only of a 3 × 5 LED matrix as an information display and an additional touch slider component for controlling user inputs. Wiping enables scrolling the characters of the alphabet and a long touch ($\geq$ 2 second) selects the denoted character in the display. First user tests have shown that a matrix display with less then 3 × 5 LEDs reduze the readability of characters seriously. So from a readability perspective a 3 × 5 LED matrix display seems to be a minimum requirement. To power a real contactless card with a 3 × 5 LED matrix display and an additional touch component is technically possible but from an energy perspective still a challenge. Figure 6 describes the card layout of the analyzed card.

Here we analyse the described user interface especially for performing an user authentication process on-card. The user authentication is performed based on a shared secret (classical password which consists of a sequence of 4 or 6 digits, e.g. *5839*). The reference password is already configured and securely stored in the card memory. Here only the user authentication process itself without additional services, like changing the user passwords, etc. is regarded. Further security aspects of smart cards and smart card application like secure storage of the password and other secrets, used cryptographic protocols or side channel free implementation of cryptographic algorithms are not subject of this paper. Here we refer to [16] or [17]. The main issue of the use case is the demonstration

 ISBN: 978-1-61208-356-8

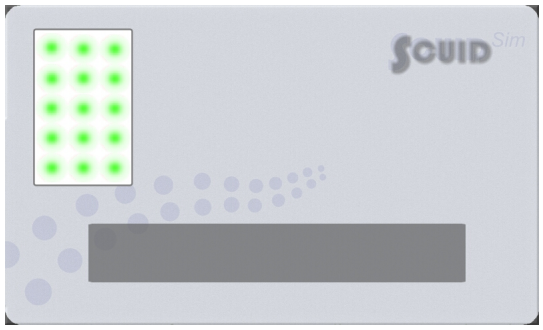


Figure 6. Layout of the smart card used in the case study (3 × 5 LED matrix display and slider component)

of the capabilities of SCUID^{Sim} to analysis user interfaces.

Concerning the user interface following questions arise:

- Which alphabet is useable with a 3 × 5 LED matrix display ?
- How can status information of the application and necessary feedbacks be given to the user, e.g. feedback of a character selection or a successful respective failed password authentication ?

A. On-card User Authentication Process

The authentication process has three security states: *in authentication — authenticated — locked*. If the card is wattless or the authentication process is still running the card is in the security state *in authentication*. If the user authentication is performed successfully the card switches to state *authenticated* and further card applications (which are not described here) can be performed. After card processing the security state switches to *in authentication* again. If the user exceeds the maximum number of authentication retries during the authentication process the card is blocked and switches to security state *locked*. In Figure 7 the whole authentication process is depicted. The security states are highlighted as trapeziums and have to be visualized to the user by an adaquate shown symbol or text.

The user authentication process starts with displaying the security state *in authentication*.

Next, one character of the used alphabet is displayed. Each selected character of the user (long user touch ≥ 2 seconds) is shortly displayed (as user feedback) followed by illustrating the number of already inserted password characters depicted as understandable symbols. If the live password has completely been entered (e.g. 4 digits) the password verification (live password $\stackrel{?}{=}$ reference password) is performed automatically. A successful user authentication has to be depicted clearly to the user. If the password authentication fails the second authentication attempt starts automatically. This has to be shown to the user. If the second authentication attempt fails again the last attempt (assumption: three password attempts) is performed which has to be denoted to the user as well. If the final authentication fails again the card switches to the security state *locked* and no user operation is possible any more. This security state has also to be displayed to the user. The white boxes in the Figure 7 represent position of the authentication process where feedback information have to be given to the user by text outputs or specific symbols.

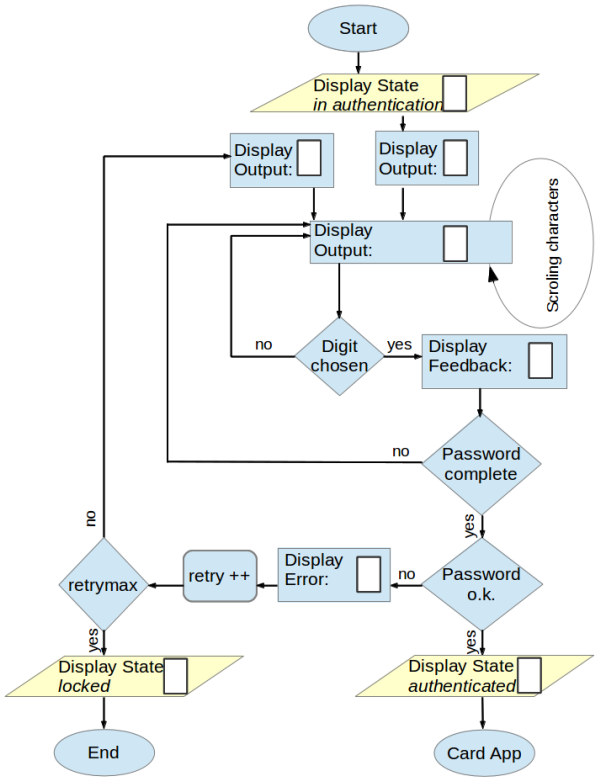


Figure 7. Authentication process

Summing up, following situations and security states have to be shown to the user:

- current security state *in authentication — authenticated — locked*
- number of already entered characters
- successful resp. failed user authentication
- current authentication round (status password retry counter)

In addition following information like: application start, application end, card in processing, etc. seems to be important information for the user and has to be displayed, too.

B. General Display Illustration Facilities

First, the principle illustration facilities of a 3 × 5 LED matrix display are presented. Static characters:

1) Characters, e.g. alphabet shown in Figure 8
2) Special characters, e.g. dice symbols shown in Figure 9
3) Symbols, e.g. arrows, rectangle, box, horizontal and vertical lines, ...

Animated symbols:

4) Special characters, e.g. falling dice symbols
5) Symbols, like falling arrow (picture frequency 200 ms), curtain up (picture frequency 200 ms), curtain

 ISBN: 978-1-61208-356-8

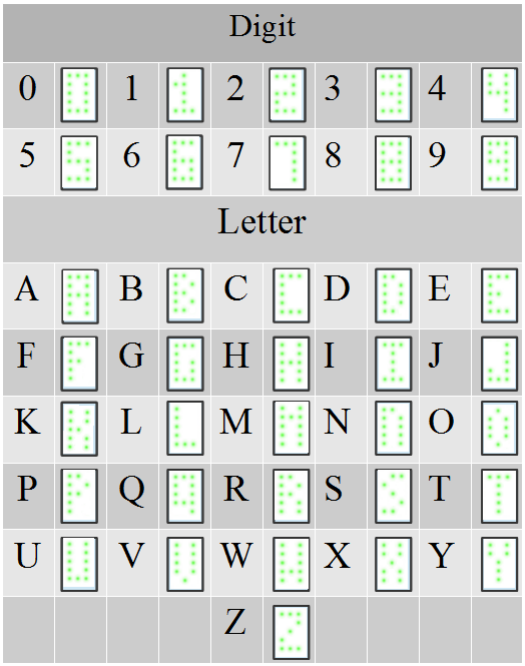


Figure 8. Alphabet

Figure 9. Dice symbols

down (picture frequency 200 ms), and rotary dots (dot frequency 200 ms) shown in the first row from left to right in Figure 10 and helix construction (sequentially build up dot by dot with dot frequency 200 ms), helix destruction (sequentially build up dot by dot with dot frequency 200 ms), o.k. symbol (sequentially build up dot by dot with dot frequency 200 ms), and fail symbol (sequentially build up dot by dot with dot frequency 200 ms) shown in the second row from left to right in Figure 10.

This listing is not complete. But it shows that even the very restricted display enables the presentation of a large range of characters and symbols especially when static and dynamic (animations) effects are exploited.

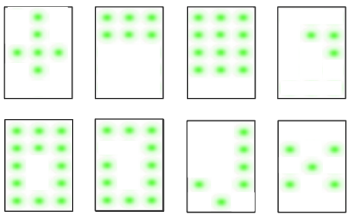


Figure 10. Animated symbols

C. User Test

First user test with only fourteen persons (students in age 20 - 30: 2 females, 12 males) have been performed. Obviously that is no adequate user cross section and no statistical relevant number of attendees. However first interesting user perceptions can be given.

The used digits of the alphabet shown in Figure 8 are distinguishable. But as seen in Figure 8 some characters are only poorly distinguishable in a 3×5 LED matrix display setting, e.g. B, G, N, O, P, Q, X. Lower letters worsen the problem dramatically. This means if only digits are processed a 3×5 LED matrix seems to be sufficient. But if letters should be processed too higher resolution displays e.g. a 5×5 LED matrix display or a 4×7 LED matrix display is needed to achieve better letter readability.

Additionally we tried to output short words (e.g. on, off, o.k., ...) as feedback to the user by sequentially diplaying the characters of the word. The users have enormous problems to read and identify even very short words depicted as sequence of letters when they do not know the displayed word before. The consequence is that this approach for displaying words can not be followed anyway in a 3×5 LED matrix display setting.

On the contrary animated symbols like, falling arrows and rectangles, dynamic curtain, circling dots etc. seem to be very intelligable to the user. Animated symbols seem to be a suitable alternative to text output to indicate card states and to give feedback information to the user. So we performed additional user tests concerning animated symbols.

We used our sample card (see Figure 6) to show the participants sequentially animations of the symbols displayed in Figure 10 in an unsorted order to link the symbols to predefined meanings. We applied following procedure: Each symbol is animated first and afterwards shown for 5 seconds to the participant as a static symbol before the animation of the next symbol starts. This procedure was performed for 5 minutes. During this test the attendees should match an animated symbol to one of the given meanings: authenticated — locked — in authentication — start application — end application — card in processing — password check o.k. — password check fails — first authentication attempt. The result was very heterogeneous. Apart from password check o.k. (o.k. symbol) and password check fails (fail symbol) with nearly 40 % correct assignments there was no significant occurency of any symbol meaning. The mappings (meaning $\leftrightarrow$ symbol) of the participants were very scattered. The finding is that even symbols need to be chosen very carefully and have to be explained to the user in detail. If it is possible to use symbols which are intelligible to all they should be applied in any case.

The consequences for our application with a 3×5 LED matrix setting are therefore:

- only digits should be used as alphabet
- text outputs as user guidance is not possible instead (animated) symbols should be used
- we prefer dice symbols to denote the number of already inserted characters of the password and
- animated symbols for indicating the security state and for arbitrary feedback to the user.

