EuroVR Conference 2015
15th and 16th October, 2015
Campus Politecnico Milano – CNR
Lecco (Italy)
INDEX

Euro VR President preface I
Euro VR 2015 Sponsors II
Keynote presentation: Richard Tisdall III
Keynote presentation: Andrea Gaggioli IV
Organizing Committee V

Proceedings

Cortechs: A Device-less Smartphone 1
Physical Impacts of Gestural Interactions Towards a New Assessment Method 8
Towards ubiquitous tracking: Presenting a scalable marker-less tracking approach using multiple depth cameras 14
3D Tune In: 3D sound visuals and gamification to facilitate the use of hearing aids 21
A prototype of an automatically assisted and person focused rehabilitation system for post stroke patients at home 23
Augmented Notepad: an augmented reality application to support elderly people in daily activities 29
GOJI an advanced virtual environment for supporting training of physical and cognitive activities for preventing the occurrence of dementia in normally living elderly with minor cognitive disorders 33
Immersive virtual reality in rehabilitation: a preliminary efficacy assessment on children with acquired brain injury 38
ZebraX: sonification of road crossing guidance data for visually impaired individuals 44
A Comparative Study on High and Low Cost IMU Drift for an Augmented Reality Application 46
Distributed Architecture of Secure Semantic AR Services 51
Framework for Creating Interactive Browser based Virtual Worlds 57
Framework for Transparent Execution of CUDA and OpenCL 63
The Effects of Multisensory Cues on the Sense of Presence and Task Performance in a Virtual Reality Environment 69
Towards New Interconnected VR and 8K Technology 75

A Layer based 3D Virtual Environment for Architectural Collaboration 79

A Multi-view and Multi-interaction System for Digital-mock up’s collaborative environment 85

A Tool for Collaborative Decision Making on Service Information Linked to 3D Geometry of Complex Hierarchical Products 94

Qualitative assessment of a beaming environment for collaborative professional activities 100

Virtual Reality for Collaborative System Engineering supporting ESA experimental vehicles 108

Ergonomic Validation of Manual Processes through Posture Detection Using Optical Sensors 113

The presentation of procedures in virtual training of assembly tasks: segmented vs whole task approaches 118

Towards a Low Cost High Immersive Virtual Environments for Training Heavy Machinery Operators 127

Virtual reality environment for disassembly sequences generation and evaluation 133

AR in support to helicopter maintenance activities 139

AR/VR enabled Architecture for Technology Enhanced Learning in Manufacturing Environments 144

Augmented Reality authoring by blue collar workers as key enabler for AR adoption by manufacturing companies 150

Exploiting ARgh! On job learning tool in Mixed reality training environment 154

A New Approach for Haptic relevant Simulation of Soft Bodies 159

A Virtual Environment for mission planning and supervision of Remotely Piloted Aircraft Systems 164

Augmented Reality using Point Clouds for Survey and Design in Outdoor Public Construction 170
Dear EuroVR 2015 Conference participants,

First of all let me thank you all for your contribution and active participation in EuroVR 2015. I would also like to thank the organisers, our EuroVR partners ITIA-CNR for hosting EuroVR 2015 in the very beautiful city of Lecco. Lecco, following EuroVR2014 in Bremen, was a step forward in a series of conferences leading back to the INTUITION Network of Excellence, starting in 2004. EuroVR is aiming to bring the European VR/MR/AR community together by bridging the gap amongst the research, academic and industrial worlds. Nowadays, with VR/MR/AR becoming more and more a commodity due to investments by large companies, it is increasingly important to get acquainted with the latest developments in the field and the EuroVR Conference is a perfect place to doing so.

EuroVR 2015 had a very interested program featuring two keynote speeches. The first one was from Richard Tisdall about what the future holds and where further developments are likely to take us in the VR/AR world. The second one was from Andrea Gaggioli which talked about the potential offered by the combination of VR and neurotechnologies for altering and transforming human experience. Further to our keynote speakers a very interesting program featuring 27 papers and several poster presentations provided opportunities for thinking and networking. Most importantly a vivid exhibition with 12 exhibitors gave the opportunity for hands-on experience with new products and systems.

Last but not least I want personally to thank our sponsors: Space4Agri, our gold sponsor, Fideas and Riprendo@Home projects, our silver sponsors which provided significant support for the organisation of the Conference.

Closing I would invite you all to continue support EuroVR and its activities and join the Association be becoming an individual or corporate member!!! Join us and let’s work together around Virtual/Mixed/Augmented Reality

Sincerely yours

Angelos Amditis
Sponsors & support
Keynote presentation: Richard Tisdall

HAPPY FINISH Director

Happy Finish is a global creative production studio of digital artists and interactive specialists of which Richard has been at the forefront of since it’s set up just over 10 years ago. Working directly with brands, agencies & photographers, the company crafts creative visual content across a multitude of medias including Print, Digital, Interactive and now Virtual and Augmented Reality. This new technology is now at the heart of what we do, with a team of artists that focus on designing and building engaging user interfaces to enrich your VR/AR experience.

Richard brings his many years of creative industry knowledge and unrivalled passion to discuss and demonstrate first-hand how brands can leverage these new platforms to engage with consumers on a whole new level.

Richard’s speech will be: AR & VR – The New Narrative

In this talk by Richard Tisdall, Business Development Director for Happy Finish, he will be exploring not only what Virtual and Augmented Reality is currently up to, but what the future holds and where further developments are likely to take us.

It’s a great talk around VR & AR and the future of advertising with these technologies in mind, and the opportunities they bring to brands. The talk will examine the advertising opportunities that VR & AR bring can create, and will showcase work Happy Finish has created for brands including Ted Baker, River Island, Lufthansa, Lamborghini, Mercedes and Honeywell. As well as this Happy Finish have just completed groundbreaking and world first projects in the fields of Virtual and Augmented Reality and Richard will be showcasing these projects and more whilst discussing how we go about taking some truly amazing briefs to Market.
Keynote presentation: Andrea Gaggioli

Ph.D. Psych., Psychology dept., Università Cattolica di Milano

Prof. Andrea Gaggioli, Ph.D. received a M.Sc. degree in Psychology from the University of Bologna and a Ph.D. degree in Psychobiology from the Faculty of Medicine of the Public University of Milan. He is currently Associate Professor of General Psychology at the Catholic University of the Sacred Hearth in Milano, Italy; Senior Researcher of the Interactive Communication and Ergonomics of NEW Technologies (ICE-NET Lab) at the same University; and Deputy Head Researcher at the Applied Technology for Neuro-Psychology Lab (ATN-P Lab), I.R.C.C.S. Istituto Auxologico Italiano, Milan, Italy. For over fifteen years, Professor Gaggioli has investigated the potential role of virtual reality and interactive technologies in promoting mental and physical wellbeing. His involvement with those research areas has led to the co-authoring of over hundred articles in refereed journals. For his scientific work, Professor Gaggioli received several international acknowledgements, including the 2005 Annual Prize of the European Academy of Rehabilitation Medicine. In addition, he has recently completed coordinating the European Commission funded INTERSTRESS project (Interreality in the Management and Treatment of Stress-Related Disorders).

Web site: www.positivetechnology.it

Andrea’s speech will be: Transformative Technology: Using Virtual Reality and Neurotechnologies for Transforming Human Experience

The emerging symbiotic relation between human beings and machines has the potential to generate new forms of perception, interaction and cognition. Examples of such interfaces include virtual/augmented/mixed reality, wearable displays, smart apparel and ambient intelligence devices. A further class of symbiotic technologies, such as neuroprosthetic devices and neuro-biofeedback, will enable a direct connection between the computer and the brain.

The goal of this talk is to describe the potential offered by the combination of VR and neurotechnologies for altering and transforming human experience. The transformative potential of VR range from the simulation of “plausible” possible worlds and possible selves to the simulation of realities that break the laws of nature and even of logic. Researchers are already looking at ways in which VR can be used to hack our ordinary perception of self and reality in order to observe what happens to specific brain or psychological processes when a person is exposed to alterations of the bodily self using multisensory conflicts. For the present discussion, I will focus on three kinds of transformative potentials that are unique to VR: (i) manipulating bodily self-consciousness; (ii) embodying another person’s subjective experience; and (iii) altering the laws of logic and nature. By virtue of this manipulation researchers hope to cast light on the neurobiological process underlying self-consciousness.
Commitee members:

Angelos Amditis, ICCS Institute of Communications and Computer Systems (Greece)

Dimitris Mavrikios, Laboratory for Manufacturing Systems & Automation – LMS (Greece)

Francesco Ferrise, Polimi (Italy)

Gabriel Zachmann, University of Bremen (Germany)

Gianluigi Reni, IRCCS E. Medea, Bioengineering Lab Bosisio Parini (Italy)

James Ritchie, Heriot-Watt University (UK)

Jerome Perret, Haption SA (France)

Kaj Helin, VTT Technical Research Centre (Finland)

Krzysztof Walczak, Poznan University of Economics (Poland)

Marco Sacco, ITIA-CNR (Italy)

Marco Schumann, Fraunhofer (Germany)

Mehdi Ammi, LIMSI (France)

Monica Bordegoni, Polimi (Italy)

Patrick Bourdot, LIMSI (France)

Philippe Gravez, CEA (France)

Robert Iacob

Roland Blach, Fraunhofer (Germany)

Stefano Mottura, ITIA-CNR (Italy)

Uwe Woessner, High Performance Computing Center Stuttgart – HLRS (Germany)

Valter Basso, Alenia Space (Italy)
PROCEEDINGS
Cortechs: A Device-less Smartphone

Melanie Buset
School of Information Technology, Human-Computer Interaction
Carleton University
melanie.buset@carleton.ca

Elias Fares
School of Information Technology, Human-Computer Interaction
Carleton University
elias.fares@carleton.ca

Jed Looker
School of Industrial Design
Carleton University
jed.looker@carleton.ca

Paden Shorey
School of Information Technology, Human-Computer Interaction
Carleton University
paden.shorey@carleton.ca

ABSTRACT
In this paper we discuss the concept user interface (UI) that was created to solve the everyday problem of carrying a smartphone device. The concept UI aims to relieve the user from a handheld device and focuses on inputs using a Brain Controlled Interface (BCI) and hand gestures. This was approached with the mindset of futuristic technologies that are capable of bringing this system to life. An experiment was conducted using the Leap Motion device and Unity to create an implementation of the concept UI. The experiment consisted of users completing a set of tasks pertaining to only one aspect of the system, the camera. The methods and results will be discussed and used to determine if the UI was feasible and easy to interact with.

Author Keywords
User Interface; Hand Gesture; Smartphone; Brain Controlled Interface.

ACM Classification Keywords
Design; Documentation; Experimentation; Human Factors.

INTRODUCTION
Today’s smartphone users are constantly anxious about losing their phone or damaging its screen. A cellphone is no longer a simple device used to call one’s mother to tell her where you are. A cellphone is now a lifestyle containment system that keeps track of every aspect of one’s life. Losing, forgetting, or damaging this device can ruin a person’s day, month, or even year if all their information is lost. In addition, the current traditional cell phone use restricts users in many ways in terms of how they interact with their devices. We constantly see people walking and looking down at their phones and missing out on life and what is in front of them in the real world.

So how do we solve the problem of the constant potential of losing or damaging one’s phone and the restrictions the current interactions have on us? We propose to get rid of the phone all together. This means removing the hardware aspect of having a smartphone and incorporating it in a person’s everyday life. We envision being able to see the real world around you while still interacting with a smartphone-like device. Cortechs was created to achieve this goal. This futuristic concept UI is created to handle the main functionalities of a smartphone while not interrupting a person’s actions with the reality around them.

SYSTEM DESCRIPTION
The main features of the Cortechs system would be very similar to the core features of a smartphone. These features would be the ability to send and receive audio calls and video calls; have a contacts list and the ability to add, edit and remove a contact; send and receive text messages and take and view photos using a camera. It is not possible to implement all the mentioned functionalities with the technologies of today. Thus, it was decided that the camera capabilities would be the most feasible to implement and only using hand gestures as an input.

IMPLEMENTED FEATURES
Since the concept UI of Cortechs relies heavily on future technologies that do not yet exist, it was impossible to build a fully working prototype. Any interaction requiring a BCI unit was unlikely to be developed. Since it was clear that the entire system could not be built, it was decided to use a Leap Motion device [1] and the Unity [2] game engine software. With these technologies, the camera application could be built to match our concept UI as close as possible. This allowed us to create part of the system and also permitted us to make use of hand gesture interaction.

The Cortechs system was originally intended to let the user think of an action or use hand gestures to activate the action. Only the hand activation segment was implemented. Unity was used to create a virtual environment where the Cortechs application can be used and Leap Motion was used to interpret the user’s hand movements. The features that were possible to implement in terms of the camera are as follows: the ability to resize the picture frame; the ability to zoom in and out; capturing a photo, switching between camera mode and video recording mode; show and hide the album of photos; and preview the last photo that was taken.

These features and their corresponding hand gestures can be seen in Figure 1. The system implementation was built to be used through a computer monitor therefore there was
also the ability to walk (move) and look around using the
computer keyboard and mouse. However this interaction
was not part of the Cortechs system but was needed for
implementation using today’s technologies.

![Cortechs Hand Gesture Tutorial](image1.png)

**Figure 1:** Cortechs hand gesture tutorial for the camera application

**METHODS**

**Materials**
The setup consisted of a MacBook Pro laptop to navigate
through the virtual environment that was developed for the
purposes of this project as seen in Figure 2. Cortechs was
developed with the Unity software and the Leap Motion
device was used to interact with our system. Our
researchers developed the virtual environment.

![Unity Software Used to Create Virtual Environment](image2.png)

**Figure 2:** Unity software used to create virtual environment for Cortechs application.

**Participants**

Participants consisted of 15 college and university students
from Algonquin College and Carleton University in Ottawa,
Ontario ($N = 15$; age range = 18-25). Participation was done
on a voluntary basis. None of the participants reported
having previous experience using a Leap Motion device or
any other motion-sensing device that tracked hand gestures.
All participants had unlimited time to learn the hand
gestures until they felt comfortable and were then asked to
complete each task in the same order.

**Procedure**
The experiment consisted of users familiarizing themselves
with the hand gestures, completing 5 different tasks and
reporting their experiences with the tasks using the NASA
Task Load Index (TLX) (See appendix A). The five tasks
were as follows: (1) Take a photo of the shrine (2) Take a
zoomed in photo of the crown (3) Switch the camera mode
to video (4) Open the image preview then close it (5) Open
the photo album then close it. Participants were also asked
to fill out a user questionnaire that was created by our
researchers (see appendix A) to report the ease of use and
feasibility of the Cortechs system and its user interface.

Prior to performing the 5 tasks, participants were given
unlimited time to review and understand the Cortechs hand
gesture tutorial as shown in Figure 1 and time to play freely
with the Leap Motion device. Afterwards, each user was
instructed by a researcher to perform a task. Participants
were timed on how long it took to complete each of the 5
tasks. Additionally, notable comments were taken such as
the user’s reaction and thoughts. Tasks 1 & 2 pertained to
the use of the camera whereas tasks 3, 4, & 5 explored the
overall user interface. After completing tasks 1 & 2 users
filled out the (TLX) and again after tasks 3, 4, & 5. We
asked participants to fill this out twice because of the
difference in the nature of the two tasks. Figure 3 shows
how the experiments were conducted and how the
participants interacted with the Leap Motion and Cortechs
UI.
ANALYSIS

We were not able to appropriately evaluate the entire system due to lack of technologies that exist today. This means that we were only able to evaluate the hand gestures that we came up with and none of the BCI portion of the system. To evaluate the system, we used a questionnaire that the participants had to fill out after completing all tasks along with two NASA TLX tests that users filled out during the testing process.

The user questionnaire survey and the NASA TLX were used in order for us to determine two things: if the interface was easy to use and if the interface was pleasant to use. The user-friendliness and user experience were the top elements we wanted to test. We also wanted to learn if the gestures were intuitive and natural to the user. We ignored the learnability of the system UI and ignored any issues that came up from using the Leap Motion device.

The Survey Questions can be seen in Appendix A and we felt this was an appropriate method of evaluation because this allowed the users to express their thoughts and any comments. Participants were able to describe their experience using the UI hand gestures and their feeling towards the system. Surveys and questionnaires are things that many people are familiar with and they allow the ease of sharing and discussing the responses. We also thought it was appropriate to time how long it took users to complete each task giving us an idea of what the user struggled with and what they understood right away.

RESULTS

We asked participants to report their interpretation of using the device on two separate occasions because of the two different uses we were exploring. Tasks 1 and 2 tested the demands of taking photos whereas tasks 3, 4, & 5 tested the ease of controlling the overall user interface.

Although the Leap Motion device was not part of the Cortechs concept UI, it was essential to implement the hand gesture aspect of the system. It was originally hypothesized that this device will decay our final results due to its inconsistencies in performance. Nonetheless, participants seemed to be able to use it without any difficulty. For our results, we ignored the walking around and looking around since this was also not part of the Cortechs concept UI.

Gesture Control Evaluation

Based on the provided hand gesture tutorial as shown in Figure 1, participants were asked if the hand motions were suitable to the paired action. Participants found taking a photo by closing the right hand and opening their left hand to be the most suitable hand-gesture-to-action pair. The least suitable was found to be resizing the frame using the left hand. The suitability of other actions is shown in Figure 3 with the right side of the scale (4) representing the highest level of suitability, and the left side of the scale (0) representing the lowest level of suitability.

When users were asked which tasks were easiest to complete, it was found to be switching between camera mode and video mode. The most difficult task to complete was looking at the image preview, meaning viewing the last image that was taken as shown in Figure 4.

The hand gesture to do this is to close the left hand and use the right hand to drag or swipe the image from the right
hand side to the center of the screen view. The ease of use for other actions is shown in Figure 5.

Figure 6: Results from asking users to "Rank the tasks from easiest (1) to hardest (5)"

BCI Evaluation

Finally, the participants were asked if the technology that existed would allow the participants to control the interface using their thoughts, would they still use hand gestures. The majority of the responses were between yes or maybe. This means that humans would rather depend on their hands and motion rather than on their direct thoughts to manipulate an interface. One assumption would be because we are very accustomed to using our hands and seeing actions take place as a result of hand gestures and finger swiping has become a norm. So accustomed that we would rather put up with the delay of moving our hands to see an action than to use our thoughts to instantly perform the action.

Task Evaluation

The total time that it took for each participant to complete each tasks was noted down. The total time of all five tasks combined were also tallied up. The longest time it took for a participant to finish all five tasks was 106 seconds. The fastest time for a participant to finish all five tasks was 36 seconds. Table 1 shows the average time of the participants’ time to complete each task.

Between the two different aspects we were testing, overall, participants reported the control of the user interface ($M = 25$) to be easier than the actual functionality of taking photos ($M = 32.6$). However, both statistics are positive as the total error mean is 120 and the two reported means from the users are very low.

![Table 1: Average time of all the participants' timings of each task](image)

**DISCUSSION**

Although some participants did comment on the sluggishness of the Leap Motion device’s performance, it did not affect their user experience and overall feeling towards the prototype. Perhaps people were forgiving of the actual system delay because of its novel aspect. The *Novelty Effect* [4] suggests that when something is new, people report enjoying their experience with it just because it is something new and not necessarily because of its real functionality. People tend to neglect and ignore any issues with the system entirely. This effect could have been seen in our study because all of the users reported never having used any type of motion tracking device like the Leap Motion. Furthermore, *The Hawthorne Effect* [5] could have been exercised due to researchers being present as participants interacted with the system. *The Hawthorne Effect* suggests that participants report only positive feedback when researchers are present because they feel that they don’t want to insult or disappoint the researcher(s) present. When testing for our system was done, the researchers informed the participants that they created the Cortechs system so that may have affected the user’s responses.

In order to somewhat control the *Hawthorne Effect*, researchers took observational notes when users were interacting with the system. Some observations that were noted were: the system had trouble picking up hand positioning if the users hand positions were not firm and clearly distinguishable; it seemed that users could only concentrate on one hand position at a time; users had difficulty determining the “sweet spot” of the Leap Motion and where to keep their hands at all times so that their input isn’t lost; the speed of interacting with the Leap Motion has to be consistent and somewhat slow; there should be more feedback given to the users to indicate that their input isn’t being detected; the zoom function was difficult and didn’t run very smooth; similar hand gestures need to have a delay after each other so the Leap Motion can determine that there is a new input; and lastly, the system crashed a lot and
researchers were not able to determine why. All of the mentioned observations are very difficult to avoid at this stage in development as they all heavily rely on the hardware and not on the Concept UI. Researchers were able to successfully build the software for such system but without a strong enough technology to handle the system. It will be difficult to improve our system without advanced motion capturing devices that are capable of such inputs.

Tasks 1 & 2, which pertain to the actual photo capturing, were reported as being more difficult overall in comparison to tasks 3, 4 & 5 which focus more on the functionality of the user interface. This finding concludes that the program itself makes sense to users and must be somewhat intuitive. The fact that tasks 1 & 2 were reported as more difficult in comparison to tasks 3, 4, & 5 could mean that this is because the technology to make a seamless experience in this context has not yet been developed. We hypothesize that in the future; these difficulty reports will also be lower with the advancement of motion sensing hardware such as the Leap Motion. Participants reported that they would still prefer to use their hands when interacting with a cell phone instead of just relying solely on a BCI. This could be due to the opposite of The Novelty Effect whereby people are comfortable doing the same thing and don’t feel comfortable exploring unknown territory. BCI’s are a fairly new technology and the abundance of unknown that comes with them could be why users reported some hesitation.

Future research could focus on the redesign of some of the gestures that were given negative feedback. Apart from fixing issues with the current prototype, means of navigating through the environment using gestures would be a step towards the initial concept and the reality of using this futuristic device in real life (without a mouse and keyboard). Finally, when technology allows for it, proper BCI could be implemented to make certain tasks are easier and gives users multiple ways of handling actions within the application. It could also be interesting to ask users more demographic questions to see if there is any evidence for sex differences in how people interact with Cortechs or just motion detecting hardware in general. Our study lacked consistency in asking users if they played video games as only some participants were asked to report video game frequency. This information would have been useful to see if video game playing is correlated to how well someone performed the tasks asked in our study.

ACKNOWLEDGMENTS
We would like to thank all the volunteers, who willingly accepted to use and test our Leap Motion implementation of Cortechs. We appreciate the feedback they have provided and the helpful comments they have given on the UI interactions. Special thanks to Prof. Anthony Whitehead for consistently replying to questions and concerns we had and keeping his line of communication open.

REFERENCES
[1] Leap Motion Controller. https://www.leapmotion.com
Appendix A
Figure 8.6

**NASA Task Load Index**

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

<table>
<thead>
<tr>
<th>Name</th>
<th>Task</th>
<th>Date</th>
</tr>
</thead>
</table>

**Mental Demand**
How mentally demanding was the task?

| Very Low | Very High |

**Physical Demand**
How physically demanding was the task?

| Very Low | Very High |

**Temporal Demand**
How hurried or rushed was the pace of the task?

| Very Low | Very High |

**Performance**
How successful were you in accomplishing what you were asked to do?

| Perfect | Failure |

**Effort**
How hard did you have to work to accomplish your level of performance?

| Very Low | Very High |

**Frustration**
How insecure, discouraged, irritated, stressed, and annoyed were you?

| Very Low | Very High |

**User Questionnaire**

1. How do you feel about the interaction with the Leap Motion device? Provide comments on what you thought of the Leap Motion device.
2. Rank each hand gesture based on its suitability to the task. (list gestures, likert scale from very suitable to not at all suitable)
3. Can you rank the tasks from easiest (5) to hardest (1)? What made the hardest task so hard to complete?
4. How would you describe your experience with the interface?
5. If you could complete the tasks using your thoughts, would you still use the hand gestures to interact with the interface?
Physical Impacts of Gestural Interactions: Towards a New Assessment Method

Sobhi AHMED, Laure LEROY, Ari BOUANICHE
Paragraphe - Université Paris 8, France

Abstract
Despite the advantages of gestural interactions, they involve several drawbacks. One major drawback is their negative physical impacts. To reduce them, it is important to go through a process of assessing risk factors to determine the interactions' level of acceptability and comfort so as to make them more ergonomic and less tiring. We propose a method for assessing the risk factors of gestures based on the methods of posture assessment in the workplace and the instructions given by various standards. The goal is to improve interaction in virtual environments and make it less stressful and more effortless.

Categories and Subject Descriptors (according to ACM CCS): H.1.2 [Models and Principles]: User/Machine Systems - Human factors. H.5.2 [Information interfaces and presentation]: User Interfaces - Evaluation/methodology; Ergonomics; Interaction styles (e.g., commands, menus, forms, direct manipulation)

1. Introduction

One of the purposes of gestural interactions is to facilitate interaction with virtual environments. They aim at being intuitive, easy to use and learn, since lots of them are based on the emulation of natural gestures [RAU99]. Some can fulfill specific needs (such as those of physically disabled people [JPF13], etc.). These interactions are supposed to entail less cognitive and physical effort than traditional interactions: for example, the use of a mouse, which demands a physical effort because of its distance from the user, calls for the user’s arm to be outstretched while requiring a very accurate gesture when pointing [LC99].

However, musculoskeletal disorders associated with gestural interactions can be caused by movements requiring substantial physical effort. What is more, the extended and/or frequent use of such systems can result in an overuse of the muscles in charge of performing said gestures [SCC11].

There exists stressful, tiring, illogical gestures and some might be impossible to perform for certain people. For instance, the interaction with some gestured-controlled TV sets is considered stressful [FW95] because of the high position of the hand during use. Interaction with touchscreens also affects user comfort negatively because of the need to keep one’s arm outstretched [LC99]. The use of big screens is sometimes considered stressful to the neck because of frequent movements of the head and eyes [BCW*06].

Few studies have been conducted on how to reduce the physical impact of gestural interactions on the human body, and this has sometimes resulted in the creation of non-ergonomic, stressful gestures that are difficult to use [AY11]. Interaction with such systems can lead to various musculoskeletal injuries.

To the goal of analyzing and assessing the health risks associated with gestures, we have studied task assessment methods in the workplace. Just like gestural interactions, those tasks consist of movements repeated frequently.

The physical impact of gestures is affected, for example, by the angle of the joint used in the gesture, the gesture’s duration, its repetition, etc. The evaluation of such factors allows the assessment of gesture quality and consequently of their physical impact. This allows the design and implementation of ergonomic gestures that will cause neither pain nor stress, and that will be easier to use. We aim to implement a gesture assessment method based on certain criteria and factors stated in current studies.

In a first part, we present the medical problems related to gestures used in videogames and the workplace. The second part studies the existing assessment methods of physical movements. In a last part, we propose a synthesis and an analysis of said methods as well as our own approach to gestural interactions.

2. Negative gestures impacts in other contexts

As mentioned previously (cf. Introduction), gestural interfaces are used more and more frequently in numerous domains. The use of such interfaces implies the performance of certain types of movements, sometimes repeatedly and/or for a long time, necessitating some effort. The overuse of the muscles in charge of these gestures can cause musculoskeletal disorders (MSDs).

“The term MSD groups some fifteen diseases acknowledged as work-related pathologies. These pathologies represent more than 70% of known work-related pathologies." [ACA11]

MSDs affect the muscles, tendons and nerves of upper and lower limbs, at the level of wrists, shoulders, elbows or
knees. A lot of MSDs have resulted from the frequent use of gestural interactions, such as those included with the Wii® gaming console [JH09].

2.1. Painful gestures

Painful gestures are often caused by being subjected to an external or internal force and by exceeding the standard angle range at which joints are normally used. Those out-of-range angle values can be occasioned by numerous movements such as extension, flexion, abduction, adduction, pronation, etc. The movement range determines whether the joint is overly used and if the gestures resulting from the movement is potentially painful. Besides, static and dynamic constraints on some parts of the human body impact movement range and interdependence. [NSMG03] [CHA97], for example, adduction (moving a body part towards the median axis of the body), abduction (moving a body part outwards from the median axis), and pronation and supination which designate limb rotations.

2.2. Injuries related to videogames based on gestural interactions

The repeated use of videogames can cause musculoskeletal injuries: for example, the use of the Wii® gaming console has occasioned sore muscles and knee, shoulder and heel injuries (DOMS: Delayed Onset Muscle Soreness) [SCC11].

Videogame-related injuries can be classified in four categories:
1. Tendinopathy: tendon injuries.
2. Bursitis: swelling and irritation of one or several bursa.
3. Enthesitis: inflammation of the sites where tendons and ligaments are inserted into the bone.
4. Epicondylitis (tennis elbow): painful inflammation of the tendon on the outside of the elbow.

The main cause for such injuries and inflammations is the repeated stress undergone by involved muscles. According to the National Electronic Injury Surveillance System (NEISS), a high percentage of MSDs (67%) involve the use of Wii® in playing virtual sports [JH09].

2.3. Work-related injuries

The movements used during gesture interactions are extremely similar to those performed in the completion of some work-related tasks at the level of repetitions, extended time span, involved muscles, postures and the force exerted [SCC11] [MP11]. These movements could occasion injuries called “Repetitive Strain Injuries” (RSIs). Several diseases have been associated with RSIs such as tendinitis, bursitis, tenosynovitis, carpal tunnel syndrome, etc. [SSC96]. Symptoms such as pain, discomfort and a sensation of localized fatigue in an overused joint can all point to RSIs.

The risk factors associated with the onset of RSIs and their level of severity depend on time span, frequency and intensity, and have been classified in six categories: awkward postures, force, effort and musculoskeletal load, static muscular work, exposure to certain physical stressors, repetition and the unvarying nature of the work, and organizational factors.

Effort depends on the joints involved, movement direction, posture and individual characteristics [ACA11].

In gestural interactions, most gestures are deemed natural (natural user interface) [RAU99], and require certain spatial movements, which in turn demand some effort as well as an internal or external force which can over-exert muscles and tendons affected by these activities [SCC11]. What is more, these movements are repetitive, and occur over a long time span [AY11]. It is therefore possible to speculate that videogame- and work-related injuries are similar to those resulting from gestural interactions. It is rather clear that movements with extended arms, device vibrations and activities involving one’s arm are very similar.

According to Nielsen [NSMG03] the basic principles of gesture ergonomics are: avoiding external positions, avoiding repetition, muscle rest, favoring neutral, relaxed positions, avoiding static positions as well as avoiding internal and external forces on joints and the interruption of the natural flow of bodily fluids.

3. Related studies

It is crucial to find gesture assessment methods to devise gestures which do not lead to fatigue and health hazards.

3.1. Gesture assessment

The reduction of the negative physical impact of gestures requires an assessment procedure. This procedure would allow determining the level of comfort and the stress they cause by measuring risk factors related to said movements. Assessment methods are classified in two categories:

3.1.1. Subjective methods. Most studies on the assessment of the negative impact of gestures and physical movements in general resort to subjective methods [NSMG03] [MP11]. Amongst those, one can find:

a. the Body Discomfort Diagram method (BDD), which assesses the level of discomfort in different parts of the body using a diagram of the body and an assessment scale. The diagram allows identifying and assessing the places and sources of discomfort by marking the affected areas [CAM96].

b. scoring methods, where a number of points is assigned to each single movement and criterion, resulting in a final score which determines the gesture’s level of comfort. Each single score is decided either by the users [NSMG03] or by experts (ergonomists, etc.) [MC93].

c. other methods are used, such as questionnaires [HIMW06], interviews, open-ended questions [MP11].

3.1.2. Objective methods and angle measurements. There exist methods and standards which allow the assessment of physical movements in a more objective way:

a. Electromyogram: The electromyogram is a tool which measures muscle activity through the detection and recording of electric signals sent by muscle motor cells used during activity. The electric signal is amplified and
processed to determine the level of muscle force exerted. [LCHF70] [FRE04]. This technique is used by [MP11] to measure muscle activity pertaining to the gestures and effort when interacting with touch-enabled devices.

b. RULA (Rapid Upper Limb Assessment): RULA is a risk-factor assessment technique for upper limbs, geared towards individuals subjected to postures, forces and muscle loads potentially leading to MSDs [MC93]. The assessed factors are: number of movements, static work, force, work posture and working time.

RULA allows the attribution of a final assessment score for each posture ranging from 1 to 7. This score indicates the level of discomfort for the posture. The higher the score, the higher the risk. It follows diagrams specifying the ranges of joint angles for various body parts. In these diagrams, a score is given to each movement depending on its angle. The farther the angle from a neutral position, the higher the score. This numbering system is also used to specify the level of force exerted as well as static and repetitive muscular activity. To calculate the scores, three score charts —defined by ergonomists—are used [MC93].

The use of RULA is manual and the assessment is only possible for one side of the body at once (left or right).

c. The ISO 11226 standard: The ISO 11226 standard [ISO00] aims at assessing health hazards for workers involved in manual labor. The assessment process involves specifying and classifying posture conditions for each body part as acceptable or not. These conditions comprise joint angle, time-related aspects and movement repetition. The classification is based on experimental studies as well as the current knowledge in ergonomics.

The assessment procedure is a one- or two-step process. The first step measures joint angles. If said angles do not exceed a given limit, the posture is deemed ‘acceptable’. If not, the second step focuses on the time span for which the posture is sustained. Extreme angles are never recommended. There exist several methods to recognize postures, such as observation, video, etc. Other factors are considered while assessing static postures, such as support (or its absence), sitting or standing position, etc.

d. The AFNOR NF EN 1005-4 standard (Safety of machinery - Human physical performance): NF EN 1005-4 is an AFNOR standard [CEN98] aiming to improve machine design in order to decrease health risks by avoiding postures and stressful movements leading to MSDs. This is done through the specification of various recommendations as well as a posture- and movement-related risks assessment method.

It defines a posture and movement assessment procedure related to working with machinery. The assessment can either be ‘acceptable’, ‘acceptable under conditions’ or ‘unacceptable’. The assessed risk factors are: movement angle, gesture time, frequency, etc. In situations determined as ‘acceptable under conditions’, other risk factors must be considered as duration, repetition, period of recovery, the presence of a support to the body, etc.

3.2. Creating non-stressful gestures

Gesture creation by the user results in gestural interfaces taking user preferences and needs into account. Approaches to creating gestural interfaces are based on the concept of interface adaptability [BCTV05]. One way is to use predefined (standard) gestures, where standard gestures are conceived from natural human gestures. A set of gesture vocabulary is derived by observing, collecting and assessing natural gestures performed by operators during scenarios [NSMG03] [RLL11] [WMW09]. The assessment is used to select the final gestures that will be used. Only few studies take physical factors into account during gesture assessment. Another way is to let the user define the gestures he wants to use in a preliminary step before starting to use the system [JPF13]. However, the physical impact of the resulting gestures is not assessed.

4. Analysis and comparison of approaches, proposition

We aim to design an assessment method for gestures used during interaction that would minimize their negative physical impacts. A complete gesture consists of a set of single gestures whose assessment results in an overall assessment of the gesture. The assessment of these gestures is done through the assessment of certain conditions and variables of the postures and physical movements effected. These conditions are: joint angles, posture duration, frequency, muscle load and external force.

Variables will be assessed based on specifications for acceptable and unacceptable movements in various studies and standards ([CEN98] [ISO00] [AGC00] [MC93]). These specifications assess movement variables, thereby evaluating the quality of the gesture.

The data related to each joint is organized in tables specifying all possible movement types for said joint and giving acceptable or unacceptable values for the various criteria and variables of movement. The angle of movement is a key factor in the assessment process, since it indicates the level of joint stress and, consequently, the potential discomfort to which that stress could lead.

![Figure 1: Shoulder movements [AGC00], modified.](image)

The various levels of acceptability and comfort for shoulder movements (Figure 1) are shown in Table 1. In this table, the acceptability of postures and gestures is mainly determined with joint angles. What is more, gesture duration, movement frequency and other factors potentially affecting the level of comfort are assessed, such as supports for the body, an even distribution of weight on legs and feet, etc. Joint ranges are classified in ‘acceptable’, ‘acceptable under conditions’ or ‘unacceptable’ categories. The acceptability of movements is always connected to tasks with enough variation at the mental and physical levels [ISO00]. Similar tables for each joint have been compiled and are not printed here for want of space.

The measurement of time is crucial in the assessment of the acceptability of work postures: the longer the gesture and the higher number of repetitions, the more stressful the movement is. The different approaches use various strategies to measure time. Some measure movement
The gesture. The process which aims to determine the level of acceptability of collected and organized so as to be used in the assessment muscle load, etc.) defined in various approaches are duration and other risk factors (such as repetition, force, tissue damage, etc.) are used to assess pre-existing gestural interfaces and find out which gestures are best. What is more, it could be used in the design phase of gestural interactions to support the body, movement and posture style (weight distribution on both feet, rotation, etc.) This application allows detecting the conditions and variables of users’ freeform empty-handed gestures, assess them, and determine their level of acceptability automatically according to various pre-existing methods and standards. The variables are mainly joint angles, duration, frequency, supports for the body, movement and posture style (weight distribution on both feet, rotation, etc.) This application could be used in the design phase of gestural interactions to decide which gestures are best. What is more, it could be used to assess pre-existing gestural interfaces and find out whether they are stressful.

5. Software structure

5.1. Approach

Our goal was to develop a computer application that would allow detecting the conditions and variables of users’ freeform empty-handed gestures, assess them, and determine their level of acceptability automatically according to various pre-existing methods and standards. The variables are mainly joint angles, duration, frequency, supports for the body, movement and posture style (weight distribution on both feet, rotation, etc.) This application could be used in the design phase of gestural interactions to decide which gestures are best. What is more, it could be used to assess pre-existing gestural interfaces and find out whether they are stressful.