 ISBN: 978-1-61208-356-8

But which animated symbol should be used for indicating the current security state *in authentication — authenticated — locked*, already performed authentication attempts . . . is still an open issue and is subject of further examinations.

V. CONCLUSION

In this paper we present our smart card user interface development and simulation tool SCUID^{Sim}. This tool enables the rapid development and simulation of smart card interfaces and applications. It can be used for user interface research, easy prototyping and performing of tests.

Therefore this tools is useable for early consideration of user handling requirements and overall user acceptance of user interfaces before a time consuming and costly prototype development starts. Especially, card designs and application modifications are performed very quickly in software without any hardware modification. This reduces the development of smart card prototypes and speeds up the whole development process. Moreover the tool is very useful for the design and exploration of new usable security concepts and algorithms for contactless cards and enables further application specific research in this direction.

SCUID^{Sim} is available in version 0.3. After further useability studies it is planned to implement additional functionalities. At the moment microphones, cameras, initial sensoric and biometrics are not supported. Here, new card component classes have to be developed and integrated in SCUID^{Sim}.

Furthermore, SCUID^{Sim} is a whole framework. Card user studies do not require the whole functionality of the framework. A specific SCUID^{Sim}-subset is sufficient. This subset of SCUID^{Sim} is currently not available, too.

The test of new interface concepts requires the consideration of different user groups (range of user ages, sex, . . .). To achieve a good test performance and test analysis a central logging of the test data and user feedbacks is desirable. Furthermore an automatic code update for all smartphones involved in an evaluation is preferable. That is subject of a further enhancement, too.

Within this paper a case study for a very restricted user interface for smart cards is described together with first user test results. This case study depiced in Section IV presents the potential of SCUID^{Sim} to analyse new user interface approaches. The analyzed user interface consists of a 3×5 LED matrix display and a slider component. This setting was used due to energy restrictions for additional components in real contactless cards. The test results show that a 3×5 LED matrix display is not adequate do display letters of an alphabet. It is sufficient to display digits. But it show that even the very restricted display enables the presentation of a large range of symbols when static and dynamic effects are exploited. This first results are used to guide further studies of smart user interfaces for contactless smart cards.

VI. ACKNOWLEDGEMENT

The authors would like to thank our students Anton Buzik, Alexander Kreth and David Sosnitza for their support implementing SCUID^{Sim}, and our colleague Christian Wieschebrink for valuable remarks. Also thanks to the anonymous reviewers for the valuable comments.

REFERENCES

[1] Gerald V. Piasenka, and Thomas M. Fox, and Kenneth H. Schmidt, "Solar cell powered smart card with integrated display and interface keypad," 1998, US patent US5777903.

[2] J. Fischer, F. Fritze, M. Tietke and M. Paeschke, "Prospects and Callenges for ID Documents with Integrated Display," in Proceedings of Printed Electronics Europe Conference, 2009.

[3] Bundesdruckerei, "RFID Security Card with a One-Time Password and LED Display," 2013, www.rfidjournal.com/articles/10512.

[4] M. Ullmann, "Flexible visual display unit as security enforcing component for contactless smart card systems," in Firth International EURASIP Workshop on RFID Technology (RFID 2007), 2007, pp. 87–90.

[5] M. Ullmann, R. Breithaupt, and F. Gehring, "On-card user authentication for contactless smart cards based on gesture recognition," in Proceedings GI Sicherheit 2012, ser. Lecture Notes In Informatiks, no. 108, 2012, pp. 223–234.

[6] BDR, "Secudis Project," 2012, http://www.bundesdruckerei.de/en/684-innovative-high-security-solutions.

[7] "Thin chips for document security," in Ultra-thin Chip Technology and Applications, J. Burghartz, Ed., 2011.

[8] Taybet Bilkay, and Kerstin Schulze, and Tatjana Egorov-Brening, and Andreas Bohn, and Silvia Janietz, "Copolythiophenes with Hydrophilic and Hydrophobic Side Chains: Synthesis, Characterization, and Performance in Organic Field Effect Transistors," Macromolecular Chemistry and Physics, vol. 213, September, 26 2012, pp. pp. 1970–1978.

[9] P. Andersson, R. Forchheimer, P. Tehrani, and M. Berggren, "Printable all-organic electrochromic active-matrix displays," Advanced Functional Materials, vol. 17, no. 16, 2007, pp. pp. 3074–3082. [Online]. Available: http://dx.doi.org/10.1002/adfm.200601241

[10] PaperDisplay, "Printed Display Products," 2013, http://www.paperdisplay.se.

[11] K. Beilke and V. Roth, "Flexcos: An open smartcard platform for research and education," in Proceedings of the 6th International Conference on Network and System Security, NSS 2012, ser. Lecture Notes In Computer Science, no. 7645, 2012, pp. pp. 277–290.

[12] R. Ballagas, M. Ringel, M. Stone, and J. Borchers, "istuff: a physical user interface toolkit for ubiquitous computing environments," in Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 2003, pp. 537–544.

[13] J. Borchers, M. Ringel, J. Tyler, and A. Fox, "Stanford interactive workspaces: a framework for physical and graphical user interface prototyping," Wireless Communications, IEEE, vol. 9, no. 6, 2002, pp. 64–69.

[14] BSI, "$SCUID^{Sim}$ manual, version 0.3," 2013.

[15] ISO/IEC, "ISO/IEC 144443 contactless Integrated Circuits Cards, Part 1-4: Physical Characteristics (1), Radio Frequency Power and Signal Interface (2), Initialization and Anticollision (3) and, Transmission Protocol (4)," 2000.

[16] W. Rankl and W. Effing, Smart card handbook. John Wiley & Sons, 2010.

[17] S. Mangard, E. Oswald, and T. Popp, Power analysis attacks: Revealing the secrets of smart cards. Springer, 2008, vol. 31.

 ISBN: 978-1-61208-356-8

Recognition of Simple Head Gestures Based on Head Pose Estimation Analysis

George Galanakis
Institute of Computer Science,
Foundation for Research and Technology Hellas
Computer Science Department,
University of Crete, Greece
e-mail: ggalan@ics.forth.gr

Pavlos Katsifarakis
Computer Science Department,
College of Science
Swansea University, UK
e-mail: 750183@swansea.ac.uk

Xenophon Zabulis, Ilia Adami
Institute of Computer Science,
Foundation for Research and Technology Hellas
e-mail: {zabulis, iadami}@ics.forth.gr

***Abstract*—A recognition method for simple gestures is proposed and evaluated. Such gestures are of interest as they are the primitive elements of more complex gestures utilized in natural communication and human computer interaction. The input to the recognition method is obtained from a head tracker that is based on images acquired from a depth camera. Candidate gestures are detected within continuous head motion and recognized, acknowledging that head pose estimates might be inaccurate. The proposed method is evaluated within the context of human-computer dialog.The reported results show that the proposed approach yields competitive recognition results to state-of-the-art approaches.**

***Index Terms*—head gesture recognition; head gesture detection.**

I. Introduction

The ability to recognize purposeful head motions, or gestures, is a special problem both in computer vision and human-computer interaction. Solving this problem accurately and robustly is of particular interest, because such head motions convey information that can be used in the natural communication of a person with a computer, or an intelligent environment. In this work, head gesture recognition targets purposeful head motions that are responses to a user interface dialog.

A central component of any head gesture recognition system is the estimation of head pose ([1]–[3]) and motion. Head pose information is of particular importance in a variety of applications and has received considerable attention in the recent years [4]. The selection of the sensory modality is important, as it relates to the reliability of this estimation which can, in turn, affect the performance of recognition. Practicality and applicability through cost-efficient and off-the-shelf hardware is also of concern and, thus, this work employs a state-of-the-art head tracker that is based on commodity depth cameras, nowadays widely available as Red Green Blue Depth (RGBD) sensors [5].

In most cases, natural gestures can be analyzed in simpler motions. For example, a horizontal shaking of the head to express negation, is usually repetitive. Moreover, each one of the repeated motions can be further analyzed. In the aforementioned example, the gesture can be regarded as a leftward and a rightward head rotation (or vice versa). In this work, we focus on the recognition of such simple motions, which we call primitive gestures. Our interest is twofold. First, due to their simplicity, these gestures are suitable for use in human computer dialogues. Second, primitive gestures are elements of higher order gestures and, thereby, their robust recognition is relevant to the recognition of more complex gestures.

In the context of this work, we use the notion of a reference head pose which, in our case, is the frontal (or, "looking straight ahead") pose. We also parameterize, human head 3D orientation upon the natural head rotations, which are called yaw, pitch, and roll (see Fig. 2). In this reference, primitive gestures correspond to a peak in the values of an angular component, while no significant rotation occurs in the remaining two angular components.

To determine the extent that the proposed method can be useful in human computer interaction we evaluate it through quantitative evaluation, in which recognition performance is measured. At the same time, this evaluation serves a secondary goal. By observing and profiling the way that subjects perform primitive gestures (i.e., how fast or how steep is a head rotation), information regarding the corresponding user motions is collected. In turn, this information can be exploited in the better recognition of these gestures and the design of systems that utilized them.

The rest of this paper is organized as follows. Section II presents related work on head gesture recognition methods and applications, Section III includes implementation details, Section IV discusses experiments and results, Section V briefly presents the applicability of head gestures within a specific example application, and Section VI concludes the presented work and suggests further applications in which head gestures can be employed.

II. Related Work

Work on the recognition of head gestures has started to emerge as long as two decades ago, but has been recently reinforced after the wide availability of depth cameras, which facilitate the pose estimation of the human head.

Some approaches to head gesture recognition capitalize on a special type of sensor (i.e., inertial [6] or pupil tracking [7]) and setup, which provides confidence to the input signal from head pose estimation. In turn, this input signal exhibits increased continuity and reduced noise and its processing is, thereby, simpler. At the other end, some approaches employ a fully passive (RGB or monochrome) camera to estimate head pose. Pertinent methods rely on facial feature detection (i.e., mouth, nose, eyebrows) and tracking to acquire head pose and motion [7]–[9]. In [10], direct measurements (pixel intensities) are utilized, but resorting to assumptions about the facial appearance of the subject and providing less accurate

 ISBN: 978-1-61208-356-8

results. In terms of sensory input, this work falls in the middle of the above range, utilizing a commodity RGBD sensor, as in [11]. Only the depth information is utilized to avoid sensitivity to illumination. However, although depth information is much more robust than color/intensity, input cannot be considered neither noise nor error free. In this context, this work accounts for poor, erroneous, or missing estimates provided as input.