5.2. Input

The application’s inputs are: The physical movements detected by a Kinect device (which will probably be replaced by a more accurate device in the future). From the

<table>
<thead>
<tr>
<th>Movement</th>
<th>Source</th>
<th>Acceptable limit (1)</th>
<th>Acceptable under conditions – Not recommended (2)</th>
<th>Unacceptable limit (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antepulsion (Flexion- front)</td>
<td>AFNOR 1005-4</td>
<td>0° - 20°</td>
<td>- 20° - 60° if static: (supported arm) or (short duration + recovery time). - 20° - 60° if frequent. - 20° - 60° if: - frequency &lt; 10 per min - short duration - &gt; 60° if short duration and not frequent</td>
<td>- &gt; 60° if static - &gt; 60° if frequent</td>
</tr>
<tr>
<td>Retropulsion (Extension- back)</td>
<td>INRS</td>
<td>0° - 20°</td>
<td>&gt; 60°</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tab Reg G</td>
<td>20° (1 pt)</td>
<td>20° - 45° (2 pts)</td>
<td>45° - 90° (3 pts)</td>
</tr>
<tr>
<td></td>
<td>RULA</td>
<td>0° - 20°(1 pt)</td>
<td>&gt; 20° (2 pts)</td>
<td></td>
</tr>
<tr>
<td>Adduction (Abduction)</td>
<td>AFNOR 1005-4</td>
<td>0°</td>
<td>- &gt; 0° if: - not frequent - short duration - &gt; 0° if static - &gt; 0° if frequent</td>
<td></td>
</tr>
<tr>
<td></td>
<td>INRS</td>
<td>0°</td>
<td>&gt; 0°</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RULA</td>
<td>0° - 20°</td>
<td>&gt; 20° (2 pts)</td>
<td></td>
</tr>
<tr>
<td>Elevated shoulder</td>
<td>AFNOR 1005-4</td>
<td>stressful</td>
<td>stressul if not frequent</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ISO 11226</td>
<td>20°</td>
<td>stressul if frequent</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RULA</td>
<td>stressul (1 pt)</td>
<td>stressul (1 pt)</td>
<td></td>
</tr>
<tr>
<td>Hyperadduction of the arm (with shoulder antepulsion)</td>
<td>ISO 11226</td>
<td>stressul</td>
<td>stressul</td>
<td></td>
</tr>
<tr>
<td>Extreme external rotation</td>
<td>Arm support</td>
<td>stressul (1 pt)</td>
<td>stressul (1 pt)</td>
<td></td>
</tr>
<tr>
<td>Trunk leaning forward</td>
<td>RULA</td>
<td>- 1 pt</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Recommendations for shoulder joint angles ([CEN98] [ISO00] [AGC00] [MC93])
ISO application: RULA, ISO 11226, AFNOR 1005-4, INRS. acceptable values for the methods that are integrated in the for the time being. This in addition to the tables of possible presence of a rotation which are manually entered second is the presence of supports for the body and the capture, we deduce joint angles and gesture duration. The of movements.

- Trunk: for a leaning movement ranging between 20°- 60°, time acceptability is calculated with this equation: 
  \[ y = -\frac{3}{40} x + \frac{11}{2} \]

- Head: For a supported bowing movement ranging between 25°-85°, time acceptability \( y \) is: 
  \[ y = -\frac{2}{40} x + \frac{131}{12} \]

- Shoulder: For a supported abduction (elevation) ranging between 20°-60°, time acceptability \( y \) is: 
  \[ y = -\frac{1}{20} x + 4 \]

- Time is incorporated in the static load (calculate load versus time separately).
  - Scores A or B is increased by 1 point if the posture is static
  - Scores A or B is increased by 1 point if repeated (more than 4 times per minute).

<table>
<thead>
<tr>
<th>Source</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISO11226</td>
<td>- Trunk: for a leaning movement ranging between 20°-60°, time acceptability is calculated with this equation: ( y = -\frac{3}{40} x + \frac{11}{2} )</td>
</tr>
<tr>
<td>INRS</td>
<td>- Head: For a supported bowing movement ranging between 25°-85°, time acceptability ( y ) is: ( y = -\frac{2}{40} x + \frac{131}{12} )</td>
</tr>
<tr>
<td>AFNOR</td>
<td>- Shoulder: For a supported abduction (elevation) ranging between 20°-60°, time acceptability ( y ) is: ( y = -\frac{1}{20} x + 4 )</td>
</tr>
</tbody>
</table>

Table 2: Recommendations for the duration and frequency of movements.

The application’s outputs are:
- A binary assessment (acceptable or not) of body postures. It depends on the data in the different approaches used (RULA, INRS, ISO, AFNOR). The result obtained through these methods is only ‘acceptable’ if all the standards and methods are deemed ‘acceptable’. We have adopted such an approach to ensure a maximum level of safety.
- Acceptability results for all body parts will also be displayed to easily locate the stressful areas.

5.3. Outputs

The application’s outputs are:
- A binary assessment (acceptable or not) of body postures. It depends on the data in the different approaches used (RULA, INRS, ISO, AFNOR). The result obtained through these methods is only ‘acceptable’ if all the standards and methods are deemed ‘acceptable’. We have adopted such an approach to ensure a maximum level of safety.
- Acceptability results for all body parts will also be displayed to easily locate the stressful areas.

5.4. Architecture

The application’s design emphasizes clarity, modularity and revisability. It was indeed essential, when envisioning a basis which could be adapted to various uses and custom applications, that designers of gestural interactions could modify the software as they see fit without causing the whole program architecture to fall apart.

To that effect, each joint has been placed in a separate class, thus allowing for the quick addition of attributes and specific methods, but also inherits from an abstract parent class which defines characteristics and generic methods common to all joints.

In the same way, test methods have been placed in separate classes, so as to make their potential evolution possible, following progress in the field or, in the case of a custom application, specific constraints or in-house assessment methods. It is thus very easy to modify or add tests to those aforementioned standards and objective methods included in the application. It is interesting to note that, in the application, gesture assessment by the ISO11226 and AFNOR NF EN 1005-4 standards have been merged into a single test. Indeed, AFNOR 1005-4 uses the exact same angle measurements as ISO 11226 and adds to them a frequency factor.

In the perspective of maximum safety, the software was designed to detect the maximum angles reached in the course of a gestural interaction, and it computes its assessments from these maximums, according to the standards and methods stated above. The measurements of time and frequency, used for the assessment of gestures with AFNOR 1005-4 are only triggered when the threshold value (which requires time span to be taken into account) is exceeded: for instance, a 20° arm abduction (see Figure 2).

6. Perspectives

- Assessment:
  The goal of the assessment is to define less stressful standard gestural interactions. We will test a gestural interaction using certain joints (for instance shoulder, elbow, wrist, etc.) to decide on whether it is stressful. If it is, we will be able to point to the problematic joint(s) and the reasons for the stress (extreme angle, repetition, etc.) A subject will perform gestural interactions and the software will assess those gestures and display the assessment result (acceptable / unacceptable). It will also be possible to test several gestures and compare them to find the least stressful. We also plan to collect additional subjective data from users to incorporate them into the assessment process for a better appreciation of user stress.

- Validation of the application (Method):
  We are aiming at validating our method through performing an experiment where subjects manipulate a gestural interface through performing certain tasks in different conditions. The application then evaluates the physical stress and gives results about the level of fatigue for each gesture and each joint. Furthermore, subjective results are collected from the subjects through a questionnaire about the level of the fatigue they felt in each condition and each joint. The results given by the subjects and those given by the application will be analyzed and compared to find whether they correlate to each other or not and by consequence whether the method is valid or not.
  1. Subjects: 20 potential users of virtual reality systems (students, museum visitors, videogame players, etc.).
  2. Tasks: We are currently developing several elementary and composite tasks to test our approach. For example, the tasks of selecting and moving an object, exploring a scene, etc. The first one is to arrange items, that is to select an object among several in the stock box, move and drop it in the corresponding box. Each task can be performed in different conditions (box height, number of times, time required, accuracy, etc.) Each subject is asked to arrange objects in various boxes using gestural interaction.
  3. Physical devices: Microsoft Kinect for Xbox motion sensor and a computer screen showing the movements and assessment results.

- Improving detection accuracy:
  We are for the moment using the Microsoft Kinect for Xbox motion sensor to detect the movements. We plan to use more accurate movement detection techniques, such as a multi-Kinect system and / or an ART-Tracking movement detection system. We are also thinking of using an EMG to detect the level of physical effort exerted, thereby making the method even more objective.
7. Conclusion

In spite of the undeniable advantages of gestural interactions, the latter still exhibit several weaknesses, amongst which their negative physical impact on the subject performing them. In order to reduce that impact, it is important to implement a risk-factor assessment procedure to determine the levels of acceptability and comfort of the suggested gestures. This will ensure that the interactions created are more ergonomic and less stressful.

We propose a semi-objective assessment method of the gestures’ risk factors based on the assessment of work-related tasks and the specifications found in certain standards.

Our objective is to try to improve interaction in virtual environments and make them easier and less detrimental to subjects.

Acknowledgements

The authors would like to thank Sафwan CHENDEB for his support and his sound advice.

Reference


Towards ubiquitous tracking: Presenting a scalable, markerless tracking approach using multiple depth cameras

MICHAEL M. OTTO¹, PHILIPP AGETHEN¹, FLORIAN GEISELHART², ENRICO RUKZIO²

¹Daimler AG, Wilhelm-Runge-Str. 11, D-89081 Ulm, [firstname.lastname@daimler.com]
²Ulm University, James-Franck-Ring, D-89081 Ulm, [firstname.lastname@uni-ulm.de]

Abstract

Even though there is promising technological progress, input is currently still one of virtual reality's biggest issues. Off-the-shelf depth cameras have the potential to resolve these tracking problems. These sensors have become common in several application areas due to their availability and affordability. However, various applications in industry and research still require large-scale tracking systems e.g. for interaction with virtual environments. As single depth-cameras have limited performance in this context, we propose a novel set of methods for multiple depth-camera registration and heuristics-based sensor fusion using skeletal tracking. Based on a distributed, service-oriented and scalable system architecture, a markerless tracking system consisting of multiple Kinect v2 sensors has been developed for real-time interaction with virtual environments. Evaluation showed that a system based on the proposed techniques help in increasing tracking areas, resolving occlusions and improving human posture analysis. This system is used for ergonomic assessments in production planning workshops and it was shown that performance and applicability of the system is suitable for the use in automotive industry and may replace conventional high-end marker-based systems partially in this domain.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [User Interfaces]: Input devices and strategies—scalable, markerless tracking and full body motion capture

1. Introduction

Interactive virtual and augmented reality assessments rely on robust, real-time tracking. With the rise of affordable depth cameras, marker-less body tracking has become a feasible option for a number of application areas, not only for gaming but also in research and industry. Being an alternative to more expensive and cumbersome marker-based motion capture systems, depth cameras are used for gestural interaction, natural user interfaces and motion capture for film making. In industry, where e.g. interaction with virtual product models and process simulations have already been common using conventional motion capture systems, depth camera based systems soon also became an appealing alternative for marker-based full-body motion capture. However, considering spatially large use cases like car assembly, the limitations of single depth cameras impede their use. Limited sensing range, a high susceptibility to self and external occlusions and a greatly varying sensing performance depending on the user’s posture and position are some of the major drawbacks that need to be faced in order to use such systems in the mentioned scenario. One way to overcome these limitations is the use of multiple depth cameras which extend sensing range and improve tracking performance. But with this approach, there are also a number of new challenges which need to be addressed in order to successfully implement such a system. First of all, it is necessary to establish a common coordinate frame for the cameras by registering them to each other. Afterwards, the data coming from different cameras need to be combined in a meaningful way to actually gain improvements in tracking performance and range.

In this paper, we propose a novel system consisting of multiple Kinect v2 cameras for the use in ergonomic assessments. We present a concept of a distributed multi-depth-camera system, whose improvements were also quantified in a systematic evaluation. The remainder of the paper is structured as follows: We start with a review of the current state of the art on multi-depth-camera systems. Then we propose
a set of registration and fusion techniques and extend those to a complete, ready-to-use tracking system. The evaluation of this system in the last section shows spatial accuracy of registration performance. Subsequently an evaluation of a concrete use case is described. The paper concludes with an overall assessment and outlook on further optimizations.

2. Related Work

Various research has already been carried out in the field of multi-depth-camera systems, however mostly focusing either on certain applications or aspects of such systems, thus leaving others unspecified to a great amount. As already lined out, different challenges have to be faced in order to successfully implement such a system: Architecture, interference, synchronization, registration and fusion. Depending on the use case, it is often also necessary to handle additional application specific issues like user identification or world coordinates registration, not being further considered within this work.

2.1. Architecture

Most of the previous work is based on two or more Kinect cameras (1st gen.) which can be connected to a single computer, thus simplifying the required amount of infrastructure to a moderate level. However, works presented by Schö- nauer [SK13] or Martínez-Zarzuela et al. [MZPHDP*14] implement distributed systems, in which skeletal and depth data is gathered on camera nodes and being sent to a central fusion unit. This component handles the creation of a common view of the tracking space. Additionally, solutions have emerged in early states, which allow to stream Kinect data via network, e.g. by Wilson [Wil15]. In our system, a different approach has been chosen. All information can be requested via service-oriented and scalable RESTful tracking services as presented by Keppmann et al. [KKS*14].

2.2. Interference

As time-of-flight (ToF) and structured light depth cameras actively illuminate the scene, interferences can occur as soon as tracking frustums overlap, since any camera also receives light emitted from other cameras. There are two main approaches to interference handling which can be found in literature, (1) optical multiplexing (e.g. presented by Butler [BIH*12] or Faion et al. [FRZH12]) and (2) post-processing algorithms e.g. hole-filling as in Maimone and Fuchs [MF11]. Often it is also possible to simply ignore interferences when using certain camera types and setups, especially in skeletal tracking applications. As the proposed system uses ToF depth cameras, which generate negligible interference noise due to their modulation, no countermeasures against interference have been implemented.

2.3. Registration

One of the main challenges in multi-depth-camera systems lies in establishing a common coordinate frame by determining rotation and translation of the cameras to each other. Various approaches have been used for this, ranging from methods adopted from the 2D computer vision domain, horn-based methods like presented by Wilson and Benko in [WB10] or checkerboard-based approaches like those presented by Berger et al. [BRB*11] or Zhang et al. [ZSCL12], over iterative closest point (ICP, see [RL01]) approaches [PMC*11] to skeleton based (ICP-like) methods in more recent publications by Faion et al. [FRZH12], Asteriadis et al. [ACZ*13] and Baek und Kim [BK15]. Most of the methods yield comparable results, however strongly differing in the ease-of-use and setup time with different approaches. The proposed approach focuses on reduced setup times and an easy setup. Thus, we decided to implement a combination of skeleton-, regression plane- and ICP-based registration.

2.4. Fusion

After establishing a valid registration, skeletal tracking data from different cameras exist in a common coordinate space; nevertheless, body tracking skeletons are still individual and separate. To gain advantages of such a setup data fusion methods can be employed to gather an improved view on the tracking space. The possible methods range from simple best-skeleton approaches, over joint-counting approaches (see Caon et al. [CYT*11]), weighted averaging methods [FRZH12], to dedicated fusion algorithms e.g. by Yeung et al. [YKW13] or Asteriadis et al. [ACZ*13], which respect data quality and the specific tracking situation. This helps in dealing with occlusion and sensing limitations. Combining the advantages of each mentioned previous works, a set of novel fusion heuristics will be presented and analytically evaluated.
2.5. Assessment of related work

While covering many of the relevant aspects, most of the previous works leave out important factors of a multiple depth-camera system for universal use. Generally, registration and fusion approaches lack end-user optimization as well as comprehensive evaluation of underlying assumptions, e.g. for factors influencing registration and fusion methods and quality. With this work, some currently missing insights and concepts will be provided, which have proven to be useful for implementing a multiple, scalable depth-camera system.

4. Software

There are two main software components: The tracking software and the fusion software which is described in the following.

4.1. Service-oriented tracking software

Implementing a service oriented RESTful tracking service instead of conventional streaming architecture has several advantages: Third party integrators have the possibility of easily reusing the services for implementing clients. Additionally, using standardized and publicly available tracking vocabulary and Resource Description Framework (RDF) one can achieve interoperability between tracking devices which is also the goal of the ARVIDA project. In this context, the presented tracking services are using a RESTful polling-based approach with linked data which is conforming to the ARVIDA standard. It has been shown by Keppmann et al. [KKS*14] that RESTful Linked Data resources can be applied for virtual reality environments.

In the tracking service, information is gathered by the event-based Kinect SDK. The web service offers all skeletal information, the status of each skeleton, the floor plane estimation and color and depth camera views as RESTful resources. RDF datagrams are serialized using Turtle format. Each datagram contains time stamps for synchronization afterwards.

4.2. Fusion and multi sensor tracking service

The fusion and multi-sensor service is running as a central component on a high performance computer and handles registration, fusion and data input/output in the tracking environment. Figure 3 depicts the architecture of the registration and fusion service.

The fusion service polls the data of the tracking services. This data is used for calculating extrinsic transformations between the cameras for the subsequent fusion process. Several pre-processing steps have to take place in advance (see Figure 3) which are described in detail in the following paragraphs. Whereas the fusion component processes all joints of the skeletal data, the registration process only uses the neck joint information. This joint was chosen, due to its advantages compared to the remaining joints: front/rear invariance, orientation-independent position and low overall...
jitter. During registration, neck joint data is being captured over time from each camera and is used as an input point cloud for the ICP algorithm. The algorithm then iteratively minimizes the difference between two point clouds gathered by the sensors. The result of the algorithm is the refined extrinsic transformation between a pair of cameras.

After the registration, the heuristics-based fusion component is able to combine skeletal data from all registered cameras and provides them as an output to possible domain-specific application components.

4.2.1. ICP extension

Gathering only the neck joint information has an drawback which has to be compensated: Since the user’s movement takes place on the flat floor plane and the height of the user’s neck joint does not vary a lot, the gathered point cloud data lie almost on a single plane. To compensate this lack of variance, additional information is used. The floor plane estimation compensates the missing information. The floor plane is a rough approximation of the distance to the floor and the angle of the sensor. Fusing this information with the ICP data offers an improved transformation for extrinsic registration between one sensor relative to the master sensor. In addition to that, we propose to use a regression plane to further precise the ICP results, if enough feature points have been gathered during the user’s movement.

4.2.2. Front/rear detection

In order to achieve maximum flexibility for the hardware sensor setup, the fusion service has to recognize, whether the user is facing the Kinect or if he is turning his back on the corresponding sensor. The SDK always presents the skeleton data as if the user were facing the camera directly. Even in a rear view, the skeleton is recognized robustly but data being presented laterally reversed. Additionally, a robust indicator if the user is turning his back towards the camera, is to evaluate the angle between the shoulder joints. Evaluating the discrete skeletal states of the collar joints, one can determine the user’s orientation to the camera in each frame.

4.2.3. Scalability

To achieve a fully scalable system with a common coordinate frame, extrinsic transformation chains have to be built (see Figure 4). The above described method is used to calculate the transformation matrix for each camera pair with an overlapping tracking frustum. For N sensors sharing an overlapping tracking area there are \( (N-1)! \) transformation matrices. Having more than two sensors sharing the same tracking area, the system is over-determined and a cross-validation of transformation chains has to be carried out with regards to the absolute transformation precision. Therefore an error metric is introduced which consists of the summed up and normalized Euclidean distances of the reprojection error. Based on this error value, the best interlinked transformation chain between master and each client can be determined.

\[
w(d) = \begin{cases} 
1 - (d - 2.5m)^2 & \text{for } 1.5m < d \leq 3.5m \\
0 & \text{for } d \leq 1.5m \cup d > 3.5m 
\end{cases}
\]

(1)

4.2.4. Time Alignment and Interpolation

The time synchronization algorithm is crucial for interpolating the asynchronously captured body tracking information generated by the depth camera sensors. Since the user has to move during registration process a worst case offset of several centimeters is induced just by event-based, non-synchronized image acquisition. To generate synchronized timestamps within the whole sensor network NTP protocol is utilized. Based on these precise timestamps, skeletal body tracking frames are virtually synchronized within the fusion software through interpolation. The depth camera’s skeleton acquisition time is assumed to be constant over all sensors. Since the user’s body has a certain inertia and the refresh rate is approximately 30Hz, the inter-frame trajectory between two skeleton datagrams can be assumed as linear movement.

4.2.5. Fusion Process with quality heuristics

Having registered all cameras via extrinsic transformation chains, the tracked skeletons generated from different views are in the same coordinate frame and need to be fused. For large-scale human tracking and posture analysis we propose a set of quality heuristics for the skeletal fusion process. Each skeleton within each sensor is given a certain weight. The higher the weight the higher the influence of the certain sensor on the user’s fused skeleton. A comprehensive set of quality measures will be presented for real time skeletal fusion:

First, we propose the distance between the user and a sensor as the distance quality measure. At a distance of approximately 2.5m tracking results are most reliable. This quality measure weights the user’s skeleton over the distance to the neck joint respectively.

\[
w(d) = \begin{cases} 
1 - (d - 2.5m)^2 & \text{for } 1.5m < d \leq 3.5m \\
0 & \text{for } d \leq 1.5m \cup d > 3.5m 
\end{cases}
\]

(1)

Second, we introduce rotation quality heuristics for robust human activity analysis. If there is multiple data on the
user’s posture we propose to weight the front facing skeletons highly and to set all rear views to weight zero. The user has to stand as orthogonal to the sensor as possible, since 30° has been found to be the maximum vertical user orientation for reliably tracking limbs:

\[ w(\phi) = \begin{cases} 1 - |\phi|/30^\circ & \text{for } |\phi| \leq 30^\circ \\ 0 & \text{for } |\phi| > 30^\circ \end{cases} \] (2)

Lastly the lateral frustum quality heuristic limits the tracking frustum to a horizontal field of view of 50° so that the limbs are still probable to be within the tracking area of the sensor (70°). We propose zero weight if the user’s center axis joints exceed 50° in horizontal Axis of the local camera coordinate frame:

\[ w(\alpha) = \begin{cases} 1 - |\alpha|/25^\circ & \text{for } |\alpha| \leq 25^\circ \\ 0 & \text{for } |\alpha| > 25^\circ \end{cases} \] (3)

5. Evaluation of registration accuracy and validity

To determine the accuracy of extrinsic transformations and therefore the spatial registration error, a series of experiments has been carried out.

5.1. Experimental setup

Since an absolute accuracy evaluation is needed, a high precision marker-based tracking system was chosen as a ground truth. The system consists of 16 ‘OptiTrack Flex 13’ cameras which reported a residual mean error of 0.624mm for the whole tracking volume. On the Kinect sensors, rigid body markers were applied on the top of the sensor. The pivot point translation of the rigid body markers was defined to be in the Kinect’s depth camera focal point to match the origins of Kinect body tracking and the Optitrack rigid body markers.

5.2. Design of Experiments

All registration scenarios were conducted using two Kinect. The registration process has been recorded 100 times; for each of the five scenarios 20 measurements were performed. During each experiment point cloud movement data was gathered for 10 seconds within the overlapping tracking area. The scenarios differed by the angles around the vertical axis: 0°, 45°, 90°, 135° and 180°. No outliers were removed for the following evaluation.

5.3. Results

Figure 5 highlights the registration performance of the fusion service. Circles depict the calculated ideal Optitrack positions. For these scenarios the Euclidean distance in the floor plane is always less than 15mm to the ground truth position. The vertical axis reveals maximum deviations of 1.5° for the sensor’s pitch axis. The body tracking estimator within the SDK reveals uncertainties especially in the vertical axis. The uncertainty of the joints can vary more than 20mm, depending on the angle of the user.

6. Ergonomic assessments

One specific use case within the automotive industry where full body motion capture data is used are ergonomic assessments of workplaces. Ergonomics experts are using motion capture technology to virtually audit end-assembly workplaces. While being tracked a worker is performing the pre-planned assembly routines in the virtual environment whereas the ergonomics expert is evaluating the movements, weights and resulting forces. The following three experiments have been performed during real production planning workshops:

- Reachability check for mounting an antenna on the roof
- Posture definition for screwing work tasks
- Stress screening for battery assembly

6.1. Experimental hardware & software setup

During these assessment workshop six Kinect sensors are utilized which are all facing the center of the workplace and are evenly distributed on the edges of the tracking area. This area covers approximately 6m x 6m since movements within real workplaces in automotive end-assembly lines have equal dimensions.

Having registered all cameras to a common world coordinate frame, the presented system architecture in combination
with the fusion heuristics enable the worker to be constantly tracked regardless of his position and his orientation within the concatenated tracking frustum.

Standardized tracking protocols have been implemented to connect to commercial VR software used: A.R.Tracking protocol, VRPN and ARVIDA linked data protocol. To carry out the mentioned ergonomic assessments in real time the virtual manufacturing software Delmia V5-6 R23 has been used in combination with Haption RTID plugin. The following pipeline was used in this case: The fusion service exposes all fused tracking joints via the A.R.Tracking protocol as 6DoF tracking data. The Haption suit and configuration maps this tracking data onto the fully flexible virtual human. 20 tracking joints are used to modify the DHM interactively.

As depicted in Figure 6 the virtual scene in Delmia V5 included a car body in the assembly status for the respective station. Dynamic parts have been simulated and attached to the right hand joint. The anthropometry of the virtual human was adjusted to the real worker’s size and weight.

6.2. Results

All mentioned manufacturing tasks could be carried out without having any prior physical mock-ups. Limitations of the pre-planned process and unfavorable ergonomic situations could be identified for all three experiments with this virtual methodology. Additionally the results gathered could be verified by subsequent traditional hardware workshops.

Comparing common marker-based tracking systems to this novel approach, ergonomic experts pointed out several effects: First of all, the users do not have to put on a special suit with retro-reflective markers. This is time-consuming and cumbersome for the tracked persons. User’s movements may be influenced by the marker suits and seem not as natural as in regular working clothes. Secondly, users can swap immediately without any preparation time, so that multiple users can test the process without any prior work. Lastly, the markerless system induces more latency and jitter to the tracking data than the marker-based tracking system. Ergonomic experts pointed out that the motion capture data quality is still sufficient to identify and solve the issues related to ergonomic assessments. Latency of several frames is considered to be irrelevant, since there is no immersive feedback to the user causing motion sickness. It became apparent that the registration and fusion precision are sufficient for human posture analysis, for profound ergonomic simulations and for large-scale view point control applications in virtual environments.

Additionally for an automatic recognition of digital human postures, the ErgoToolkit was utilized in this pilot case that was presented by Alexopoulos et al. in 2013 [AMC13]. With this additional plugin a rough stress screening could be carried out automatically and critical postures could be detected reliably. Furthermore, experts appreciated the side benefits of this tracking approach like visibility checks through interactive viewpoint control and validation of assembly and disassembly routines for dynamic virtual objects via hand joint tracking. Follow-up processing times like documentation can be reduced significantly, by pre-filling assessment sheets automatically. All of these use cases will directly profit of advances in multi depth-camera tracking technologies.

7. Conclusion

In our effort to improve multiple depth-camera systems, we developed a novel large-scale multi depth camera system, which supports scalable setups and different use cases through its service-oriented architecture. The number of possible tracking nodes is limited by computing power and network throughput. Setups up to ten tracking services have been successfully tested but more sensors should be possible as long as network throughput is sufficient. The fusion service itself can be addressed transparently and acts externally as if it was a single sensor tracking service. Standardized tracking protocols have been implemented in order to achieve interoperability. Furthermore, several novel registration-relevant techniques have been presented and evaluated like time-synchronous interpolation, front/rear detection and error measures. Additionally, a comprehensive set of quality heuristics has been derived for the skeletal fusion process, which showed to improve skeletal tracking.

Three pilot test cases within the automotive industry have been carried out to evaluate the system’s performance with real ergonomic use cases. The requirements in terms of oc-
clusion robustness (e.g. when working with car bodies in the tracking area), tracking range and tracking precision could be fulfilled in each of the pilot test cases. Since the novel system proved its applicability, reduced costs and the ease-of-use, it will complement the variety of existing industrial tracking systems.

In future work, we plan to refine the fusion process by extending heuristics with additional criteria and more fine-grained weighting, e.g. on a per-joint or per-bone level instead of the current per-body approach. Additional measurements and analysis may also broaden the insight on the behavior of the proprietary Kinect technology and thus lead to further improvements in this approach.

8. Acknowledgments

We would like to thank Brian Moore, Microsoft Germany, for supporting this work. This work was additionally supported by the German Federal Ministry of Education and Research (ARVIDA project, grant no. 01IM13001N).

References


3D-Tune-In: 3D sound, visuals and gamification to facilitate the use of hearing aids

Lorenzo Picinali¹, Mirabelle D’Cruz, Luca Simeone on behalf of 3D Tune-In consortium

Abstract
This is a short paper introducing an innovative approach using 3D sound, visuals and gamification techniques to support people with hearing devices to understand their many features and how to calibrate them in different real world situations (e.g. at a concert, in a restaurant, on a street, at a train station, in a classroom). 3D-Tune-In brings together the relevant stakeholders from traditional gaming industries (Reactify, UK; Vianet, Italy; XTeam, Italy; Nerlaska, Spain); academic institutions (Imperial College London, UK; De Montfort University, UK; the University of Nottingham, UK; University of Malaga, Spain); a large European hearing aid manufacturer (GN Hearing, Italy); and hearing Associations (Extra Care, UK; Hearing Link, UK; Action Deafness, UK; Accesibilidad y Personas Sordas, Spain and Ente Nazionale Sordi, Italy) to create 3D environments which will greatly improve people’s quality of life, generate new markets and create job opportunities.

Categories and Subject Descriptors (according to ACM CCS): 3D, sound, visualisation, virtual environments, digital games, gamification

1. Introduction
3D Tune-In “3D-games for TUNing and lEarnNg about hearing aids” is a three-year European project funded under the Horizon 2020 ICT work programme. Coordinated by the recently established Dyson School of Design Engineering of Imperial College, UK, it has nine university and industry partners from Italy, Spain and the UK, and began on 1 May, 2015.

2. Background to the project
Over 90 million people in Europe currently suffer from hearing loss, and due to an ageing population this number is likely to continue to increase. While hearing aid technologies have dramatically advanced in the last 25 years, people’s perception and use of these devices have changed very little. Hearing aids are now sensibly smaller, but incorporate several functions that go far beyond the simple amplification and equalization operation performed by the analogue devices (see Figure 1 below).

Nevertheless, this technological advancement is not always accessible or accessed by the hearing impaired population. The majority of individuals with hearing aids use the device as if it was a standard analogue hearing aid, i.e. only for its amplification and equalization features, and new algorithms are under-used or not exploited to their full potential. Hearing impairment in older adults can lead to frustration, low esteem, withdrawal and social inclusion [W99]. Furthermore, in children hearing loss affects speech and language development that impacts on academic achievement and future vocational choices [Y03].

Figure 1: Miniaturization from an analogue hearing aid (on the left) to a modern digital one (on the right)

Traditional gaming technologies have been successfully employed in non-leisure scenarios for learning and skill acquisition, empowerment and social inclusion [DDD12]. In these scenarios, mechanisms such as game dynamics to ensure an adequate level of competition among the players, an effective reward system and a captivating storyline have proven to be effective elements to engage and motivate users [MG11]. The challenge is to facilitate the successful exploitation of existing, overlooked or neglected functionalities of hearing devices to optimize their potential thus greatly improving people’s quality of life, and their interactions with other people and their surrounding environment.

¹ Image from www.hearinglink.org, last accessed 06/2015
3. The 3D Tune-In Project

The 3D Tune-In project will create an innovative toolkit based on 3D sound, visuals and gamification techniques tailored to different target audiences (e.g. older users and children). The consortium consists of relevant stakeholders from traditional gaming industries (Reactify, UK; Vianet, Italy; XTeam, Italy; Nerlaska, Spain); academic institutions (Imperial College London, De Montfort University and the University of Nottingham, UK; University of Malaga, Spain); a large European hearing aid manufacturer (GN Hearing; Italy); and hearing Associations (Extra Care, UK; Hearing Link, UK; Action Deafness, UK; Accesibilidad y Personas Sordas, Spain; and Ente Nazionale Sordi, Italy).

A participatory approach will be adopted to create virtual auditory and visual scenes that will:

- Enable end users to explore, review and customize hearing aid devices for different scenarios (e.g. at a concert, in a restaurant, on a street, at a train station, in a classroom, see Figure 2 below)

![Figure 2: Possible scenarios for 3D-Tune-In depicting indoor and outdoor environments](image)

- Enable individuals with no hearing impairment to understand how hearing loss can compromise everyday activities, and how a hearing aid can improve this situation
- Enable Small-Medium-Enterprises (SMEs) in digital gaming to explore new non-leisure applications in the area of hearing loss and hearing aid technology with support from the scientific community
- Enable hearing aid providers to evaluate and demonstrate the various functionalities of their products using 3D technologies to improve their services and increase sales

The expected outcome includes: (i) Technology transfer between traditional SME game developers and broader research and industrial communities in 3D sound and virtual reality; (ii) 3D Tune-In Toolkit for development of further hearing aid-related technologies; (iii) 3D game applications; and (iv) Guidelines on the effectiveness of 3D and digital games on hearing loss and hearing aid technologies.

4. 3D Tune-In and VR

The 3D Tune-In project will employ acoustic and visual VR technologies in the attempt to transform the approach towards HA technologies of both stakeholders and manufacturers/sellers. A custom binaural 3D audio engine will be developed and implemented, including functionalities such as binaural reverberation, Head Related Transfer Function selection and hearing loss and hearing aid simulation. The 3D Tune-In binaural engine will be released with an open source license.

5. Preliminary results

The project kick-off took place at Imperial College London on 11-12 May, 2015. [WP15]. The project consortium discussed the state of the art in hearing devices and the potential 3D technologies and challenges within the project. In addition to the 24 participants from the project consortium an initial ‘User Requirements Workshop’ was organised by the University of Nottingham with representatives from the UK and Italian Hearing Associations. This workshop enabled us to identify relevant contacts, user groups and appropriate tools to select for the initial phase of requirements capture. The next 6 months will include questionnaires, interviews and focus groups with relevant stakeholders to support the appropriate design of the 3D-Tune-In.

6. Conclusion

3D-Tune-In provides a great opportunity, not only to make a real difference to many people’s lives now and in the future, but also opens up new markets to the virtual reality community. For more information please contact the 3D Tune-In coordinator Dr Lorenzo Picinali (l.picinali@imperial.ac.uk), or visit: www.3d-tune-in.eu.

7. Acknowledgements

The research leading to this contribution has received funding from the European Community’s H2020 Programme under Grant Agreement No.644051. This article reflects the authors' views. The European Commission is not liable for any use that may be made of the information contained therein.

References


A prototype of an automatically assisted and person-focused rehabilitation system for post-stroke patients at home

Stefano Mottura1 Sara Arlati1 Luca Fontana1

1National Research Council, Institute of Industrial Technologies and Automation; Milano, Italy; {stefano.mottura, sara.arlati, luca.fontana}@itiia.cnr.it

Abstract
Rehabilitation after stroke supported by mechatronic devices and digital environments is a matter of research since many years. It has been discovered that these technologies bring effective improvements in patients’ life but clearer results are expected to come. The developed systems often focus their attention on games with the main aim of engaging the patients. In this paper a prototype of a robotic device coupled with a digital environment for rehabilitating at home the upper limb by focusing the attention both on the patient and on the quality of the performance is presented.

Categories and Subject Descriptors (according to ACM CCS): H.1.2 [User/Machine Systems]: Human Factors, Software Psychology; H.5.1 [Multimedia Information Systems]; H.5.2 [User Interfaces]: Auditory (non-speech) Feedback, GUI; J.3 [Life and Medical Sciences]: Health.

1. Introduction
Nowadays stroke is a major cause of disabilities, since cerebral vascular accidents often lead to various motor, functional and cognitive impairments [Men13]. In this context, the ultimate role of rehabilitation is helping patients in recovering not only at physical and sensory level but also at intellectual, psychological and social functional levels [Fli90]. The rehabilitation activities are useful for promoting recovery because of different reasons and, among them, motor training was proven to be effective in exciting the neuroplasticity by inducing the creation of alternative pathways for sensory perception and motor skills [APV*11]. In this context, one of the aims of the research is to enhance the rehabilitation processes by integrating, or substituting, the “classical” exercises with the help of technological means that can support patients while they perform the rehabilitation tasks. The use of specifically designed robotic devices and interactive dedicated Virtual Environments (VEs) has been proven to be effective in such an aim [LHC*14][NNK*14][FD13]. Robotics, and more generally mechatronic systems, address and constrain the patient rehabilitation movements by allowing a high level of repeatability. The devices can be also equipped with sensors for recording additional data for further investigations on the activity of the patients [LBS*02]. The device-assisted rehabilitation is sometimes integrated with “games for rehabilitation”, designed to increase the engagement of the user [BPL*13]. In many occasions, a disadvantage of games and VEs is that no feedback is provided to the patients: the goal of the game can often be reached using different strategies, such as for example compensation movements, that are not corrected by the system as they would be by a real therapist [PHB12]. Consequently, even if the patients show improvements in their physical functionalities, they are left unaware of some aspects of the rehabilitation, such as the quality of the movements they are performing, their progress over time and other information that a human therapist would provide. In this paper we present a prototype of a system for the rehabilitation of the upper limb in post-stroke patients for domestic use. Such a system is called VA_LINarm and it is made up of a low-cost robotic variable-stiffness device for the rehabilitation of the upper limb, namely LINarm [MCL*14], interfaced with a software application which provides several features to help patients in their rehabilitation process with the LINarm robotic device. The acronym “VA” in the VA_LINarm stands for Virtual Assisted LINarm, because a sort of virtual assistant, which keeps track of the actions of the patients and gives him/her back the proper comments and suggestions, is integrated in the application. The purpose of this prototype is to provide an instrument able to allow patients to do an effective rehabilitation at home.