The methods employed for head gesture recognition can be classified into two main categories, those which employ a Finite State Machine (FSM) and those which are based on learning, typically through an instantiation of Hidden Markov Models (HMMs).

Simple gestures, such as the one of interest in this work, have been recognized by a number of methods that employ FSMs. FSMs are simple to formulate but, in the other hand, do not scale with ease. In [7], an FSM recognizes nodding and shaking gestures, which are then used in the context of a dialog-based user interface. The same FSM is used by [8] in a self-portrait camera which is controlled by nodding and shaking gestures. In [12], an FSM is introduced for detecting nodding and shaking gestures, useful for interacting with avatars on mobile devices. In [13], FSM-detected head gestures have been used along with hand gestures in order to achieve interaction within a multi-modal user interface. The above methods lead to the use of rather complex FSMs in order to accommodate multiple gestures, while still support a smaller vocabulary of gestures (typically 4, based on up, down, left, and right motions).

Methods based on machine learning and recognition of temporal patterns techniques are also present in the literature. Recently, in [11], two HMMs are trained to recognize nod and shake gestures; "other" gestures are recognized by a third HMM as fallback. In [9], a HMM is trained for each of three different head movement gestures; right, left, left-forward, which are used in the context of sign language sentences. In [14], shaking, neutral and nodding gestures are detected by continuous HMMs and then provided to a dialog manager which operates a coffee machine. Similarly, in [15], Ordered means models (OMMs) are trained to recognize nod, shake, tilt and look gestures among two participants in a conversation; OMMs, are described as "rigorously reduced versions of HMMs. In [16], a multi-class Support Vector Machine (SVM) is augmented with contextual features, to recognize nod and shake gestures. These gestures are evaluated in the context of document browsing and dialog box confirmation. Finally, in [17], a multi-class SVM is trained to detect "Yes" and "No" head gestures, along with other hand gestures. In the same context, some methods learn gestures directly from posture data such as [6] which operates on head orientation readings to detect nodding and shaking gestures. In [10], a set of ten gestures is recognized by Continuous Dynamic Programming which compares live images with previously trained image sequences, annotated respectively.

The works above employ HMMs to treat gestures that contain multiple more simple gestures, resulting in complex gesture models, while considerable effort is required for the training of the system. In comparison to HMMs and SVMs the proposed work does not require a preceding training phase, but capitalizes upon the examination of each rotational component of the head pose. Moreover, it is concluded that the results of the proposed work, as shown in Section IV-A1, are not only comparable but, in most cases, outperform recognition rates in the literature.

Fig. 1. The raw input depth from the tracker (left) and the head pose estimate superimposed on the input color image.

III. METHOD

A. Sensory input

A head tracker [1], that receives input from an RGBD camera is employed to sense the current pose of the subject's head, in real-time. Fig. 1 illustrates the result of the tracker for a given input image. It is noted that any other head tracker (i.e., [2], [3]) could be used instead of this one, however the particular one was selected due to its reported increased accuracy and execution speed. The input is either the estimated 3D pose or null (in case of tracking failure) and is received multiple times per second. Acknowledging that erroneous or inaccurate estimates may be provided, as well as, that tracking may exhibit transient failures this information stream is adopted as the "sensory input" to the proposed system.

As the head tracker and the recognition system which we developed are implemented in different programming environments, their communication was achieved through a service interoperability platform [18].

3D pose is defined as the 3D translation and 3D rotation of the head from a reference pose and is, thereby, represented by 6 degrees of freedom (6 DoF). These DoF correspond to a translation 3D vector and a 3D rotation which is parameterized as in terms of Euler rotations, that is, as a rotation of the head about the xx', yy', and zz' axes. These rotations are referred as P_i, Y_i and R_i respectively (see Fig. 2).

In the context of this work, translation does not play a primary role as we assume the expression of gestures to be invariant to the translational motion of the head and that they can, also, be expressed while the subject is in motion. We also assume that rapid and large motions of the subject's head, which would influence the comprehension of a gesture do not occur as they are not typically performed by subjects.

B. Parsing of candidate gestures

Before describing recognition approach, we model the pursued primitive gesture, as a motion which starts and ends at the reference pose and, in between, a single peak of significant

 ISBN: 978-1-61208-356-8

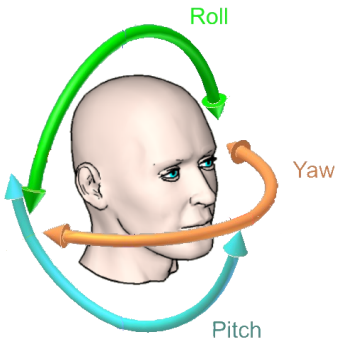


Fig. 2. Head rotation axes.

amplitude in the value of an orientation component occurs. Candidate gestures consist of a sequence of three head states;

1) postured and approximately motionless for a brief time interval at the reference pose,
2) performing a rotational motion and, possibly, a mild translation motion, and
3) postured again at the reference pose for a brief time interval,

Using this description we are able to "parse" the continuous sensory input into constituent, discrete elements. Each such element is then considered as candidate gesture. A candidate gesture is a head motion that might be expressing a gesture, or not. Each such candidate, is attempted to be, correctly, recognized as a gesture or as non-gesture. For candidates recognized as expressing a known gesture, labeling of the particular gesture is also attempted.

Depending on the type of motion of the second state, the gesture may be recognized as an instance of the known gestures, or not. The reference pose is defined as the pose of the head at approximately zero rotation in all the 3 axes. We have extrinsically calibrated our camera and estimated its relevant posture to the ground plane through, conventional, grid-based calibration [19]. In this way, we performed a change of reference coordinate and poses so that the reference pose, in our setup, this corresponds to the user's head facing frontally without any inclination of the head. The reference pose is defined to occur when $\forall\ p \in \{P_i, Y_i, R_i\},\ |p| < \tau_r$, where τ_r is a configurable threshold relaxing the requirement for exact frontal posture and is in the order of a few degrees ($10°$).

To parse candidate gestures we defined a simple state-machine, with parameterization in the transition of the states. We call it Buffered State Machine because the transition from a state to another is performed when a buffer is completed by a number n of valid tokens. This means that we have to acquire n consecutive poses in reference position to start identifying the gesture. This stabilizes the system against small estimation errors. The value of n is configurable with respect to the frame rate that the head tracker operates. In our implementation the value $n = 5$ was selected, based on preliminary observations of user behavior, adjusting the head rest at the reference pose

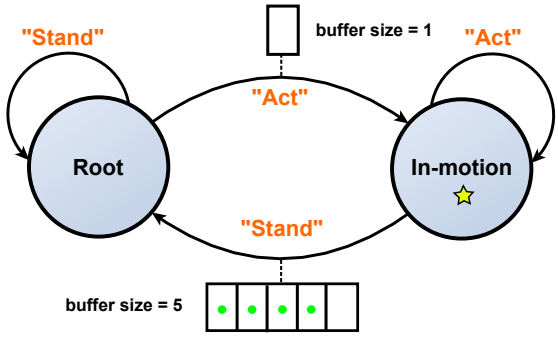


Fig. 3. The state machine which was employed to parse candidate gestures.

to have an (approximate) duration between $1/3 - 1/2$ $sec.$ Fig. 3 illustrates the described state-machine, being in "In-motion" state. In the particular case, if we receive one more "Stand" token (command), a transition to "Root" will be performed. Otherwise, the buffer will be invalidated, since we want 5 consequent "Stand" tokens.

When the subject's head is detected by the sensor, the pose estimation is continuous and the recognition component stores the estimations in a double buffer. Whenever the transition to the reference pose occurs, the current buffer is "parsed" and passed to the next stage for recognition, while the other buffer stores the more recent poses. The above is feasible because poses are received as events via the interoperability platform and are handled by a different thread. In cases of head pose estimation failures, a null result (estimate) is produced. In such cases, the recognition will stop receiving events until the tracker resumes operation. If such an event occurs during the expression of a gesture then, typically, the gesture fails to be recognized.

C. Gesture detection and recognition

Upon parsing of the gesture, the signal segment acquired during the "In-motion" state is assessed, in order to reason whether the candidate is indeed a primitive gesture and, if so, recognize which one it is.

To detect a gesture we investigate the content of the rotational components of this signal segment. We test for two conditions, for this purpose. The first is that the motion in the rotational component corresponding to a particular gesture matches the prototype of the gesture. Fig. 4 illustrates a prototype motion as assumed above. The second is that the remaining 2 rotational components do not correspond to a significant motion.

To test for the first condition, we consider the values of the 3 rotational components (pitch, yaw, roll) of the pose estimates. Each component is independently processed and its input is treated as a stream. Prior to its consideration, each stream is passed through a low-pass, Gaussian filter to eliminate tracking jitter. Henceforth, we call the signal of a rotational component within the time interval $[t_A, t_B]$ as dominant, if it is the sole one exhibiting significant motion. For example, Fig. 5(a) illustrates the acquired sensory input for the Y rotational component, at a time interval which is

 ISBN: 978-1-61208-356-8

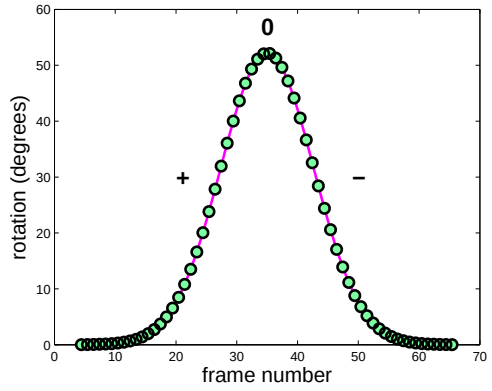


Fig. 4. Prototype of head motion as a function of the rotational component corresponding to the axis of the particular rotation.

segmented (dashed blue lines) by the Buffered State Machine, while Fig. 5(b) shows the output from its low-pass filtering. Note that, in Fig. 5(b), only the part between the two dashed blue lines is shown and, thus, they are omitted.

The second condition implies that, in addition to detecting a dominant motion we need to determine that motions in the remaining 2 components are insignificant, or henceforth "neutral" motions. For this purpose, "soft" thresholds (s_T) are defined. A "soft" threshold specifies the accepted amount of motion in a rotational component, when it is considered irrelevant to a gesture. For instance, when a "Head Up" gesture is performed, we do not expect significant motion in the yaw component. Henceforth, we call the signal of a rotational component within the time interval $[t_A, t_B]$ as neutral if it does not surpass the soft threshold s_T.

In order to recognize a gesture, each rotational component is investigated separately. Let f be a function of time which represents the value of the rotational component in consideration (pitch, yaw, roll). Let also $[t_A, t_B]$ the time interval for which the signal of the above component was acquired. As the primitive gestures to be recognized have the form of a peak, in the dominant rotational axis, candidate gestures are first tested as to whether they exhibit the potential of containing such a peak. This consists of the fulfillment of the following three conditions:

1) a single peak of f occurs during the entire interval
2) f advances in a strictly positive (negative) followed by a strictly negative (positive) manner around the peak
3) a threshold h_T is overcome, so that the peak exhibits significant amplitude to be attributed to an intensional gesture rather than an unintentional head motion.

In Fig. 5, characteristic data are shown for the Y component of rotation for the ideal model of a Head Down gesture, the acquired sensory input, and the output from its low-pass filtering. The implementation of these three conditions is as follows.

First, the peak has to be single; $\forall t_i \in [t_A, t_B]$, f has a single peak e. The reason is that recurring motions during the "In-motion" state should be omitted. Thereby, the zero-crossings

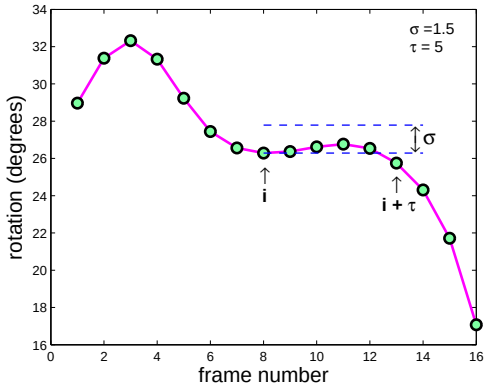


Fig. 6. A close-up of the special case displayed in Fig.5(d).

of the first derivative of f, f', are detected and counted. Detection of zero-crossings is performed by a, conventional rule, which is that if f' exhibits a zero-crossing within $[t_{i-1}, t_i]$ then $f'(t_{i-1}) \cdot f'(t_i) < 0$ should hold for exactly one i. Fig. 5(c) shows a filtered signal in which two peeks occur and recognition fails.

Second, the sign of the peak is specified. When the sign of the peak is positive the following condition must hold:

$$\forall t_i \in [t_A, e), sgn(f) = 1 \wedge \forall t_i \in (e, t_B], sgn(f) = -1, \quad (1)$$

while when it is negative the corresponding condition becomes:

$$\forall t_i \in [t_A, e), sgn(f) = -1 \wedge \forall t_i \in (e, t_B], sgn(f) = 1 \quad (2)$$

In the above, $sgn()$ denotes the sign function. Two options regulate how literate the requirement for the function f being strictly positive or negative. The accepted jitter is specified by threshold σ in rotation axis and τ in time axis, so that the following should hold:

$$\begin{cases} f(t_i) - f(t_i - 1) \leq \sigma & \forall t_i \in [t_A, e) \\ f(t_i) > f(t_j),\ t_j < t_i + \tau,\ j \in [i, i + \tau] & \\ & \\ f(t_i - 1) - f(t_i) \leq \sigma & \forall t_i \in (e, t_B] \\ f(t_i) < f(t_j),\ t_j < t_i + \tau,\ j \in [i, i + \tau] & \end{cases} \quad (3)$$

Fig. 5(d) illustrates such case of a permitted peak, that is treated as jitter. The segment of interest is presented along with the fulfilled requirements σ and τ in Fig. 6.

Third, a "hard" threshold (h_T) has to be overcome, such that $|e| > h_T$. Thresholds may be different across rotation axes, due to anatomical differences in head rotation about each axis. In particular for the pitch axis, the positive and negative thresholds are different as well; $|h_T^+| \neq |h_T^-|$ and $|s_T^+| \neq |s_T^-|$.

Both thresholds h_T, s_T, are empirically adjusted, based on the experimental user studies of Section IV. In total, 12 different thresholds were adjusted, based on the following combinations of h_T, s_T with each rotational component and the sign of the peak as note by the Cartesian product of the corresponding sets: $\{h_T, s_T\} \times \{pitch, yaw, roll\} \times \{positive, negative\}$.

 ISBN: 978-1-61208-356-8

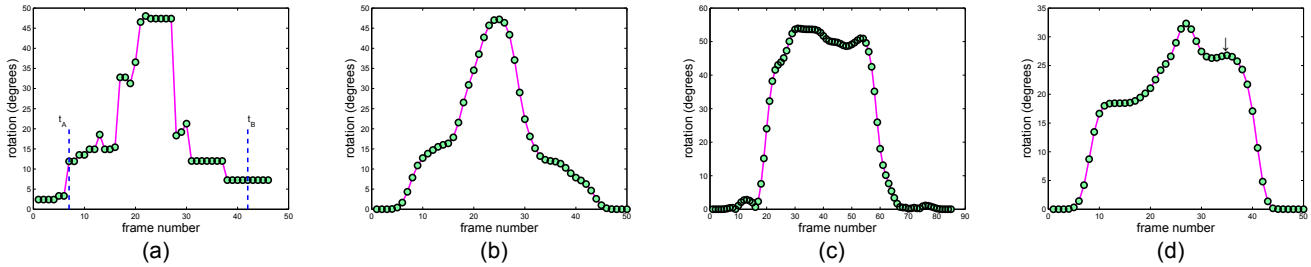


Fig. 5. Characteristic cases of acquired data, showing head motion as a function of the Y rotational component, for expressed "Head Down" gestures. See Section III-C.

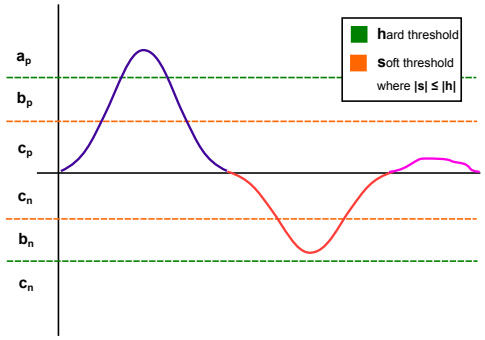


Fig. 7. A demonstration of the thresholds. See Section III-C.

The combination of the the three rotational axes with the sign of the peak results into 6 primitive gestures. Thereby, following the examination of the motion, a gesture is recognized as follows:

- "Head Up" when pitch is negative and dominant, while yaw and roll are neutral
- "Head Down" when pitch is positive and dominant, while yaw and roll are neutral
- "Head Right" when yaw is positive and dominant, while pitch and roll are neutral
- "Head Left" when yaw is negative and dominant, while pitch and roll are neutral
- "Roll Right" when roll is negative and dominant, while pitch and yaw are neutral
- "Roll Left" when roll is positive and dominant, while pitch and yaw are neutral

Fig. 7 illustrates the recognition processing by providing an example of a rotational component, in this case pitch. The signal has been already parsed in three segments, indicated by the corresponding three colors of the curve, by virtue of the process described in Section III-B.

The blue segment is positive and dominant, so we have to examine yaw and roll; if they are neutral then a "Head Down" is recognized. The red segment is negative. Its peek is below h_T but above s_T, which means none of the gestures will be recognized. The magenta segment is positive and neutral on the shown axis; a gesture might be recognized if one of the other rotational components (not shown) is dominant during this time interval.

IV. Experiments

The system was run on a personal computer (PC) with an Intel Core i7, at $2.67\,GHz$ with $6\,GB$ of random access memory and an NVIDIA GTX680 graphics processing unit (GPU). The head tracker was executed on the GPU while gesture recognition on the central processing unit of that PC. The head tracker offered estimates at a rate of $15\,Hz$.

The system was evaluated with the help of 13 test users, all naive to the experimental hypotheses. All test users had normal hearing and did not experience any kinetic problems.

The setup of the experiment included a 480×640 depth camera (an RGBD Kinect sensor) adjusted to a floor mount, and a chair in front of the mount at a distance of $\approx 1\,m$ (see Fig. 8). The sensor was adjusted so that it was at a height comfortable for each individual user. To avoid visual disruption during the experiment, the monitor of the PC was not present.

The evaluation task was enabled by a software module that was developed for the purposes of this evaluation. The system employed a speech synthesizer to prompt the user to perform a gesture and to provide feedback regarding its recognition. During an evaluation session, the system attempted to recognize gestures performed by the user, in individual trials. Each evaluation session, was comprised of 18 trials, testing the recognition of the 6 studied head gestures; 3 trials were dedicated for each gesture type. The execution order of the trials was decided randomly at each session, by the system.

The evaluation task was the following. The system would prompt the user to perform a particular gesture. Upon announcement of this prompt, the system monitored the user. If a gesture was recognized thereafter, the user was informed of the occurrence of the recognition event and the label of recognition. If a gesture was not recognized or if a different than the prompted gesture was recognized the system provided feedback. This feedback pointed out the unexpected outcome and, also, prompted the user to repeat the trial, up to two additional times. A trial was complete upon recognition of the prompted gesture or if three recognition failures occurred. When a trial was complete the system proceeded to the next trial. During the evaluation, the user had the option to pause the process in order to rest and continue later.

 ISBN: 978-1-61208-356-8

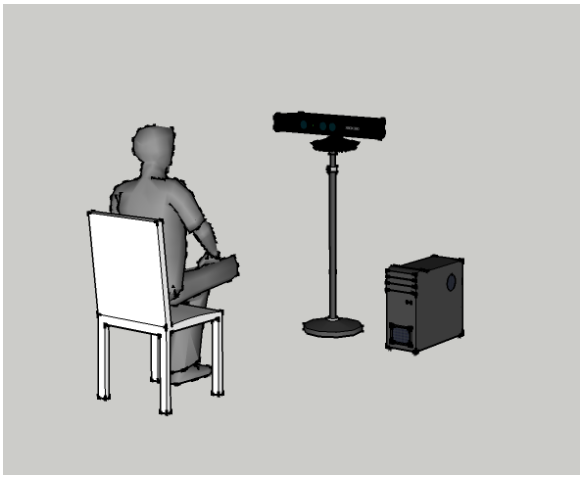


Fig. 8. The experiment setup. A user is sitting in front of a sensor which is adjusted to his height.

Before each session the test user was informed about the required gestures, and was asked to rehearse, in order to validate their comprehension. Also, in the beginning of each session, the system was initialized by acquiring a frontal head pose that was the reference pose for the individual user.

During the evaluation, we kept notes of the events. At the end of each session, the test users were interviewed about their experience, whether they had any difficulties achieving the task, whether they fully comprehended the provided feedback etc.