2. System overview and way of operation
The VA_LINarm is composed by the following main elements (Fig. 1): the LINarm device (Fig. 2), a Microsoft Kinect v2 sensor and a graphical user interface (GUI)
shown on a screen placed in front of the patient. The VA application controls the entire system by handling all the input and the output data coming from and going to the elements mentioned above. The LINarm is a 1-degree of freedom actuated device: a linear track is provided with a movable handle that has to be hold by patient (as shown in Fig. 1). The motion of the handle along the track makes the impaired upper limb to be exercised. Depending on the impairment level of the patient, the device can be used either in a passive way, meaning that the motion of the handle is provided by the two motors of the device, or in an active way, which implies an active contribution of the patient. At low level, the motion parameters of the device are set and managed by the VA application which is interfaced with the LINarm firmware. The Kinect v2 sensor is used for tracking in real time the body of the patient with the aim of checking if each body segment is in the right position while doing the rehabilitation exercises.

The GUI is aimed to constantly keep the patient aware about what he/she is doing and to show him/her, when needed, a virtual assistant character (called Jimmy) to engage, correct and encourage the patient. The contents of the GUI are the outcomes of different functional modules running in the VA application that collect and elaborate runtime data from the physical devices and eventually perform specific elaborations to evaluate the performance (as mentioned in Paragraph 4). Four main logical operative units in the overall process can be identified (Fig. 3): the Exercises Module, the Evaluation Module, the Kinect Module and the GUI Module. While the patient is performing the exercise, according to the parameters set in the Exercise Module, both his/her interaction with the LINarm device and his/her body position are monitored with the processing done by the Kinect Module. These data are used for an immediate visual feedback in the GUI Module and constitute the input for the Evaluation Module, which estimates the performance and drives the virtual assistant in its proper lines. The application has been developed using C# language, with the use of Microsoft Windows Presentation Foundation (WPF) libraries [Wpf15].

![Figure 1. The complete VA_LINarm system.](image1)

![Figure 2. The robotic device LINarm [MCL*14].](image2)

3. The Exercises Module

The Exercises Module (EXM) is dedicated to the management of the physical exercises behaviour and execution with the robotic device. As already mentioned, the device has one linear degree of freedom, thus the handle can be moved from a start point to an end point, driven by the motors and/or by the action of the patient. There are different parameters that have to be set in the VA application to model the specific exercise behaviour: the control mode (passive or active), the start/end points, the device stiffness, the maximum velocity and acceleration. All these parameters can be modified even at runtime. Instead, the device values monitored while the patient is performing the exercise are: the actual position and the velocity of the handle and the force applied on it, both in modulus and direction. The combination of different sets of parameters and different checks on the incoming data allows implementing a large number of exercises that are all implemented with different derivations of a BaseExercise class. The basic function logic for the exercises (implemented in the BaseExercise class) has been modelled as a sequence of instructions to be sent to the Arduino controller for driving the LINarm initialization (I) and the exercise runtime (R).
In I, the setup of all the parameters defining the specific kinematic and dynamic behaviour of the device is done. In R, the sequence of instructions is sent step by step to the device so that it acts as required. The phases I and R are implemented with virtual methods allowing to maintain the same programming paradigm for all the implemented exercises. The specific exercise is implemented in a derived class, where the methods related to I and R are overridden with their own proper implementation. All the operations related to the low-level management of the LINarm device are implemented in another specific software module (a class called LINormManager) that plays as a driver for the hardware and whose methods are called, when needed, by the exercises classes. The LINormManager uses the event-driven model [Eeh15] for notifying what is happening with the device. The most used event is for notifying that new kinematic and dynamic data are available. The EXM receives this kind of event, thus it gets notified whenever new data are coming from the device and it acquires these data to know in real time which is the current position of the handle, its velocity and how much force is applied on it, and this allows the EXM to determine the completion of the current single operation. The events model is also used by EXM for notifying the other modules the things happening during the runtime (as, for example: target point reached, required force threshold reached, first half of the required movement has been done). Fig. 4 summarizes the whole process.

Figure 4. The flow of information during the execution of the k-th instruction of an exercise. The red arrows indicate the event-driven communications, whereas the blue arrows indicate communication with the physical device, while the black arrows represent the logical flow.

4. The Evaluation Module

The Evaluation Module (EM) is the VA software module that takes as input the device raw data and that uses them to estimate the current performance of the patient and, at the end, to select the proper lines of virtual assistant Jimmy. The evaluation of the patient performance occurs in two steps (see Fig. 5 below) that come in succession inside a runtime loop. In the first step, the raw data are elaborated to calculate how severe the possible errors are. In the second step, the results of the first step are combined to define what has to be signalled to the patient.

Before describing how these two steps work, it is necessary to mention that the EM evaluates the patient performance by checking if certain physical quantities, derived from body and device data, are in specific acceptable ranges. These quantities, their ranges and their relevance in the evaluation process must be defined by a competent figure, as for example a therapist, before the entire system is released to the patient for using it at home. Only in this way, in fact, the patients will be put in the condition to face exercises and corrections that are customized on their actual residual capabilities.

The information set by the therapist are mainly used within the first step of the EM. In this step, in fact, the collection of all these values of interest and their comparison with the pre-set ranges occur. The comparison takes place either each time a new set of raw data is ready or at the end of a complete repetition of the rehabilitation movement. In the latter case, the data corresponding to the entire movement are stored and their evaluation takes place only when the EM signals that the movement has been completed (see also Paragraph 3). The result of this first step of evaluation is a real number, in the [0, 1] range, for each single physical quantity considered. This formalization has been chosen to normalize all the possible errors: a value of 0 means that no error was committed, while values comprised in (0, 1] indicate that the evaluated quantity is outside of its acceptable range. The higher the normalized error is, the farther is the evaluated quantity from its range of acceptability.

The second step of the EM is devoted to the combination of all the normalized errors resulting from the first step, with the aim of selecting what has to be told to the patient by the virtual assistant. In particular, the selection of the message is performed by going through all the normalized errors and by evaluating if there is at least one that exceeds a certain threshold (\(T\)). If so, an erroneous state is identified, whereas if no values exceed \(T\), a successful state is determined. If a certain state persists over a certain time \(T\), an application event, whose arguments contain the piece of information to select a specific Jimmy line, is raised. Since more than one error can occur at the same time, the way to discriminate which one has to be signalled is determined by assigning to each error a priority value. With the priority assignment, different errors assume different relevancies into the evaluation process: the higher the priority is, the earlier that error has to be corrected. It is important that the priorities definition reflects the “cause/effect chain” of errors: an error causing another one must have a higher priority. For instance, if the patient is sit with a wrong posture, it is likely that he/she will perform the exercise by keeping the exercised arm in a wrong position, too. To properly correct this situation, it is necessary to communicate to the patient that his/her posture is wrong and then,
eventually, tell him/her about other body segments. To make the VA application treat this situation in the right way, the priority of all the errors related to posture must be higher than the priority associated to the errors related to the wrong positioning of the other body segments.

Another important aspect that is handled in this second step of the EM is the timing of the messages. A predefined value defines the time \( T_{\text{wait}} \) that must elapse from an appearance of Jimmy and the following one. This is done to avoid overloading the patient, who may also have some (mild) cognitive impairments, with too much information [Rik11]. The value of \( T_{\text{wait}} \), as well as \( t_{\text{th}} \), may be varied during the execution of the exercise by someone who assists the patient, but does not necessarily have medical competences (i.e. a caregiver). In this way, the rate of the appearances of Jimmy may be increased or decreased, depending on the patient’s levels of attention, tiredness, concentration, etc. This aspect provides a further level of software customization, allowing the patient to have a rehabilitation treatment personalized not only on his/her residual capabilities but also on his/her actual physical and emotional state.

Figure 5. How the Evaluation Module works. In the first step, raw data are collected and evaluated singularly. In the second step, the results of the first one are combined to select the message to be communicated to the patient.

It is important to notice that the EM allows the entire system to be modular and easily expandable/modifyable according to the patient capabilities or the exercise to be performed. This flexibility is allowed thanks to the software architecture with which the EM and, in particular, the first step of its processing, were implemented. In the first step, in fact, the evaluation of each quantity of interest occurs in a different software submodule, called Evaluation Unit (EU). All the EUs have the same general architecture so that their main functionalities can be recalled in the same way inside a runtime loop. To do this, it was chosen to make each EU class implement the same interface (a reference type of C#, [Int15]), called IEvaluation (column I in Fig. 6). The definition of an interface, in fact, allows giving different implementations to methods that have the same signature. In this case, this software design choice allowed to implement different Update methods, so that each derived class could be characterized for collecting different raw data to compute different exercise performance parameters. Moreover, it also allows having different Evaluate methods implementations so that the previous values can be compared as needed to the pre-set thresholds to return a normalized error (column II in Fig. 6).

The VA application contains classes that are able to compute and evaluate anatomical angles, leaning values, force values and the time elapsed between two events. These classes are instantiated by passing them parameters defining the specific physical quantity that has to be evaluated but also the type of measure to perform and the allowed range for that measure (column III in Fig. 6). For instance, to evaluate the elbow flexion-extension (and to instantiate the corresponding class), it is necessary to tell the software: that the specific EU will be an instance of the AngleEvaluation class, that the angle to evaluate is comprised between the arm and the forearm and that the acceptable range is between 10 and 60 degrees. All these pieces of information required for proper object allocation are given to the software at runtime thanks to the use of a runtime plug-in framework [Srn15]. This allows to have a customized evaluation model for each patient, as mentioned before, but also to expand the original application without modifying the original source code.

5. The GUI Module

The Graphical User Interface Module is the VA software module that manages the multimedia contents addressed to the patient during the exercise execution, through the screen and the loudspeakers. This module is very important since it is the communication channel for conveying to the patient the information related to his/her performance. The GUI shows to the patient two types of contents (Fig. 7): continuous feedback about what he/she is actually doing (type A contents) and comments, suggestions, warnings about the performance only when needed (type B contents). The widgets of the type A are: an avatar, a gauge indicator showing the force exerted on the handle and a monitor for showing the force exerted on the handle and a monitor for the handle position. The avatar is the means that in real time gives the patient an augmented feedback on each body segment position. Its use allows the patient to monitor continuously his/her movements and provides a way for gaining awareness of the self during the exercises.
The avatar was intentionally chosen to avoid cognitive biases in the patients related to perception of the self with respect to their physical impaired bodies. In addition, the avatar is able to provide an immediate feedback when a certain body part is not in the right position. In fact, when a bad posture is detected by the system, the avatar becomes semi-transparent and the joints out of the optimal range blink in red. The avatar is driven with the data stream provided incoming from the Kinect v2 device, as also shown in Fig. 3. The 3D body has been modelled with proper skeleton information [JB00] in a tree of kinematic chains for moving the limbs and the head independently. The current version of the avatar’s model is just a simple silhouette that only reflects body parts movements but its features will be improved in the future to increase the patient embodiment. The avatar body will have a general male/female aspect, while its face will be customizable by selecting among different types of hair, eyes, beards, glasses, etc.

As mentioned above, the GUI module, as all the Windows-related contents of the application, has been developed within the WPF framework [Wpf15]. While the VA application is a C# application, the functionalities aimed to the management of the avatar and its real time 3D rendering have been developed in a C++ DLL. The developed classes devoted to the avatar representation are in fact based on the OGRE3D C++ libraries (with OpenGL rendering system) [Ogr15]. These libraries were properly wrapped to be used in the VA application.

The gauge indicator (in Fig. 7) is a widget that indicates if the force applied by the patient on the handle is in an acceptable range (green area). It is an indication of whether the patient is exerting the right amount of force while performing the exercise to keep the needle in the green area. This indicator provides a continuous feedback about how much the patient is pushing or pulling on the handle. This kind of feedback is helpful because the patient may (partially) lack the complete control on the functionalities of his/her impaired limb. As last widget of type A, shown alternatively to the gauge indicator, there is the hand-track monitor (in Fig. 7). It is aimed at showing in real-time the position of the handle along the track on the physical device. It is a vertical bar (the track) with a hand (the handle, held by the patient) moving on it according to the actual position on the device. This widget can be used also for showing some types of targets to be reached (in Fig. 7, a donut) in order to make the upper limb flex and extend to cover a certain range of motion. To animate these last two widgets, the data coming from the device, that are also addressed to the Evaluation Module, are used.

The only interface element of type B is the virtual assistant Jimmy. It is the communication channel for reproducing in the digital world the positive relationship patient-caregiver established during a "traditional" rehabilitation path [APV*11][CV12]. Jimmy is an animated smiley face with hands (in Fig. 7) and it appears only when something has to be signalled in a way that has to be clearer than the little invasive feedback provided by type A widgets. On one side, Jimmy recalls the patient’s attention on possible wrong actions, with the aim of making him/her aware and correcting certain specific wrong behaviours; on the other side, Jimmy encourages and motivates the patient by appearing after a successful series of movement repetitions to congratulate the patient. The speech of Jimmy is synthesized and filtered in order to dissimulate the specific gender male/female for avoiding possible unease in the patient. The Jimmy character, as a whole, acts with eyes expressions, head and hands movements in concert with the meaning of the current speech to simulate empathy and to better emphasize the message to be conveyed (Fig. 8) [Bat94]. The virtual assistant is a 3D model that has been properly animated into a set of short audio-visual clips that are reproduced using the VLC player libraries for WPF. The Evaluation Module which estimates the performance, as described in Paragraph 4, and returns a final result (a numerical index) identifying which specific animation of Jimmy has to be played.
6. Remarks and conclusions

In this paper the main functionalities of a low-cost, at-home rehabilitation system prototype have been presented together with the software architecture and design. Research typically carries on studies in the field of rehabilitation mainly focusing on the innovation in IT and mechatronics. The work presented here joins the concept that the patient is the key subject in the rehabilitation process. This concept is naturally deeply rooted in the “usual” rehabilitation processes performed by humans (medical staff) but it seems to be not so well defined when the patients have to face (semi)-automated systems, where the human interaction between two subjects (patient and medical staff) is almost lost. The entire system will be soon tested on healthy subjects with the aim to identify possible critical aspects, especially with regard to the quantity and the quality of the feedback given to the user. After these first trials and the consequent modifications of the software, the system will be tested with chronic post-stroke patients. From this second session of tests it will be possible to analyse the patients’ response toward the avatar, the virtual assistant and the general approach to this new technological mean to perform rehabilitation.

7. Acknowledgments

The work presented in this paper has been carried out within the Riprendo@HOME Project, Regional Research Project founded inside the framework agreement between Regione Lombardia (Italy) and National Research Council, D.G.R. n. 3728-11 July 2012. We would like to mention our colleagues Andrea Zangiacomi and Claudia Redaelli for the work about the perceptive and cognitive requirements of the patients and our colleague Nicola Colosio that has designed, developed and rendered Jimmy and its animations.

References

Augmented Notepad: an augmented reality application to support elderly people in daily activities

MARIA DI SUMMA, MARCO CAMPANALE
Institute of Industrial Technologies and Automation National Research Council of Italy, ITIA-CNR; Bari, Italy; maria.disumma@itia.cnr.it, marco.campanale@itia.cnr.it

Abstract

Modern society is witnessing a rapid and inexorable aging of the population, as a result of which there have been significant scientific advances in the medical field. It is also important, however for the scientific community to find ways to ensure that global aging is economically sustainable.

The purpose of this work is to propose an augmented reality tool that can provide support to elderly people in their daily lives. This is an important topic for an aging society seeking ways of enabling older people to live independently for as long as possible. This paper presents a framework designed to offer a certain degree of independence to the elderly in their homes. The first prototype, which was developed as part of the regional @Monitech project, has allowed us to evaluate the feasibility of the idea and its potential.

Categories and Subject Descriptors (according to ACM CCS): H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities

1. Introduction

Analysis of the annual reports released by the World Health Organization (WHO) since 2005 has shown that life expectancy has increased significantly in recent years, and all the evidence suggests that there will be further increases in the future. Another pertinent statistic is that the yearly increase in the elderly population is twice that of the total population. The data collected by the World Health Organization comes from 194 countries and is based on a series of indicators of mortality, diseases and related health systems, including life expectancy, deaths due to disease, health services and treatments, and investment in health, as well as behaviours and risk factors for health. Obviously, the data collected must be considered in the context of its regional origins. In fact, in the most economically developed countries life expectancy tends to increase annually and, as a consequence, greater funding is required to tackle the increase in health problems that inevitably arise from the phenomenon. The data collected by the WHO has been published in World Health Statistics 2015. Among the various results in the document, it is worth focusing on the index of mortality for adults, or, more specifically, the probability of death between 15 and 60 per 1000 inhabitants. In Italy alone between 1990 and 2013 the rate has almost halved: for men it has fallen from 129 to 69, and for women from 60 to 38 [who].

This confirms that the global population is aging and emphasizes the consequent need to explore new ways to cope with the phenomenon and its obvious economic implications. An increase in the number of elderly often means a rise in admissions to hospitals or nursing homes and this inevitably imposes a huge strain on community resources. Ambient Assisted Living (aAL) (AAL) is a research area which seeks to provide concrete solutions to the problem. AAL is a new approach that combines information and communications technology (ICT) with the human and social dimension of individuals in order to improve quality of life as much as possible.

The application of technology to ensure that the wellbeing of the population becomes an increasingly important priority is therefore at the core of this new research area, which can provide solutions to support the elderly by equipping them with the tools they require to maintain their independence. One of the goals is to find technological solutions that are inexpensive and more easily accepted by the elderly so that they can be assisted at home for as long as possible, thus reducing the need for hospitalization, improving their emotional wellbeing and obtaining a clear economic benefit for
2. Framework

This article shows the design of a framework that will be developed through an augmented reality system to support the elderly in their daily activities. Augmented reality is a technology that has already been available for a number of years [A’97], but it has only recently achieved success in fields other than that of scientific research, thanks to the development of low-cost tools and to the spread of smartphones and tablets. The progress made in areas that might be described as ‘historic augmented reality’, such as the video game industry [PT02] [Whi03], mechanical maintenance [HF11] [TGW’95] and medicine [BFO92] [ABB*01], have led the scientific community to introduce augmented reality into many other areas such as education [SR10] or the newer field of ambient assisted living [ALVMGMPM09]. The framework was inspired in the work of Aviles-Lopez [ALVMGMPM09], which describes a very complex set-up involving sentient visors coupled with a sensorial infrastructure that are able to interact with older people and understand their needs without requiring the use of any software application. As an alternative to the system proposed by Aviles-Lopez [ALVMGMPM09], it is possible to create a simple, versatile tool that is gradually customizable and adaptable to different demands in different scenarios, and does not require a sensorial infrastructure. This work is targeted at the elderly people of tomorrow, who, on account of their familiarity with smartphones and tablets, will have no problems interacting with simple applications; augmented reality can thus become a means of support in the daily life of everyone.

Our application is an attempt to provide a simple, inexpensive solution to the problems of those elderly people who live alone. A diagram of the framework for the application is shown in Figure 1. The application offers different functions depending on who performs the log-in. The caregiver logs in via a web browser and enters data that will be later viewed by the user. Once entered, the data is saved in the database. Now the user can use Vuforia features such as marker recognition and the display of augmented reality images. Through the camera of his tablet or smartphone he frames a symbol that will provide him with support in tackling a specific task, using augmented reality images or captions; whenever he clicks on an augmented reality image or a caption on the screen, the system updates data on the mobile application and continues to do so until the completion of the task. When he finishes using the application, the database is updated and the new data is made available to the caregivers. An application of this kind must be able to be installed on smartphones and tablets. Obviously, from a technical point of view this framework requires the implementation of a mechanism to prevent simultaneous access to data by the caregiver and the user. The system has to be simple and non-invasive; these are both fundamental characteristics of any tool intended to provide constant support to older people performing their daily activities. The peculiarity of the system is the fun aspect of the application, which should make it more easily acceptable to the user.

2.1. Medication

One of the features of the application is to support the user when he is taking medication by displaying a green or red symbol in augmented reality to inform him if the medicine he is going to take is the right one or not.

2.2. Monitoring

As regards monitoring, augmented reality might provide support in the measurement of blood pressure (BP) by offering a step-by-step description of how to obtain readings correctly: this involves the user framing the sphygmomanometer with his smartphone and receiving an initial instruction on what to do; once he has read this, he clicks to receive and read the second instruction, then repeats the process as often as required until the end of the procedure.
2.3. Diet

Elderly people often have to follow a controlled diet that may require them to provide daily or weekly answers to questions like:
1. Have you drunk the recommended 2 litres of water today?
2. Have you walked for at least 30 minutes?
3. Have you eaten at least 5 portions of fruit and vegetables?
4. Have you eaten fish?

The proposed Augmented Reality system tries to provide support for the performance of these tasks.

Case one (water)

The aim is to make the experience of using the application enjoyable and stimulating for older people. We assume that 2 litres of water corresponds to 8 glasses. First, the elderly person uses his device to frame a marker on a bottle specially prepared for the purpose. Initially he will see 8 glasses of water in augmented reality. Subsequently, whenever he drinks a glass of water, he clicks on the image of a glass and it disappears. Once all eight glasses have been drunk, a symbolic image of the objective attained will provide him with a sense of achievement Figure 2.

Case two (walking)

In this case the goal is to remind a user who is about to leave the house that a half-hour walk would be beneficial. This is done by displaying either an image in augmented reality or a caption when he frames the front door of his house in his device. If he does the walk and then clicks on the image or the caption, he will see a smile replace the image.

Case three (fruit and vegetables)

This feature is similar to that outlined in Case One. When the user frames the refrigerator with his device, he receives a visual signal reminding him how many portions of fruit and vegetables he still needs to eat that day. If he eats all the portions, he will be rewarded symbolically with the display of a smiling face Figure 3. The smiling face represents the approval of the doctor who gave him the instructions; his concrete reward is an improvement in his quality of life.

Case four (fish)

For his weekly consumption of fish, the user may be shown a question, an image, a caption or all three. If he eats all the portions recommended, a funny picture will be displayed to encourage him to follow a healthy diet.

3. Prototype

The prototype of the augmented reality application was developed as part of @Monitech, a Pre Commercial Procurement (PCP) project funded by the Puglia region. It is an android application that allows the user to recognize a medicine from its container. The augmented reality application is activated when a sound is emitted to attract the user’s attention and remind him that he must take a specific medicine. When the alarm sounds on the user device, a popup window - like the one shown in Figure 4 (a) - appears. The user, by clicking on Help me to recognize the medicine, activates the augmented reality application. He then sees a caption on his device reminding him what medicine he should take, as in Figure 4 (b), and frames the medicine containers on his shelf, until he encounters the right one, at which point the application recognizes the medicine container and superimposes a green circle with a tick that provides the user with confirmation that this is the right medicine to be taken at this time (Figure 4 (c)). If the correct medicine is found but it is not yet time for the user to take it, he will see a red circle with a cross, as shown in Figure 4 (d). This feature is useful...
for informing the user that he is not properly following the treatment plan.

![Image](a)

![Image](b)

![Image](c)

![Image](d)

**Figure 4:** This prototype image shows an example of operation system.

### 4. Conclusion

The proposed application attracted the interest of many physicians and biologists involved in the project. They expressed willingness to provide all the necessary support for the application by performing a series of tests designed to assess how easily this new tool might be accepted by users and to evaluate its effectiveness in terms of improving the quality of life of older people through more careful observance of doctors’ advice. The goal is to develop a complete application and distribute it to a target group of individuals who will use it for a period during which we will monitor the benefits and any possible drawbacks. This will allow us to understand how to go about system optimization and also in what other research areas it might - with appropriate modifications - be used. Our long-term aim is to produce an application that does not use marker and to manage the massive amount of data. Since the proposed framework lends itself to subsequent modifications, after the testing phase we believe it will be possible to make it more efficient. It will increase the autonomy of the elderly and contribute to the creation of a more sustainable economic system in a world in which economic resources are increasingly limited.

### References


GOJI an advanced virtual environment for supporting training of physical and cognitive activities for preventing the occurrence of dementia in normally living elderly with minor cognitive disorders

Luca Greci

Abstract

Alzheimer’s disease (AD) is the most frequent cause of dementia. The incidence of dementia in populations of 65 years or over is 9% and doubles every 5 years. Recent epidemic studies predict that in 2020 there will be over 48 million cases of dementia (Italian Health Ministry). Visual–spatial impairments are often among the first symptoms noted in AD and in the course of clinical evolution this function can become severely impaired. Aim of the Goji project is to provide a preventive program, for asymptomatic or early mild symptomatic individuals, through the development of a Virtual Environment for visuospatial (VSP) training and a physiological evaluation of its efficacy.

The paper presents the virtual environment developed, focusing on the choice of the development tools and on the scenario’s development and functionality.


1. Introduction: the Goji project

Cognitive impairment and dementia, due to neurodegenerative disorders, are main manifestations of senescence, the negative process that parallels aging.

Dementia is the third leading cause of death in people over 75 years and the second one in those of 80 or more. In Italy about one million persons suffer from this disease. Different neurodegenerative disorders causing dementias are observed in elderly but Alzheimer’s disease (AD) is the most frequent accounting for 60-70% of all the disorders [SFG11].

The pathogenesis of AD is related to the so-called amyloid cascade, the process of aggregation of beta-amyloid peptides into fibrils. The neuro-pathological hallmarks of AD are the deposition of beta-amyloid into the extracellular brain space and the intra-neuronal neurofibrillary tangles due to the degeneration of the cytoskeleton tau protein [MT10].

The large majority of AD are sporadic but 5-10% of the diseases are familiar and due to autosomic dominant mutations most frequently of the genes of amyloid protein precursor or of presenilin 1 and 2 [SFG11].

The pathological process of AD precedes of decades the clinical severe manifestations of dementia, suggesting that early interventions in asymptomatic or early mild symptomatic individuals might slow the disease progression [SAB_11].

Many factors have been reported to increase the risk of occurrence of sporadic AD and many of them are theoretically modifiable for preventive purposes.

Observational studies have identified a wide range of potentially modifiable risk factors for AD [BY11] and dementia, including cardiovascular risk factors (e.g., hypertension, diabetes, obesity), psychosocial factors (e.g., depression) and health behaviours (e.g., low level of physical or mental activity, smoking).

There are evidences from randomized-controlled trials (RCTs) that few of these factors can affect AD incidence. Physical inactivity seems a risk factor of particular relevance for dementia development/progression due to its prevalence in the population at large and in particular among old people [SSR_07]. Many studies have shown that high level of education, intelligence/QI, and occupational attainment, mentally stimulating leisure activities are associated with a lower risk of dementia and sporadic AD. Recently, a reduction in the incidence of dementia in different birth cohort of elderly in UK has been reported most likely as a consequence of adopting more safe life styles [MAB_13].

Some RCTs have found that healthy, sedentary elders who begin exercise programs experience show significant improvements in cognitive function, particularly mental processing speed [SM13].
Cognitive activity is believed to protect from dementia by increasing the ‘brain reserve’, which refers generally to the capacity of the brain to withstand the effects of pathology by recruiting alternative neurological processes or pathways. Randomized clinical trial have shown that active cognitive stimulation can improve cognition of elderly persons but it is not clear whether this improvement might translate into better performances of Activities of Daily Living (ADL). Measuring the person’s ability to perform ADL is a diagnostic criterion for MCI and AD [TMS13] [OPa12]. The ADL has been classified with the International Classification of Functioning Disability and Health (ICF) and Activities of Daily Living to provide a comprehensive framework of definitions and structures for rehabilitation. Elderly persons not demented with cognitive impairments (minor cognitive impairment [DSM V], MCI mild cognitive impairment, prodromal AD) represent an identifiable group at high risk of dementia and seem a reasonable target for a comprehensive preventive program. Among many possible procedures, cognitive stimulation has demonstrated some preventive effectiveness. Recently, a study showed that cognitive stimulation with video game training could enhance cognitive control in older adults [ABR_13], also simple computer based virtual environments proved to be effective in enhancing cognition of older adults. The use of Virtual Reality (VR) for rehabilitation and neuropsychological assessment has been investigated for over a decade. Several pioneer research groups have already demonstrated improved clinical performance using VR imaging, planning and control techniques [GP00]. VR applications today deal with the challenge of identification of disease and cognitive training of mild cognitive impairment (MCI) and dementia patients, focusing on navigation and orientation, face recognition, cognitive functionality, and other activities of daily living (IADL). VR provides the possibility of performing activities, tasks, and tests in a virtual environment able to change to different characteristics and needs of individual patients [GAF_15]. Virtual reality training has been shown to be an effective, motivating, and safe training tool [MTB_14].

The Goji project is aimed to define a comprehensive preventive program of dementia in elderly with minor cognitive disorders, probably due to neurodegenerative diseases. A cohort of persons impaired in one/two of four cognitive domains (memory, visuo-spatial capacities, language, executive functions) will be identified. A comprehensive preventive program will be tailored on the individual and group characteristics of the cohort. The use of VR coupled with physical exercise enhances the cognitive function more than traditional training and has a great potential for preventing cognitive decline [AAB_12].

A virtual environment (VE) simulating ADL will be designed with the purpose of constructing an effective and easy accessible technological tool for cognitive stimulation. For the Goji, one-year project, the study limited the attention to visuo-spatial capacities, since the impairment in this domain limits physical activity and eventually the activities of daily living.

2. Goji environment

The Goji project tries to answer to both the visual-spatial cognitive training as well as evaluating the impact of physical activities on cognitive impairment. The Goji’s virtual environment is composed by two different tasks simulating ADL: going to the Supermarket (GtoS) and Virtual Shopping (VS). The two tasks can work standalone or in sequence.

In the GtoS scenario the patient performs a session of 30 minutes on a cyclo-ergometer. During the session a virtual environment simulating a route is shown to the patient both to increase the user’s engagement and to administer some sub task as choose where to go (turn left or right) and stop and go at cross. At the end of the session the patient can stop the exercise or enter to the Virtual Shopping.

The VS is a digital environment of a supermarket. Inside the VS, a five products to buy shopping list is given to the user. The user has to face two different tasks: task 1, to find the lane containing the highlighted product of the list by choosing, from the top of each lane, the sign containing the same word of the list (figure 1); task 2 to identify the product on a virtual shelf (figure 2).
bottle of wine) or same objects but with special issue (discount). The levels of the two tasks can be mixed depending on the user’s skills or level of impairment. For both the tasks, a report of the user’s choice and the time to perform the tasks is saved in the form of XML.

3. Scenarios development

3.1. The Going to Shopping scenario

The aim of the GtoS environment is to engage the users while they are on the bike exercise. Each users will perform at least three sessions of 30 minutes every/per week.

To increase the engagement and the sense of presence of the user, an urban route and a path in the park were chosen as scenarios and they were developed most realistic as possible. Both the scenarios have real-time global illumination, the wind affects the movement of the foliage of the trees, the grass and the water. Realistic 3D sounds are present in the scenarios. The objects populating the scenarios (animals, cars) are not static but move along the user.

Due to the high number of trials, each scenario (the park and the city) has been developed with more routes inside the same scenario.

Each path is defined by the system at the beginning of the trial session. The path is made by a set of check points, and the point of view is subjective. A camera moves with the user and looks towards each point of the set. When the route has tight curves the camera moves jerky thus possibly making the user uncomfortable.

For this reason, we are also investigating the use of a third person point of view; in this case the camera will follow from a certain distance and height a 3D model of a bicycle along the path. The movement of the camera will be smoother than with the first solution, but unfortunately the sense of presence will be decreased [DN15].

Two scripts have been implemented to manage the user’s interaction with the scenarios. The first allows the user to follow the scenarios’ path while the other allows the user to read the data coming from the DLL (developed by a partner of the project) to set the bicycle virtual speed, coherently with the cycling speed (figure 3).

3.2. The Virtual Shopping scenario

The Vs scenario wants to reproduce a daily life activity like shopping at the grocery store. The scenario has been developed to be projected on a big surface so as to simulate the real dimension of the objects to choose from the shelf. The choice is driven by the need, coming from the medical team, to provide a good immersion of the user in the environment enhancing the sense of presence.

The application flow is the following: the user or the caregiver creates a user’s profile or logs in. The caregiver, through a GUI, chooses the level for the lane and the shelf task. A shopping list with five object to buy is shown to the user. The user starts the exercise, first he has to find the lane where the highlighted object of the shopping list is available. Then the user has to find the object on the shelf.

The VS scenario is composed by a main script, Scene Manager (SM), which manages the 3D scene and all the events of the virtual environment. The SM calls different scripts in order to retrieve the data required to generate the scenes: data storage, lane generator, shelf generator, speech and error manager (figure 4).

The lane generator script gives to the SM the data to set the 3D scene of the supermarket: number of the lane, the path to follow when the user makes the right choice, the type, position and number of element to put on the signs above the lane.

The shelf generator script provides to the SM the arrangement of the object on the shelf and the type of object to load. The error manager script is called by the SM during the lane and the shelf task to verify the choice made by the user respect to element to be selected. In case of a right answer the script tells to the SM to load a new step otherwise it notifies...
to the SM that an error has occurred and what kind of help gives to the user.

![Diagram showing the Virtual Shopping scripts architecture]

### Figure 4: Virtual Shopping scripts architecture

The speech script is called by the error manager to provide an audio help about the name of the object to choose both in the lane and shelf task.

The user profile, the date, the level chosen both for the lane and shelf task, the time used to perform the exercise and the errors made by the user are saved in the data storage script, on an XML file. This script is called from the SM all the time an action is done.

### 4. Development platform and hardware

Different aspects have been taken into account choosing the development platform as well as the hardware for the interaction between the user and the virtual environment.

Unity [Uni15] was chosen as the development platform considering that the virtual environments had to be as realistic as possible in order to enhance the sense of presence and engagement of the user. This platform provides a set of tools that match with the environments’ needs and manages code created outside Unity in the form of a Plugin. The code is written under the form of script in C#.

Different aspects have been taken into account during the design of the virtual environment: immersion, interaction and sense of presence. Technology selection has been driven by the four-factor solution for presence: physical space, engagement, naturalness and “negative effects” [BBA04].

There are three levels of immersive systems: “fully immersive” (360° information space), “semi-immersive” (the extent of display is less than 360°) and “non-immersive” (desktop pc) [Kal00]. Because one of the requirements of the Virtual shopping is having the environment in scale 1:1, the “non-immersive” system has been discarded. The Head Mounted Display (HMD) and a projected video wall have been compared. In the literature, is reported that the difference between the two systems was in the “negative effects”:

- The use of HMD in an interactive environment may provoke different undesirable effects like dizziness, disorientation and nausea during and up to 60 minutes after use [HVP02].
- For this “negative aspects” the HMD technology has been discarded and the projection system has been chosen as visualization solution for both the scenarios.

For the interaction between the user and the environment, different touch systems have been evaluated: Kinect, Leap Motion, Touchless Touch, mouse, interactive whiteboard and infrared interactive board. Some tests on usability, sense of presence, accuracy and fatigue have been done. As a result, we have decided to use an infrared interactive board. The nature of the interaction (tapping) between the user and the infrared interactive board involves the user’s body enhancing the sense of presence [MV505]. No arm fatigue was reported by the users. The precision in the touch recognition was the same of the mouse. Moreover it works coupled with a short lens projector, the same technology chosen for the visualization, capable to provide an interactive area of 100°.

The final set-up is made by an Ultra-Short Throw Interactive [Eps15] projector for both the scenarios.

### 5. Conclusion and further development

The paper has presented the VR environment developed for the Goji project. The project is still running, the medical equip, partner of the project, is now defining the protocol to administer the VE to a cohort of patients and recruit patients for the validation of the VE. The group will be composed by: 5 subjects for the whole protocol (VE training and movement analysis), 5 subjects just within the VE, and 5 subjects as control group.

An internal test on 20 persons with different ages (20-65) has been performed in order to validate the VE.

The VS has worked has expected, no software’s bugs have been reported during the test. Regarding the usability of the VS, the tests have reported some problems. The users had difficulties to read the shopping list because its position has been changed in the different tasks (lane, shelf). In the task 1, the users tried to find the target object of the shopping list not on the sign above the lane but on the shelf of the 3D environment. The transition between the lane and the shelf task, a camera animation moving the users in front of the shelf, for some users have resulted dizzy.

Some improvement are scheduled to overcome the problems coming from the internal test, before the release of the VE for the final validation.

The GtoS scenario is still under development, in the next two months the modelling task will end. In the same period the DLL to connect the cyclo-ergometer with the environment will be released by the project partner and it will be integrated with the environment.

### Acknowledge

Goji project, Strategic project on Active Ageing, Ministry of Instruction, University and Research.
References


Immersive virtual reality in rehabilitation: a preliminary efficacy assessment on children with acquired brain injury

EMILIA BIFFI1, ELENA BERETTA1, ELEONORA DIELLA1, DANIELE PANZERI1, CRISTINA MAGHINI1, ANNA CARLA TURCONI1, SANDRA STRAZZER1, GIANLUIGI RENI1

1Scientific Institute Eugenio Medea, Bosisio Parini (Lecco), Italy
*Corresponding author; email: emilia.biffi@bp.lnf.it

Abstract
Virtual reality has been consolidated as an effective mean to foster motor and cognitive therapy and rehabilitation in adult and paediatric patients. The Gait Real-time Analysis Interactive Lab (GRAIL) is a dedicated solution for gait analysis and gait training in motivating environments. Few research groups developed applications for the gait treatment in post-stroke patients as well as for the fall risk prevention in amputees and elders while no evidence on its use for the rehabilitation of children is available. This system may enhance therapy outcomes over conventional approaches, particularly with respect to motor learning, thanks to the high plasticity of neuronal circuits in children. As a proof of concept, a pilot study was performed involving 4 children suffering from acquired brain injury (ABI), who underwent a 5 session treatment with GRAIL to improve walking and balance ability. Results are promising: improvements were recorded at the ankle level, selectively at the affected side, and at the pelvic level, while small changes were measured at the hip and knee joints, which were already comparable to healthy subjects. All these changes also conveyed advances in the symmetry of the walking pattern. In the next future, a longer intervention will be proposed and more children will be enrolled to strongly prove the effectiveness of GRAIL in the rehabilitation of children with ABI.

I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism – Virtual reality

1. Introduction
In the last few years, traditional physiotherapy have been flanked and supported by innovative technologies for rehabilitation in the field of robotics and virtual reality (VR). Currently, VR has been consolidated as an effective mean to foster motor and cognitive therapy and rehabilitation in adult patients suffering from stroke [4], Multiple Sclerosis [9] and acquired brain injury [18]. Recently, it has been proven the efficacy of VR devices based on videogames in supporting the conventional rehabilitative treatment of children affected by several diseases (e.g. cerebral palsy, autism, as well as pediatric headache), both at hospital [1, 22, 27] and in home or school environment [15, 20]. Currently, there are few commercial platforms based on VR which can be used to flank and support traditional therapy in clinical institutes. Among them, the CAREN (Computer Assisted Rehabilitation Environment) and the GRAIL (Gait Real-time Analysis Interactive Lab) by Motek Medical (the Netherlands) have shown exciting prospects both as evaluative and rehabilitative systems. The first Italian GRAIL lab has been recently installed at the Scientific Institute Eugenio Medea (Bosisio Parini, Italy) and launched in June 2014, thanks to funds of the Italian Ministry of Health. The GRAIL can be used with different patients affected by both neurologic and orthopedic pathologies. The literature about the clinical use of this system for the evaluation of movement and control strategies is recent and mainly tackles adult subjects. The main studies described in the literature address the response to the VR environment as well as the strategies used for stability maintenance and for perturbation compensation. It has been demonstrated that gait velocity can be modulated by the modification of the optical flow of VR environments [14], that medio-lateral gait perturbations affect kinematics more than antero-posterior ones [17] and that gait velocity, step length and step frequency act on the (un)stability and the resulting fall probability [2, 10-12]. Some researchers are evaluating the comparability between gait analysis performed with GRAIL and traditional overground acquisitions, in both healthy subjects and patients with transtibial amputation [7], and the effect that the VR environment has on gait kinematic and kinetic [23]. Overall these studies demonstrate that the VR environment synchronized with the treadmill guarantees the use of the GRAIL as a system for gait evaluation.
The possibility to generate perturbations using both the treadmill and the immersive VR encourages the development of protocols for gait treatment. Few research groups developed applications which imitate different ground surfaces [16], or specific exergames for the gait treatment in post-stroke patients [6] as well as for the fall risk prevention in amputees and elders [21]. Recent results demonstrate improvements in gait performance, the recovery of trunk movement and the muscular reinforcement in amputees [5, 13]. The use of VR system in pediatric patients is promising thanks to the high plasticity of neuronal circuits in children, but a scientific evidence of its usability and benefit still lacks. The first study which compares the gait evaluation performed in standard laboratory with the one performed with the GRAIL system in both healthy pediatric subjects and Cerebral Palsy subjects was published only few months ago [26]. Moreover, no rehabilitative treatments with GRAIL have been proposed in the pediatric population.

The aim of this study is to preliminary evaluate the efficacy of a short rehabilitative treatment on a GRAIL system in a small group of children affected by acquired brain injury (ABI).

2. Materials and Methods

2.1. The GRAIL system

The GRAIL is a dedicated solution for gait analysis and gait training in motivating environments [23]. It is an integrated platform made up of an instrumented dual-belt treadmill, a two degrees of freedom motion frame and integrated force plates (16 channels, sample frequency 1000 Hz) to capture useful gait data during the trial. The system is equipped with 10 optoelectronic cameras (sample frequency 100 Hz) for kinematic data acquisition, a motion-capture system and 3 video cameras. The integration with synchronized VR environments which are projected on a 180° cylindrical projection screen allows the subject to walk and move in natural and attractive settings. The motion frame can translate in longitudinal and lateral direction to assess compensatory strategies and investigate dynamic stability. Moreover, the self-paced functionality of the treadmill is designed to simulate a more realistic walking environment in which the subject determines his own walking speed. The system automatically adjusts the speed of the treadmill and of the VR scenery to the walking ability of the subject allowing more realistic gait training in a safe environment.

GRAIL acquires kinematic and kinetic data and then processes them in real time by means of the 25-marker Human Body Model [25]. This feature gives the possibility of using gait and movement parameters directly within virtual games and for the creation of multisensory stimulations. Furthermore, it makes available useful information during the session for a personalized training of static and dynamic postural control. Real-time filtering is performed with a 2nd order Butterworth filter, with a cut-off frequency equal to 6 Hz and then gait traces are saved in a csv file. It is possible to compute the mean and standard deviation of every gait parameters by acquiring many steps during the trial, which also extremely reduces data acquisition and processing time.

2.2. The D-Flow software

The whole system is controlled by the D-flow, a software that oversees the relationship between the subject, the scenario and the interactive feedbacks and stimulations.

The D-Flow runs on Microsoft Windows and it is designed for the development of interactive and immersive virtual reality applications, for the purpose of clinical research and rehabilitation. The subject is a central part of a real-time feedback loop, in which multi-sensory input devices (e.g. motion capture systems, force plates and electromyography) measure the behavior of the subject, while output devices (e.g. motion platforms, treadmills, audio devices and displays) return motor sensory, visual and auditory feedback to the subject. The system operator defines feedback strategies through a flexible and extensible application development framework, based on visual programming. The D-Flow programming, indeed, is based on the concept of modules, that are components with a specific functionality, which can be combined to create complex, interactive virtual reality applications. Some modules directly control specific hardware devices, such as a treadmill or a motion base. Other modules provide access to real-time data streams from live input devices. Others manipulate virtual objects or detect collisions between them, thus allowing the interaction between the subject and the virtual environment. Finally, D-Flow also contains a general purpose scripting module and a module for expression parsing. In addition to data-based communication, the D-Flow kernel framework allows for event-based communication between modules. The operator can define a set of global events, which can be broadcast by modules at specific occurrences, while each module exposes a set of module actions that affect the behavior of the module in a specific way, enabling maximum flexibility in event-based communication [8].

2.3. Participants

Participants to the study group (SG) were recruited from the in-patient setting of the Acquired Brain Injury Unit, Scientific Institute IRCCS Eugenio Medea, Bosissio Parini, Italy. The principal selection criteria were as follows: diagnosis of ABI in paediatric age and adolescence (2–20 years); severity of motor impairment equal to level I and II, classified according to Gross Motor Function Classification System (GMFCS) [19]; ability to follow the instructions. The exclusion criteria were as follows: severe muscle spasticity and/or contracture, a diagnosis of severe learning disability, behavioural problems, and visual or hearing difficulties that would impact on function and participation. Concerning the control group (CG), we used reference gait data about 10 young healthy adults which were previously enrolled in a study on GRAIL (data not yet published).

The study protocol was approved by the Ethics Committee of Scientific Institute, IRCCS Eugenio Medea, Bosissio Parini, Italy. Written consent was obtained from the parents of each patient.
2.4. Rehabilitation protocol

The rehabilitation protocol included 5 30-minute sessions on GRAIL, twice a week, for a total duration of 3 weeks. The treatment included exercises to improve walking and balance ability in engaging VR environments (e.g. skiing and sailing by changing posture and balance, walking in a forest). The first (T0) and the last (T1) sessions were partially devoted to the kinematic and kinetic evaluation. The procedure required a 5 minute adaptation period, then subjects were asked to walk at fixed-speed and in self-paced configuration for 2 minutes each condition. Gait analysis was performed barefoot.

Reference kinematic and kinetic data were collected in a single session where the healthy subjects were asked to walk at fixed-speed and in self-paced configuration for 10 minutes each setting, after a 5 minute familiarization period; consecutive measurements were performed.

Gait parameters describing spatio-temporal information, kinematic and kinetic data were extracted as previously done [24, 26]. Mean and standard deviation were computed across the collected steps.

The paired non-parametric Wilcoxon test was performed to compare gait data of patients before and after the rehabilitation treatment and to compare the left and the right side. The Mann-Whitney test was used to evaluate changes of patients’ data with respect to healthy subjects. The significance level was established at $p<0.05$.

3. Results

Considering inclusion and exclusion criteria, 4 children with ABI (one boy and three girls) were recruited (mean age 13.7 ±3.7 years old) in the SG. Patients’ characteristics and clinical information were collected during the study recruitment (see Table 1 for same details). Clinical information comprised etiology, age at injury, previous neurosurgery, type of motor impairment, severity of motor impairment classified according to Gross Motor Function Classification System (GMFCS) [19]. All the patients were affected by mild motor impairment, with a prevalence of left hemiplegia (three patients). CG was composed of 10 young healthy adults (mean age 26±1.7 years old, 1 male, 9 females).

All the participants completed the rehabilitation program. Changes in their walking pattern were detected and all of them tended towards healthy subjects’ features. Moreover step-by-step variability (i.e. the standard deviation of gait data across steps) was reduced after the intervention. The biggest improvements between T0 and T1 were at the ankle level: data report significant improvements of the ankle joint dorsiflexion in stance (symbol * in Figure 1, left) and a positive trend of the maximum plantar-flexion in swing (Figure 1, right) at the left side. These changes made SG kinematic data, that initially differed from the CG (symbol # in Figure 1), comparable to healthy subjects at T1. Moreover the initial lack of walking symmetry (mark x in Figure 1) was restored after the intervention with GRAIL. From a kinetic point of view, the left ankle showed a promising increase of the maximum power at the left side (T0 0.71 (1.13) vs T1 1.28 (1.04) W/kg, symbol * in Table 2) even if it remained significantly different from healthy subjects (2.63 (1.22) W/kg, symbol # Table 2) and asymmetrical with respect to the right side (symbol x in Table 2).

No significant changes were observed in the kinematic data at the hip and knee level which were already comparable to healthy subjects, while significant improvements in the maximum values of the moment in extension and of the power in flexion at the hip were observed, thus reducing walking asymmetry. In contrast, no changes were measured in the kinetic at the knee level.

Finally the mean pelvic tilt and its value at the IC significantly improved after the treatment, thus getting comparable to the CG. The pelvic obliquity range of motion (ROM) also significantly improved but still remained smaller than the one of the controls. All these results are reported in Table 2.

4. Discussion

The prolonged rehabilitation of children affected by ABI takes advantage of new devices which can engage young patients. Indeed, interventions based on robotics have been previously demonstrated to convey improvements of hip kinematics in these subjects [3]. GRAIL is an integrated technological platform which has been validated to be a reliable device to perform movement evaluations in the paediatric population [26].

Its use with a rehabilitation perspective is still under investigation and no data involving children are available at our knowledge. In this work we demonstrated that a short training of walking and balance on a GRAIL system of children suffering from acquired brain injury is effective. Indeed, our results showed changes in kinematics and/or kinetics at the ankle, hip and pelvic level. These changes were in the direction of the healthy subjects, thus suggesting benefits of this new platform for the walking recovery.
**TABLE II. REHABILITATION TREATMENT OUTCOMES.**

<table>
<thead>
<tr>
<th>Gait parameter</th>
<th>side</th>
<th>SG T0 *</th>
<th>SG T1</th>
<th>CG</th>
<th>T0 L vs R</th>
<th>T1 L vs R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max ankle flexion power (W/kg)</td>
<td>L</td>
<td>0.71 (1.13) #</td>
<td>1.28 (1.04) #</td>
<td>2.63 (1.22)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>0.72 (0.38) #</td>
<td>0.74 (0.12) #</td>
<td>1.17 (0.92)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range of hip flexion (°)</td>
<td>L</td>
<td>41.69 (3.62)</td>
<td>42.93 (3.48)</td>
<td>42.20 (6.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>39.47 (6.01)</td>
<td>39.96 (5.69)</td>
<td>43.30 (6.95)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max hip extension moment (Nm/kg) b</td>
<td>L</td>
<td>0.27 (0.16) #</td>
<td>0.37 (0.19) * #</td>
<td>0.54 (0.24)</td>
<td>1.11 (0.54)</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>0.34 (0.14) #</td>
<td>0.42 (0.14) *</td>
<td>0.50 (0.26)</td>
<td>1.17 (0.92)</td>
<td></td>
</tr>
<tr>
<td>Max hip flexion power (W/kg)</td>
<td>L</td>
<td>0.69 (0.32) #</td>
<td>0.72 (0.18) #</td>
<td>1.11 (0.68)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>0.72 (0.38) #</td>
<td>0.74 (0.12) #</td>
<td>1.17 (0.43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range of knee flexion (°)</td>
<td>L</td>
<td>60.63 (13.71)</td>
<td>63.79 (7.63)</td>
<td>65.51 (2.98)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>66.07 (7.25)</td>
<td>66.06 (5.02)</td>
<td>65.89 (3.38)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max knee flexion power (W/kg)</td>
<td>L</td>
<td>0.48 (0.64)</td>
<td>0.63 (0.38)</td>
<td>0.70 (0.68)</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>0.44 (0.15) #</td>
<td>0.42 (0.11) #</td>
<td>0.75 (0.43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pelvic tilt at initial contact (°)</td>
<td>L</td>
<td>20.71 (47.04) #</td>
<td>10.34 (6.54) *</td>
<td>10.94 (7.56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>20.58 (46.32) #</td>
<td>11.51 (8.15) *</td>
<td>10.75 (7.99)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean pelvic tilt (°)</td>
<td>L</td>
<td>21.19 (46.42) #</td>
<td>11.35 (6.06) *</td>
<td>10.39 (7.34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>21.19 (45.99) #</td>
<td>11.40 (5.99) *</td>
<td>10.40 (7.29)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range of pelvic obliquity (°)</td>
<td>L</td>
<td>7.02 (2.21) #</td>
<td>8.22 (2.09) * #</td>
<td>11.02 (1.87)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>7.20 (2.66) #</td>
<td>8.37 (2.08) * #</td>
<td>11.05 (1.98)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Median and interquartile range. SG refers to patients, CG to healthy subjects. T0 and T1 are the measurements before and after the training respectively. * identifies significant differences between patients at T0 and at T1 (p<0.05, Wilcoxon test). # indicates significant differences between patients and healthy subjects (p<0.05, Mann-Whitney test). x in the last two columns denotes significant differences between the left and the right side (p<0.05, Wilcoxon test).

b. Extension is here reported with positive values.

**Figure 1:** Peak ankle dorsi-flexion during stance (left) and peak ankle plantar-flexion during swing (right). The x symbol identifies differences between the left and right leg (p<0.05); the * symbol defines differences between SG at T0 and SG at T1 (p<0.05); the # symbol identifies differences between SG and CG (p<0.05).

The main improvements were recorded at the ankle level, selectively at the left side. Therefore the intervention induced changes at the most impaired side (three over four patients suffered from left hemiplegia), also conveying improvements in the symmetry of the walking pattern. Changes were also measured at the pelvic level, which moved towards the controls. Small changes were observed at the hip and knee joints, which can be justified by the mild motor damage of the enrolled patients. These data, where selective changes at the affected side are recorded, also make us confident that, beside a predictable learning effect due to the adaptation of patients to the platform, GRAIL conveys a strong and clear rehabilitative effect due to relearning.
The main drawbacks of this study are the size of the treated group, which will be enlarged in the next months, and the features of the control group, which was not age-matched. However, the study aimed to be a preliminary explorative evaluation of GRAIL efficacy on children rehabilitation, which at our knowledge lacks in the literature. Moreover, the enrolled patients were mainly adolescents (3 over 4 patients) and therefore with a gait pattern similar to young adults.

To conclude, these promising results suggest that children with acquired brain injury may take advantage of a rehabilitation therapy involving GRAIL, and that longer intervention may have even bigger improvements.

References


ZebraX: sonification of road crossing guidance data for visually impaired individuals

Lorenzo Picinali1, Andrea Gerino2, Sergio Mascetti2, Dragan Ahmetovic2 and Cristian Bernareggi2

1Dyson School of Design Engineering, Imperial College London, SW7 2AZ, UK; l.picinali@imperial.ac.uk
2Università degli Studi di Milano, Dept. of Computer Science, Via Comelico 39, 20135 Milano, IT

Abstract
In the last years several solutions were proposed to support people with visual impairment in road crossing; the focus is on the computer vision technique that recognises the pedestrian crossing and that computes the relative position between the user and the crossing. This contribution addresses instead a different problem: the design of an auditory interface that effectively guides the user during pedestrian crossing. Two original auditory guiding modes based on data sonification have been developed compared with a guiding mode based on speech messages. A demo of all three guiding modes will be given at the conference venue.

Categories and Subject Descriptors (according to ACM CCS): H.3.3 [User Interfaces]: Auditory (non-speech) feedback—K.4.2 [Social Issues]: Assistive technologies for persons with disabilities—H.5.1 [Multimedia Information Systems]: Audio input/output—

1 Introduction
Mobile devices provide new exciting opportunities to people with Visual Impairments or Blindness (VIB). Indeed, most commercial devices (e.g., iOS and Android) are accessible to users with VIB.

On one side, this allows people with VIB to use most of the applications available on mobile devices like, for example, web browsers and email clients. On the other side, accessible mobile devices can be used to implement assistive technologies, with great advantages for both developers and users.

This demo presents ZebraX, an iPhone application that adopts a state-of-the-art algorithm to detect zebra crossing, and guides individuals with VIB to cross the road. Guidance information is delivered through three guidance modes: one based on text-to-speech instructions, and two based on data sonification. The latter are similar, with the main difference being that one produces mono sound (i.e. one single sound signal) and the other produces stereo sound (i.e. two different sound signals, one for the left and one for the right ear). From the applicative point of view, a major difference can be noted; stereo sonification requires the user to wear headphones, while mono sonification can also be reproduced from the internal speaker on the device itself.

The topics presented in this demo have been recently published in [Ahmetovic, 2014] and [Mascetti, 2015].

2 ZebraX architecture

ZebraX is divided into three main modules: Recognizer, Logic and Navigator. The Recognizer module implements the ZebraRecognizer algorithm [Ahmetovic, 2014] which computes the relative distance between the user and the zebra crossing. The Logic module uses the positioning data computed by the Recognizer module to compute the messages that are to be conveyed to the user. A total of 7 messages are computed about the relative position of the crosswalk: ‘rotate left’, ‘rotate right’, ‘step left’, ‘step right’, ‘not found’, ‘crosswalk ahead’, and ‘cross’. Two additional messages are also employed for helping the user to hold the device in the correct position: ‘rotate up’ and ‘rotate down’. These two messages are related to the ‘vertical rotation angle’ (i.e., the device pitch angle), computed by the Logic module through the accelerometer data. Finally, the Navigator module is in charge of the interaction with the user.
by acquiring input through the touchscreen, and delivering audio and haptic (vibro-tactile) feedback.

3 Auditory guiding modes

Employing a user-centered approach, three auditory guiding modes have been implemented: speech, mono and stereo; the first based on speech messages, the second and third based on mono and stereo sonification, respectively.

3.1 Speech guiding mode

Referring to the instructions computed by the Logic module, the Navigator module delivers to the user a set of speech messages generated by the iOS on-board text-to-speech synthesizer. Since the subjects who participated to the evaluation were all Italian mother-tongue, the messages were delivered in Italian (English translation between brackets).

- Abbassa/alza il dispositivo (Rise/lower the phone)
- Ruota a sinistra/destra (Rotate left/right)
- Passo a sinistra/destra (Step left/right)
- Non trovato (Crosswalk not found)
- Strisce davanti (Crosswalk ahead)
- Attraversa (Cross)

Each message is reproduced once, as soon as the Logic module computes an instruction different from the previous one.

3.2 Sonification guiding modes

Sonification is the use of non-speech audio to convey information. For ZebraX, a parameter mapping sonification approach was employed based on the creation of a link between the data to be rendered and the parameters of a synthesizer (or of any other device which generates or plays back sound).

Two solutions have been designed: the mono sonification delivers one monaural audio signal, which is suitable to be played by the device speaker. Vice-versa, the stereo sonification employs sound spatialization in order to allow the user to clearly perceive certain sounds as coming from the left or from the right, therefore to convey information using an additional cue. Left-right spatialization has been achieved through modifying the inter-aural (i.e. between the ears) level and time differences.

Both mono and stereo modes employ highly-localizable impulsive sounds. Modifying their pitch, repetitions, location and volume, a set of auditory signatures was created in order to deliver each of the 7 instructions (see Section 3.1) to the user. As an example, in order to deliver the Rotate left/right instruction, an impulsive sound with fast transients and in-harmonic spectrum (similar to a percussive sound on metal) is employed. For the stereo mode the frequency of the sound is fixed at 500 Hz, and the left-right information is delivered spatializing the sound on the left or on the right, accordingly. For the mono version, the left-right information is delivered modifying the frequency of the sound: 300 Hz for the left rotation and 1200 Hz for the right rotation. For both modes, the repetition rate of the sound is modified linearly from 1.6 Hz (large rotation) to 3.3 Hz (small rotation), varying continuously until the user reaches the target angle.

More information about the sonification modes can be found in [Mascetti, 2015].

4 Evaluation

An extensive evaluation stage was performed, involving 11 sighted (blindfolded) and 15 visually impaired individual. The evaluation included both a quantitative (road crossing task) and a qualitative (questionnaire) stage. The experimental results show that the ZebraX prototype can effectively guide people with VI in road crossing with any of the three auditory guiding modes. Most test subjects (75%) declared to prefer one of the two sonification guiding modes with respect to the speech mode. This result supports the usefulness of the two sonifications, also considering that the test subjects preferred the sonifications despite these being less immediate to use. At the same time, no guiding mode was found to be best suited for all test subjects. While the mean align and cross time for all subjects was slightly shorter for the speech guiding mode (not statistically significant), a large number of test subjects were faster to align and cross with the sonification guiding modes.

5 Future works

The results of this study allow to start working along several directions in research and development. The three guiding modes find their direct application in commercial software to support road crossing. Further investigations should be carried out on the effects of training on test subjects’ performances, in particular using the sonifications. Another possible further development could consist in the design of a new guiding mode incorporating both speech and data sonification for rendering different messages and quantities.

6 Demo

During the ZebraX demo, blindfolded visitors will be invited to complete the road crossing task using the application, and choosing between one of the three auditory guiding modes. In order to limit the possibility of hazards, the task will be completed indoor, using a plastic sheet with a 1:1 representation of a zebra crossing. An assistant will be on the side of each visitor at all time.

References


A Comparative Study on High and Low Cost IMU Drift for an Augmented Reality Application

Lakshmiprabha N. S.1, Alexandre Santos1, Olga Beltramello1

1European Organization for Nuclear Research, CERN, Switzerland
2University of Rome Tor Vergata, Rome, Italy

Abstract
An Augmented Reality (AR) application greatly depends on the pose estimation algorithms to know the desired object location in a real world scene. It is also very important to have a jitter-less and smooth virtual content rendering on the AR display device. Keeping the intent of accuracy and fast pose estimation in mind, the camera and the Inertial measurement unit (IMU) are fused using Extended Kalman filter (EKF). However, the fundamental problem with an IMU is that it drifts over time and this greatly depends on the cost, size and weight of an IMU. The commercially available head mounted displays (HMDs) have in-built IMUs that are hobbyist (low cost) grade with high frequency noise and low frequency drift. The objective of this work is to analyse the registration error caused by the high and low cost IMU drift in the context of an AR application. This study will give us an idea on the registration error which can be considered in future to develop algorithms and techniques that will effectively diminish the jitter in rendering. The preliminary results obtained under static and simulated head (pan and tilt) movements show that the low cost IMU drifts more than the high cost IMU.

Categories and Subject Descriptors (according to ACM CCS): Information Interfaces and Presentation [H.5.1]: Multimedia Information Systems—Artificial, Augmented, and Virtual Realities

Keywords: Augmented reality, Pose estimation, Inertial measurement unit, Sensor fusion, Extended Kalman filter, Registration error, Head mounted displays.

1. Introduction
In an Augmented Reality (AR) application, it is critical to know the position and orientation of an object in the scene in order to overlay the virtual AR content at the correct location. Humans do this pose estimation using a vision and vestibular system [CLD07]. It is possible to mimic the human vision and vestibular system using a camera and an inertial measurement unit (IMU) for an AR application. Although a camera based vision algorithm gives an accurate pose estimation, it is computationally expensive and slow whereas the IMU gives inertial (angular velocity and acceleration) data at a faster rate but it is noisy. Hence the fusion of these two sensors can complement each other for accuracy and speed. The major problem with the IMUs is the accumulated noise over time that results in drift and this can cause registration errors in an AR application. This error has to be kept very low for better perception of information conveyed using virtual augmentation as this has a direct relation with the usability of AR technology.

One way to overcome the IMU drift is by periodically correcting it with an accurate vision algorithm pose using sensor fusion techniques [HSea06]. Most available research work uses Xsens IMU [HSea06, MMea12, BS08, GSea09] for demonstrating the sensor fusion results for different AR applications. However, the Xsens IMU costs equally or higher [Xse] than the commercially available optical see-through head mounted displays (OST-HMDs). Xsens has many inertial sensing products such as MTi-10 costs €900/unit, MTi-100 costs €1400/unit and MTi-30 costs €1800/unit, whereas the OST-HMD namely Moverio BT-200 from Epson costs €699/unit [Eps], Lumus DK-40 costs $2000/unit [Lum] and Vuzix Wrap 1200 DX AR costs £1299/unit [Vuz]. The IMU in this OST-HMD cannot be compared in terms of cost and quality to the Xsens IMU (MTi-10 for example). This also complicates the assumption of the research work in [HSea06, MMea12, BS08, GSea09] as possible fusion techniques and its performance becomes difficult to achieve for an AR application using OST-HMD. Therefore, it is very
important to compare and analyse the effect of high and low cost IMU drift in terms of the registration error. These results will help to better design the fusion techniques for low cost IMUs which can also be adopted for AR applications using OST-HMD. Thus the main focus of this paper is to investigate and compare the registration error caused by the high and low cost IMU drift. The inconsistency or misalignment in rendering the AR virtual content as a result of IMU drift is measured under static and controlled movements using a pan and tilt set-up. Similar study was performed in [MMea12] where they used the tripod for spherical motion to quantitatively analyse their fusion method. Whereas, we use the pan and tilt set-up to simulate the controlled head movement at different speeds. In this work, Xsens MTi-10 and CHRobotics UM6 are used as the high and low cost IMUs and we share the preliminary results in terms of registration error from these IMUs.

The above mentioned fusion of camera and IMU data is used for AR in a maintenance application in an extreme Atlas particle physics detector environment. The implementation details of the basic building blocks required for this study are explained in section 2. Further, the preliminary results and the quantitative evaluation of this registration error are discussed in the experimental results (see section 3).

2. AR Registration Error
Augmented reality is a user centric technology, where every application is evaluated using user experiences. The major requirement in an AR application is fast reactive response to the user movement and minimal registration error (misalignment) between the AR virtual content and the desired real-world object. The lag between user movements and virtual content updates on a display device can be minimized by using different sources of information to know the desired object pose. Thus, sensor fused information from different sources like camera and IMU [HSea06] can result in fast, accurate, robust and reliable pose estimation. The noise introduced in the sensor measurement can also result in the registration error. Typical examples are noisy or blurred images from the camera and the drift developed in IMU data over time.

The quality of the IMU data greatly depends on the cost, size and weight. In the field of robotics [Hcea10], they categorised IMUs in four major groups namely, navigation, tactile, industrial and hobbyist grade IMUs. In these groups, industrial and hobbyist grade IMUs are possibly suitable for an AR application. Due to the size, cost and weight the commercially available OST-HMD includes the hobbyist grade IMUs. These low cost IMUs data are more susceptible to noise and increased drift than the high cost IMUs. The objective of this paper is to study the effect of IMU drift from a high cost Xsens MTi-10 and a low cost CHRobotics UM6 IMU in an AR application. The implementation of different modules required for the quantitative analysis of registration errors caused by the IMU drift is explained in the next section.

2.1. Pose Estimation and Sensor Fusion
The basic building blocks for evaluating the registration errors are 1) pose estimation from camera, 2) Camera-IMU spatial and temporal calibration, 3) sensor fusion and 4) AR virtual content rendering. Any or all of these above mentioned blocks can introduce the registration error such as the accuracy of pose estimation from camera images, incorrect camera-IMU calibration, parameters in sensor fusion and lag in rendering the virtual content. Proper implementation and tuning of the above mentioned blocks is a must for measuring the registration error introduced by the IMU drift.

The pose estimation from camera images is performed using marker tracking from ubitrack library [NWea04]. This accurate marker tracking pose can serve as a ground truth for computing the registration error. As there are two set-ups (as shown in Figure 1(a) and 1(c)), under static conditions single marker tracking is enough to measure the pose whereas in the controlled pan and tilt movement, we use multi-marker tracking to accurately estimate the pose.

There are different methods proposed for the camera-IMU calibration purpose, namely the methods in [TL89, LD07, JDH08, PF13]. In this paper, the calibration was performed using [PF13], as this method does not require any extra hardware or manual intervention. A minute error in the calibration parameters can create a huge deviation in rendering AR virtual content. The spatial and temporal calibration parameters obtained from [PF13] are used for transforming the IMU data from a body coordinate to a camera coordinate system.

This work falls under the inside-out tracking category [MMea12] and the camera and IMU data are fused using Bayesian estimation framework classically called as Extended Kalman filter (EKF) [BW01, HSea06, NSL15]. We use a tightly coupled approach as detailed in [MMea12] for fusing data in contrast to the loosely coupled approach due to its better performance. The fusion and noise parameters are tuned to perfection and kept constant for both the MTi-10 and UM6 IMUs.

AR virtual content rendering was done using OpenGL implementation in ubitrack library [NWea04]. The rendering is very fast since, there is only a primitive virtual coordinate frame (see Figure 1(c) and 1(d)). All these blocks are optimized to perfection and the same procedure is used for both the IMUs. The only difference in the two IMUs experiments is the data acquisition from the sensors which was performed using the API provided from the manufacturers. Thus the only source for the registration error is the IMU drift over time which can be measured by comparing with the ground truth marker tracking pose as detailed in the experimental results section 3.
3. Experimental Results

The experiments were carried out using a logitech C920 camera, a Xsens IMU MTi-10 [Xse], a CHRobotics UM6 [um6] and a FLIR pan and tilt set-up. In order to quantify the registration error developed due to the IMU drift, accurate marker tracking from the camera images is used as a ground truth pose. At time t₀, the pose from the marker tracking is given as an input to the sensor fusion. At time t₀+x, the pose from the sensor fusion is computed from the marker tracking pose at t₀ together with the IMU data at t₀+x. The registration error calculated using this sensor fused pose and the ground truth marker pose at t₀+x (used only for comparison and not for fusion) is investigated under static and dynamic conditions.

As shown in Figure 1(a), the camera and the IMU were mounted on a tripod and the center marker (highlighted with yellow square) was used as the ground truth for static conditions. The drift developed in the IMU is clearly evident in the Figure 1(b), where the virtual coordinate frame (in RGB) has moved from the marker center to the marker corner. In Figure 2(a) and (b), the ground truth marker pose (RGB) along with the IMU data fused pose (CMY) at different time periods (500ms, 1s, 2s, 5s, 10s) from the MTi-10 and UM6 are shown respectively. These figures clearly indicate that the registration error produced is different in the two different sensors and it is more in the case of UM6 IMU (low cost). The quantitative analysis of the error measured is discussed later in this section.

Furthermore, to analyse the effect of the IMU drift under controlled movements, a pan and tilt set-up was used as shown in Figure 1(c) to simulate the head movements. This set-up was configured in a monitoring mode which spins from left to right (pan) for a prescribed number of times and returns to the home position. It was programmed to spin at different speeds in terms of positions per second (P/S). Since the set-up is panning and tilting, multi-marker tracking that computes poses with respect to the center marker (i.e. all other markers are related and bundle adjusted to the center marker) is used as a ground truth (see Figure 1(d)). Similar to a static test, one marker pose at t₀ was used to fuse with the IMU data and then the pose is calculated only using the IMU data continuously for 10 seconds. This fused pose from IMU data over time is compared with the multi-marker tracking pose at a particular time to measure the registration error introduced by the IMU drift. This error can be clearly noticed in Figure 1(d) that the virtual coordinate frame has wobbled significantly from its ground truth pose. The measured registration error at 500 and 2000 P/S using MTi-10 is shown in Figure 3(a) and 3(b) respectively and using UM6 is shown in Figure 3(c) and 3(d) respectively. In the Figure 3, the green plot is from the multi-marker tracking, red is from the sensor fusion and the black lines give the error in three axis with respect to the ground truth. The panning speed beyond 2000 P/S causes motion blur to the camera images and hence the multi-marker tracking pose becomes erroneous. For this rea-
Figure 3: The IMU drift and its impact on registration error, (a) and (b) Controlled pan movement at 500 and 2000 P/S using Xsens MTi-10, (c) and (d) Controlled pan movement at 500 and 2000 P/S using CHRobotics UM6.

Figure 4: Minimum, maximum and mean registration error measured under static and simulated head movement using (a) Xsens MTi-10 and (b) CHRobotics UM6.

Table 1: The standard deviation of error measured along three axis using MTi-10 and UM6 IMU at different speed over 10 seconds.

<table>
<thead>
<tr>
<th>Speed (P/S)</th>
<th>MTi-10</th>
<th>UM6</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Y</td>
<td>Z</td>
</tr>
<tr>
<td>Static</td>
<td>9.6</td>
<td>17.6</td>
</tr>
<tr>
<td>500</td>
<td>19.6</td>
<td>2.4</td>
</tr>
<tr>
<td>1000</td>
<td>18.9</td>
<td>8.9</td>
</tr>
<tr>
<td>2000</td>
<td>45.0</td>
<td>16.2</td>
</tr>
</tbody>
</table>

The maximum and minimum error range measured over 10 seconds under static and dynamic conditions using MTi-10 and UM6 are shown in Figure 4(a) and Figure 4(b). These plots quantitatively show the increase in error from static to controlled head movements at different speeds for each IMU type. Also, the increase in error from high to low cost IMUs is illustrated from these plots. Further, the standard deviation of the registration error from MTi-10 and UM6 IMU was calculated along the three axis and it is presented in Table 1. The registration error along X and Z axis in case of MTi-10 IMU is increasing with speed whereas in case of UM6 it is more fluctuating due to the random noise. The standard deviation of error along Y axis from UM6 IMU is approximately 10 times that of the MTi-10 IMU. Also, there is a decrease in error along Y axis in both the IMUs from static to controlled movement of 500 P/S. This could be due to the panning motion of the set-up, the drift got compensated. Further at increased speed, the registration error along Y axis is also increasing eventually. Nevertheless, a point to be noted is that the registration error caused by the IMU drift at static condition can be perceived much more easier by the user than during the head movement (say at 500 P/S). As a result of this, Xsens MTi-10 IMU is very well suitable for AR application then the low cost IMUs. For AR on OST-HMD with low cost IMUs need a special data processing and fusion technique which will reduce this effect of drift. Additionally, a careful evaluation at each stage such as camera pose estimation, spatial and temporal calibration between camera-IMU, IMU data processing and sensor fusion has to be considered for jitter-less AR applications.

4. Conclusion

In this paper, the high and low cost IMU drift with its effect in terms of registration errors in the context of an AR application was analysed. The experiments were carried out both under static and simulated head movements. The registration error was steadily increasing from static to controlled movements at different speeds and in case of UM6 it is more random due to noise. The preliminary results also indicate that the low cost IMUs drift atleast ten times more than the high cost IMUs. Hence it is essential to design sensor fusion methods for the low cost IMUs to minimise the registration error.
error for its application in AR field. Further, in future the option of using an adaptive filtering technique to reduce the low cost IMU drift before sensor fusion step can be considered.

Acknowledgements

The authors wish to thank all other members of the EDUSAFE consortium. This research has been supported by the European Commission FP7 Programme under contract number PITN-GA-2012-316919-EDUSAFE.

References


Distributed Architecture of Secure Semantic AR Services

Adam Wójtowicz, Rafał Wojciechowski, Dariusz Rumiński, Krzysztof Walczak

Department of Information Technology, Poznań University of Economics, Poland

Abstract

This paper describes a new approach to building ubiquitous secure augmented reality environments using semantic techniques. The approach, called USARE – Ubiquitous Semantic Augmented Reality Environments, enables creation of AR environments consisting of AR presentations, which are dynamically composed in real time based on multiple distributed data sources and the usage context. Creation of AR presentations in open architectures, where new services can be dynamically added by third parties, raises serious security concerns related to access control and users’ privacy. To address this problem, a generic architecture of secure semantic AR services and a new access control protocol, which enforces appropriate usage of AR content according to semantically described access policies, are proposed.

Categories and Subject Descriptors (according to ACM CCS): [Security and privacy]: Software and application security—Domain-specific security and privacy architectures; I.3.2 [Computer Graphics]: Graphics Systems—Distributed/network graphics; I.3.6 [Computer Graphics]: Methodology and Techniques—Graphics data structures and data types

1. Introduction

Augmented reality (AR) technology enables superimposing computer generated content, such as interactive 2D and 3D multimedia objects, in real time, on a view of real objects [Woj12]. Widespread use of AR technology has been enabled in the recent years by remarkable progress in consumer-level hardware performance, in particular, in the computational and graphical performance of mobile devices and quickly growing bandwidth of mobile networks. Since in AR systems, synthetic content can be superimposed directly on a view of real objects, AR offers a powerful tool for visualization of different kinds of contextual information. Education, entertainment, e-commerce and tourism are examples of application domains, in which AR-based systems become increasingly used [WC13].