A. Quantitative analysis

1) Recognition accuracy: The recognition accuracy of the method was measured in terms of percentage of correct recognitions. The results are shown in Tab. I and in Tab. II. The first column of Tab. I shows the percentage of correct detection in the first gesture attempt, while the second column shows the percentage of correct detection after the third attempt. The high percentages of the second column indicate that users adjusted their gestures after the first or second failure to match the user expected recognition requirements of the system.

The first column of Tab. II, shows the sum of misses in all trials and the second column shows the proportion of the recognition errors, or otherwise, how many of the misses were recognized as another gesture. In all cases, recognition errors occurred due to pose estimation failures.

It is important to note that no false positives detections of gestures occurred in any of the experiments and this is due to mainly two reasons: the head tracker is very accurate in the pose estimates that it provides and the time interval for the detection of gesture was constrained by the experiment task (i.e., the time the user had to perform the gesture was guided by the system).

In further analysis of the results shown in Tab. I and II, more conclusions can be drawn for the gestures that received lower recognition accuracy scores. For example, in the case of the ”Head Down” gesture, it was concluded that the lower score is due to the inability of the head tracker to calculate the position

TABLE I
RECOGNITION ACCURACY

Gesture	First time recognized	Any time recognized
Head Up	95%	100%
Head Down	72%	100%
Head Right	92%	100%
Head Left	64%	90%
Roll Left	74%	100%
Roll Right	74%	97%

TABLE II
MISSES DURING THE EXPERIMENTS

Gesture	Sum of misses in all trials	Recognition errors
Head Up	2	0%
Head Down	12	16.67%
Head Right	3	0%
Head Left	22	9.09%
Roll Left	13	0%
Roll Right	13	7.69%

of the head because the face becomes self-occluded and the image avails less facial information. In the case of the ”Roll left” and ”Roll Right” gestures, the lower recognition accuracy percentages are not due to any shortcomings of the tracker, but rather due to the fact that since this gesture is not a commonly performed gesture, its execution range varies from person to person. Finally, an interesting result is related to the accuracy of the “Head Left” gesture in contrast to its relevant “Head Right”. From our investigation, half of the failures occurred because s_T^- of the pitch component was surpassed, meaning that participants unexpectedly inclined their head to the up direction while turning to the left. Such behavior should be investigated in further experiments though.

Some of the proposed works mentioned in Section II provide accuracy evaluations in order to prove the reliability of their systems. Though they are not directly comparable due to differences in gestures and head pose estimation method, we discuss the relationship of the proposed work to the state of the art. In [10], where a training phase is preceded, a ratio of 97% of the gestures are successfully recognized when the test user is the same with the person used for the training, while this ratio falls to 80% when they are different persons. In [14], accuracy depends on the states of the trained HMM, and it varies from 88% to 100%. In [17], “yes” and “no” gestures are recognized with a ratio of 88% and 77% respectively, while most of the other hand-based gestures are recognized at higher ratios. In [13], the recognition rate on the head gestures is over 92%, while in [6], 76.4% of the “nodding” and 80% of the “shaking” gestures are recognized. In [11], a recognition average ratio of 86% is reported. The

 ISBN: 978-1-61208-356-8

context-based approach in [16], increases the recognition ratio of the "nod" gestures which reaches 91%. Finally, in [15], the classification rates range from 75.95% to 98.4% when the training subjects are different from the testing. For a particular gesture, the ratio is 44.84%, though. It is noted that all classification-based methods include mismatches in recognitions, because a decision is made among all classes, but in most cases such a ratio is acceptable. Furthermore, all recognition methods with support of natural interaction have a failure rate. In general, the recognition accuracy depends on the number of the recognized gestures, on their complexity, but are also related to the proposed method. As it was presented in Section IV-A1, the average recognition ratio of our method ranges from 78.5% average, to 97.8% average when users familiarize with the system. We conclude that the proposed work offers results that are not only comparable but, in most cases, outperform reported recognition rates.

In our case, the results indicate that the proposed method can be reliably employed in human-computer dialog applications. As shown, false positive recognitions are rare, but are also undesirable in many cases. In order to overtake such situations, a dialog could expect from the user an extra confirmation. For example if a "Head Down" gesture is utilized as a "yes", then the dialog could expect it twice. Alternatively, the dialog could inform the user about the recognized gesture, permitting a period of cancellation which will be triggered by a gesture or by a simple posture outside the reference position; an invalid gesture. A different option for reducing false positives is to place a restricted time interval for gesture expression, as discussed below.

2) Gesture execution time: Another measurement we acquired, was the execution time of each gesture. Fig. 9 shows the distribution of the recognized or non-recognized gestures at each time-slot. The chart shows that, for the majority of gestures, execution time was below $2\,sec$. As we noticed during the experiments, large execution times were sometimes present due to pose estimation failures and thereby measured execution times were greater than actual (that is, due to a recognition failure the system kept waiting for a gesture to occur but to no avail).

We conclude that as gestures typically occur during a $2\,sec$ limited interval, it is for the benefit of an application that uses such gesture to avail a similar time interval for gesture expression, during a user interface dialog (and, in case of recognition failure, prompt the user to execute the gesture again). In this way, gesture recognition becomes more reliable as potential false positive recognitions are avoided. In addition, in cases of recognition failure, the system becomes more responsive, quickly prompting the user to execute the unrecognized gesture again, instead of letting the user wait for an unnecessary longer timeout.

3) User investigation: In preliminary experiments, the hard and soft thresholds h_T and s_T were initially fixed at the same values for all of the rotational components. However, we noticed that subjects did not perform rotations of the same magnitude on each axis, due to anatomical reasons (i.e., users typically do not lower the head as many degrees as they rotate it horizontally). We also observed that when performing different rotations, amplitude of the neutral (irrelevant) rotations differed. Hence we analyzed the behavior of the subjects in order to see if the thresholds required for the recognition could be adapted for each rotational axis to the benefit of recognition rates.

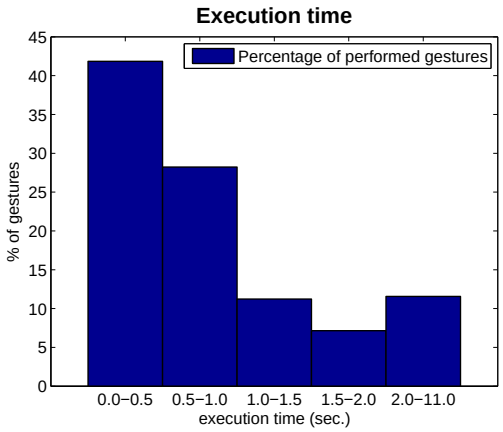


Fig. 9. Distribution of the gestures depending on the execution time. Most of them performed at below $2\,sec$..

The resulting per-axis maximum angle for each single gesture of the evaluation was stored and two types of diagrams were formed (Fig. 10 & Fig. 11). The dominant angle diagram (Fig. 10) displays the distribution of the angle on an axis when a gesture related to this axis was required. Fig. 10 shows the performed angles when a "Head Left" gesture was required; that is the graphs shows the values of the dominant rotational component. The yellow line depicts the h_T. The neutral angle diagram (Fig. 11) shows the distribution of the same angle for the neutral rotational components. Fig. 11 depicts the distribution in the yaw axis, when gestures different than "Head Left" were prompted. The yellow line shows the s_T, which is equal to h_T in this case. Both diagrams show additional information about the first attempt to perform the gesture, which is marked by the green dots, while magenta dots mark the repeats.

Following the preliminary experiments, the thresholds were tuned. The tuning accomplished for both h_T and s_T in the following ranges;

- $[10°, 25°]$ for the h_T^+ and s_T^+ of the pitch component
- $[15°, 25°]$ for the h_T^+ and s_T^+ of the yaw and roll components
- $[-15°, -25°]$ for the h_T^- and s_T^- of the pitch, yaw and roll components

Eventually, all the thresholds were adjusted as Tab III shows. We notice that the h_T^+ and s_T^+ of the pitch component, which are related to the "Head Down" gesture, have a lower absolute value than the others. This can be explained by the anatomy of the neck, which allows a smaller inclination of the head when it is directed down.

For the similar reasons as above, we measured also the ranges of rotational motions for the recognized gestures. As

 ISBN: 978-1-61208-356-8

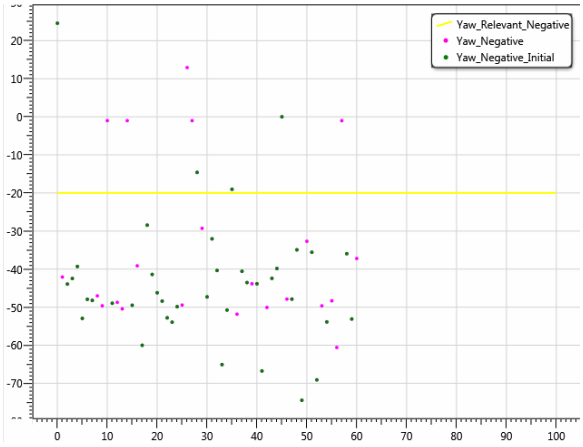


Fig. 10. Distribution of the peaks of the yaw rotational component, when a dominant to negative yaw gesture was expressed.

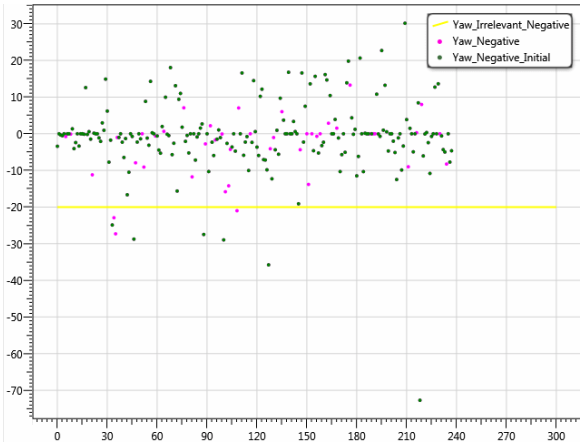


Fig. 11. Distribution of the peaks of the yaw rotational component, when an neutral to yaw gesture was expressed.

mentioned in SectionIII-C, each gesture is related with an axis of rotation. Tab.IV shows the ranges of the per gesture relevant axis, of the truly recognized gestures, and the mean and standard deviation, as well. These observations complement the analysis for tuning the thresholds and besides the general interest as a user study, can be used to tune parameters of the head tracker for optimization of its performance (i.e., by bounding the head pose estimation search space).

B. Qualitative analysis of head gestures

At the end of the experiment, the participants were asked to express any thoughts they had during the execution of the gestures, i.e., if they felt that the tracker was recognizing them easily, if they experienced any fatigue, and so on. Eight of

TABLE III
RECOGNITION THRESHOLDS

	h_T^+	h_T^-	s_T^+	s_T^-
pitch	15°	-20°	15°	-20°
yaw	20°	-20°	20°	-20°
roll	20°	-20°	20°	-20°

TABLE IV
RANGES OF RECOGNIZED GESTURES

	Min	Max	mean	stdev
Head Up	-51.44°	-21.49°	-35.41°	9.55°
Head Down	17.10°	58.49°	30.92°	8.30°
Head Right	26.65°	57.22°	42.82°	8.38°
Head Left	-65.08°	-28.45°	-43.94°	8.24°
Roll Right	-44.72°	-21.56°	-31.21°	6.09°
Roll Left	20.15°	54.01°	35.44°	8.89°

the participants said that the instructions given were clear and that the tracker behaved as expected. One participant said he was uncertain of the 'required' speed the gesture had to be performed in, in order to be recognized by the system. Another participant said that he intentionally performed the gestures in a wider than usual range in order to facilitate the system in recognizing it. The above two comments indicate that some users are just not aware or familiar with the hardware system capabilities. In addition, three participants mentioned neck fatigue especially caused by the "Roll" gestures, another thought that "Roll" is an unnatural motion, suggesting diagonal ones instead. Finally, two participants named issues with the "Head Down" gesture, but as of our observation during the evaluation, these caused by estimation errors due to the self-occlusion of largely bent head relative to the camera.

V. PILOT APPLICATION

The evaluation discussed in Section IV targeted the interaction with dialogs. In a dialog application the system prompts the user to provide input in the form of gestures. Other applications though, let the user interact with the system in a continuous manner; they handle events which are emitted by the available input devices (i.e., mouse clicks or keyboard strokes). Considering this, every gesture recognition system can be regarded as an input device. A primary difference of these systems with an everyday input device, is that the user is supposed to concentrate in the interaction, with limited habitual or natural movements, in favor of preventing false recognitions. In a spectrum of applications this user cooperation can be assumed, as gesture interaction is an essential communication modality for people with mobility difficulties.

For our demo the publicly available labyrinth/puzzle game called Bloxorz [20] was adopted. The recognized gestures were associated with keyboard events, which were then operated the subject of the game, which is a box. The box has two degrees of freedom, controlled with "Head{Up,Down}" and "Roll{Left,Right}" gestures, forming a natural mapping. Moreover the game's puzzle nature doesn't expect successive fast movements, qualifying the head gestures modality as suitable for the interaction. In Fig. 12 a user is shown using the system.

Following the employment of the application, it is concluded that the utilized gestured recognition system provided the ability to fully control it. However, it is noted that further

 ISBN: 978-1-61208-356-8

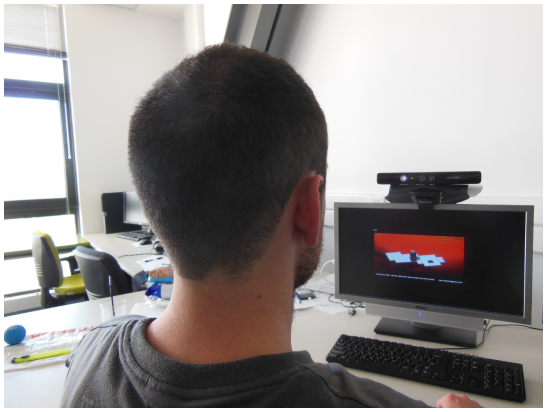


Fig. 12. The game application is shown in the left screen. Right screen shows the output window from the pose estimation.

work is required in order to use head gestures as the sole method of user interaction with an application.

VI. Conclusion

A method for simple gesture detection and recognition that is based on 3D head tracking was presented along with its evaluation. The proposed work explores the potential of recognizing robustly primitive head motions as a means for natural human computer interaction.

In this context, the proposed method was evaluated indicating that recognition provides sufficiently reliable recognition rates for employment in human-computer dialogs. The proposed method has been, also, utilized beyond the context of such a dialog. We concluded that the detection and parsing of gestures from continuous head motion of the proposed method, is a property that sets the foundations for the generic use of these gestures in human computer interaction. In that respect, investigation of usability issues is the topic of future work.

The evaluation of the proposed method indicated that advances in head tracking accuracy are the most important topic of future work, as recognition failures are mainly due to shortcomings of the underlying head pose estimation technology. Based on this finding, we conclude that the proposed technology is suitable to be applied at the spatial range of operation of the corresponding head trackers. In turn, this range is determined by the accuracy of the utilized depth sensor, which is in the order of $.5\,m$ to $1.5\,m$. As a consequence, in the context of an intelligent environment one could envisage utilization of head gesture at special locations, such as when the user is situated at location related to a particular activity.

In the evaluation, a study of user behavior in terms of gesture execution time and steepness of head rotation was performed. We have observed that downward head rotations are, usually, performed in smaller amount of rotation, with reasons that can be probably traced to the head rotation ergonomy and anatomy. Given a constant, with respect to axis of rotation, accuracy of head pose estimation this indicates the increased vulnerability of downward gestures, which can be of interest in the design of pertinent applications. Alternatively, the purposeful placement of the imaging sensor may be considered, so as to better image corresponding head motions.

Acknowledgment

This work has been supported by the FORTH-ICS RTD Programme “Ambient Intelligence and Smart Environments”.

References

[1] P. Padeleris, X. Zabulis, and A. A. Argyros, “Head pose estimation on depth data based on particle swarm optimization,” in CVPR Workshops, 2012, pp. 42–49.

[2] Q. Cai, D. Gallup, C. Zhang, and Z. Zhang, “3d deformable face tracking with a commodity depth camera,” in 11th European Conference on Computer Vision: Part III. Springer-Verlag, 2010, pp. 229–242.

[3] L. Morency, A. Rahimi, N. Checka, and T. Darrell, “Fast stereo-based head tracking for interactive environments,” in Fifth IEEE International Conference on Automatic Face and Gesture Recognition, May 2002, pp. 390–395.

[4] A. Riener and A. Sippl, “Head-pose-based attention recognition on large public displays,” Computer Graphics and Applications, IEEE, vol. 34, no. 1, 2014, pp. 32–41.

[5] Z. Zhang, “Microsoft kinect sensor and its effect,” MultiMedia, IEEE, vol. 19, no. 2, Feb 2012, pp. 4–10.

[6] Z. Yu, Z. Yu, H. Aoyama, M. Ozeki, and Y. Nakamura, “Capture, recognition, and visualization of human semantic interactions in meetings,” in IEEE International Conference on Pervasive Computing and Communications. IEEE, 2010, pp. 107–115.

[7] J. W. Davis and S. Vaks, “A perceptual user interface for recognizing head gesture acknowledgements,” in Workshop on perceptive user interfaces. ACM, 2001, pp. 1–7.

[8] S. Chu and J. Tanaka, “Head nod and shake gesture interface for a self-portrait camera,” in International Conference on Advances in Computer-Human Interactions, 2012, pp. 112–117.

[9] D. Kelly, J. Reilly Delannoy, J. McDonald, and C. Markham, “Automatic recognition of head movement gestures in sign language sentences,” in China Ireland International Conference on Information and Communication Technology. Dept. of Computer Science, National University of Ireland, Maynooth, 2009, pp. 142–145.

[10] H. Wu, T. Shioyama, and H. Kobayashi, “Spotting recognition of head gestures from color image series,” vol. 1, 1998, pp. 83–85.

[11] H. Wei, P. Scanlon, Y. Li, D. S. Monaghan, and N. E. O’Connor, “Real-time head nod and shake detection for continuous human affect recognition,” in International Workshop on Image Analysis for Multimedia Interactive Services. IEEE, 2013, pp. 1–4.

[12] R. Li, C. Taskiran, and M. Danielsen, “Head pose tracking and gesture detection using block motion vectors on mobile devices,” in International conference on mobile technology, applications, and systems. ACM, 2007, pp. 572–575.

[13] A. Agrawal, R. Raj, and S. Porwal, “Vision-based multimodal human-computer interaction using hand and head gestures,” in IEEE Conference on Information & Communication Technologies. IEEE, 2013, pp. 1288–1292.

[14] J. Gast et al., “Did I get it right: Head gestures analysis for human-machine interactions,” in Human-Computer Interaction. Novel Interaction Methods and Techniques. Springer, 2009, pp. 170–177.

[15] N. Wohler et al., “A calibration-free head gesture recognition system with online capability,” in International Conference on Pattern Recognition. IEEE, 2010, pp. 3814–3817.

[16] L.-P. Morency and T. Darrell, “Head gesture recognition in intelligent interfaces: the role of context in improving recognition,” in International conference on Intelligent user interfaces. ACM, 2006, pp. 32–38.

[17] K. Biswas and S. K. Basu, “Gesture recognition using Microsoft Kinect®,” in Automation, Robotics and Applications (ICARA), 2011 5th International Conference on. IEEE, 2011, pp. 100–103.

[18] Y. Georgalis, D. Grammenos, and C. Stephanidis, “Middleware for ambient intelligence environments: Reviewing requirements and communication technologies.” in HCI (6), ser. Lecture Notes in Computer Science, C. Stephanidis, Ed., vol. 5615. Springer, 2009, pp. 168–177.

[19] Z. Zhang, “Flexible camera calibration by viewing a plane from unknown orientations,” in IEEE International Conference on Computer Vision, vol. 1. IEEE, 1999, pp. 666–673.

[20] “Bloxorz puzzle game,” URL: http://www.miniclip.com/games/bloxorz [retrieved: June, 2014].

 ISBN: 978-1-61208-356-8

Toward a Multi-Domain Platform for Live Sensor Data Visualisation and Collaborative Analysis

Marco Forin, Paolo Sacco, Alessio Pierluigi Placitelli
Vitrociset S.p.A.
Via Tiburtina 1020, 00156 Rome, Italy
Emails: {m.forin, p.sacco, a.placitelli.cons}@vitrociset.it

***Abstract*—In this work, we present a web platform that seeks to tackle the challenges that come from the real-time visualisation of georeferenced sensor data in a multi-user, multi-touch environment. We introduce an input device agnostic user interface and the concept of realistic input reaction. We discuss the implemented components and the presentation system as a whole. The demonstrated platform also provides a set of building blocks to personalise the visualisation, easing its reuse in different monitoring scenarios. Finally, we show how the platform can be used to assist collaborative data visualisation and analysis.**

Keywords–environmental monitoring; real-time GIS data analysis; collaborative analysis; tabletops; natural user interfaces; multi touch.