Existing AR platforms support mainly two forms of augmentation: directional augmentation – based on relative geographical position and orientation of the user’s device and recorded coordinates of specific points of interest, and image-based augmentation – based on image matching and tracking. Image-based augmentation is more powerful as the synthetic content is directly aligned with a view of real objects. However, due to limitations of the available image matching algorithms, image-based AR applications are built independently for specific purposes, which reduces the number of images that can be matched. The possibility of sharing AR content between such applications is limited. Yet, taking into account the diversity of application domains and information, which can be presented using AR technology, the most promising are open social environments, in which different kinds of augmentation content can be contributed by different users and providers. In such AR systems, AR presentations can be created dynamically based on the available data sources and the current context, through selection of data and automatic composition of AR scenes.

To enable creation of open social AR environments, the concept of Ubiquitous Semantic Augmented Reality Environments (USARE) is proposed. USARE is a new approach to building AR presentations, which combines the advantages of both directional and image-based augmentation. In USARE, AR presentations use image-based augmentation, but they are dynamically composed in real time based on the current context and multiple distributed data sources. Contextual creation of AR presentations enables access to a variety of data sources, guarantees scalability and provides for seamless operation. In USARE, AR presentations are accessed through mobile devices equipped with a camera and are visualized using a single mobile application, called USARE Browser, which is capable of presenting rich image-based augmentations coming from different service-providers.

USARE presentations are dynamically composed based on the user’s context (data describing user’s preferences, privileges, location, time, device’s capabilities) and four types of semantically described elements. The first is a set of visual markers (trackables) representing the real world objects that can be augmented. The second are content objects representing 2D and 3D multimedia content to be presented in relation to the markers. The third are business
data to be presented in AR presentations. The last one is a description of scenarios of presentations indicating the composition of AR scenes and interaction available to a user. These elements are in general independent and may be offered by different service providers in a distributed architecture. Discovery and matching between the elements of AR presentations are possible based on semantic web techniques.

Development of distributed AR presentations based on open architectures such as USARE, where new services can be dynamically added by third parties, triggers synergy related to integration of federated communities of users and service providers, but at the same time raises serious security concerns related to access control and users’ privacy.

In this paper, we describe a generic architecture of secure semantic AR services and an access control protocol which enforces appropriate usage of AR content according to semantically described access policies. The remainder of this paper is structured as follows. Section 2 presents the current state of the art in AR content security. Section 3 introduces the concept of USARE – Ubiquitous Semantic Augmented Reality Environments. Section 4 describes the protocol for decentralized semantic access control to AR services. Section 5 describes an example application of the proposed approach in a city environment. Finally, Section 6 concludes the paper.

2. Content security in AR systems

There are many techniques that aim at protecting multimedia data. DRM (Digital Rights Management) is a technique to control access and usage of digital content, including multimedia data as described by Zeng et al. in [ZYI06]. Modern DRM techniques are designed to maintain control over the content during a large part of the content lifecycle. However, constantly developing AR techniques in conjunction with the development of mobile infrastructures challenge the existing systems. In particular, content processed by multiple AR services interacting dynamically with each other in a mobile environment and created by distributed service providers cannot be sufficiently protected by current DRM solutions.

The most distinguished standardization effort in the domain of protecting the usage of multimedia content is MPEG-21 REL – a rule-based access control language [WDB05]. Unfortunately, the Digital Item representation, which is the base for this model, is not expressive enough to support interactive AR presentations with spatially-sensitive composite content contributed by different service providers. Generic standards developed to allow modelling of rule-based access control, such as XACML [Mos05], despite their usefulness in many multimedia protection scenarios, do not support spatial constraints. XACML has spatial extension called GeoXACML [Geo15], however, mainly due to its two-dimensional nature and lack of AR interaction protection, GeoXACML is not sufficient for AR frameworks. The same applies to GEO-RBAC [BCD05].

Data and service security in interactive, collaborative and federated 3D environments is a wide and growing topic. Fine-grained access control mechanisms designed for 3D collaborative environments that are based on the analysis of call graphs, such as [Wój12][Wój14] could set a base for further research. However, they cannot be applied directly since they do not take into account the specifics of AR, i.e., mobile spatial interaction and dynamic service composition. Similarly, security models for large-scale distribution of structured 3D content proposed in [Wój13], that represent an attempt to mitigate the malicious host problem, cannot be applied directly.

There is also AR-specific research on data security focused on user privacy protection. Due to the novelty of the problem, many articles only point to forthcoming research directions, but do not present any solutions. An example is [DDJ13], in which OS-level access control to AR objects, such as human face or skeleton, is discussed. Other researchers focus on providing AR-specific location privacy [AS10], however, they propose anonymization-based approach only.

The existing approaches do not enable building ubiquitous AR environments, which would enable context-dependent augmentation based on semantics of both the real and the virtual objects, while taking security concerns into account appropriately.

3. Ubiquitous semantic AR environments

3.1. AR presentations and AR services

Ubiquitous semantic AR environments (USARE) consist of AR presentations presented in the context of real environments. AR presentations are composed of elements retrieved from AR services. Those services are divided into four categories: trackable AR services, content object AR services, business data AR services, and scenario AR services, as described below.

- **Trackable AR services** – provide binary representations of physical objects, which are used for tracking objects and augmenting them with multimedia content. These services are usually provided by owners of physical objects represented by the trackable markers, e.g., museums may share photos of the cultural objects held in their collections, advertising agencies may publish images of advertising posters.

- **Content object AR services** – provide multimedia content used for augmenting physical objects and enabling user interaction (3D models, 2D images, video and audio). Providers of these services can be entities which are either the owners of physical objects being augmented or entities independent from the owners of physical objects providing third-party content.

- **Business data AR services** – provide structured data (text, numbers) that are used to create AR presentations. Data retrieved from business data AR services usually are not directly visualized within AR environments, but they can be transformed into a multimedia form for visualization (e.g., image, 3D object). Furthermore, AR data can influence visual, spatial, temporal and behavioral features of AR presentations. Business data AR services can be provided by various entities: the owners of physical objects, multimedia content providers or other independent entities, such as educational institutions, tourist agencies, municipalities, etc.
1. Scenario AR services – provide definitions of AR scenarios. Each AR scenario definition describes visual, spatial, temporal and behavioral features of an AR presentation. Scenario AR services can be offered by different entities, but most often they are business entities or public entities that want to offer AR interfaces to their services and business processes.

2. AR scenarios

The main element enabling amalgamation of different content and data sources into AR presentations are AR scenarios [RW14][WRF14]. An AR scenario specifies the content and the flow of an AR presentation. In particular, it describes:

- AR objects – trackable objects as well as content objects and business data to be presented on them;
- spatial and temporal relationships between the AR objects;
- behavior of the AR objects (presentation and interaction properties).

An AR scenario is composed of a number of AR scene definitions. Each AR scene definition contains semantic rules for selection of AR objects. AR scenes comprising a particular AR scenario can be presented sequentially or simultaneously. Also, AR scenes may embed other AR scenarios.

In the runtime, AR presentations are created based on an AR scenario. To this end, the AR scenario definition is evaluated and executed by the AR scenario execution engine in the USARE browser. During the execution, semantic rules included in AR scene definitions are triggered and appropriate AR objects are retrieved from AR services. Composition and behavior of an AR presentation may change as a result of the progress of the AR scenario.

While an AR scene is active, its components can be activated and deactivated. Thus, the AR objects constituting an AR scene can be active or non-active.

- Active objects may be presented in an AR environment if they contain a visual, aural or haptic representation. Behavior logic of active objects is executed, i.e., the objects can interact with other objects, a user can interact with the objects.
- Non-active objects are not presented in an AR environment and their behavior logic is not executed.

3. Protocol of semantic access control for AR services

To ensure an appropriate level of security in ubiquitous semantic AR environments, the access control protocol for AR services has been developed. The protocol is based on a shared AR Ontology, which is used by the USARE Browser, and the following software entities, as shown in Figure 1:

- Semantic AR Service Catalog,
- User Assertion Provider,
- Semantic Policy Decision Point.

Semantic access control policies for AR scenarios, trackables, business data and content objects based on the same AR Ontology are also provided.

AR scenarios are designed and built based on metadata obtained from the Semantic AR Service Catalog that describes trackables, business data, content objects, and other AR scenarios. The procedure of the secure scenario authoring is regulated by a separate protocol, which is out of the scope of this work. The USARE Browser registers itself at the User Assertion Provider and proves its attributes through a secure channel.

4. Protocol steps

The steps comprising the protocol for secure access to AR services consists of a number of steps – as presented in Figure 1. For the clarity of presentation, the diagram does not show technical messages, which do not influence the logic of the data flow, e.g., signature verification steps are intentionally omitted.

Figure 1: Protocol of semantic access control for AR

The protocol of semantic access control for AR consists of the following steps:

1. The USARE Browser requests an AR scenario from an AR Scenario Provider (this request is denoted as the initial request).
2. The AR Scenario Provider responds either with an AR scenario in the case of a publicly available scenario (and then it ends the message interchange process) or with a request for a list of selected user attributes that are necessary to evaluate access control policy.
3. The USARE Browser requests the Semantic Policy Decision Point for decision whether providing the AR Scenario Provider with information about itself is allowed according to the Semantic Privacy Policy for AR User. The requested attributes and the privacy policy are sent to the Semantic Policy Decision Point.
4. The Semantic Policy Decision Point evaluates the privacy policy with respect to the initial request.
5. The client, after authenticating himself with a client certificate, requests from the User Assertion Provider digitally signed user assertions confirming values of the attributes.

6. The User Assertion Provider responds with digitally signed user assertions proving authenticity of the values of the requested attributes.

7. The USARE Browser requests the Semantic Policy Decision Point for decision regarding the initial request. In the request message, the USARE Browser sends also user assertions obtained in the previous step.

8. The Semantic Policy Decision Point requests the AR Scenario Provider for the access control (AC) policy to be evaluated.

9. The Semantic Policy Decision Point receives the AC policy from the AR Scenario Provider.

10. In some cases, knowing the semantic policy and the user attributes is enough for the Semantic Policy Decision Point to evaluate the policy with respect to the initial request. However, generally it is required to query the trusted Domain Knowledgebase.

11. The trusted Domain Knowledgebase sends back the facts required by evaluation process of the policy with respect to the initial request.

12. The Semantic Policy Decision Point evaluates the AC policy with respect to the initial request. The digitally signed result (“allow” or “deny” assertion) is sent back to the USARE Browser.

13. Having the “allow” assertion, the USARE Browser passes it to the AR Scenario Provider.

14. In the response message, the AR Scenario Provider sends the requested AR scenario back to the USARE Browser.

15. Usually, an AR scenario contains semantic rules that specify trackables, business data and content objects (also other AR scenarios that are either linked or embedded within the original scenario) that are to be used within the scenario at the given context. Therefore, the USARE Browser sends these semantic parameters to the AR Semantic Catalog.

16. The AR Semantic Catalog performs reasoning (based on semantic knowledgebase data) and responds with URIs of the required trackables, business data, and content objects.

17. In the subsequent steps, based on the protocol pattern described above (steps from #1 to #16), the USARE Browser obtains required data from Trackable Providers, Business Data Providers, and Content Object Providers.

The most important difference of the data flow in the 2nd course of the protocol concerns access control to the trackables, business data, and content objects according to their AC policies. Before the evaluation of such policy in the step #7 (in the 2nd or subsequent courses of the protocol), the USARE Browser sends to the Semantic Policy Decision Point not only user assertions, but also scenario “allow” proofs (obtained in the step #12 of the 1st course). Therefore, the Semantic Policy Decision Point “knows” the context of the trackables / business data / content objects usage (knows the scenario in which it is going to be used), and the trackables / business data / content objects AC Policies constraining its usage in specified scenarios can be evaluated.

5. Example use of USARE approach

In this section, an application example demonstrating the practical use of the proposed approach to creation of a secure ubiquitous AR environment is described.

5.1. City-wide AR exploration

The use case concerns AR-based exploration of cultural events in a city. The exploration is realized as a two-stage process. In the first stage, a city-wide AR scenario, published by a municipal culture service provider, is used. The scenario uses common markers visible and recognizable from a far distance: images of bus stop signs, bookshop signs and other cultural institutions’ signs, which are augmented with interactive multimedia content designed to attract user attention. When a user activates an interaction element, a second-stage AR scenario is loaded to the browser.

For example, on a bus-stop shelter several city light boxes are installed. The boxes are maintained by an advertising agency, which uses them for showing advertising posters of their clients. When a user points a smartphone towards a bus stop sign, a new scenario, specific to the advertising agency, is started.

One of the city light boxes contains a movie poster downloaded as a trackable object in the second-stage scenario. In the USARE Browser, the poster is augmented with additional data, which allows users to view information and multimedia content related to the movie (e.g., a trailer, a storyline, user reviews, cast), as well as to buy a ticket for a movie in a cinema (Figure 2). The AR presentation is created using trackables, content objects and business data contributed by several AR service providers:

- **The trackable AR service providers:**
  - municipal information services providing first-stage scenario trackables;
  - a movie distributor or an advertising agency that can create a trackable object based on the poster image provided by the distributor.

- **The content object AR service providers:**
  - a digital art agency (a movie distribution company) providing multimedia data related to the movie (e.g., a trailer, photos);
  - an advertising agency or other entities providing multimedia content required to build a user interface for AR scenarios.

- **The scenario AR service providers:**
  - municipal information services for the city-wide exploration scenario;
  - an advertising agency or other entity involved in the development of interactive multimedia presentations.
• The business data AR service providers:
  o popular movie websites providing reviews and user opinions on movies;
  o cinemas providing information on tickets for movies and the services for buying tickets online.

The AR presentation (Figure 2) is composed of four elements: the movie trailer, ticket price, buy ticket button, and movie rating. A user can interact with the AR presentation, e.g., he/she can play a movie trailer by tapping on the play button. An electronic ticket to a cinema can be bought by tapping on the buy ticket button.

![Figure 2: Example of a USARE environment](image)

5.2. Usage of the access control protocol

The protocol described in Section 4 is employed multiple times in the presented use case – it is used as a basic building-block for message interchange. This section illustrates how it proceeds in the case of AR-based exploration of movie events in a city.

Initially, the USARE Browser – based on recorded user preferences – requests a generic, city-wide exploration AR scenario from a municipal information service provider. The requested scenario is public, so it is sent back directly to the client (cf. protocol step #2 – public case). The scenario requires trackables, i.e., images of bus stop signs and content objects, i.e., interaction elements enabling activation of the second-stage scenario. The required trackables and content objects are obtained by the browser from municipal trackable and content object providers, based on public policies (simple two-step protocol version).

After collecting the required trackables and content objects, the generic scenario is executed by the USARE Browser. When the user points the phone towards a bus stop, an interaction element (a button) is overlaid enabling activation of the second-stage scenario. If the user clicks on the button, a new AR scenario, specific for bus stop city light boxes, is requested. However, in this example the new scenario requires age proof and non-anonymity proof as a part of the access control policy (step #2). The browser verifies whether disclosing these attributes violates the user’s privacy policy (steps #3-#4), and, if they do not, obtains assertions proving authenticity of the requested attributes (steps #5-#6). Next, the browser requests for the authorization decision from the decision point (step #7), which – after reasoning based on the user assertions and the access control policy (steps #8-#9) – sends the decision assertion back to the browser (step #12). In the presented protocol course, the domain knowledgebase querying (steps #10-#11) can be omitted. Then, the decision is passed to the scenario provider, which can be the advertising agency maintaining the city light boxes located at bus stops (steps #13-#14).

The AR scenario requires trackables representing posters, and for each poster a number of content objects, such as trailers, photos, cinema ticket prices and user reviews – in the case of a movie poster. These objects are not hardcoded in the scenario as URIs of the pre-selected services of service providers, but they are selected using semantic rules. These rules are evaluated by the AR Semantic Catalog (steps #15-#16) that either has the required knowledge regarding the service providers and services identification or uses external knowledge sources from the domain (e.g., movie) knowledgebase. In the subsequent courses of the protocol, a trackable of a movie poster is obtained from a movie distributor, content objects (trailers, photos) are obtained from the digital art agency, and business data (cinema ticket prices and user reviews) are obtained from cinema services and movie critics portals.

In the presented example, a movie distributor, as a part of its semantic access control policy, restricts the use of its trackable markers within scenarios to advertising agencies which are bound by a legal agreement (this constraint contains no hardcoded service URIs, but is again expressed semantically). Therefore, the knowledgebase querying precedes the policy evaluation (steps #10-#11 are not omitted in this protocol course). Similarly, a movie critics portal, in its semantic access control policy, allows the use of its data (user reviews regarding a particular movie) only within scenarios that are directly referenced by scenarios provided by municipal information services (anti data harvesting policy), which is also expressed semantically. Thus, in the case of business data access control policy, the knowledgebase querying precedes the policy evaluation as well, in the steps #10-#11 of its protocol course.

When a user points a smartphone camera towards the movie poster, due to the scenario running on the smartphone, the poster is augmented with trailers, photos, user reviews and ticket prices. Also, the user can tap on the augmenting buttons in order to buy cinema tickets using appropriate business data services.

6. Conclusions

The presented USARE approach enables development of a new class of augmented reality applications in which appropriate security constraints are applied in the process of dynamic creation of interactive AR presentations.
Main advantages of the proposed protocol and architecture for the AR access control are:

1. Fine-grained access control – it is based on the custom and precise attributes (assertions).
2. Broad range of access control. Service providers can limit the usage of their services and content according to specified attributes of end users as well as specific attributes of other service providers (e.g., trackable provider vs. scenario provider conflict of interests).
3. Ability to express access control policy rules semantically (dynamically evaluated). There is no need to use low-level values or parameters. Interoperability of the access control policies is provided.
4. Access control model suitable for decentralized processing of large scale and decentralized (e.g., citywide) AR systems:
   - AR scenario provider is separated from trackable providers, business data providers and content object providers.
   - Service providers can take advantage of a trusted AR semantic catalog created as an element of the trusted infrastructure.
5. User anonymity and privacy preservation. Due to enforcement of the semantic privacy policy by specialized infrastructure, end users can control what data can be (indirectly) obtained by service providers – when, by whom, in what context.

Acknowledgements

This research work has been supported by the Polish National Science Centre (NCN) Grant No. DEC-2012/07/B/ST6/01523.

References


Framework for Creating Interactive Browser-Based Virtual Worlds

René Apitzsch\textsuperscript{1}, Felix Pfeiffer\textsuperscript{1}, Tino Riedel\textsuperscript{1}, Andreas Golle\textsuperscript{1}, Mario Lorenz\textsuperscript{1}, Philipp Klimant\textsuperscript{1}, Matthias Putz\textsuperscript{1,2}

\textsuperscript{1}Professorship for Machine Tools and Forming Technology, Institute for Machine Tools and Production Processes, Technische Universität Chemnitz Chemnitz, Germany
\textsuperscript{2}Fraunhofer Institute for Machine Tools and Forming Technology IWU, Chemnitz, Germany

Abstract
Virtual Reality (VR) provides many opportunities for various kinds of applications. It can be used for design reviews, to exercise certain tasks that are too expensive or risky in reality and to transfer information by gaining a much deeper understanding of complex interrelations. Especially in the last case Virtual Reality or three dimensional worlds enriched with additional information offer crucial benefits compared to plain texts or images. However, the downside is that these kind of enriched 3D worlds are not accessible to the public at large. To solve this problem a framework was developed, combining 2D (text, image, video etc.) and 3D content to create enriched 3D worlds accessible through a web browser. Thus anybody can access educational Virtual Reality scenarios for information transfer using current and upcoming web technologies. This paper presents the developed framework which can be used for a wide range of applications. Furthermore the applicability of the framework is shown and the advantages of this approach are confirmed by describing a practical implementation.

Categories and Subject Descriptors (according to ACM CCS): H.5.1 [INFORMATION INTERFACES AND PRESENTATION (e.g., HCI)]: Multimedia Information Systems; H.5.2 [INFORMATION INTERFACES AND PRESENTATION (e.g., HCI)]: User Interfaces

1. Introduction and motivation

As certain studies by Radvansky have shown it is easier to memorize or remember pictures than words [Rad11]. Therefore the process of information transfer, or more precisely the process of learning, can be vastly improved by turning text-based information into two-dimensional pictures or rather into three-dimensional worlds.

Nowadays on the one hand interactive 3D content is widely used and on the other hand information is easily accessible over the Internet. However, the combination of both, which means actual 3D worlds enriched with information, is rare on the Internet.

The proposed solution described in this paper is a framework which provides an easily accessible and user-oriented enriched virtual reality scenario for information transfer and supports content creators to add and modify content in a 3D world.

2. State of the art

2.1. Current information transfer in the World Wide Web

To comprehensively transfer information such as texts, sounds, (animated) graphics or videos over the Internet, the information has to be organized in a certain semantic structure. Currently this is achieved by the generally used markup language HTML and the style sheet language CSS. In their latest versions (HTML5 and CSS3) these two languages allow the creation of 3D objects and scenes, which also implies 3D semantic structures. Although these
features are well supported in modern Internet browsers, they are currently rarely used. For this reason the presentation of information on the Internet commonly remains on a two-dimensional level. The same is true for the interaction.

2.2. Distribution of 3D content in the World Wide Web

Interactive 3D content is widely used outside the Internet. Its main field of application are games as well as training simulators. In contrast, 3D content has a shadowy existence on the Internet. Although the technology exists with HTML5 and CSS3 the frequency of use is relatively low. However, 3D content on the Internet is primarily used for miscellaneous showcases and examples, games, advertisement or car configurators, but especially complex plugin-free accessible 3D worlds enriched with information with the main purpose of information transfer do not exist.

3. The framework – virtual reality for the Web

The most basic distinction the developed framework makes is between user interface (presentation layer) and author interface (data access and handling layer). The main purpose of the user interface is the navigation inside the canvas whereas the author interface serves the content creator to populate the framework with content and create a 3D web presenter.

3.1. User interface

The user interface is mainly based on features introduced with HTML5 and CSS3. In addition JavaScript is used, which allows client-side scripts to interact with the user, control the browser and alter the document content [Fln06]. Based on that JavaScript also allows ensuring downward compatibility to older web browsers which do not support HTML5 and CSS3 features. The user interface consists of the following components (see Figure 1).

Figure 1: User interface of the framework: start page (top left); 3D view (top right); media container view (bottom left); sidebar view (bottom right)

The canvas contains the 3D models and provides a visual overview of the 3D scene. It is intended to be the main interaction interface for the user to gain access to the desired information and shall be an invitation to explore and discover the 3D scene emphasizing a geographic and/or semantic structure of information.

The head-up display (HUD) has multiple purposes. The first one is to provide information about the current navigation level. Another purpose is to show corresponding buttons if the current navigation level provides further information such as a slideshow of video. Moreover the HUD provides a clear opportunity to navigate upwards, downwards and sideways in the hierarchic structure, whereas the canvas depends on designated 3D objects to trigger navigation events. Thus an advantage of the HUD is to not unnecessarily clutter the 3D scene with 3D objects. Furthermore, the HUD or more precisely the handles on the left side can provide access to a usually hidden sidebar. In addition to the expressive icons of the HUD tooltips are shown to improve comprehension if the mouse pointer hovers over
designated elements. Due to technical shortcomings this feature is disabled on touch devices.

The media container is universally usable and embeds various types of additional information in form of media files such as images, videos, PDFs and websites. The media container also shows the current and total slide number and includes control elements in order to change the slide or close the container.

The sidebar is optional, but can consist of a freely definable number of menus providing supplementary information, e.g. navigation menu, contact information or imprint. The most important component is the navigation menu. It reflects the hierarchy of the 3D scene in form of a 2D tree structure and gives quick access to the content of the web presenter.

Also optional is the footer as an overlay at the bottom of the canvas. It provides permanently visible information such as logos or legal advice.

A key aspect of the user interface is the ability of cross-device interaction. This means the user interface is applicable to different screen sizes, resolutions and aspect ratios. Moreover the user interface also seamlessly supports the two expectable ways of interaction, meaning touch and pointing device interaction.

3.2. Author interface

Similar to the user interface the author interface is intended to be as simple and intuitive as possible, whereby almost everyone can add (3D) content and supplementary information or extend and maintain functions with just little effort. To reach this aim different measures were taken. First of all the framework was built on a modular structure. This means that all PHP and JavaScript functions were pooled into separated files regarding their purpose (see Table 1). Furthermore all CSS rule sets and PHP functions concerning the template of the user interface were gathered in as many files as necessary, but as few as possible, which can be reused to prevent code duplications and to improve the clarity of source code. The popular JavaScript library jQuery was also used to simplify the programming and to be able to easily access and manipulate DOM elements [Lin06].

Besides the self-written functions third-party JavaScript libraries were also used. PDF.js is a JavaScript library based on web standards to parse and render PDF files in the web browser. -prefix-free.js is ensuring cross-browser compatibility by adding browser-specific vendor prefixes to the CSS rule sets. X3DOM should be particularly mentioned, which is an open-source framework and runtime for 3D graphics on the Web. It allows X3D files to be natively integrated into the DOM tree without requiring any browser plugin. The browser just has to support WebGL or Adobe Flash used as the designated internal X3DOM fallback. Adobe Flash requires the Adobe Flash Player, which is proprietary and not officially supported by Android, iOS or Windows Phone. This missing cross-platform compatibility is the main reason why a specific 2D fallback solution has been implemented.

A key concept of the framework concerns the so called field of action (FOA) representing the state of the menu, i.e. the canvas [SF1G]. Each field of action has an abstract FOA-ID being defined as an entry of the navigation menu. The definition also assigns the media related to the FOA_ID. The hierarchical structure of the navigation menu declares when a certain FOA-ID (state) can be accessed. To provide multi-language-support the abstract FOA-ID is referring to the respective entry in a language file.

To create a new 3D web presenter or to fill an existing with (new) content the author has to carry out the following tasks:

1) Defining the hierarchical structure of information: defining entries (FOA-IDs) in the navigation menu depending on the geographic or semantic structure of the 3D scene.

2) Building the canvas containing the 3D scene: placing X3D files into a predefined folder.

3) Defining the user interface actions: defining which (3D) elements are hoverable and clickable and what happens due to the interaction (change FOA-ID, start animation, change camera, open sidebar etc.).

4) Inserting content: placing slides, videos or pdfs into predefined folders.

The 2D fallback can be created in two different ways. One way is to pre-render the background images of the 2D version in the 3D graphic program where the 3D models were created and place them in the predefined folders as well as their corresponding clickable icons, which also have to be manually aligned over the background image. The other possibility consists in using commands coming with the X3DOM framework to directly render images in the browser out of the 3D scene. This reduces the effort and keeps the 2D and 3D version as similar as possible. This procedure was automated by using the following steps:

1) Render background images: out of every camera perspective (view) a base64-encoded PNG-file is generated.

<table>
<thead>
<tr>
<th>Table 1: Important JavaScript files and their purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>File Name</strong></td>
</tr>
<tr>
<td>autoscale.js</td>
</tr>
<tr>
<td>caching.js</td>
</tr>
<tr>
<td>detectmobilebrowser.js</td>
</tr>
<tr>
<td>interaction.js</td>
</tr>
<tr>
<td>media.js</td>
</tr>
</tbody>
</table>
2) Render clickable areas: in every view every interactive object is separately rendered to a base64-encoded PNG-file with an alpha channel.

3) Convert PNG to SVG: every PNG-file is converted to an SVG-file provided with specific tags using Potrace [Sel03].

4) Merging SVG: all SVG-files related to one view are merged into a single file considering the z-indices of the 3D scene using Imgagick [Sit05].

5) Composing 2D view: all generated images are placed into pre-defined folders.

3.3. Drawbacks

X3DOM does not implement all the specifications of X3D/VRML defined by the Web3D Consortium. The developers of X3DOM intend that the user deploys JavaScript to implement higher, more complex functions by themselves [BEJZ09]. In X3DOM ScriptNodes, some sensors and prototypes are not supported and will not be in the future. For this reason the behavior considering user interaction cannot be implemented within the X3D files and therefore has to be outsourced into separate files.

Despite all the implemented features the interactions are very individual and still have to be made by the author of the web presenter himself, which requires programming skills.

Not every browser or platform supports 3D rendering out of the box or is powerful enough to ensure a fluid experience. Therefore it had to be ensured that the user still gets access to the information provided by the presentation engine. As a solution a 2D version was implemented as a fallback. If the script detects an unsupported browser or platform the 3D canvas is automatically replaced by, for example, pre-rendered images of the 3D scene or actual photos. Interactive 3D elements will optionally be replaced by 2D elements overlaying these pictures, but even without these elements, access is still granted by the navigation menu located in the sidebar.

3.4. Benefits

All the previously described features of the developed framework lead to several advantages. Using 3D worlds enriched with information, the framework especially emphasizes the geographic (location related information) or semantic (context related information) structure of this information. The 3D world also allows non-linear presentation of content with no constraints in interaction or significance of content. Due to several implemented functions and the use of modern browser features and an adaptable user interface, the framework is compatible across several devices and browsers. Furthermore the framework can be multilingual. Based on web technologies the framework is usable online and offline. For both versions a (portable) web server is required because the framework uses PHP as the server-sided scripting language. The framework offers many functions to simplify the creation of 3D web presenters such as easy adding and modifying of content due to predefined folders for media files. Furthermore the authors can easily change the appearance of the web presenter because the 3D and 2D version use the same HUD.

4. Use case implementation

4.1. General information

After showing the key features and the drawbacks, a use case implementation of the framework will be presented in the following. On behalf of the Saxon State Ministry for Science and the Arts (SMKW) a web presenter was created for the first time with the proposed framework. As a showcase project serves The Saxon Excellence Initiative serves as a showcase project. It is an initiative to strengthen promising fields of research and to increase the quality of research in the state of Saxony. The task was to create a 3D web presenter containing all five participating clusters of excellence (eniPROD, LIFE, OncoRay, ECEMP and ADDE). eniPROD considers energy-efficient product and process innovations in production engineering. LIFE researches civilization diseases. OncoRay targets an improved cure of cancer through biologically individualized and technologically optimized radiotherapy. ECEMP focuses on innovative materials and technologies for energy management, environmental engineering and lightweight engineering. ADDE deals with functional structure design of new high performance materials via atomic design and defect engineering.

Each cluster has its own project structure. Moreover some of the clusters have a geographic structure of information (LIFE, OncoRay), others have a semantic structure (eniPROD, ECEMP, ADDE). The challenge was to find an overall unifying structure to accommodate them all. The result can be seen in Figure 2.

The intro welcomes the user or the visitor to the web presenter. It shows achievements and first impressions of the Excellence Initiative, and it raises questions teasing the user to take a closer look to find the answers. The intro is followed by the actual start page of the web presenter. Its purpose is to provide general information (slides, videos) about each cluster and the Excellence Initiative itself. It also grants the user access to the following cluster level, which is the first level showing actual 3D worlds enriched with information. In contrast, the first two levels intentionally show only 2D content to keep data traffic down and to ensure high compatibility in order to keep the bounce rate down, to show every visitor a uniform and recognizable design [Kin13]. As mentioned before, inside the cluster level each cluster has its own geographic or semantic structure. Depending on the kind of structure each project has its own 3D model as a metaphor.

4.2. Implementation

The main part of the implementation concerns the creation of the 3D elements, which implies considering many aspects. At the beginning a consistent naming scheme has to be specified in order to have a defined interface to access the 3D objects per JavaScript. Furthermore it is important to sort the created 3D elements into groups. A suitable
scene organization is characterized by a compromise between function and performance and leads to an efficient workflow. This facilitates change requests, especially regarding the development of project metaphors in coordination with the officials of each cluster. It is also essential to define a threshold value for the model complexity, regarding the amount of object and polygons, to ensure a satisfying 3D performance even on commonly slower mobile devices. Moreover the technical feasibility as well as the ease of use has to be guaranteed. Yet not only the 3D elements have to be created but also the viewports and tracking shots have to be defined.

The export of the 3D scene from 3dsMax to X3D files takes place with a self-written exporter into predefined folders. Thus the scene can directly be watched in the web presenter. Based on this 3D scene shown in the canvas the images of the 2D version were directly generated and saved by self-written JavaScript functions.

To simplify the workflow of getting such a great amount of media files into the web presenter a self-written PowerPoint macro was used to generate JPG files out of the slides in the desired resolution. In addition, the provided video files have to be converted into the MP4 file format, so that they become conform with HTML5 and can be natively played by modern web browsers.

![Diagram of the web presenter and the top level virtual worlds](image)

**Figure 2:** Structure of the web presenter and the top level virtual worlds
5. Conclusion and outlook

The introduced framework enables content creators to build interactive browser-based virtual worlds for a wide range of applications. A successfully working and worldwide accessible implementation is the presentation of research results as shown in section 4. Another use case could be a 3D guidance on a touchscreen computer at the entrance of large complex buildings or a 3D web-based virtual educational system.

During the development many drawbacks were overcome, but also potentials were discovered for improvement. A server-free and offline version independent from operating systems could be imaginable. The media files could be integrated directly into the 3D scene. Videos could be shown on virtual 3D monitors or slides as the different pages of a virtual 3D notepad. This approach makes the use of media overlays obsolete and takes advantage of the virtual reality.

References


A Framework for Transparent Execution of Massively-Parallel Applications on CUDA and OpenCL

Jörn Teuber¹, Rene Weller¹, Gabriel Zachmann¹

¹University of Bremen, Germany; {jteuber,weller,zach}@cs.uni-bremen.de

Abstract
We present a novel framework for the simultaneous development for different massively parallel platforms. Currently, our framework supports CUDA and OpenCL but it can be easily adapted to other programming languages. The main idea is to provide an easy-to-use abstraction layer that encapsulates the calls of own parallel device code as well as library functions. With our framework the code has to be written only once and can then be used transparently for CUDA and OpenCL. The output is a single binary file and the application can decide during run-time which particular GPU-method it will use. This enables us to support new features of specific platforms while maintaining compatibility. We have applied our framework to a typical project using CUDA and ported it easily to OpenCL. Furthermore we present a comparison of the running times of the ported library on the different supported platforms.

Categories and Subject Descriptors (according to ACM CCS): I.3.6 [Computer Graphics]: Methodology and Techniques—Device independence

1. Introduction
The single core performance of CPUs stagnates widely since many years. On the other hand, the demand for larger scenes with increasing details, accurate physics simulation, marker less tracking devices and many other expectations on modern VR applications requires more and more computational power. Obviously, the solution is an increasing degree of parallelization. While the parallelization of CPUs progresses relatively slowly, current GPUs offer an exhaustive number of parallel processing units. Unfortunately, it is often not trivial to parallelize tasks or algorithms. Hence, a lot of research has been spent in the development of new algorithms that are better suited for parallel execution.

Especially in the academic world, such algorithms are often implemented only as a proof-of-concept. To do this, researches often choose NVIDIA’s CUDA programming language [NBGS08] because it is easy to use, it offers sophisticated tools for debugging and profiling and it has great community support. Moreover, a lot of libraries exist that allow rapid application development and the features of new GPU generations are integrated immediately. Unfortunately, programs developed in CUDA are restricted to work only on NVIDIA GPUs [LNOM08]. Obviously, software companies that sell their products to a broader audience need a different API.

Here, OpenCL [SGS10], a platform-independent standard for parallel programming, defined by the Khronos Group, is usually the method of choice. However, porting code from CUDA to OpenCL requires the replacement of standard libraries, for instance for parallel sorting or image processing, and the reimplementation of large parts of the code.

We present a novel method to overcome these, often time consuming, limitations. Our new wrapper framework for CUDA/OpenCL supports an easy work-flow for porting CUDA to OpenCL code and supports platform independent development from the start. To do that we implemented an easy to use common interface that encapsulates the calls of parallel device code as well as library functions. The resulting single binary includes both CUDA as well as OpenCL code, and the application can decide during run-time which particular GPU-method it will use. This enables us to support new features that are supported much earlier by
CUDA while maintaining compatibility. In addition, it allows us to perform comparisons between CUDA and OpenCL, and between execution on the GPU and on the CPU regarding their respective performance very easily. We have applied our framework to an extensive CUDA library for the parallel computation of sphere packings [WZ10].

2. Related Work

In the past there were several projects to enable developers to port CUDA to OpenCL. The most recent one, CU2CL [GSFM13], is an attempt at an automated translator for CUDA code to OpenCL code. It translates given .cu files, the .cpp file equivalent for CUDA code, into several C++ and OpenCL files to use with OpenCL. This automatic conversion requires all CUDA API calls to be encapsulated into functions in .cu files, which is not the common way of using CUDA. Also the intended result of CU2CL is just the code for OpenCL, CUDA can not directly be used side-by-side. Hence, the programmers have to avoid the most recent CUDA features that are not yet supported by OpenCL and moreover, it does not support external library calls.

A project more similar to ours was Swan [HDF11], which combined a translator for CUDA kernels to OpenCL kernel code with a common API encapsulating the CUDA and OpenCL APIs. Unfortunately this project was discontinued in 2010 and not updated since then. To our knowledge there are no other current projects providing transparent access to CUDA and OpenCL.

There are several papers comparing the performance of CUDA and OpenCL, mostly for very specific use cases. [KDH10] for instance investigated a Monte Carlo simulation of a quantum spin system while [MS14] considered text encryption using the GPU, to name a few. Both papers observed that OpenCL is slightly slower than CUDA. [LCMH14] on the other hand used CUDA, OpenCL and GLSL to accelerate skeletal animations and found that all three methods yielded a comparable performance.

3. Our Approach

The basic idea of our approach is to maintain a single wrapper class that encapsulates all CUDA/OpenCL related functions. This class has a simple and general interface that allows the handling of both parallel programming languages simultaneously. The output is a single binary for both CUDA and OpenCL. The application can decide at runtime, which library to use. Also it is reusable and may enable other programs to provide both OpenCL and CUDA support.

We start with a short recap of the GPU programming terminology. GPGPU is the abbreviation for general purpose computation on graphics processing units. It is used to refer to any kind of computing on the GPU outside of computer graphics, from re-purposed shaders to CUDA and OpenCL. The processor on which the program is started, i.e. the CPU, is called the Host while the processor that is used for the massively parallel algorithm is called device. This can be a GPU, but also a CPU or an accelerator card in the case of OpenCL. Kernels are the programs that are being run in parallel on the device.
void main()
{
    // initialize with CUDA as GPGPU method if available, or OpenCL if not
    if( GPUWrapper::getCUDADeviceCount() > 0 )
        GPUWrapper::init( GPCUDA );
    else
        GPUWrapper::init( GPOpenCL );

    GPUWrapper* gpu = GPUWrapper::getSingletonPtr();

    // generate random numbers on the host
    float host_array[1000000];
    std::generate(host_array, host_array+1000000, rand);

    // create a new array on the device and copy the array to the device
    DevMem<float>* device_array = gpu->copyToDev( host_array, 1000000 );

    // sort the array and download it from the device
    gpu->sort( device_array );
    device_array->copyToHost( host_array );

    return 0;
}

Figure 2: Simple example program using our wrapper for sorting a random array.

Our approach consists of three main components (See Figure 1).

- **GPUWrapper class**: This class builds the core of our approach. It realizes the main interface for the developers, including initialization and platform-independent access to the massively parallel functions and libraries.
- **DevMem class**: A template class to handle the massively parallel data structures.
- **GPUUser class**: This class implements the real mapping to the particular platform. Actually, this class is used only internally and not directly accessed by the developer.

In the following, we will describe the functionality of all these components in more details. The GPUWrapper class is the central wrapper for both OpenCL and CUDA, the functionality of which are implemented in the same-named classes. The GPUWrapper class is a singleton which has to be initialized by the user with the information about the GPGPU method and device to use. We use the singleton pattern here to provide universal access to the initialized GPUWrapper. Actually, all CUDA API functions can be accessed everywhere in the code but the access to OpenCL devices is restricted to only a single device at the same time.

The DevMem template encapsulates the different models of pointer to device memory. It also provides the basic member functions like memSet and copyToHost. Other important methods for GPGPU like scan, sort or reduce are provided by GPUWrapper and than delegated to the Thrust library for CUDA or boost::compute for OpenCL.

The GPUUser class consists simply of static pointers to objects of the CUDA and the OpenCL classes, which are set by GPUWrapper during initialization. This provides all instances of DevMem access to CUDA and OpenCL. A template inheriting from a class with static variables is a common pattern to provide all instances of the template with access to those static member variables.

Figure 2 shows a typical example how to use our approach: We start with the initialization of an GPUUser object, define some device data using our DevMem template class and finally, do our massively parallel computations on this data.

So far, our wrapper can automatically handle only code on the host side, but not the kernel code. This means, the kernel code has to be converted manually. Luckily, the porting of the kernels from CUDA to OpenCL is mostly straight-forward text replacement as long as no advanced CUDA specific methods are used. In our case, we just had to make minor adjustments, which were mostly due to the different memory addressing models of CUDA and OpenCL and the fact that OpenCL uses plain C while CUDA allows more C++-like code.

The method of kernel calls differs significantly be-
between CUDA and OpenCL. NVIDIA uses a compiler extension to make CUDA kernel calls to be as similar to normal function calls as possible. OpenCL kernels on the other hand are runtime-compiled for every device (with the option to save and load compiled kernels at runtime) and called by supplying an OpenCL function with the handle of the kernel and other information. This dissimilarity leads to a necessary deviation from good wrapper design, which is the Kernel class, objects of which can only be created by the OpenCL class. Fortunately, NVIDIA suggests to use small wrapper functions for the kernel calls anyway, so by just extending those, the main host code does not change. You can see one of those wrapper functions in Figure 3.

4. Application

We applied our method to a library that computes space filling sphere packings for arbitrary 3D objects [WZ10]. Originally, the library was developed in CUDA but industrial partners have the demand to run it also on systems without NVIDIA GPUs and even without dedicated graphics cards at all. The library relies on typical library calls for sorting and scanning but it additionally contains different individual kernels and different data structures.

We will provide a short recap on the basic idea of the sphere packing algorithm to enable a better judgment of its complexity: The main idea is to generate a greedy sphere packing, i.e., to insert successively the largest possible sphere into the object considering the already inserted spheres. The major challenge is to find the position of this largest spheres. Here, the authors provided a simple heuristic that is similar to techniques that are often used in machine-learning algorithms: A so-called prototype is initially inserted at an arbitrary position inside the object and then this prototype is iteratively moved, depending on its minimum distance to the surface (see Fig. 4). In order to avoid local minima, the authors do not use a single prototype but many of them that move independently and hence, can be easily parallelized. In order to accelerate the distance computations the library contains different implementations of discrete distance fields (see [TWZG13] for more details). Algorithm 1 provides a short overview on the main steps in pseudo code.

5. Results

We have applied our semi-automatic CUDA/OpenCL version to the above mentioned sphere packing library. The porting of the complete library took only a half day for the adjustment of the CUDA kernels code. Additionally, we tested the performance of the compiled dual-executable on an PC with Intel I7 CPU and a NVIDIA GTX 680 GPU with 8 GByte of memory. The CUDA code was executed on the GPU exclusively whereas the OpenCL code was compiled for the CPU as well as for the NVIDIA GPU and the built in Intel graphics adapter. We tested different 3D objects and filled them with different numbers of spheres.

Figure 5 shows the results of our tests. In all our test cases the CUDA and the OpenCL-GPU version have a very similar performance. In our first test case, the OpenCL-CPU version is slower than all GPU versions, which is expected. We recognized a speed-up of about an order of magnitude with the GPU version for the NVIDIA GPU. Even the built-in Intel GPU runs up to a factor of two faster than the CPU version.

Surprisingly, in our second scenario (see Fig. 5, right), the CPU outperforms both GPU versions at the beginning of the algorithm, when the number of spheres is small. In this scenario we changed the reso-
void countMemory(int nrTriangles, ObjectOnDevice *object, 
    DevMem<unsigned int> *cellsPerTriangle)
{
    if (GPUWrapper::getSingletonPtr()->getType() == GPCuda)
    {
        uint numThreadsPerBlock, numBlocks;
        computeGridSize(nrTriangles, blockSize, &numBlocks, &numThreadsPerBlock);

        countMemoryKernel<<<numBlocks, numThreads>>>(object->m_uiNumTriangles, 
            object->m_dVertices->getCUDA(), object->m_dVertexIndices->getCUDA());
    }
    else if (GPUWrapper::getSingletonPtr()->getType() == GPOpenCL)
    {
        static Kernel* spKernel = NULL;
        if (spKernel == NULL)
            spKernel = GPUWrapper::getSingletonPtr()->getRawOpenCL()->createKernel("countMemoryKernel", "ExplicitGrid_kernels.cl");

        spKernel->execute(nrTriangles, blockSize, object->m_uiNumTriangles, getGridConfigOCL(), 
            object->m_dVertices->getOCL(), object->m_dVertexIndices->getOCL());
    }
}

Figure 3: Example of a user-defined wrapper function for custom kernel calls. This has to be implemented by the application’s programmer for every custom kernel. If CUDA is used, the kernel is called with the respective CUDA syntax. In the case of OpenCL, the kernel is loaded from a file into a static pointer and executed using a variadic method of the kernel class.

olution of the discrete distance map while maintaining
the number of prototypes. Basically, in the convergence
step of Algorithm 1, each prototype checks all grid cells
in the discrete distance field that could possibly con-
tain the closest point on the surface. If we increase the
resolution of the discrete distance field, this results in
a higher number of cells that has to be visited and con-
sequently, to an increasing workload per prototype, i.e.
per thread. On the GPU, the threads on a warp are exe-
cuted in lock-step, whereas the CPU can schedule new
threads at any time. Consequently, we have a worse
degree of parallelization on the GPU and therefore a
worse core utilization than on the CPU.

Actually, we were already aware of this problem but
we did not anticipate the magnitude of the impact on
the timings. So, the porting of the sphere packing to
OpenCL provided us also new insights about oppor-
tunities to further optimize the algorithm.

6. Conclusion and Future Work

We presented a new wrapper approach which enables
programmers to develop massively parallel algorithms
for CUDA and OpenCL simultaneously or to port their
CUDA algorithms to OpenCL or vice versa very eas-
ily. Our semi-automatic wrapper is easy to use and it
supports platform specific libraries as well as kernel
code written by the application programmers. We used
our wrapper to port a complex library from CUDA to
OpenCL. In addition, we present the running times
of that library that allow a comparison of the respec-
tive performances of massively-parallel implementa-
tions on CUDA, OpenCL/GPU, and OpenCL/CPU.
Our measurements show a very similar running time
of CUDA vs. OpenCL on the GPU with OpenCL on the
CPU being slower by about an order of magnitude in
case of optimized parallelization.

In conclusion, the additional work of porting a pro-
gram from CUDA to OpenCL can be held low. In this
paper, we presented such a software architecture that
is easy to implement. Having thus the opportunity to
run your code on different platforms and parallelization
APIs at any time during the development process helps a lot to gain insights into how your code per-
forms on those vastly different platforms.

In the future we would like to extend our wrap-
per to support more external libraries and we will re-
lease it under an open-source license as we believe
that this might be useful to other people. Furthermore,
some parts of the kernel code still have to be ported
manually. A complete automatic conversion between
CUDA and OpenCL kernel code would be an interesting
project for the future.
Figure 5: Performance of sphere packing algorithm with different resolutions of the discrete distance map for the dragon model (bottom right). In case of a low resolution we have a high degree of parallelization and both GPU versions outperform the CPU version by a factor of 5 (upper left). In case of a high resolution, the degree of parallelization decreases and the CPU version is faster, at least in early stages of the algorithm (upper right).

References


*Natalia Cooper, **Ferdinando Millella, **Fain Cant, **Carlo Pinto, *Mark D White and *Georg F Meyer
*University of Liverpool, Liverpool, UK **Virtual Engineering Centre, Daresbury, UK.

Abstract

The aim of this study was to evaluate the effect of visual, haptic and audio sensory cues on participant’s sense of presence and task performance in a highly immersive virtual environment. Participants were required to change a wheel of a (virtual) racing car in the 3D environment. Subjective ratings of presence and comfort were recorded using the Immersive Tendencies Questionnaire (ITQ), [WS98], the Presence Questionnaire (PQ) [WS98] and Simulator Sickness Questionnaire (SSQ), [KLB* 93]. The time taken to complete the task was used as an objective performance measure. Auditory, haptic and visual cues signalling critical events in the simulation were manipulated in a factorial design. Participants wore 3D glasses for visual cues, headphones for audio feedback and vibration gloves for tactile feedback. Participants held a physical pneumatic tool. Events, such as the full extraction of a bolt were signalled by haptic (vibration frequency change), acoustic (change in tool sound) and visual (colour change of bolt) cues or combinations of cues. Data was collected in two blocks containing all eight sensory cue combinations: the task was once performed in a normal VR environment (control) and once (motion) in an environment where the position of the virtual environment was sinusoidally modulated by 2 cm in the depth plane at 0.5 Hz to simulate inaccurate participant tracking. All participants completed all 16 conditions in a pseudorandom sequence to control for order and learning effects. Subjective ratings for presence, discomfort and perceived cues effectiveness were recorded after each condition. Participants performed best when all cues were present. Significant main effects of audio and tactile cue presentation on task performance and also on participants' presence ratings were found. We also found a significant negative effect of environment motion on task performance and participants' discomfort ratings.

Categories and subject descriptors (according to ACM CCS): performance measures, auditory feedback, haptic I/O, virtual reality.

1. Introduction

Virtual reality (VR) environments are useful as tools for training, perceptual motor research, interpersonal communication, data visualisation and many other purposes that depend on accurate presentation of visual stimuli and recording on user interaction with VR. Users evaluation of these systems often includes all qualitative experience a user has whilst engaging and interacting with a given system [PR14]. It is generally believed that high levels of immersion can cause an increased sense of presence that can make some applications more effective [BM07; LSS*. 12]. Here, immersion refers to the objective description of what any particular system does provide, whilst presence is associated to the state of consciousness of the user and the sense of ‘being’ in the virtual world [SLU*. 96]. Many presence-evoking media technologies were designed so that people can accomplish a task with greater efficiency and previous studies show that greater immersion in VR cause subjects to perform better. This was confirmed in studies of target localisation and acquisition [Oak09; BM07; AMH95] spatial attention [SS09] as well as interaction with a VR system [BIL* 02]. It has been suggested that illusion of self-motion can also make a positive contribution to the overall experience and effectiveness of VR systems [RSA* 06] however, undesirable side effects of self-motion have been identified mainly in the large scale collaborative VR environments where one user controls the (shared) view of a number of users in the simulation. The presentation of anchors in VR and avoidance of non-informative signals has been suggested to minimise these undesired side effects of self-motion [MSW*. 13]. Future research into multisensory cuing and self-motion coordination should further investigate which factors contribute the most to the desirable effects of VR systems.

1.1. Multimodal cues

In order to support training and performance in VR it is essential to provide necessary sensory cues that are required for the task [PR14]. It has been generally believed that the greater number of human senses stimulated, greater the capability of the stimulus to produce a sense of presence. In studies where uni-modal, bimodal or multisensory stimuli are provided it has been found that the subject reported a greater sense of presence when more sensory cues were provided [LS08; BM07; LSS*.12] and especially when they were presented ‘matching’ location and direction [MWR*. 05]. However, it has been pointed out that the importance of all sensory cues may not be equal. Typically, simulation in the VR environment relied mostly on visual cues as these were found to be most important. Previous research has shown that addition of auditory cues can enhance human performance as well as the perceived sense of presence at minimal increase in cost [JDN*.09; JEE*.04; MWT*.12]. Some studies have suggested that
there is a lack of haptic cuing in VR environments mostly due to expensive devices, associated with technology and difficulties in achieving a realistic haptic interaction [Edw00]. Studies that address haptic cuing in VR have found that tactile feedback can potentially be a promising type of additional feedback as it can contribute to effective interaction when the visual or auditory modalities are compromised, engaged or overwhelmed [AMH95; HSC*_05; JEE*_04; VJE03]. However, others have found the opposite when they reported that tactile and audio feedback can also be perceived as distracting and annoying, decreased overall performance as well as having negative effects on accuracy [Oak09; Bre03; VJE03].

Overall, previous studies have argued that additional multisensory information can enhance the interaction with a system through providing more salient stimuli [BIL*_02], however it has been also identified that some of the cues, mostly tactile cues, are much more difficult to present due to technological constrains and cost. The main aim of this study was to investigate whether the presentation of the relevant information in the different domains will have any subjective or objective difference to task performance. Our research focuses on influences of unimodal, bimodal and multimodal sensory information and its resulting effects on perceived sense of presence and individual task performance.

2. Methods

2.1. Participants

For this study we recruited 16 participants via opportunity sampling. There were 11 males and 5 females, with the age ranging from 18-48. All participants reported normal to corrected to normal vision and normal hearing.

2.2. Virtual reality set up

The experiment was conducted at the Virtual engineering centre (VEC) facility located in Science and Technologies Facilities Council (STFC) in Daresbury. VEC is part of the School of Engineering at the University of Liverpool. It contains virtual laboratories with High Performance Computing (HPC) and provides facilities for advanced modelling, simulation and immersive visualisation.

2.3. Apparatus

The laboratory consists of a planar display screen of length 6.0 m and height 2.1 m behind which are two active stereo projectors that create 3390 x 1200 resolution images at a rate of 120 Hz. 3D stereo images are produced by an NVIDIA Quadro K6000 GPU. Observers wear wireless LCD shutter glasses that are synchronized with the projectors to provide stereoscopic images. Object position is tracked using 16 high-spec infrared cameras (VICON Bonita B10, 250 fps capture speed, motion resolution of 0.5mm of translation and 0.5 degrees of rotation in a 4m x 4m volume using 9mm markers). Position data, computed using VICON Tracker software, is broadcast in real-time across the internal network using a VRPN protocol at a rate of 200 Hz and used to update the virtual environment.

The following objects are tracked in order to provide the required interaction within the virtual immersive environment: glasses (for head tracking and POV adjustment), subject hands (to drive subject’s virtual hands) and the impact wrench, the tool used to remove wheel bolts. A faithful digital mock-up of the impact wrench is used to interact with the bolts. Through accurate calibration both hands and impact wrench overlap with their virtual counterparts from the subject’s perspective. In this way the subject has the perception that (s)he interacts with virtual objects (wheels and bolts) using his/her real hands and the real power tool. The wheel change simulations runs at a constant speed of 15 fps across all possible combinations of cues to ensure an accurate time recording in all experiment conditions (i.e. times are not affected by enabled/disabled cues).

Tactile stimulus is provided by two “tactile gloves” realised by adding to the VICON hand tracking kit a vibration motor attached to the palm (Fig. xx). The motor is actuated by PWM drives receiving information on collision detection, level of vibration, etc. by a device wirelessly connected to the CPU running the immersive scenario. The vibration occurs with variable intensity, based on the specific task. For example, the subject can feel an intermediate level of vibration when screwing a bolt out or back in place, which steps up to the maximum level as soon as the bolt is completely screwed in or reduced to zero when is completely removed. In this way we mimic the intensity of vibrations generated by the impact wrench when performing the real task.
(ITQ)[WS98], Presence Questionnaire (PQ) [WS98] and Simulation Sickness Questionnaire (SSQ) [KLB* .93].

Figure 2. Participant wear headphones, vibration gloves and holding impact wrench whilst performing the task.

2.5. Procedure

Before participants started the experiment they were required to fill in the Immersive Tendencies Questionnaire and Simulation Sickness Questionnaire to provide a baseline measure. The room was darkened during all experiments. Participants wore 3D shutter glasses, vibration gloves and headphones that played continues white noise to mask vibration noise from the gloves (see figure 2). The task was to change the wheel on the virtual racing car in the 3D environment as fast as possible. During the task participants were provided with additional visual, tactile and audio cues. The cues were presented as unimodal, bimodal and multimodal feedback in randomized order (A, V, T, AV, AT, TV, AVT, NONE). The virtual environment was manipulated in two experimental blocks containing either static or lateral motion (0.2Hz) of the whole visual scene. Within each block participants performed the task 8 times in randomized order of conditions. Each block lasted approximately 15 minutes and participants had at least 15 minutes break between the two successive blocks.

2.6. Task

Each participant started with two practice trials. This was followed by experimental conditions in each block in counterbalanced order. The time started when the participants got in contact with the physical tool. First, they had to unscrew 5 bolts from the wheel on the virtual racing car (see figure 1). After this they had to pick the wheel up and put it on the stand located next to the racing car. Then they had to go and grab another wheel from the stand on the other side, attach it on the racing car and screw the bolts back in. The overall recording stopped when the participants placed the tool back on the table.

2.7. Multisensory cues

During the task participant were presented with different sensory cues. The visual cues presented during the task included the bolts turning yellow when in contact with the tool and red when the bolts were completely out; the wheel turned yellow when in contact and red when in the right position; the virtual hands of the participant turn yellow when in contact with virtual parts. The tactile cues presented during the task included a vibration sensation when the tool was in contact with the bolt following a more intense vibration when the bolt was completely out; and when the virtual hands were in contact with the wheel. The audio cues presented during the task included a drilling noise when in contact with the bolt and a ‘snap’ sound when the wheel was placed on the stand and on the racing car. After each condition participants were asked to rate their sense of presence on a short questionnaire (7 questions). After each experimental block participants had a short break and were asked to rate their sense of presence on PQ, their feeling of discomfort on SSQ and two sensory questions asking them which cue or a combination of cues they found most useful in the bolt screwing task and in the wheel position task. After this, participants performed the task in a second experimental block (8 times) whilst filling the short questionnaire between conditions. Then participants were asked to fill in a second set of the same questionnaires as before.

3. Results

Descriptive statistics for each condition are presented in Table 1. Overall, participants rated the multisensory feedback most favourably, which is confirmed in their overall task performance.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Objective data</th>
<th>Subjective data</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>49744.91 (9266.17)</td>
<td>5.7 (1.4)</td>
</tr>
<tr>
<td>V</td>
<td>52395.53 (13474.73)</td>
<td>5.3 (1.5)</td>
</tr>
<tr>
<td>T</td>
<td>49927.91 (8765.34)</td>
<td>5.6 (1.2)</td>
</tr>
<tr>
<td>AV</td>
<td>49184.60 (6390.88)</td>
<td>5.9 (1.4)</td>
</tr>
<tr>
<td>AT</td>
<td>49501.02 (12552.18)</td>
<td>6.1 (1.5)</td>
</tr>
<tr>
<td>TV</td>
<td>50274.69 (10051.76)</td>
<td>5.8 (1.4)</td>
</tr>
<tr>
<td>ATV</td>
<td>46916.04 (8109.92)</td>
<td>6.4 (1.2)</td>
</tr>
<tr>
<td>NONE</td>
<td>55168.88 (14464.42)</td>
<td>4.2 (1.3)</td>
</tr>
</tbody>
</table>

Table1. Descriptive statistics Mean (SD) for each condition

We analysed our experimental results with 2x8 repeated measures ANOVA on overall task performance. We found a marginally significant main effect of condition (F(7,112) =1.977,p=0.06). To investigate this further we grouped together conditions where each of the sensory cues was on and off. Overall mean times when cues were on and off can be seen in Figure 3. After performing a paired sample t-test we found a significant effect of audio (p < 0.05) and tactile cues (p < 0.05). This suggests that participants performed significantly better when audio and tactile cues were on as oppose to off.
Significant main effects of three factors used in the factorial design: Tactile, Audio and Visual. Error bars represent standard error of the mean. Furthermore, we found a significant correlation between objective and subjective data (r = -0.979, p < 0.001). This suggests that when participants reported an increased sense of presence they completed the task faster (see figure 4.).

Figure 4. Correlation between subjective and objective measures

3.1. Modulation of the environment

To investigate the effect of the modulation of the environment we compared objective and subjective ratings and we found a significant negative correlation between the feelings of discomfort and perceived sense of presence (r = -0.613, p < 0.05) (see figure 5). This suggests that when participants reported increased feelings of discomfort their perceived feeling of immersion and presence decreased.

Figure 5. Correlation between discomfort and sense of presence

4. Discussion

The presented study was designed to investigate the potential beneficial effects of multisensory feedback on human performance in association with unimodal, bimodal and trimodal sensory cues. Our results show that trimodal feedback (AVT) was the most preferred type of feedback followed by bimodal (AV, AT, TV) and then unimodal feedback, which is in line with previous research [LS08; BM07; KHJ*12; JEE04; AMH95]. We also investigated favourable effects of multisensory feedback on perceived sense of presence in virtual reality environments. Our results clearly show that objective and subjective measures were enhanced by presentation of multimodal feedback. As previous studies have suggested these results may reflect the fact that multimodal feedback can maximise human physical abilities as well as enhance users’ sense of presence and immersion in the VR environment [LS08; BM07; LSS*12]. The main findings of our study is to suggests that we need to include user experience when investigating the usability of feedback signals [KHJ*12]. We argue that the auditory, tactile and visual cues are important additional cues that add to the objective performance as well as subjective evaluation of VR environments.

5. Conclusion

In order to support training and performance in VR it is essential to provide necessary sensory cues that are required for the task. These cues can be presented in uni-modal, bimodal or multimodal modalities, including different viewing perspective and stereoscopic presentation. The results from our study show that multimodal feedback can have the most favourable effects on users’ perceived sense of presence and task performance. Future implications of our research suggest that even though the additional multisensory cues are pseudo-realistic i.e. they provide relevant information in an unrealistic fashion; they are still enhancing the user’s sense of presence and helping in task performance. For our future research we propose to investigate whether the multisensory cuing will support the transfer-of-learning between real and virtual settings. An understanding of conditions and multisensory cuing under which VR users experience a enhance sense of presence and performance...
give us a valuable insights into human cognition and psychology [KHJ* 12; BH95; PR14]. Furthermore, it can help designers to allocate proportionally computational resources when building future designs of the virtual systems with multimodal feedback.

References:


Towards New Interconnected VR and 8K Technology

ARTUR BINCZEWSKI¹, MACIEJ GLOWIAK², MACIEJ STROZYK³

¹PSNC; Poznan, Poland; artur@man.poznan.pl
²PSNC; Poznan, Poland; mac@man.poznan.pl
³PSNC; Poznan, Poland; mackosir@man.poznan.pl

Abstract
The number of VR installations increases every year, new systems are created at universities, research centres as well as in industry laboratories. The quality of images and immersive feeling of each standalone system is almost perfect but there is one feature missing – cooperation and interaction between VRs. Nowadays high speed networks are powerful enough to provide efficient and real-time collaboration between remote VR installations. This combination of two cutting edge technologies – 100Gbps network on one hand and Ultra High Definition (4K, 8K) VR systems on the other hand, opens up new area of research and is able to provide new dimension of VR experience in integrated environments. Network integrated VR technology brings also another benefit, wide remote access to the most advanced installation available only in the biggest visualisation centres.

The article provides state-of-the-art in video network transmission and UHD visualisation systems as well as some initial ideas and examples of the first experiments showing the possibility and power of integrated solutions. It also shortly describes new 8K VR installations at PSNC and research direction towards integrated VR environment.

Categories and Subject Descriptors (according to ACM CCS): I.3.2 [Computer Graphics]: Graphics Systems — Distributed/network graphics

1. Introduction
Network integrated VR environment increase the competitiveness of the European creative industries by offering innovating cutting-edge technology for video transmission over high speed networks solving theirs two important needs: “Move data, Not People” by facilitating remote production and collaboration and “More, faster and better pixels” by easing the access to new 8K format, which provides higher frames rates and higher dynamic range with better colour and brightness. There are many advanced appliances and opportunities of Ultra-High resolution network interconnected visualisation infrastructures such as: scientific visualisation of large-scale, complex models or remote collaboration in combination of low-latency uncompressed video streaming. Many research projects and experiments require an integrated networked visualisation facility and immediate video data transfer especially if special interaction is required. Astronomical and medical branch is a good example where the innovative technologies of advanced images streaming and processing are strongly required. Another interesting example may be E-Learning, where medical schools can take huge advantage of UHD video streaming for low-latency live or rendered content. Considering that for many teaching hospitals it is really important to teach the students about a rare or complex operations, creating a networked media system between hospitals and streaming operations from one hospital to audience in another location for teaching purposes is essential. Live streaming of the surgery often needs a combination with rendered models of human organs in order to introduce students or other medical stuff to the details of operation. Ultra-High video resolutions and perfect colour mapping are strongly recommended in such scenarios, as the details of the picture are key for understanding the medical surgery.
The live UHD video streaming and associated advanced interaction require using low-latency networks and the huge amount of data to be transferred needs – depending on the actual resolution – high speed optical networks. This usually means multiple 10G or 100G network connectivity between end-users. Such connectivity is enabled by GEANT network or Global Lambda Integrated Facility (GLIF) infrastructure and locally by National Research and Educational Networks such as PIONIER network in Poland.

In this article, authors shortly describe the Ultra-High definition technology and present the advantages of fast optical network interconnection between advanced visualisation infrastructures such as immersive caves or novel 8K displays giving some examples from own research and demonstrations.

2. The state-of-the-art of high resolution display and streaming technology

In the last 10 years there was rapid adoption of HD technology for professional and consumer market – ranging from TV’s, computer displays, DVDs to digital cameras. New technology beyond HD, called Ultra-High Definition TV, has been adopted from laboratory, by prototype phase, to the public testing and market products. UHDTV-1 resolution and technology was originally used in Digital Cinema as “4K” and finally became new professional and consumer standard.

Today’s networks are capable of transport hundreds of parallel compressed 4K streams, or at least a few uncompressed at the same time. The technology is adopted by television, movie and entertainment industry, but also by scientific visualisation for various appliances (biology, chemistry as well as telemedicine, e-learning and telepresence). However 4K offers very high image quality, scientists and researchers keep thinking about new steps beyond 4K. New technology, called 8K or UHDTV-2 stands for 4 times larger image than the 4K and the resolution is 7680x4320 (8K or 4320p). A single 8K frame consists of 33 million pixels. UHDTV-2 originally developed by NHK, the Japanese public broadcasting organization, became SMPTE standard in 2012 [SMPTE12]. Since then European research laboratories and universities try to be active in new technology, just to mention 8K visualisation installation at University of Essex or the new PSNC laboratory (8K 3D). The research required to deploy of stereoscopic 8K in PSNC is beyond the state-of-the-art and concentrates on integration, testing and development of new applications for encoding, decoding and streaming. However uncompressed 8K network transmission would be great in context of VR visualisation where the low latency is the key aspect, it is extremely demanding regarding network bandwidth and processing. 8K 3D stream would fill or exceed even a 100Gbit link. That’s the reason, the network transmission of big volumes of 8K visual data requires high quality and real-time codecs. JPEG2000 codec is considered as the real alternative to uncompressed video streaming, as the codec is visually lossless, low latency and robustness which is necessary for VR remote visualisation. [NRM*_13]

In 2013 PSNC compared visual differences between 4K video streaming with JPEG2000 compression and uncompressed. Two images were displayed side by side on two same LCD displays. JPEG2000 compressed video of a single 4K stream required 20 times less bandwidth and reached 500Mbit/s. There was no visual degradation of the quality, as JPEG2000 at such bitrates is perceived as a visually lossless codec. The only visible side effect of introducing compression to the chain was additional latency of 4-6 additional frames, what was caused by internal input/output queues of the encoding and decoding devices. Using a high frame rate of 60fps it means less than 100 ms which makes the fast and good quality video compression system still possible in real-time appliances [BGI*13].

3. Joint VR and UHD research in VISIONAIR

Anyway, the combination of Ultra-High Definition and immersive installations is not impossible. Both network-enabled Ultra-High Definition infrastructures and Virtual and Augmented Reality installations were connected in a common project called VISIONAIR (EC funded in FP7). This project focused on building pan-European community around high level scientific visualisation and interaction including Virtual and Augmented Reality, Ultra-High Definition, Scientific Visualisation and Collaborative Environments. VISIONAIR opened key national and regional visualisation infrastructures of 25 partners in Europe for European scientists and enabled research teams to work with world-class infrastructure. VISIONAIR granted effective and convenient access to the best European visualisation infrastructures such as VR, AR, UHD, caves, holographic displays, etc. for more than 120 external projects and experiments in scope of Trans-National Access (TNA). These projects represented various inter-disciplinary areas and domains such as: Biological and Medical sciences, Environmental and Earth sciences, Mathematics and ICT, Engineering, Material Sciences and Analytical facilities, Physical Sciences and Social Sciences and Humanities. There were a few lessons learned from the VISIONAIR project, but one of the main, in context of combination of UHD and VR, was a demand from users on having more pixels in VR installations in order to achieve more natural experience and – on the other hand – to have more interaction and immersive perception in high-resolution displays. Both types of VR and UHD installations seems to be complementary, therefore there is a need of combination of various types of visualisation installations and displays into one platform connected by network. This will provide users and researchers with different types of complementary possibilities in scientific visualisation. Good examples of such combination are two projects conducted as a research in scope of the VISIONAIR project.

3.1. Robot interaction with low-latency live video streaming

The first example of Ultra-High Definition and natural interaction was the demonstration during TERENA Net-
working Conference 2013 in Maastricht. It was performed by three European institutes: SURFNet, PSNC, University of Twente together with Ciena company as network technology vendor. The demonstration has been focused on connection various types of displays and send UHD 3D video streams on the large distance with 40G network with very low latency. This meant use of uncompressed streaming technology.

All the partners devised a demonstration to show what is required to control a remote environment. The main goal was to reach out to researchers, teachers and students and create an interactive show using the newest network and visualization technologies, that would be easily understandable for all participants, even those not related to ICT. On the other hand, the experiment should use cutting edge visualization, haptic and network technologies.

Local participants of the conference were able to control a robot arm to place small remote objects of different size and weight on the tower construction. This tower construction was located remotely in Poznan, Poland. The participants had several visual feedbacks such as: UHDTV 60p, holographic illusion display and 2D and 3D displays through various compressed and uncompressed video streams. Users were able to feel the haptic response from the robot arm.

To control a remote environment, a specific machine interface was required. A haptic robot arm gave humans the opportunity to make a virtual world tangible. Haptic devices were able to perceive and transmit movements very precisely, and provide force-feedback. The person that controls one of the arms was able to feel weights, movements and pressure of the other arm. Both arms were network connected to each other and enable users to control precisely the position and movement of remote objects at a long distance with a very low delay.

Since the robot arms were in different countries, more than 2000 km of light-path away from each other, it was necessary to provide the users with visual feedback. By supplying different types of displays, resolutions and a low latency transmission of live video streams, users were able to watch the remote environment and interact with it. Participants could choose between 5-meter-wide stereoscopic 60p high frame rate UHDTV projection, 4K displays as well as a holographic illusion screen.

For demonstration needs, SURFNet (Netherlands) and PIONIER (Poland) optical networks were connected together via Hamburg without regeneration of optical signal, which could introduce additional delay in transmission. Live video streams were transmitted from the PSNC studio in Poznan, where two 4K cameras were connected to special optical transmission system delivered by Ciena. The total bit stream of transmitted video data reached about 30Gbit/s constantly over 5 days of the conference, what means that about 1.45 PB of data transferred during the conference. The total bi-directional latency between moving haptic device and get the visual feedback was about 60 ms in total, which meant only 3-4 frames of delay. This was perceived as almost no delay by users. The partners of the demonstration successfully combined multiple research and education aspects and the experience collected can be used in various collaboration scenarios for more complex use cases using UHD and VR. [WDD13]. Especially interesting appliance using live streaming with low-latency network is cave-2-cave transmission, where two or more Virtual Reality installations are connected with very fast network (100Gb) in order to ease remote teams to collaborate in real time.

3.2. Collaboration through various visualisation technologies

Another good example of the combination of various types of displays for collaboration was the subject of the research actions carried out in the VISIONAIR project. They aimed in implementation and testing the remote collaboration of multiple designers in locations distributed across the Europe. Research conducted by PSNC and its partners (KTH from Sweden, University of Bristol from U.K., University of Twente from Netherlands, I2CAT from Spain and University of Kaiserslautern from Germany) focused on enabling designers to simultaneously edit complex 3D models and allowing them the live observation of their work results using advanced visualization systems, just to mention holographic table, 3D UHD displays and CAVE. As the basic platform for rendering PSNC proposed VITRALL software run on dedicated distributed computing infrastructure, which was designed and developed by PSNC. For high resolution visualization requirements, the software was modified and adjusted during the VISIONAIR project. [SBK*_12]

The results of development, integration and testing of the visualization and network technologies have allowed to perform two demonstrations: the first one took place during the VISIONAIR Open Forum conference in Poznan in February 2014 and the second one during VISIONAIR workshops at KTH in Stockholm in October 2014. Both of them presented the remote collaboration scenario, where 3D models were edited by multiple creators and visualized across multiple displays – ranging from holographic table, through stereoscopic HD and 4K TVs up to the 3D UHD and CAVE displays in PSNC. Participants from all the locations were able to control locally displayed video image of the edited models and comment all the modifications made by the editor. This scenario allowed people located in different places to collaborate on editing advanced 3D models simultaneously.

4. New visualisation laboratory for UHD and VR in PSNC

In 2015, Poznan Supercomputing and Networking Center, which has rich past experience in deployment 4K and construction VR prototype caves decided to build two new installations using cutting-edge technologies.

The 8K 3D wall (more than 33 MegaPixels) and immersive cave for scientific visualisation will be constructed until the end of year and connected using optical network. Both
installations will be available for collaboration and live streaming from remote laboratories and institutes all over the world. The joint VR and UHD laboratories are depicted on Fig.1.

![Figure 1. Integrated UHD VR laboratory in PSNC](image)

The 8K 3D laboratory in PSNC, as a platform for testing and development 8K technology and codecs, will be composed of three major installations connected with 40Gb/100Gb network and will form an integrated multimedia environment for recording, processing, transmission and visualisation of UHDTV-2 quality content. These three essential elements are: 8K 3D 60p display wall for advanced scientific visualization, uncompressed and compressed streaming system by 10/40/100Gb network and content acquisition system build of two advanced 8K live cameras placed in 3D rig.

Besides the UHD set up, PSNC is also deploying new cave installation with total number of pixels reaching 18 Million, which makes the installation similar to UHD display, but introduces immersive and interaction principles. The resolution of the cave walls will be 2765x1920 pixels (for larger walls and a floor) and 1920x1920 pixels (for smaller walls).

Both laboratories will be interconnected with 40/100Gbps network in order to provide integrated VR environment supporting scientists and researchers from various areas of science such as Biological and Medical sciences, Environmental and Earth sciences, Mathematics and ICT, Engineering, Material Sciences and Analytical facilities, Physical Sciences and Social Sciences and Humanities. PSNC which itself is affiliated to the Institute of Bioorganic Chemistry has rich experience in collaboration with external scientists and researchers, just to mention VISIONAIR project, where it offered about 12 Trans-National Access.