I. INTRODUCTION

In the recent years, sensor networks have been used to solve a variety of problems, ranging from environmental monitoring [1] to fine grained structural health monitoring [2]. Such deployed sensors might be used to ensure public safety and provide a steady flow of information to higher level decision makers, support systems and crisis first responders [3][4]. The amount of data collected by geographically distributed sensors, independently of their function, can be of a considerable volume and present a challenge to interactive visualisation. In this work, we present a platform for sensor data presentation and collaborative analysis. The platform aims to integrate and visualise in a clear and understandable way live data feeds coming from deployed sensors, geographical information systems and the result of higher level reasoning coming from data fusion engines or complex event processing. Moreover, it is designed to allow its use on Multi-Touch Tables (MTT), thus facilitating the collaboration and analysis through the means of natural user interfaces. Practical use cases of the system comprise real-time pollutant agents detection and warning system, nuclear waste monitoring and tracking or situation awareness and emergency monitoring in control rooms.

This paper is organized as follows. In Section II, we describe the previously proposed methods for web-based sensor data visualisation and discuss the limitations of these methods. In Section III, we describe the design elements and key principles we based our platform on. In Section IV, we describe our platform and explain how it tries to address the limitation of the approaches in the literature. Finally, in Section V, we present our conclusions and the direction of our future work.

II. RELATED WORK

In recent years, many largely different approaches have been proposed to interactively visualise considerable amounts of data coming from geographically distributed sensors. In this section, we introduce some of them, specifically focusing on web-based solutions. The SenseWeb project [5] demonstrates a web-based data gathering and visualisation infrastructure relying on Microsoft SensorMap for the visualisation, although not taking advantage of open geospatial standards. The National Oceanic and Atmospheric Administration's (NOAA) nowCOAST [6] aggregates heterogeneous informations, such as meteorological, oceanographic and hydrological data into a single, web-based visualisation platform, only partly based on open source technologies. Previous literature on multi-user, multi-touch interactions mainly focused on researching novel interaction techniques to mediate the issues involved in the collaborative interaction on a shared surface, such as content orientation, occlusion and reach [7]. The reacTable project demonstrated a multi-user, collaborative, electro-acoustic musical instrument on a multi-touch table [8]. In DTLens [9], a set of consistent interaction was investigated to allow multi-user exploration of geographical data on tabletops. In [10], a multi-touch system which allows multiple users to interact on a touch sensitive surface. Even though many sensor data visualisation systems have been demonstrated and implemented by the research community, they do not support multi-touch interactions out of the box thus not allowing collaborative multi-user touch interactions. The ones that support this kind of interaction lack real-time sensor live data integration with Geographic Information Systems (GIS). Moreover, an additional shortcoming of the aforementioned approaches is that their systems are tied to a specific domain or hardware platform and do not provide enough flexibility to be reused in different scenarios.

III. DESIGN PRINCIPLES

In this section, we describe the design principles we followed in the development of the platform.

A. Easier deployment, scalability and easier maintainability

Given the power of modern consumer hardware and the increasing efficiency of web browsing software, with the rise of emergent technologies like HTML5 and WebGL, a number of obstacles to the development of truly interactive web applications have been removed [11]. As a consequence, web applications can be considered as a feasible alternative to native applications for interactive software. Our platform (see Figure 1) is entirely based on state-of-the-art open technologies, exploiting the potential of the latest HTML5 draft [12] and the latest Javascript and WebGL [13] specifications. The strict adherence to open web standards and technologies allows to have a platform-independent software system which is:

 ISBN: 978-1-61208-356-8

1) *simple to deploy*: the platform has to be deployed on a single machine and is automatically accessible to all the devices with a network connection;
2) *simple to maintain*: updates have to be delivered to a single machine;
3) *simple to scale*: a variety of open source, enterprise grade, components are already available for this purpose.

B. Multi-user, touch environment

The platform needs to exploit the potential benefits given by the use of MTTs. Each element of the platform has to react to touch inputs. The platform also has to abstract the user away from the challenges involved in the collaborative use of MTT including, but not limited to, content orientation, gestural interaction and group interaction [7]. Moreover, the platform has to provide support for legacy input devices like the mouse and be easily extensible to support new input paradigms (i.e., touch-less interaction).

C. Extensible widgets

The platform User Interface is made up by reusable modules called *widgets*. Each widget has to be replaceable. The developer has to be able to write new widget, either extending available ones or starting from scratch.

D. Standard communication protocols

Modules within the platform have to communicate using standard communication protocols and specifications. Since the platform works within a web browser, protocols like WebSocket [14], Server-Sent Events [15] and HTTP [16] are used. Geospatial data is delivered through the main protocols defined by the Open Geospatial Consortium (OGC): Web Feature Service (WFS), Web Coverage Service (WCS) and Web Map Service (WMS) protocols.

E. Open source stack

The platform has to integrate the most commonly used, widely tested, open source, third party libraries and encapsulate them into self-contained components, whenever this is possible. In our platform the jQuery [17] library, a small and fast Javascript library, is used to simplify web document manipulation. The platform also makes extensive use of the doT.js library [18] to provide template based presentation of live sensor data. The platform web pages are served through an instance of the Apache HTTP Server [19]. Furthermore, geospatial data is served using GeoServer [20], which implements the OGC standards. The OpenLayers [21] library is used to visualise data layers on bi-dimensional cartography while the CesiumJS [22] is used for data visualisation in a three-dimensional representation of the region of interest.

IV. Platform Overview

Our platform allows geographically distributed sensor data visualisation. Its main strengths, which have been the focus of our research activities, are its flexibility and the multi-user, multi-touch capabilities. User interactions are characterized by the ability to use different input paradigms and devices, such as touch-based and mouse-based commands. Depending on the device used to access the platform, it can either be used completely through touch or mouse inputs, or both of them. Touch gestures are designed to maximize user action throughput when using the system: commands are triggered with the detection of a different number of touches or based on the kinematic parameters of the touch points. The platform user interface is designed as a desktop environment built within a web application, benefiting of the cross-device availability given by the latest web technologies. Furthermore, structuring the web application as desktop environment, helps reducing the learning time as the operator should be already familiar with native desktop environments which are commonly available on commodity personal computers.

A. Multi-touch, collaborative analysis

MTTs enable interaction with the hands and the fingers, providing each user in a multi-user scenario with the ability to manipulate virtual objects as if they were physical. Moreover, two-handed, multi-fingered input is more natural and flexible than mouse and stylus input devices [23]. The multiple points of contact in MTTs enable novel interaction flows, enhancing multiple user parallel reasoning and collaboration on the same interface. The multi-user interaction on a MTT can be used to exploit collaborative analysis and visualisation in different scenarios [24], as well as improve the decision making process in military [25] or clinical [26] settings. Due to the increasing use of mobile touch devices such as phones and tablets, interacting with touch surfaces has become a common practice, not dependent on the age of the user. As a result of this, the domain experts which are more resistant on using new software or technologies have a more positive attitude toward MTTs which aids the quicker learning of platform functionalities compared to the use of legacy devices. Furthermore, since a MTT allows a display to provide a common informative context as a shared workspace, parallel and collaborative analysis can be easily exploited. For example, if a user is examining the live data coming from sensors deployed on an extended geographical area detects an event of interest, he can send the relative data to another user on the MTT for further analysis. In another scenario, different users could concurrently analyse different sensor feeds coming from different geographical areas to resolve a common problem. The parallelism and the quick data/information exchange in face-to-face settings around a MTT may foster the collaboration among these individuals and consequently give an advantage during decision making processes and the analysis of emergency situations.

B. Widgets

A widget is a graphical user interface component which is part of the presentation platform, consisting of a title bar, a content area and input-reactive corners. To foster collaboration [27] among multiple users around a multi touch table, each widget in our platform can be repositioned and oriented freely, since a widget presented right-side up to one user might be upside-down for another. Widget repositioning is achieved by dragging the title bar with a single finger, the size can be varied by dragging its corners toward or away from each other while its orientation is changed by performing a clockwise or counter-clockwise rotation while holding down the two

 ISBN: 978-1-61208-356-8

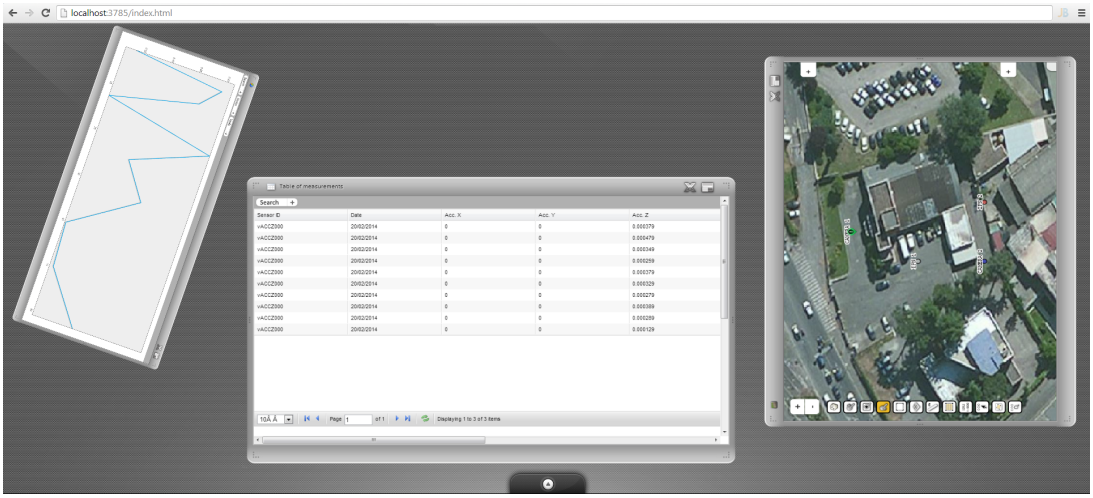


Figure 1: An overview of the sensor data presentation platform

fingers on the aforementioned corners, thus allowing users to organise the personal and group workspace on the table. In our platform, we introduced the concept realistic input reaction: the motion of a widget produced by user input is modelled by taking in consideration the laws of the motion of points and bodies or kinematics. Widgets behave as if they were physical objects reacting to an applied force. This feature makes touch manipulation of the widgets more natural and allows to transfer the motion of the fingers to the fingers, thus enabling users to drag widgets toward other users without moving around the table, simply by dragging it with the appropriate speed and then lifting the finger. We will briefly review the main widgets of the proposed platform.