5. Conclusions

The development of the visualisation technology allows to construct novel UHD displays and immersive caves with lower cost than it was required some years ago. Both the visualisation and network technology is getting cheaper. Therefore the live exchange of high resolution and high frame rate video content between various types of visualisation infrastructures is easier nowadays and in the future will support more complex and advanced research. Conducted experiments and demonstrations proved that even now it is possible, however requires a lot of technology integration. This is possible area for further collaboration in international projects or European initiatives such as Euro VR.

References


A Layer-based 3D Virtual Environment for Architectural Collaboration

Susanne Schmidt, Gerd Bruder, Frank Steinicke

Department of Computer Science, University of Hamburg, Germany, {schmidt, bruder, steinicke}@informatik.uni-hamburg.de

Abstract
Architectural design processes involve a variety of users with different levels of expertise such as architects, engineers, investors or end customers. An efficient process requires all involved parties to obtain a common understanding of the architectural models and problems to be discussed. This is an ambitious task as architects as well as other involved parties often need to work with two-dimensional (2D) floor plans. While these plans are meaningful and easy to interpret for professionals, ordinary users often face problems when deducing three-dimensional (3D) properties of a building.

In this paper we address this problem by introducing an immersive virtual environment (VE) for collaborative exploration of virtual architectural models. We explore a layer-based visualization method, which stacks 2D floor plans in space providing a simple 3D impression without actually using a 3D model. Based on architectural work processes we developed a user interface including two registered representations of the same building. Our user interface allows an architect to specify a region of interest within a 3D overall view while other participants can follow his perspective in a second 2D view. In our setup the virtual building is displayed on two separate walls of an L-shaped projection system.

Categories and Subject Descriptors (according to ACM CCS): H.5.1 [Information Interfaces and Presentation (e.g. HCI)]: Multimedia Information Systems—Artificial, augmented, and virtual realities; J.5 [Computer Applications]: Arts and Humanities—Architecture

1. Introduction

Visualization is an important instrument in architectural review and design processes, and various computer-aided design (CAD) tools are available to support architects. Since design proposals and revisions have to be communicated between multiple people with different roles and objectives, collaboration is an essential aspect of architectural design processes. In order to get a better understanding how collaborative 3D user interfaces can support this process, we briefly summarize the typical building design process and its different stages [PS02, Tho63]:

1. Before a first architectural design can be developed, the architects have to gather information about the upcoming project. This includes interviews to identify the client’s expectations and requirements while external constraints must be considered, for example as specified by a set of regulations of the urban planning departments.

2. The analysis of the collected data provides a basis for the subsequent schematic design phase. In this stage architects develop initial ideas about basic shapes and proportions of the building as well as its appearance in the cityscape. In addition to rough sketches, physical 3D block models are used in order to facilitate a better communication and evaluation of design proposals.

3. If the conceptual proposals are approved by all decision-makers, the architects work out more details. The exact partitioning of floors into rooms is defined and interior elements are incorporated, including windows, doors, lights, power points, cable channels and so on. Furthermore, the used materials are also discussed at this stage.

4. Finally, the architects document all design decisions in a full specification, which is used as the basis for the real construction work.

Throughout all stages several iterations are required to integrate the feedback of the involved stakeholders such as building owner, investors, occupants, maintenance engineers...
or other authorities. Thus, the communication between all parties plays an essential role in successfully completing the project.

We are partners in such an urban planning process for the construction of a complex government building. Within this process, we perform focus groups and field studies with the involved parties to gain a better understanding of the requirements and constraints, and to simultaneously develop solutions, which support the entire process.

As virtual reality (VR) researchers, who have cooperated several times with architects in the past, the most surprising observation that we made right from the beginning was the limited usage of 3D visualization or VR-based exploration during these complex processes. Indeed, virtual design tools such as 3D modeling, simulation as well as game engines are becoming increasingly sophisticated, and building information modeling (BIM) is becoming a better known established collaboration process in the construction industry [Azh11]. However, many of the construction managers, architects and engineering firms still rely on and carry out most processes with 2D representations. The reasons are manifold. For instance, stakeholders usually involved in the complex construction projects such as structural engineers, electricians or other technical firms still rely on and plan their work on 2D floor plans. The plans can easily be viewed on desktops, laptops, projection walls or even printouts. Furthermore, during architectural education often two-dimensional drawings, illustrations and plans are used [Yee07]. As a matter of fact, these drafts are usually subject to many changes in the early planning stages, and adjustments in a 2D plan are less time-consuming and expensive than in their 3D counterparts.

For these reasons, 3D modeling is often outsourced to external partners. However, more and more owners are increasingly requiring BIM services from construction managers, architects and engineering firms, because non-professionals in the field of architecture often face difficulties when deducing 3D properties of a building from 2D views [BSVH10]. Miscommunication and misunderstandings at the early stages are often fatal and expensive, and hence it is crucial to identify and communicate problems and open issues in an early stage of the development process.

Putting these considerations together, in this paper we present a layer-based 3D interface prototype, which we developed in the scope of an architectural development process to allow multiple users to explore and review a building design collaboratively. The approach simply uses stacked 2D floor plans with the goal to receive a realistic impression of its 3D properties even if no actual 3D building model is available.

The remainder of this paper is structured as follows. Section 2 resumes projects that are related to the use of virtual environments (VEs) in architecture and construction. Section 3 describes the developed hardware and software setup, the visualization approach as well as concepts on how to interact with the virtual building. In Section 4 we present a pilot study that we conducted to evaluate the usage of the layer-based VE in the scope of an architectural design process that we are involved in. Section 5 concludes the paper.

2. Related Work

Architectural collaborative 3D user interfaces based on virtual reality [BF97, Why02] or augmented reality [Wan09] technologies have a variety of applications in the fields of architectural design, review, presentation, decision making and construction. The goal of most of these projects is to improve the scheduling and coordination within the professional project team and therefore save time and cost as well as supporting interactions with clients, managers and end-users [Why03].

In [PS02] several trends using VEs within the architectural domain are discussed. In particular, a lot of effort has been spent into the conceptual design and realization of VEs that aim to support architects in designing and constructing 3D buildings [AMR06, AEI03, RFK’98]. The intention of those VEs is to provide users with the functionality of typical CAD tools within an immersive virtual environment (IVE) allowing a more natural and intuitive interaction.

During an architectural design process many decisions have to be made not solely regarding the entire building but also specific parts of it. In order to support these decisions, mixed reality setups have been introduced [DSD’02, WD13], in which physical scale models are augmented with virtual parts. The users perceive this virtual content by wearing video or optical see-through head-mounted displays (HMDs).

Another approach are immersive walkthroughs, in which virtual 3D models are explored at real scale from an ego-centric perspective. For instance, [BSVH10] developed a virtual studio system for architectural design based on chroma keying approaches, in which users could perceive real-world objects in an augmented virtuality setup. In particular, users could see their own body and use tools like rulers while immersed in the VE.

Although IVEs based on a tracking system and stereoscopic display can be used to explore architectural designs from an ego-centric perspective, natural interaction in such setups is limited due to restrictions of the range of tracking sensors or physical obstacles. A solution to this problem are redirected walking techniques [RKW01]. For instance, the ArchExplore project [BSH09] implemented a redirected walking strategy, which allowed users to explore a large virtual architectural model while remaining in a limited physical workspace. The project also included virtual portals to connect different virtual locations or VEs representing alternative design proposals.
Figure 1: Illustration of the L-Shape consisting of a front and a floor screen. The inset shows a Wiimote with magnifier lens and attached tracking markers.

Such immersive setups can provide architects and end-customers a spatial impression of a building’s room layout and interior. However, most of the existing setups only consider one user at a time. For architectural design reviews, multiple, potentially remotely distributed participants, have to be considered. The establishment of a shared interaction space enabling collaboration within such project teams is the objective of the GreenSpace project [DC96]. They provide two different interfaces: an immersive interface using a HMD and an on-screen interface. Independent of the selected interface, all participants can communicate over a network audio system. In order to support following a partner’s perspective, every user is represented by an avatar in the shared interaction space.

A comparable approach for multi-functional co-located or distributed teams is presented in [FWB13]. They describe a generalized software architecture for building a collaboration platform with different layers of abstraction. Their idea is to use two separate spaces for public team interaction and a private view respectively. In addition to a shared 3D design desktop interface, a stereoscopic real-life size digital mock-up enables users to explore the currently discussed design in an immersive way.

3. Layer-based 3D Virtual Environment

In this section we describe our layer-based 3D virtual environment and user interface including the hardware and software components as well as the visualization and exploration techniques.

3.1. Hardware Setup

For the visualization of architectural models we use an L-Shape projection setup as the main VR environment, but we also support other VR hardware setups such as a tracked HMD real walking space. The L-Shape consists of two projection screens arranged at right angles as depicted in Figure 1. Two stereo-capable ProjectionDesign F10 AS3D projectors with SXGA+ 1400×1050 pixels resolution and 60Hz per eye update rate, one for the front and one for the floor screen, are arranged behind the L-Shape at a height of about 2 meters. In order to reduce the required distance between projectors and screens, two mirrors are used reflecting the images downward to the floor and to the back of the front screen respectively.

For stereoscopic display users wear active shutter glasses that synchronize with the projectors using DLP link technology. For head tracking five retroreflective markers are attached to the glasses. We use the ARTTRACK2 system by A.R.T., which offers high accuracy within a tracking distance of up to 4.5 meters. For good tracking coverage and robust tracking under occlusions we use seven infrared cameras, which are mounted to the ceiling and corners of the L-Shape. The tracking system is accessed by a remote workstation using the DTrack2 software. The tracked six degrees of freedom of the identified targets are sent via the VRPN protocol [VRP], which is used for 3D interaction as well as head tracking. The corresponding VRPN client software runs on an Intel graphics workstation with a Core i7 3.4GHz CPU and Nvidia Quadro K5000 graphics card. Unity3D Pro is used to render the 3D scene.

We use two input devices for interacting with the architectural 3D models. First, a Wiimote controller is connected to the workstation via Bluetooth. In the current setup only the buttons of the Wiimote are used for input. The second input device is an off-the-shelf magnifier lens that is equipped with 4 additional passive markers as illustrated in Figure 1. Its position and orientation are tracked by the ARTTRACK2 system and can be assigned to an object in the Unity3D scene.

3.2. Layer Construction

As explained in the introduction, 3D data for a planned building is often not available in the early development phases. To allow users to get a spatial impression of the building even without existing 3D data, we stacked the existing 2D floor plans as illustrated in Figure 2a. Sectional views can serve as additional source for proper ceiling heights as those are essential for obtaining an intuitive understanding of proportions and dimensions.

This layer-based visualization is realized by stacking floor plans with an adjustable offset. While this step is done manually in our prototype, it can be easily automated as long as the 2D input data complies with some basic formatting rules. Besides the building itself the Unity3D scene contains a virtual representation of both the front and the floor screen. Hence, after calibration of the projectors and tracking system, the scene is an exact virtual replica of the L-Shape and therefore facilitates a proper scaling and positioning of the building.
In order to support the collaborative process, we render the 3D building model on the front screen creating the illusion of displaying a 3D block model standing on the floor (or alternatively on a pedestal in mid-air). In this setup the front screen can display additional information like a more detailed or labeled 2D plan of the currently selected floor.

3.3. Collaborative Layer-based Interaction

The goal of this project is to create a 3D user interface for exploring architectural 3D models with a strong focus on collaboration. The described L-Shape setup is a suitable environment for this purpose as it allows multiple users to share one single interaction space. The challenge is to provide interaction concepts ensuring that all involved persons concentrate their attention on the same part of the model. To avoid the emergence of disorientation all scene navigation is executed by a single person who we refer to as the operator. This position is typically held by the architect or the person currently guiding the conversation. The operator’s head is tracked and the head’s pose is coupled to the virtual camera. Other users can wear shutter glasses as well, allowing them to perceive the model stereoscopically. However, in order to get a less distorted perception of the model they stay close to the operator. In a focus group meeting with architects three different exploration modes were identified as particularly helpful. Using the buttons of the Wiimote the operator can switch between these modes:

1. In the **overview mode** the building is treated as a single uniform 3D object, which can be rotated around the y-axis (see Figure 2a). In this mode the overall appearance of the 3D building as well as its context can be examined without showing too much detail concerning the interior. Additional virtual contents like 2D face textures applied to the outer walls of the building or a road map projected onto the floor screen of the L-Shape can support this stage of exploration.

2. In the **highlight mode**, one particular floor can be highlighted while all overlying floors appear transparently. Additionally, all floors move upwards or downwards to guarantee the active floor is always on an easily accessible level. This mode is well suited for focussing on a specific floor without losing the general view of the building.

3. In the **focus mode**, the building can be folded to the currently selected floor in order to focus the user’s attention on this specific floor. As single rooms and their labels are clearly visible in this mode, it enables a more detailed exploration of the building’s interior.

The last mode provides a second 2D view of the active floor on the front screen. Thus, users can take part in the discussion of a specific floor without the need of following the operator’s movements. Furthermore, they do not have to enter the L-Shape or wear shutter glasses since the projection of the active floor plan on the front screen is displayed centered around zero parallax.

In order to guide the users’ attention on what the operator is currently focusing on we exploit a second input device, i.e., the tracked magnifier lens. The purpose of the lens is to allow the operator to specify a region of interest in the 2D floor plan. By moving the lens above the three-dimensional view of the building projected on the floor screen, a circular area is highlighted in the 2D floor plan on the front screen; all parts of the floor that are not within this region are culled out. The interaction model of the magnifier adopts the typical real-world interaction. Moving the lens in the horizontal plane results in a motion of the spot on the vertical front screen. While the 2D floor plan is magnified by an initial factor the operator can increase the zoom level by moving the lens downwards.

4. Pilot Study

In this section we describe the evaluation of the current design of the user interface in the iterative human-centered design process.

4.1. Study Design and Procedure

In order to gather information about the usability of the interface we conducted a pilot study with 8 future inhabitants (1 female, 4 males, ages 28 to 45, M=35) of a recent building planning process of the University of Hamburg, for which we received the architectural documents and, in particular, the 2D floor plans as annotated PDF documents. The building consisted of a basement as well as 11 stories as shown in Figure 2.

We performed the study with a two-stage procedure:

1. In the first stage we displayed the 2D floor plans using Adobe Acrobat on a 55 inch multi-touch tabletop, around which the participants were gathered.

2. The second stage consisted of the participants moving over to the L-Shape projection setup, in which the floors were displayed using the layer-based VE described in Section 3.

We purposely chose a simple PDF viewer to receive an impression of our prototype’s features and interaction techniques rather than comparing our system to a professional CAD software with a specific toolset.

The tasks we gave the participants consisted of finding their future office rooms in the building, and following the paths they had to take to reach the rooms from the main entry of the building. We asked the participants to discuss the observations they made about the spatial properties of the building during the collaborative process using the **think-aloud** protocol [Lew82]. While the task involved mainly collaborative touch interaction in the tabletop setup, one participant...
assumed the position of the operator in the L-Shape environment. After 20 minutes of active discussions in each phase we asked them to come to a conclusion. After the collaborative phases, we asked them to fill out NASA Task-Load-Index (TLX) questionnaires [Har06] and performed a debriefing with the participants, encouraging them to comment on the advantages and disadvantages of the user interfaces. The study took approximately one hour in total.

4.2. Results and Discussion

In the following we present the questionnaire results and subjective comments during the pilot study.

**Task Load** We analyzed the questionnaire data with Wilcoxon signed ranks tests [Ots15]. The results of the NASA TLX questionnaire show a significant difference for mean task load of 67.59 (SD = 21.29) for the table condition and 26.97 (SD = 21.69) for the L-Shape condition, Z = 2.02, p = .04. In particular, we found a significant difference for mental demand between the table (M = 68.60, SD = 23.01) and L-Shape (M = 29.00, SD = 27.63) conditions, Z = 2.02, p = .04. We found no significant difference for physical demand between the table (M = 35.20, SD = 28.35) and L-Shape (M = 24.40, SD = 25.16) conditions, Z = .944, p = .35. We found a significant difference for temporal demand between the table (M = 59.20, SD = 24.77) and L-Shape (M = 23.20, SD = 18.57) conditions, Z = 2.02, p = .04. Additionally, we found a significant difference for effort between the table (M = 69.00, SD = 15.54) and L-Shape (M = 28.40, SD = 23.27) conditions, Z = 2.02, p = .04. Moreover, we found a significant difference for frustration between the table (M = 72.20, SD = 26.25) and L-Shape (M = 21.60, SD = 21.30) conditions, Z = 2.02, p = .04. The results indicate that completing the task in the L-Shape was significantly less demanding than interpreting the floor plans when they were displayed on the table.

**Subjective Comments** We grouped the comments during the think-aloud and debriefing sessions, and identified four main topics:

1. **Misinterpretations:** Throughout the exploration process, a number of uncertainties regarding the architectural annotations occurred. In the layered setup, most of these uncertainties, e.g., concerning the role of a room, could be resolved by the participants by magnifying the region of interest. However, the zooming tool in the PDF was rarely used. The participants also approved the cut-outs in the layered visualization as they helped them to identify wall penetrations. A recurrent point of discussion in the PDF visualization was the currently selected floor, since it did not match the page number. Besides these differences, there were also some annotations in the 2D floor plans that could not be interpreted without help by an architect, e.g., the markings that indicated the direction of stairs. In the layered view these questions could partially be resolved due to contextual information or switching back and forth between different layers.

2. **Navigation:** The process of navigating through the building and finding the entrance took much longer in the PDF than in the layered visualization. Besides the learning effect that appears in the second phase of the experiment, this can be reasoned by the just mentioned misinterpretations regarding the current floor level. The navigation was further aggravated by the fact that all scrolling and zooming operations in the PDF caused a sequential reload of the page content. Additionally, multiple participants stated that switching and selecting floors in the L-Shape is much easier and faster than in the PDF and therefore improves the navigation through the building.

3. **Sense of space:** After every phase, we asked the participants to show the actual location of a specific room in a physical 3D model of the building. In both versions the participants were able to point to this location correctly. However, the ceiling height of a floor could not be inferred from the PDF plans, which was criticized by one participant in the debriefing. In general, the test group
stated that the L-Shape setup provided a better spatial impression of the building than the PDF.

4. Collaboration: Regarding the collaborative aspect of the two compared user interfaces, the opinions in our test group were divided. While some participants felt that it is easier to directly point on a specific location in the PDF, others preferred the magnifier interface for showing something to their group partners. The designation of an operator in the L-Shape setup was judged favorably as it prevents conflicting user inputs.

In conclusion, one participant remarked, that he would prefer to have an actual physical 3D model compared to both visualizations. However, in the absence of such a physical model the layered virtual view was preferred over the PDF.

5. Conclusion

In this paper we have presented a layer-based 3D VE which allows the collaborative exploration of building designs in the context of architectural review and design processes. We proposed a stacked layout based on 2D floor plans to facilitate the discussion and evaluation of designs in early development stages when 3D models are not available yet. The stereoscopic 3D view of the building was supplemented by a monoscopic representation to support interactions with multiple users. To attain a more natural interaction we also introduced different input devices such as a magnifier lens.

In the future it is important to evaluate the comfort and effectiveness of the proposed interaction concepts. For that purpose, a more extensive study involving other existing CAD tools can be conducted. A further comparison with a fully-fledged 3D model would also be useful to investigate which spatial characteristics can be explored in the layer-based visualization and which cannot. We also intend to study the effects of lifting the projection surface to a higher level, for instance by placing a purpose-built table to project the active floor onto the table with zero parallax, which means it is displayed perspective correct for all users. Finally, we will expand our work concerning hybrid shared interaction spaces.

References


A Multi-view and Multi-interaction System for Digital-mock up’s collaborative environment

BO LI\textsuperscript{1}, RUDING LOU\textsuperscript{1}, FREDERIC SEGONDS\textsuperscript{2}, FREDERIC MERIENNE\textsuperscript{1}

\textsuperscript{1}LE2I UMR6306, Arts et Métiers, CNRS, Univ. Bourgogne Franche-Comté, HeSam; \{bo.li, ruding.Lou, frederic.merienne\}@ensam.eu

\textsuperscript{2}Arts et Métiers ParisTech, LCPI, 151 bd de l’Hôpital, 75013 Paris, France; frederic.segonds@ensam.eu

Abstract
The current industrial PLM tool generally relies on Concurrent Engineering (CE), which involves conducting product design and manufacturing stages in parallel and integrating technical data for sharing among different experts in parallel. Various experts use domain-specific software to produce various data. This package of data is usually called Digital mock-up (DMU), as well as Building Information Model (BIM) in architectural engineering [SNA12]. For sharing the DMU data, many works have been done to improve the interoperability among the engineering software and among the models in domains of mechanical design [FR07] and eco-design [RRR13]. However, the computer-human interaction (CHI) currently used in the context of CE project reviews is not optimized to enhance the interoperability among various experts of different domains. Here the CHI concerns both complex DMU visualization and multi-users interaction. Since the DMU has its multiple representations according to involved domains [Par04], therefore when various experts need to work together on the DMU they may prefer their own point-of-view on the DMU and proper manner to interact with the DMU. With the development of 3D visualization and virtual reality CHI technology, it is possible to devise more intuitive tools and methods to enhance the interoperability of collaboration among experts both in multi-view and multi-interaction [NA13] for co-located synchronous collaborative design activities. In this paper, we discuss the different approaches of displaying multiple point-of-views of DMU and multiple interactions with DMU in the context of 3D visualization, virtual reality and augmented reality. A co-located collaborative environment of CHI supporting system is proposed. This collaborative environment allows the experts to see respectively the multiple point-of-view of the DMU in front of a unique display system and to interact with the DMU in using different metaphors according to their specific needs. This could be used to assist collaborative design during project review where some decision on product design solution should be made.

Categories and Subject Descriptors (according to ACM CCS): J.6 [Computer-Aided Engineering]: Computer-aided design (CAD), Computer-aided manufacturing (CAM)—Collaborative design; H.5.2 [INFORMATION INTERFACES AND PRESENTATION]: User Interfaces—Multi user interface, display, virtual reality

1. Introduction
Concurrent engineering (CE) has become a widely used approach in industry. In compare with traditional sequential engineering (SE), CE has changed the modality of Product Lifecycle Management (PLM) from sequential to parallel. PLM activities, including design, analysis, manufacturing, recycling and maintenance, are well arranged with proper overlaps at the same time [SE98]. It is an integrated product development process strategy with which everyone involved works collaboratively in parallel in order to reduce the overall product development time [SR09]. The scientific contributions of this paper are: (a) A concept of computer-human interface (CHI) supporting multi-view and multi-interaction for collaborative design. (b) A state of the art of multi-view display and multi-interaction. (c) The appropriate solutions have been described and proposed.

1.1. Point-of-views of DMU
Every product lifecycle activity of CE needs expert and domain-specific computer tools. These tools mainly contain
the computer aided design (CAD) of product, computer aided engineering (CAE) for physical analysis, computer aided manufacturing (CAM) for all operations of manufacturing and other software from different domains. Various experts use domain-specific software to produce various data [Par04]. Each expert considers his/her own contribution to the product as one point-of-view (POV) of the whole product development according to his/her expertise. Then he/she shares his/her information with other experts by sending the data produced by the domain-specific software into a global database [GD07]. The large package of data itself, together with the product structure and attributes of this data package builds up a Digital mock-up (DMU) in industrial engineering. It concerns the generation and management of digital representations of physical and functional characteristics of industrial products. Building Information Model (BIM) is similar to DMU for architectural engineering [SNA12]. BIM model is also a set of interacting policies, processes and technologies containing building design and project data in digital format throughout the building’s life cycle [Pent06].

DMU can present data with different meaning and form a series of data in different modalities. E.g. in automobile engineering, a sketch, 3D parts as well as an assembly of them, a point cloud as well as mesh model in reverse are all possible representations of DMU, as shown in Figure 1. That means a DMU has multiple representations. Meanwhile, from the expert’s position, every expert has his own POV of the DMU. The POV decides which data in the DMU will be put out by expert. If two experts focus on a same DMU representation, e.g. a mesh model representation, they will have the same data format. But their POVs may still be different because their specific requirement e.g. one expert for the whole mesh model while another for a tiny sub model. Their POVs of this representation is different. This also causes the differences of data resource and data quantity.

![Figure 1: During entire PLM, multiple experts work with one unique DMU. Each expert has a POV of DMU, such as a sketch, a single part, an assembly of a component or the whole car, CAE model for simulations, exterior design and a point cloud or mesh model in reverse engineering.](image)

As described above, DMU offers experts many representations. It usually provide geometry data to CAD tools for part design, interference examination between parts and assembly [WLG09], assembly process design, maintenance design, kinematics simulation; provide mesh and constrain data to CAE tools for Finite Element Analysis (FEA) such as structural simulation or Computational fluid dynamics (CFD) calculations in the aerodynamics or thermal simulations; provide geometry and material data to CAM tools for numerical manufacturing process management [FSLF11]. For experts, they might have various POVs of each representation. This will result in various POVs of DMU.

1.2. Collaboration

Since the concept of CE requires the simultaneous progress of all engineering aspects, each expert should communicate with others in real time on the status of the product development that he/she is working [MMOR13]. Thus, the communication among both the domain-specific softwares and the experts is increasingly important. Many works have been done to improve the interoperability among engineering software and among the models [BN08]. E.g. [FR07] and [RRR13] proposed a model-based approach for the design of mechanical products. The model exported by several expert tools can be shared as collaboration knowledge in domains of mechanical design and eco-design.

In architecture engineering, BIM software could integrate all the domain specific representations of a building along its lifecycle in one information technology (IT) platform and save them into a unique file. In compare with using several elementary domain specific softwares, BIM software overcomes the problem in terms of interoperability among them [GJG10].

Since product lifecycle activity of CE depends not only on computer tools but also on the factor of human beings as well. As interoperability is the communication among tools, collaboration is the communication among experts. Since interoperability among tools has been improved, the facilities on communication among experts have also to be enhanced.

During the product development, the activity that mostly needs the collaboration of all experts of PLM is project review. It is arranged regularly as milestones during collaborative design. Meeting support systems [KS95] can be employed to support creative activities in collaboration. The development of the tools to support design during project reviews is also important [Joh88].

Project review during product development can strongly summarize the current work and assign the work of next stage by making modifications and proposing solutions to both strategical and technical details [SE98]. The content for project review normally relies on the information generated from DMU [FSLF11]. A DMU provides different information representations and each expert can choose the one from his/her POV. Simultaneously the experts exchange
their opinions of several domains according to their specialities [PFGL08]. Then they could discuss and communicate in real time.

To enhance the collaboration among experts from different domains to communicate with DMU, a novel CHI has been taken into consideration. CHI mainly concerns both complex DMU visualization and multi-users interaction [CNM’05]. In this article, we aim at the visualization technology of DMU’s different POVs and multi-interaction technology for multiple users activities applied in a co-located synchronous project review support system.

1.3. Multi-view visualization

Visualization is a very important part of CHI in the context of collaboration. As human beings, visualization is the most effect way to accept information, to understand the intention and to take a decision.

The normal co-located synchronous project review support system is with private view devices like laptops and tablet computer. Every expert gathers together in a meeting room with their laptops in their own hands. A screen is also usually available to show shared information among the experts; one can display the content from his/her private view on the shared screen in order to diffuse the information to everyone. As we can imagine, everyone wants to show others the opinion in his/her domain. But the fact is when one sees others’ view with the information not familiar with, he/she still cannot understand those domains. This is because the experts from all other domains with all different technical, different educational and cultural level, even simply different language backgrounds [CNM’05]. They don’t have the same knowledge in their mind and cannot exchange information immediately in real time. This reduces the effect of communication and increases the difficulty of discussing and negotiating with others.

On the other hand, many commercial DMU and BIM platforms can integrate design, analysis and manufacture. However, an expert only uses a part of the platform to finish his work. Separate displays, like using single laptop or screen wall that put several separate screens together, display different domains of information separately. Expert has to exchange eyes and body to deal with the information fragments. This may reduce the expert’s concentration psychologically and increase the possibility of misunderstanding and complex of communication [ZW14]. When attending a project review, in which facial expressions and hand gestures interaction are important to express ideas among each other, experts requires more face-to-face communication. If an expert can be presented only with his own POV in a shared visual space with other experts, he can avoid switching eyes between another expert and himself. This will help the expert to communicate and collaborate [ABM’97] with others and also to overcome the sense of isolation that happens when experts use their own tool in his laptop to attend this project review. Thus, a co-located multi-view DMU representation support system is proposed.

We can imagine an ideal collaborative working status, which is presented on Figure 1.

![Figure 1: A case study of usual collaborative situation described by BIM experts working collaboratively using one screen in real time:](image)

Expert 1, field of expertise: construction, proposed a modification of the BIM model;
Expert 2, field of expertise: urban engineering, obtains the modification effect in urban POV of DMU in real time, expressed approval of Expert 1;
Expert 3, field of expertise: structure analysis, finds a conflict in the building structure POV of DMU, expressed an opposition and conducted a further discussion with Expert 1;
Expert 4, field of expertise: building design, has got not much change in his POV, chose to stay and wait for the discussion result.

Many stereoscopy technologies have been widely used to represent 3D images. For DMU and BIM, normally standard commercial tools are widely used: CATIA, Autodesk Inventor, AutoCAD, Revit, Civil 3D, MACAO (Microstation), Vianova Virtual Map, etc. With the development of virtual reality technology, the approach of representation of DMU became diverse. In related work of this article, we discuss the main stereoscopy technologies and their approach in extending to multi-view 3D.

1.4. Multi-interaction

Interaction allows human and machine to communicate with each other. Here we concern about how different experts interact with the various POVs of DMU. The multiplicity of interaction is an important criterion of CHI. Intelligent CHI will allow users to interact with multiple metaphors and interpret one metaphor to more than one single command [MHPW06].

However, multi-Interaction has two levels of meaning. From the technical level, multi-interaction means multiple interaction devices [SKV14, Her08]. As far as we could imagine, 3D visualization and vision techniques, 3D sound
technologies and haptic devices like force feedback and tactile feedback, all of these devices could give the user one or more interaction methods with the DMU. They bring the user not only the visual perception, but also the perception of immersive sound and touching effect on virtual objects [Mer10]. Multi-interaction can support a variety of creative work for group experts’ alternating activities like collaborative discussions and presentations [GJPR10].

From an interaction metaphor level, multi-interaction means that different user-defined metaphors can be conducted in real time [WMW09]. As listed in Table 1, when interacting, two experts may choose metaphors to use on objects and obtain some results. We put a “Y” in the table for the situation that various metaphors result in alternative meaning. So multi-interaction can be summarized as: One interact metaphor can be used by different experts and generate different meaning according to the experts’ domains [PU97]. Similarly, two experts interact with the same object but their interaction metaphor (gesture) may be different.

Each expert could choose interaction metaphors different from the ones chosen by the others in virtual navigation and manipulation of the model [BCC08]. A series of problems deriving from interaction metaphors will be discussed in the future work. Only a group manipulation interaction metaphor problem will be mentioned here to illustrate.

Experts manipulate the model by modifying (addition, deletion, rearrangement etc.) its parts or the elementary sub models [BISGM02, WL06, MZTZ04]. E.g. modifying airplane rivets, building or deleting pipes. One short interaction should be taken instead of repetitive interaction with relative tasks. Group manipulation for a certain category of objects according to certain rules will reduce the length of intervals of operation.

Many 3D stereoscopy technologies have been widely used to represent 3D images. For DMU and BIM, normally standard commercial tools are widely used: CATIA, Autodesk Inventor, AutoCAD, Revit, Civil 3D, MACAO (Microstation), Vianova Virtual Map, etc. With the development of virtual reality technology, the approach of representation of DMU became diverse. In related work of this article, we discuss the main 3D stereoscopy technologies and their approach in extending to multi-view 3D.

| Table 1: Various experts can choose the same metaphor to use to interact with one project. The same result or not indicates whether a collaborative multi-interaction has an alternative meaning. |
|---------------------------------|-----------------|-----------------|-----------------|
| Same Metaphor | Same Object | Same Result | Multiple Meanings |
| × | × | × | N |
| × | × | × | Y |
| × | × | × | N |
| × | × | × | N |
| × | × | × | Y |
| × | × | × | N |
| × | × | N | N |

2. Related work

2.1. Multi-view

Normal device for display provides one 2D view to the user, like television, computer and smartphone. Compared to single view display, multi-view visualisation offers more views for more users. Two slightly different 2D views for human eyes can be fused in the human brain for having stereoscopic POV [Dod05]. Since the geometry model of a DMU is usually in 3D format, it will be at least four 2D views for two users to have 3D POVs of a DMU. Many approaches and their applications have proposed multi-view solutions, as well as 3D multi-view solutions if a 3D POV is in need. They may come from the improvement of existing single view and 3D single view. But the purpose of all these approach is to increase the number of views.

[Mis09] and [NUH10] applied the glasses based stereoscopy technology to a multi-view approach. Two original shutter glasses or polarized glasses are restructured by putting two left eye lenses together and two right eye lenses together as two new pairs of glasses. Each user can see one view of the former 3D image through new glass in 2D. This approach is very practical to display a POV in 2D view to multiple users.

[MFP08] describes how to make screen-based autostereoscopic systems display two 2D views for two POVs. The naked 3D parallax-barrier or lenticular sheet screen let the user see one image with two eyes but with slight difference in vertical direction. Adding more images and resetting the parallax in vertical direction, each user works as one eye in 3D display. He/she can have a POV of 2D image in a stable position and from a fixed angle.

[TV15] and [Dis14] improved the shutter glasses technology accompanied with a screen with high refresh ratio of 240Hz. Each of the four eyes from two users is displayed in the ratio of 60Hz, which is the lowest ratio for human being to see clearly. And four eyes could be displayed in sequential separately. In total, four 2D views provide two 3D POVs to the two users for playing two games simultaneously.

[KCZT12] is based on an old style Liquid Crystal Display (LCD) screen which can display clear image only when line of sight is perpendicular to the screen or in a range of field of angle. Taking advantage of this drawback, three POVs can be realized by displaying three different images in the same time. There are two POVs from each side of the screen and one POV from the perpendicular direction in front of the screen.
An immersion CAVE-like display approach for co-localized multi-user collaboration is proposed in [MBT11] and [MB11]. This approach combines technologies of active 3D glass (shutter glass) and passive 3D glass (anaglyph or polarised glass). Two users are displayed separately in two successive time intervals. When being displayed, user’s space position is tracked in real time so that the views in each eye can be modified from the user’s position. Four 2D views provide two 3D POVs to two users. Similar approach is adopted in a co-located multi-view table [LHS’14].

A co-located multi-view system which provides six users different POVs of a virtual environment is proposed in [KKB’11]. Three high frequency (360Hz) Digital Light Processing (DLP) projectors are for six users’ left eyes. Each projector displays only one of the basic colours (red, green and blue) to one left eye with frequency of 60Hz. These three projectors can have 6 views. Adding another 3 projectors for six right eyes plus polarized glasses, totally 12 views, or we could say 6 3D POVs is realized.

As we have discussed before, experts have their own point-of-view of DMU. These different POVs can have internal relationship. They are not isolated but are interacted on each other. On one hand, these relationships are restricted by different experts’ individual requirements; on the other hand, they are links for collaboration among experts involved.

The visualizations of different point-of-view are usually described in two ways below:

- Two experts focus on the same scale of the DMU model, or the same resolution of DMU, which means the contents of these two point-of-views of DMU have the same level of data. E.g. both experts are focus on one part of a product. According to their different speciality, the POVs are different: one POV of FEA in specialty of structure analysis; one CFD POV in specialty of thermal analysis. If one change the geometry structure, both of their simulation analysis result will also change in real time. In this case, multi-view support system can play an important role for collaboration.
- Two experts are not on the same scale of the DMU. E.g. one is focuses on the architecture exterior design of a building and another is focus on electricity wiring design of a certain wall in this building. Second one’s scale requires more detailed data than the first one. For exterior design, the response which is brought by changing electricity wiring is too weak and tiny. But they still have influence with each other because they are still in the same DMU. We can image that electricity wiring design will have effect on wall construction, wall construction will have effect on building structure, and building structure will finally have effect on building exterior design. In this case, multi-view support system seems not having great effect on these two experts.

There are a lot of multi-view display research and applications. From a technical perspective, they definitely display POVs to multiple users, but from a collaboration perspective, whether the task in the application has to be accomplished by two or more users is still a problem.

If two users are having less effect, or even no effect on the working contents, the multi-view displays will not have evident results in collaboration. Sometimes this may cause negative effect. There will be no difference if two users work separately. In some application, the reason why two users work collaboratively only seems to save one device for working.

So we describe the criteria of the collaborative effect that co-located multi-view support system brings to experts.

- **Interference**: co-located multi-view support system brings conflict between users. When multiple users have similar interactions with the displaying content, e.g. by pointing at one position on a screen, two users want to complete an interacting motion but they find physical conflicts. It is better to work separately instead of interference. The multi-view contents have no relation but conflicts.
- **Unnecessary**: co-located multi-view support system brings nothing. If there is no relationship among multiple contents, users will have nothing difference with working on separated screens for each one. So it is unnecessary if the multi-view contents have no relation for collaboration.
- **Help**: co-located multi-view support system brings a lot. Multi-view contents have relations among them so that each view once gets changed, the other views will be updated in real time.

Next we will use our criteria to discuss collaborative effects of multi-view applications mentioned in 2.1.

One of the applications of [Mis09] and [NUH*10] is that two people are looking at a sentence which has been translated into two languages. Each user can understand the meaning when seeing the same screen. These multi-view devices really helped in this application case.