1) Geospatial Data Widget: One of the components available within our platform is the *Geospatial Data Widget* (see figure 2). This component is completely built using Javascript, HTML5 and WebGL without relying on third party native software or closed source libraries. It does not require the installation of any browser plug-in as it is completely based on open web standards. The visualization component is integrated within platform and provides the seamless blend of geospatial data (aerial photographs, terrain elevation data) with live georeferenced sensor information about the monitored environment coming from the deployed sensors. Moreover, the geospatial data widget is also able to display both a bi-dimensional and a three-dimensional view of the monitored environment. When a bi-dimensional view is activated, aerial photographs of the region of interest are requested using the WMS protocol. If a three-dimensional visualization is requested, the terrain surface is built by exploiting aerial photographs and terrain elevation data for the region of interest. The visualisation is further augmented with additional data layers (buildings and 3D models) and real-time data collected from the sensors deployed on the field (GPS positions, measurements, video feeds, etc.). The style of each data feed can be personalized at deployment time or runtime, thus allowing to show different icons or models for different types of data coming from the sensors. Besides, each single data feed can be independently shown, hid or displayed with a particular opacity by interacting with the relative entry in the list of

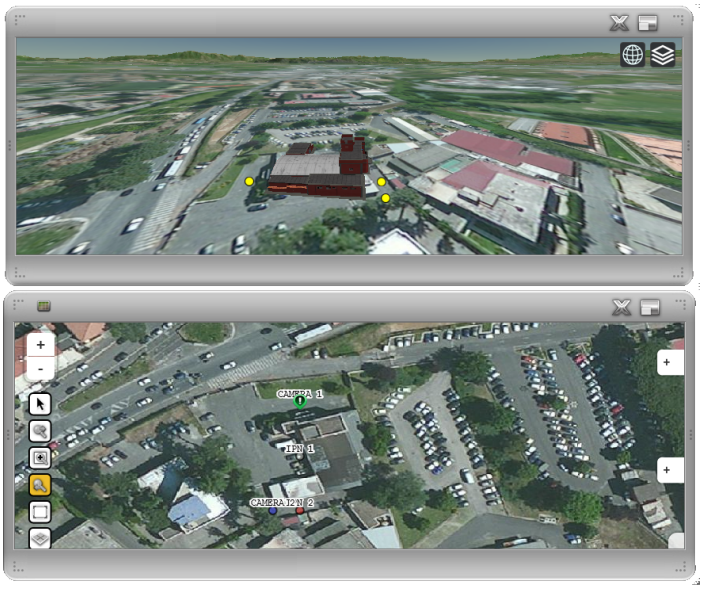


Figure 2: The geospatial data widget

available sensors. A personalised HTML page can be shown when a sensor is selected in this widget. To ease the integration with different information systems and fusion engines, real-time information feeds can be streamed to the visualization component using different open formats and protocols: JSON, XML, GeoJSON and KML over WebSockets.

2) Common Alerting Protocol Widget: the Common Alerting Protocol [28] (CAP) is an emergency alert format which allows a consistent warning message to be disseminated over heterogeneous warning systems. Our platform supports CAP alerts visualisation and analysis through the CAP widget.

3) Data Table Widget: the data table widget displays the data coming from the deployed sensors in a tabular format. It can be either connected to a real-time sensor data feed or to a database storage system. Customised queries can be used to gather specific informations.

 ISBN: 978-1-61208-356-8

4) Graph Widget: this widget produces a graphical representation of the historical trend that a particular variable, coming from the deployed sensors or a database connection, assumes over time.

5) Video Streaming Widget: the video widget allows to display a video stream within the platform. The stream can come from a deployed sensor (i.e., a camera) or from a remote server.

6) Organiser Widget: when dealing with more than one widget on a single display, visual clutter might become an issue. This control enables users to reorder the widgets within the screen area by maximising the visibility of each widget's content area while spreading them around the empty areas of the display. Classical widgets reordering functions, such as widget tiling and cascading, are available as well. The organiser also serves a tool to easily locate desired opened widgets.

V. Conclusion and future work

In this work, a platform for the presentation and collaborative analysis of real-time data coming from geographically distributed sensor has been presented. Such system, which has been entirely written as a web software without any third party browser plug-in dependency, can be viewed in any browser supporting the latest HTML5, Javascript and WebGL specifications. Its presentation layer hides the heterogeneous nature of the real-time data coming from remote sensors, thus displaying data to the end user in a consistent and homogeneous way. Moreover, the presentation layer is optimised for displaying on different devices: MTTs, mobile tablets and personal computers. In particular, the platform is tailored to be used on MTTs enabling easier, collaborative data analysis. To further enhance the multi-touch collaborative experience, we introduced the realistic input reaction for widgets, to adhere to the user mental model of physical object movements. Future work will be focused on extending the collaborative interaction metaphors on multi-touch displays and adding support for touch-less interactions. Furthermore, given the increasing availability of augmented reality devices, additional research efforts will explore the use of the platform on such devices. Moreover, we will investigate how to describe the interface in abstract fashion thus enabling its auto adjustment depending on the used input device. Besides, an experimental campaign is scheduled to assess the potential advantage of using the platform, in a multi user environment, on a multi-touch table in a command and control room scenario.

References

[1] L. Yu, N. Wang, and X. Meng, "Real-time forest fire detection with wireless sensor networks," in Wireless Communications, Networking and Mobile Computing, 2005. Proceedings. 2005 International Conference on, vol. 2. IEEE, 2005, pp. 1214–1217.

[2] S. Kim et al., "Health monitoring of civil infrastructures using wireless sensor networks," in Information Processing in Sensor Networks, 2007. IPSN 2007. 6th International Symposium on, April 2007, pp. 254–263.

[3] S. M. George et al., "Distressnet: a wireless ad hoc and sensor network architecture for situation management in disaster response," Communications Magazine, IEEE, vol. 48, no. 3, 2010, pp. 128–136.

[4] K. Lorincz et al., "Sensor networks for emergency response: challenges and opportunities," Pervasive Computing, IEEE, vol. 3, no. 4, 2004, pp. 16–23.

[5] S. Michel et al., "Environmental monitoring 2.0," in Data Engineering, 2009. ICDE'09. IEEE 25th International Conference on. IEEE, 2009, pp. 1507–1510.

[6] M. Allard, "Noaa's nowcoast: A gis-web mapping portal to discover and display real-time coastal observations, satellite imagery and noaa forecasts," in 15th Symposium on Education, 2006.

[7] C. Shen et al., "Informing the design of direct-touch tabletops," IEEE Comput. Graph. Appl., vol. 26, no. 5, Sep. 2006, pp. 36–46. [Online]. Available: http://dx.doi.org/10.1109/MCG.2006.109

[8] M. Kaltenbranner, S. Jorda, G. Geiger, and M. Alonso, "The reactable*: A collaborative musical instrument," in Enabling Technologies: Infrastructure for Collaborative Enterprises, 2006. WETICE'06. 15th IEEE International Workshops on. IEEE, 2006, pp. 406–411.

[9] C. Forlines and C. Shen, "Dtlens: multi-user tabletop spatial data exploration," in Proceedings of the 18th annual ACM symposium on User interface software and technology. ACM, 2005, pp. 119–122.

[10] P. Dietz and D. Leigh, "Diamondtouch: a multi-user touch technology," in Proceedings of the 14th annual ACM symposium on User interface software and technology. ACM, 2001, pp. 219–226.

[11] A. Taivalsaari and T. Mikkonen, "The web as an application platform: The saga continues," in Software Engineering and Advanced Applications (SEAA), 2011 37th EUROMICRO Conference on, Aug 2011, pp. 170–174.

[12] Html5 draft specification. [Online]. Available: http://www.w3.org/TR/html5/ [retrieved: Jun., 2014]

[13] Webgl 1.0 specification. [Online]. Available: http://www.khronos.org/registry/webgl/specs/latest/1.0/ [retrieved: Jun., 2014]

[14] Websocket specification (rfc6455). [Online]. Available: https://tools.ietf.org/html/rfc6455 [retrieved: Jun., 2014]

[15] Eventsource specification. [Online]. Available: http://dev.w3.org/html5/eventsource/ [retrieved: Jun., 2014]

[16] Http specification (rfc2616). [Online]. Available: http://www.w3.org/Protocols/rfc2616/rfc2616.html [retrieved: Jun., 2014]

[17] jquery library. [Online]. Available: http://jquery.com/ [retrieved: Jun., 2014]

[18] dot.js template library. [Online]. Available: http://olado.github.io/doT/ [retrieved: Jun., 2014]

[19] Apache httpd. [Online]. Available: http://httpd.apache.org/ [retrieved: Jun., 2014]

[20] Geoserver. [Online]. Available: http://geoserver.org/ [retrieved: Jun., 2014]

[21] Openlayers library. [Online]. Available: http://openlayers.org/ [retrieved: Jun., 2014]

[22] Cesiumjs library. [Online]. Available: http://cesiumjs.org [retrieved: Jun., 2014]

[23] R. Harper, Being human: Human-computer interaction in the year 2020. Microsoft Research, 2008.

[24] S. S. Krupenia and C. Aguero, "Asset distribution with a multitouch table," in Proceedings of the 9th International ISCRAM Conference. 2012, 2012.

[25] N. Wahab and H. B. Zaman, "The significance of multi-touch table in collaborative setting: How relevant this technology in military decision making," Applied Mechanics and Materials, vol. 278, 2013, pp. 1830–1833.

[26] M. Avila-Garcia, A. E. Trefethen, M. Brady, and F. Gleeson, "Using interactive and multi-touch technology to support decision making in multidisciplinary team meetings," in Computer-Based Medical Systems (CBMS), 2010 IEEE 23rd International Symposium on. IEEE, 2010, pp. 98–103.

[27] R. Kruger, S. Carpendale, S. D. Scott, and S. Greenberg, "Roles of orientation in tabletop collaboration: Comprehension, coordination and communication," Computer Supported Cooperative Work (CSCW), vol. 13, no. 5-6, 2004, pp. 501–537.

[28] OASIS, "Common alerting protocol version 1.2," 2010. [Online]. Available: http://docs.oasis-open.org/emergency/cap/v1.2/CAP-v1.2-os.html

 ISBN: 978-1-61208-356-8