[MFP08] proposed some applications that may be used to multiple 2D view devices. One is to show several layers
of Google Earth map with different geospatial information such as city name. Another application is to show a series of images that vary continuously in light density. Users can see different point-of-views standing in front of different angles with a display. These different point-of-views seem have influence inside, but not that obvious like DMU. These multi-view devices help collaboration or not depend on how users treat these series of images.

High frequency display like [TV15] and [Dis14] provided applications with which the two users play games separately on a same screen. From the point of collaboration, this is unnecessary because there is no difference if they use two screens.

[KCZT12] provided applications of seeing two pictures from different side of the screen and playing cards face-to-face with a judge in the middle. These multi-views of image displaying have no relations among them. For card games, it depends on which kind of games. If the game cannot be separated into several screens, the multi-view display will be strongly help to collaboration.

In the virtual assembly chain application [MBT11], two users collaborate to define the position of a seat by taking charge of different tasks to help each other. This multi-view system really helped the collaboration. However, in this team work application, users should have the same knowledge and specialty of assembly. Experts’ professional domains are fixed.

One application of [LHS’14] is to manipulate pictures separately and to share pictures in a specific zone on the screen. From the point of collaboration, this application can be totally done on two screens because there are no relationship among the pictures they choose, it is unnecessary to have pictures manipulated co-located. Another application is to annotate roads on a map to generate a path. Two users can see two maps of the same region in different large and with different information. One with city roads details while the other with altitude level map. In this application, multi-view of different maps is really helping the collaboration between two kinds of users. Moreover, due to the interaction of both users’ hands and the screen, an interference problem cannot be ignored.

A co-located multi-view system [KKB’11] with six views, has been applied to see and to manipulate a single model. If this approach can be applied to display six DMU point-of-views, this multi-view device will definitely help multiple users’ collaboration.

As we discussed above, for real time collaboration, multi-view support system not only enables different users to share a display device, more importantly, its content and application have a strong requirement of information relation. Unlike a lot of applications that can actually be done separately, project review is a task that experts must work together with strong collaboration. This more practical application of D-MU in industrial product and architecture design has more demand of multi-view support system.

2.3. Multi-interaction

To realize a multi-interaction CHI, many works have been done by extending the existing single interaction to a multiple way. Since touchpad has been a widely used CHI device for single interaction, each touchpad gives the user a personal scene to interact with vision and certain gestures. A lot of multi-interaction approaches developed a group scene, with which users can work together on it, to replace the remaining personal scenes of all the related devices. [HHL’07, GPH’11, ZBC’14, MLPvdH14, SKV14] are extending personal touchpads to an extra group scene or develop a new touchpad into a multi-user device, switchable from personal scene to group scene. This approach develops a device with a group scene to allow more users to work together and can still keep the user independence just like working on a touchpad. However, the metaphor of interaction is also as the former single touchpad, not varying according to different users.

[VB04] presents a CHI for different kinds of users differed in the distance away from a screen. For each of the four users in front of a gesture controlled screen, the CHI system has a special way of interaction. Not disturbing other users, this CHI can help four users interact with the content at the same time. This is a good example for giving different interaction strategies to different users with certain characters. However, if the users could choose their own way to interact, that will be much more ergonomic.

[SBL’07] and [MBT11] provide two users a CAVE based immersion environment, especially with gesture manipulation device, speech recognition device and haptic input device to interact with virtual models in multimodal mode. For different events of manipulation, users can generate a mixed rendering and multimodal feedback, which is useful in complex virtual scenes such as virtual assembly. This is a good example of multi-interaction devices utilization in virtual reality immersive environment.

[SGH’12, GWB04, WPS11] provide special working medium such as a stick or a ring to interact with the virtual object. Meanwhile special metaphors for interaction and a set of interaction principles are defined in a proper way. This might be equivalent to the creation of new device and redefinition the metaphor for the novel device, which gives the users a certain amount of freedom.

[BCC’08] lets users to choose methods to select an object. For example, a user may select an object if he/she holds their hand for more than a specific period, or if they make a rapid poking motion at the object. This approach allows user to define the interaction metaphor according to the user’s willing, which is really helpful for our multi-interaction platform.
As we discussed, the multi-interaction of this CHI system should be multiple not only in using different device, but give users the freedom of define the interaction metaphor as much as possible.

3. Proposing appropriate solutions

We proposed several solutions for our estimated system. The simplest solution is with traditional VR visualisation technology, i.e., Anaglyph and Polarization. One POV of original geometry and one POV of aerodynamics CFD analysis result of an automobile DMU can be displayed on the screen and be separated by glasses. The glasses are formed by two original anaglyph or polarization lenses from one side.

If we superimpose two kinds of glasses, four POVs appear. Obviously the disadvantages of the two devices appear at the same time, i.e., the colour distortion due to Anaglyph and the less brightness due to polarized glasses.

We also proposed another two-view solution using a collaborative polarized table. Users can be provided 3D scenes in two directions. E.g. if one virtual wall is displaying vertically on the screen, each user standing physically at one side of the table can only see one side of the wall that faces him.

Besides multi POVs devices, we proposed an improvement of current device of Holografika [BFA05] with advanced naked eye 3D displaying technology. It has over 30 laser projectors behind the screen. We are looking for a solution with Holografika to project multi-view contents to several users.

For multi-interaction, according to our current device, we propose a method utilizing Kinect in front of a certain screen. Kinect could identify several users and we could define ourselves that the same gesture of different users would react differently.

As the multi-view visualization system for DMU and multi-interaction are two parts of our whole collaborative DMU CHI. Therefore multi-view support system and multi-interaction can be modules of the entire collaborative platform. With the development of the multi-view and multi-interaction support system, these modules are of substitutability.

4. Conclusion and future work

This article describes digital mock-up’s property of multi-representation. DMU contains all the product information during product life cycle in concurrent engineering. Domain-oriented experts have different POVs of DMU so they have problems of collaboration through several professional fields. Thus, a multi-view visualization and multi-interaction system for DMU’s collaborative environment is proposed, on the scope of improving the interoperability among different experts.

For multi-view visualization, many main approaches and their applications have been discussed. Each approach has its apparent advantage and drawback and there is still room for improvement. However, most of the multi-view application has little effect on multiple users’ collaboration. Many multi-interaction technologies have been discussed on both device level and metaphor level. Both levels of multiple interactions are necessary for our ideal system.

We have proposed multi-view support systems progressively using anaglyph and polarization, 3D table and Holografika. We also proposed a Kinect solution for multi-interaction.

Multi-view and multi-interaction support system is considered as part of the entire DMU CHI in collaboration. In the future work, a prototype of proposed multi-view and multi-modal interaction approaches will be developed. Each expert has its own style to interact with DMU, then to obtain the diverse response and finally to be displayed by multi-view support system. A multi-input and multi-output platform for working with DMU more collaboratively will be realized.

References


[FR07] France R., Rumble B.: Model-driven development of


[Mis09] MISTRY P.: Thirdeye: a technique that enables multiple viewers to see different content on a single display screen. In ACM SIGGRAPH ASIA 2009 Posters (2009), ACM, p. 29. 4, 5


[Pen06] PENTILIA H.: Describing the changes in architectural information technology to understand design complexity and free-form architectural expression. 2006. 2


A Tool for Collaborative Decision Making on Service Information Linked to 3D Geometry of Complex Hierarchical Products

Daniel González-Toledo¹, María Cuevas-Rodríguez¹, Carlos Garre¹, Luis Molina-Tanco¹, Arcadio Reyes-Lecuona³

¹Universidad de Málaga; Málaga, Spain;

Abstract
During the lifecycle of many industrial products, different services are applied over the different parts or sub-assemblies of the product geometry, from design and validation tests to maintenance operations. After a service is run, a decision-making process usually starts from the analysis of the service results. This paper presents a tool for analysis and decision-making based on service results linked to the geometry of complex products organized in a hierarchy of sub-assemblies. The tool is based in the use of separate models for the product, the services and the analysis of results, and the generation of different views of the models including a virtual representation of the product with overlapped service result information. The virtual representation of the product consists in a render of a 3D scan of the product where techniques to handle visual occlusion between parts are applied. An application case is presented for collaborative discussion of inspection results of a power plant steam turbine.

Categories and Subject Descriptors (according to ACM CCS): I.3.8 [Computer Graphics]: Applications; J.2 [Physical Sciences and Engineering]: Engineering.

1. Introduction
The lifecycle of many industrial products includes situations in which several stakeholders have to analyse complex information that leads to important decisions. Examples of such situations are the evaluation of alternatives during the design stage or the maintenance actions following inspections during the operation stage. The information analysis process often involves a collaborative discussion between the different actors implied in the process. The analysis may imply decision-making regarding the potential actions to take (change or keep design, replace or repair part, etc).

In many cases, the product consists in a physical assembly of parts defining a product geometry; for complex products, this assembly is usually hierarchical, with several levels of sub-assemblies. In this type of products, most services are linked in some way with the geometry of the product. Design alternative tests are usually focused in one sub-assembly or one specific part, and different maintenance services may be run over different parts of the product or over an area or volume defined within the product geometry.

In this paper we present a tool for analysis and decision-making based on service results linked to the geometry of a complex hierarchical product. Most ideas in this paper can be generalized for many types of products and services, but the tool is presented around an application case in which the product is a power plant steam turbine and the service consists in a full inspection of the turbine for maintenance. The analysis of the inspection results is achieved through a discussion between power plant operators and inspection service engineers, leading to a final decision, which may range from re-scheduling future inspections to scheduling maintenance activities to repair or replace a part.

This application case is one of the cluster cases of EU funded project Use-it-Wisely, in collaboration with Tecnatom S.A, a Spanish company working in inspection and maintenance of power plant machinery. The preliminary work leading to these results was described in [RTG_14].
The rest of the paper is organized as follows. First, we present a summary of related work and an overview of our tool, which consists in a set of views generated from a set of models. Details on the models are given in Section 4, while Section 5 presents the detail of the views. We conclude by presenting specific implementation details of our application case and a discussion including future work plans.

2. Related work

Previous work has acknowledged the need to offer 3D mediated communication tools for industrial collaborative online decision making [SV13], [NLHK15]. The emphasis is in real time collaboration, including text chatting and video conferencing. Integrating synchronous communication in desktop applications, however, limits in practice their scope to simple 3D structures. In contrast, a power plant turbine is a complex structure with several thousand parts to which inspections can be performed. We share with such applications the goal of capturing the information that supports the decision process, and linking such information to the 3D product structure. However maintenance does not necessarily require synchronous collaboration between the actors involved, allowing us to put the focus on managing such structural complexity to support maintenance decision making.

Communication and visualization of 3D structures is fundamental for facilitating decisions regarding maintenance of complex products – e.g. what parts of the product have or have not been inspected or where exactly has a problem been detected. Information associated with complex assemblies is difficult to visualize as some parts obscure others. There are multiple approaches to facilitate 3D structure understanding via occlusion handling techniques. These include using transparency, cutaways [Bur11] or exploded views [LACS08], [KTS09]. In our tool we extend these techniques to interactive web-based visualization. To the best of our knowledge, TeamPlatform 3D online 3D viewer is the only product available with goals similar to ours [Tea15]. However their global transparency and exploded view approaches do not take into account the point of view nor the selected parts in the product. Moreover our choice of Unity 3D Web Player Plugin as 3D rendering framework provides much faster interactive performance.

3. Overview of the tool

Our tool consists in a set of interactive views generated from three different data models. The product model contains the geometrical and hierarchical representation of the product (the steam turbine). The service model contains the results of the service (turbine inspection) linked to different areas of the product geometry. The analysis model consists in a set of collaborative discussions, one for each inspection result. From these models, we generate four different views, as seen in Fig 1.

![Fig 1. Model-view architecture of the tool.](image)

Communication and visualization of 3D structures is fundamental for facilitating decisions regarding maintenance of complex products – e.g. what parts of the product have or have not been inspected or where exactly has a problem been detected. Information associated with complex assemblies is difficult to visualize as some parts obscure others. There are multiple approaches to facilitate 3D structure understanding via occlusion handling techniques. These include using transparency, cutaways [Bur11] or exploded views [LACS08], [KTS09]. In our tool we extend these techniques to interactive web-based visualization. To the best of our knowledge, TeamPlatform 3D online 3D viewer is the only product available with goals similar to ours [Tea15]. However their global transparency and exploded view approaches do not take into account the point of view nor the selected parts in the product. Moreover our choice of Unity 3D Web Player Plugin as 3D rendering framework provides much faster interactive performance.

3. Overview of the tool

Our tool consists in a set of interactive views generated from three different data models. The product model contains the geometrical and hierarchical representation of the product (the steam turbine). The service model contains the results of the service (turbine inspection) linked to different areas of the product geometry. The analysis model consists in a set of collaborative discussions, one for each inspection result.

The different views generated from the models allow the users to navigate through the product to access the service results and run the appropriate analysis. The typical use case of the tool can be seen as a three-step sequence:

1. The user navigates through the product model.
2. This navigation allows the user to access the different results contained in the service model.
3. For each accessed service result, the user may access the analysis model and start or contribute to the analysis of that result.

From these models, we generate four different views, as seen in Fig 1.

We have followed a user-centered design methodology, with two final users directly involved in the whole design, development and test process and with regular prototype test trials run with a representative group of stakeholders. This allowed us to extract the main components of the user mental model, which includes a set of relations between the
product parts: hierarchical relations (this is part of that), assembly relations (this is assembled on top of that), geometrical relations (this is beside that) and service relations (this result includes the inspection of this and that).

To navigate through the product, the user can use either the product tree or the 3D render views. The product tree view is used to navigate through the hierarchy of the product, while the 3D render view is more oriented to navigation through the geometry. The 3D render view shows basic service results information overlapped with the product geometry. To access more detailed information of a service result, when users reach an inspection area in the geometry (navigating through the product views), they may open the service result view associated to that area of the product geometry. From the service result view, the user may open the discussion view to collaborate in the analysis of that service result.

While the implemented product model with its two views can be used in many different applications involving complex physical products, the service and analysis models and their related views are application-dependent. In our case, the service model consists in a set of inspection results and the service result view shows textual information of a selection of inspection results. Our analysis model consists in a set of discussions, in which each discussion consists in a list of (textual and graphical) contributions from different actors. The discussion view shows the historic of contributions of all actors to a single discussion and allows registering new contributions.

4. Models
The tool allows exploration of information contained in the different models. In this section we present a general description of the different type of models with specific implementation details for our application case.

4.1. Product model
Our product model consists of two main parts: the product tree and the geometry data. The geometry data is a set of files containing the description of the geometry of different parts. In our case, our files contain 3D meshes generated through a 3D scan of a real steam turbine and a post-process with 3D modelling software. An alternative would be using CAD representations (either 3D or 2D).

The product tree consists in a tree structure where each node contains: a) Hierarchy information: parent and children nodes; b) Geometry information: associated geometry data and 3D transformations; and c) Assembly information: order in the assembly sequence of the product and direction in which it is assembled.

In our implementation, each leaf node is associated to one or more geometry areas (3D meshes, in our case) from the geometry data. While the product node represents a semantically atomic part of the product, we store separate meshes for different areas of a single part to allow connection with the service model. For each associated geometry area, the node stores also a 3D transform with the position and orientation of the mesh in the global (product) reference system. This allows to build products containing many instances of the same geometrical parts without the need to store separate meshes for each instance.

While the geometry information specifies where each part area is located in the final product (through the 3D transforms), it does not contain information about the assembly procedure of each part. This information is necessary for the generation of exploded-views of the product, allowing to expose internal parts which are completely occluded in the 3D render view. The type of assembly information to store depends on the type of exploded-views to generate, but typically includes an assembly sequence (sequence of explosion) and the assembly direction (explosion direction) of each part.

4.2. Service model
The service model is application-dependent. In our case, turbine inspections are organized in a tree structure based on the different types of elements and areas to inspect rather than on product geometry. Fig. 2 shows an example, in which the turbine/product tree has a node called Rotor, with children nodes corresponding to the different Stages of the rotor, and each stage node with children for the different blade Rings. Instead, the inspection/service tree has a single node for Blades inspection, with children for the different Stages and each stage node with children corresponding to the different areas of the blade to inspect (Profile, Root).

![Fig. 2. Example of link between the product and service models.](image)

The Profile node of the service tree contains a link to all profile meshes of Blade nodes in all Rings of Stage 1. This link between product areas/meshes and service tree nodes allows to access the service results when navigating through the product. There exist three types of nodes in our service model tree: a) Inspection results: leaf nodes, containing the final information we want to access. Each leaf node contains the result of one inspection technique applied over one inspection point in one date; b) Inspection points: are in the second-lowest tree level (leaf-parent) and contain the link with the product areas; and c) Other nodes: all nodes above the inspection point level are used to organize and give a structure to the inspection procedure.
4.3. Analysis model

The analysis model depends directly on the service model and thus is application-dependent. Typically, the analysis model will consist in a data structure linked to the service model, so that different analysis procedures can be run for different service results. In our case, the analysis model consists in a set of Discussion data structures, where each discussion is linked to one inspection point node of the service model. We constrain the analysis model so that one inspection point may be linked to one discussion only if there is at least one inspection result for that inspection point.

Each Discussion data structure consists of: a) Discussion state, which may be: empty (discussion for this inspection point has not yet started), open (discussion currently in progress) or closed (a decision was taken and no more contributions are accepted); b) Link to the corresponding inspection point in the service model (one-to-one); c) The discussion information itself: a set of contributions from different actors, which may include textual annotations, and/or attached images or documents; and d) List of related discussions from other inspection points. The possibility to establish relations between discussions allows the decision making process to be less dependent of the service model structure.

5. Model views

In this section we present the different views generated from our models. Most views are generated from data of a single model, while the 3D render view is generated combining data from the product and service models.

5.1. 3D render view

This view combines information coming from the product and service models. It consists in an interactive 3D render of the whole product allowing the user to navigate through the product geometry. The view allows multiple selection of product parts using the mouse. When the user makes a selection, the context menu is open presenting several options which will be described in the following paragraphs. Since some parts may be occluded by other parts, some mechanisms to deal with occlusion have been implemented.

Navigation

The choice of DoFs for navigation depends on the user mental model of the product. In our case, we found that axial symmetry is what governs the mental model on how to move around the turbine. Navigation based on cylindrical coordinates would be a natural choice for this mental model, but it does not allow all viewpoints to be accessed. We instead implemented navigation along ellipsoidal coordinates, adjusting the forward and up vectors of the camera to keep the turbine shaft horizontal.

When the user makes a selection, the context menu provides a Focus option, which makes a smooth transition to an optimal viewpoint for maximum visibility of the selected part/s. The shape of the navigation ellipsoid is then adapted to the bounding box of the selection.

Occlusion handling

Some parts cannot be selected nor even visualized in the default 3D render of the product. Visual occlusion occurs depending on viewpoint, but there may be parts which are always occluded regardless of viewpoint (as in the case of small parts inside bigger container parts). When the user makes a selection, the context menu provides an option to make that selection semi-transparent. This allows the user to see through container parts, but not to access contained parts inside for selection. To mitigate this problem, the interface provides a slider for setting transparency level with a double function: it gradually changes the amount of transparency of the selected parts until it reaches a threshold, after which the part is completely removed from the render, so that it no longer blocks access to the occluded parts.

Manual selection of the parts to make transparent is a slow procedure when users know exactly which part they want to view or access. Moreover, the selection of parts to make transparent changes depending on the viewpoint, so the user should manually change selection while navigating. To assist the user with transparency choice, when the user makes a selection, the interface presents an adaptive transparency option, which can be switched on and off. Adaptive transparency continuously changes the degree of transparency of all occluding parts while navigating (changing viewpoint) through the product (See Fig. 3).

![Fig. 3. Occlusion handling in the 3D render view. (a): in the default render, a portion of the turbine shaft is occluded by the crowns. (b): when applying transparency to the occluding parts, a portion of the turbine shaft is revealed in which the red color indicates a flaw result in the service model. (c): hiding the occluding parts allows complete access to the occluded part. (d): the exploded view allows access to the occluded part without visual removal of parts.](image-url)
The interface provides also options to generate different types of exploded views of the product. We implement the symmetric and focused layouts of [KTS09] and the interactive riffling of [LACS08]. The exploded views allow the user to see and access internal parts while keeping the spatial relations with the rest of the product, without the need to visually remove other parts. The exploded views are generated using the assembly information stored in the product nodes. In our case, the assembly information was manually introduced, but it could be automatically generated using the method of [LACS08]. We store assembly information for all product nodes, not only leaf-nodes, to obtain explosion layouts where the accumulative explosion of each tree level exhibits the hierarchical nature of the product.

Overlapping service model information

There exist different methods to overlap information from the service model with the geometry of the product, such as using labels [TKGS14] or modifying the render of specific parts or areas. In our implementation, we show most details of the service in the service result view and we just want the 3D render view to show at a glance where the most important results are located. Our intention is to focus the attention of the user rather than providing detailed information of the service results. For this, our interface provides an option to mark in red the parts having at least one inspection result with a detected flaw, which are the results on which decision-making is required.

5.2. Product tree view

As seen in Fig. 1, this view allows the user to navigate through product hierarchy with an interface similar to that of a file explorer. The nodes can be expanded and collapsed until reaching leaf nodes corresponding to atomic parts. Multiple selection can be done at any level of the product tree. When a set of nodes is selected, all corresponding areas/meshes will be highlighted in the 3D render view; in a similar fashion, selecting one mesh in the 3D render view will highlight the corresponding leaf node in the product tree view. This link between the product tree and 3D render views allows a navigation through the product combining the user knowledge on both product hierarchy and geometry.

5.3. Service result view

This view is typically open after the user access a specific area of the product using the product tree and/or 3D render views. If there exists at least one service result for that area, the user has an option to open the corresponding service result view. In our implementation, filter-based search facilities have also been implemented, allowing the user to reach a service result view without the need to navigate through the product. The service result view may show information from more than one inspection point, when the user selects multiple areas or after a filter-based search.

The service result view shows all information contained in one or more leaf nodes of the service model (inspection results) which is completely application-dependent. In our case, we show textual information regarding the inspection technique applied, the inspection date and the description of the potential flaws found (the inspection result itself). The service result view includes a link to the discussion view related with the inspection point to which the inspection result belongs.

5.4. Discussion view

The discussion view is accessed through the link present in the service result view, given the one-to-one relation between service results and discussions. When the discussion state is empty, the first access to the discussion view will automatically change its state to open.

The layout of the discussion view is different depending on the discussion state. For open discussions, the view provides a set of controls for registering new contributions (in the form of text or attached files), establishing relations with other discussions and a button to change the discussion state to closed. For closed discussions, the view consists in a read-only summary with the historic of all contributions, including the final decision if it was annotated.

6. Implementation and results

There exist different choices for data storage and access of the different models. For the product model, we have stored the product tree inside a MySQL database and the geometry data as separate .blend files. Since our geometry files are large, we store the file paths in the database rather than using a BLOB approach. A different approach to store the product tree would be to use specifically formatted JSON or XML files. The same applies for the service model, which we implement as separate tables inside the same database.

For the analysis model, we have used a customized installation of Bitnami Redmine [Les13]. All discussion data is stored inside Redmine database, except for the link with the service model, which is stored in a table inside our database, relating inspection point IDs with Redmine issue numbers.

To implement the functionality of our application, including the generation of the different views, we used the ASP.NET MVC framework. Most functionalities have been implemented from scratch in C#, except for the 3D render view and the discussion view. Fig 4 shows a final layout of our application combining different views.

The 3D Render View has been implemented using Unity Pro 5 with C# scripting and is embedded in the application using the Unity Web Player plug-in [Gol09]. To generate the Discussion View, we launch a new tab in the browser with the corresponding issue view of our customized Redmine installation (See Fig 5). The layout and aesthetics of the issue view have been highly modified, leaving only the functionalities needed by our application: registering notes, attaching
files, relating issues and following issues. Discussions (is-
sues) are created and modified from the application using
the Redmine REST API. For the Discussion View of closed
discussions, the application downloads the issue PDF gener-
ated by Redmine and stores it in the application database.

7. Conclusion

This paper presents an application which allows navigation
through a virtual representation of a complex hierarchical
product to access information regarding the results of a ser-
vice run over different parts of the product geometry. The
application provides a tool for analysis of the service results.
The paper presents general ideas applicable to different
types of analysis of different services over different products
and its implementation over a specific application case, con-
sisting in the discussion (analysis) of inspection results (ser-
vice) over a steam turbine (product).

Future work plans include augmenting the amount of ser-
vice information overlapped in the 3D render view by using
labels [TKGS14], running experimental tests with the final
users to improve the usability of the interface and improving
occlusion handling with more advanced transparency tech-
niques [Bur11] and more interactive exploded views. The
3D scanning and modelling process of the full turbine is still
in progress and we are adding more parts to the virtual rep-
resentation of the product when they are available.

Acknowledgment

The research leading to these results has received funding
from the European Community’s Seventh Framework Pro-
gramme under grant agreement No. 609027 (Project Use-it-
wisely). We would like to thank Tecnatom S.A. and the Hu-
man Factors Research Group of University of Nottingham
for their collaboration.

References

[Bur11] BURNS M.S.: Efficient and comprehensible visual-
ization of complex 3-D scenes. PhD Thesis, Department of

[Go09] GOLDSTONE W.: Unity game development essent-

[KTS09] KALKOFEN D., TATZGERN M., SCHMALSTIEG D.: Explosion diagrams in augmented re-

[LACS08] LI W., AGRAWALA M., CURLESS B., SALE-
SIN D.: Automated generation of interactive 3D explod-

[Les13] LESYUK, A.: Mastering Redmine. Packt Publish-

[TLK15] LIANG, T., KIM, T.-J. A web-based collaborative framework for fa-
cilitating decision making on a 3D design developing pro-
doi:10.1016/j.jcde.2015.02.001

[RTG_14] REYES-LECUONA A., MOLINA-TANCO L.,
GONZALEZ-TOLEDO D., FLORES S., FRUTOS E., PA-
TEL H. HOUGHTON R.: Design, Maintenance and Refur-
bishment of Turbines in a Collaborative Environment. In

[SV13] SILTANEN, P., VALLI, S. Web-based 3D Medi-
ated Communication in Manufacturing Industry, in: Concur-
rent Engineering Approaches for Sustainable Product De-
velopment in a Multi-Disciplinary Environment (2013).


[TKGS14] TATZGERN M., KALKOFEN D., GRASSET
R., SCHMALSTEIG D.: Hedgehog labelling: view manage-
techniques for external labels in 3D space. In Proc.
IEEE Virtual Reality (April 2014). , pp 27-32. doi:
10.1109/VR.2014.6802046.

![Fig. 4. Composed view layout of our turbine inspection tool. Left: product tree view; Middle: 3D render view; Right: service result view.](image-url)
Qualitative assessment of an immersive teleoperation environment for collaborative professional activities in a "beaming" experiment

Guillermo Gomez\textsuperscript{1}, Carole Plasson\textsuperscript{2}, Frédéric Elisei\textsuperscript{2}, Frédéric Noël\textsuperscript{1}, Gérard Bailly\textsuperscript{2},

\textsuperscript{1}Univ. Grenoble Alpes, G-SCOP, F-38000 Grenoble, France
CNRS, G-SCOP, F-38000 Grenoble, France
\textsuperscript{2}Univ. Grenoble Alpes, GIPSA, F-38000 Grenoble, France
CNRS, GIPSA, F-38000 Grenoble, France

Abstract
The current study assesses an immersive teleoperation platform for professional collaborative activities. This platform uses an iCub robot and virtual-reality hardware (VR headset and motion capture) to elicit an embodiment situation, where the pilot collaborates with remote actors through the robot ("beaming platform"): through the robot, the pilot can act and perceive using his natural affordances (head moving, vision and hearing) that are mapped to/from the robot effectors/sensors. One challenge is to measure how the robot expression capacity level can modulate the efficiency of the participation of the remote pilot with his collaborators. An experience where the pilot trains two subjects to assemble a mechanical system was organised: we observed the capacity of the two subjects to realize the task they discover and to collaborate with the pilot, with a more or less expressive robot. This experience clearly highlighted the importance of embodiment and the opportunities opened by beaming platforms, allowing the enhancement of the physical space by agents: i.e. social robots or conversational agents.

Categories and Subject Descriptors (according to ACM CCS): I.2.9 [Computer Graphics]: Robotics/Operator interfaces—H.1.2 [Information Systems]: User/Machine Systems/Human factors —H.5.3 [Information Systems]: Group and Organization Interfaces—

1. Introduction
The ever-growing technologies in the fields of telerobotics have led both the scientific and industrial community wonder when and how they can start to adapt and deploy these technologies in their respective fields of work. A particular application has turn out to be particularly interesting: gifting a person with the capacity to transport to a distant location, thanks to a humanoid robot that serves as a platform between the individual and another group of people on a determined physical location, for acting there as well as perceiving from there: it is a beaming [STB\textsuperscript{*}12] process where the individual is replicated by the robot, and common affordances are used (for example, moving one’s head or hand to move the robot’s head or hand) and sense the environment accordingly.

Beaming refers to technologies digitally transporting a representation of yourself to a distant place, where you can interact with the people there as if you were there (acting, perceiving). The possibilities for such a technology are tremendous, and this research focuses on the industrial engineering application field, which is often a collaborative work.

Although humanoid robots are not expected to replace humans, they could be very helpful in some particular tasks. However there is still plenty of work to design robots that can be socially and industrially competent, in order to have a satisfying and successful interaction with humans in diverse environments and domains.

It is important to study how humans react when interacting with a robot instead of a real person, and the impact on their behaviour. Humans are very sensitive to verbal and non-verbal communication, and humanoid robots are still being developed in terms of their "human-likeness" and
"sociability". The evaluation of a beaming platform implies many different features to study in order to provide a deep analysis of the strengths and weaknesses.

The current study uses a system of interaction in a multisite collaborative work, allowing collaborators to be "tele-represented" by a humanoid robot with an active and physical presence. An experience was organized to perform a remote assembly training scenario with the confrontation of different objects. It is expected to check the condition for an actor to participate intuitively from a remote location. The same experience is run while degrading the functions of the environment to identify the most important functions which are associated here with the robot degrees of freedom allowing more or less expression transmission.

The next section introduces the technical state of the art about remote teleoperation. Then a short state of the art presents the domain of application we are concerned with (collaborative design activities). At this step the experience built for the current study will be described and the final section will introduce the first level of the results from this experience.

2. Remote teleoperation issues

2.1. Beaming platforms

Telerobotics [She89, She92] is the core technology of the assessed environment. It allows to remotely drive a robotic system. Robots are an extremely good alternative to humans for multiple applications (medicine, military, nuclear plant maintenance, construction) and also for collaborative distant work.

The "Wizard of Oz" (WoZ) refers to a person remotely operating a robot [Rie12], controlling its movement, navigation, speech, gestures, for example by triggering predefined actions by pressing keys or teleoperating movements of some body segments from a mouse, a joystick or any other motion capture device. WoZ controls a robot from fully autonomous to fully tele-operated. The controlled robot serves more as a proxy for humans than as an independent entity. A valid WoZ simulation expects [Rie12, GPL∗12] "to simulate the future system, given human limitations", "to specify the future system’s behaviour" and "to make the simulation convincing".

Part of telerobotics, the beaming concept (Being in Augmented Multimodal Naturally Networked Gatherings) was introduced [STB∗12]. Beaming is the name for the process that allows someone - here the pilot - to instantaneously transfer himself from one physical place to any given distant location, thanks to the monitoring of the pilot’s actions, physiological states and emotions which can be streamed over the internet while simultaneously streaming visual, audio, spatial, and context back from the distant destination. This flow of data is synchronized and the result is an unified virtual environment representing the physical space of the destination in real time. The destination is thus inhabited by both local and virtual users.

Beaming allows to be embodied in a virtual interface. The destination/visitor interface is non-symmetrical, but beaming is defined to support symmetric social interaction between the visitor and locals. Thus an intensive attention is given to the quality of the visual immersion for the pilot, copying the physical remote destination. But indeed visualisation is just a part of a successful beaming platform, since all human senses should participate to the embodiment experience up to the "self-perception" of the avatar in the remote location.

With a virtual filter, human senses may be tricked introducing an infinite range of applications where perception may be faked [NSVW∗12] and where awareness becomes a real ethical issue. Gathering immersive virtual reality environment with tele-operated systems opens applications which otherwise would be impossible to achieve. For example, [BEFS15] uses a humanoid robot, tele-operated by a human pilot to artificially provide the robot with social skills.

Such a WoZ platform was used to experience effects of culture in human robot interaction and social robotics [BV13] highlighting the role of head nods for social interaction. A basic analysis grid is used to assess a robot performance through various subjective criteria: likeability, engaging, satisfaction, useful, human-likeness, efficiency, credibility, human-presence.

[GPL∗12] focused on the perception of gaze and head movements through an iCub robot. It allows the interaction to get closer to humans because it feels more "human-like", with the help of another sensor for capturing gaze and head motion.

2.2. Human robot interaction

With a telerobotic beaming platform, a human robot interaction is taking place at the shared physical place. Many key issues arise then [PDGA14]: mutual interaction learning, human presence undertaking, integrating social rules and protocols, as well as legible motion. There are also more specific issues like: robustness and efficiency, comfort and intentionality to specify robot behaviour, the capacity of the robot to adapt and plan according to its environment just like a human would do it, among many others.

[BCBJ∗09, SLMF06, CKH∗15] deeply analysed the effects of head nodding that can be simulated or reproduced by a robot. They analysed how people use visual feedback while conversing with robots. Humans apparently nod at the robot during conversation.

[CKH∗15] debates over social cues of interaction as well. They assume that human robot interaction comes only when the capacities of perception of the environment as much as
movement of the robot, conjugate harmoniously to attain the individual expectations. Likewise, signal and gestural responses from the robot such as head nodding along with other non-verbal behaviour, help individuals to know the intentions of their interlocutor, and so react and adapt their behaviour accordingly.

2.3. Multiparty dialog and situated interaction

The robot behaviour must integrate qualities that favour the communication between multiple parties. Gaze perception plays a crucial role in human-human interaction [Lan00, MEB12]. It has been proved that it affects several aspects of communication and dialogue. An embodied conversational avatar must integrate an accurate gaze model. Gaze remains one of the strongest nonverbal cues in human interaction and this result should be assessed for human-robot interaction.

Interaction with humans expects to consider also human factors. Primary causes of malfunctions in processes are due to human factors [Nar99]. Among all the human factors to consider that might have an impact on repairing, problem solving and effectiveness, we can cite a few: attitude, fatigue, workload, schedule, etc. For [HC98] having eyes and hands busy, drive decision about the usage or not of voice recognizing systems instead of classic systems like a keyboard. In addition, speech is a strongly personal activity which contains more information than just the text produced; then voice acts also as an important human factor.

2.4. Collaboration within design and manufacturing activities

The current study applies a beaming platform to support collaborative professional activity, specially in a design and/or manufacturing context. Design and manufacturing engineering is a cognitive activity [Cho09, GKO4]. There are several design modes: the prescriptive approach, which describes the process of design as a succession of stages. This approach masters processes, acknowledge the creative part, and is teachable. Design is also viewed [Sim96] as a “problem solving” process where it becomes a non-linear and non-deterministic process of resolution, and has multiple solutions. [Sch83] presents design as a process of reflexive conversation: designers work in virtual worlds, and their talent lies on the capacity to build and manipulate these worlds. Then design is an art in the professional activity, it is a form of intelligence that is not teachable but can be learned, where identifying a problem and improvising is a necessary “art” for designers.

To collaborate designers use intermediary objects [BB03]. Objects are a second essential component in design and manufacturing processes. Design is often supported by visual reasoning with object representations. [Kua06] built experiences to analyse the links between cognitive micro-activities during design and demonstrated the separation between design reasoning, intention of design, intention of representation, and modeling actions.

[OOCS92] observed how design often takes place in a context of a collaborative face-to-face meeting. They try to analyse how designers use their time, what kind of activities they develop, how they organise these activities and if there exist any similar “models” between different projects and teams of people. They conclude that people expose their problems, alternatives and criteria to evaluate them, and they share their expertise. Results show that 40% of the time is spent on design discussions and briefings account for 30% of the time, which implies that reunions apparently have a role of coordination, generating a discussion in order to clarify the best way to proceed, and moreover that meetings produce an important amount of clarifications, since participants make and answer questions on diverse subjects.

Design and manufacturing activities provide a good context to assess the usability and utility of a beaming platform.

3. An experience for collaborative remote training

3.1. Research question

This study expects to assess the conditions for creating a beaming context to provide an efficient support for design and manufacturing professional activities. A basic beaming platform is controlled and its functions depreciated to check their importance. Especially the following research questions are submitted:

1. What is the impact of the different grades of movement and expression replicated by a humanoid robot in the interaction with individuals on a collaborative work environment?
2. How to assess this collaboration between individuals and the robot?

Increasing the freedom and expressiveness of the robot’s movements should raise the fluency of the avatar representing the distant person, and thus improve the collaboration with a team of individuals, allowing a more meaningful interaction and a greater implication of individuals in the task. Such an outcome comes with the interest of building a platform for interaction of people in a collaborative setting, which has applications in the industrial sector within a realistic use-case.

3.2. A beaming platform

The beaming platform used for the experience integrates an iCub robot, which is 1 m tall. This robot has 53 degrees of freedom, plus 5 extra DoF for the jaw and mouth [PRM*15]. In the used beaming platform version, only a few are mapped from a real person: it can move the head using the three degrees of freedom, and it can move the jaw and lips thanks
to the generation of a model adapted to track pilot’s lips. To create the model for the lips movement, the pilot had to glue a few motion infra-red reflectors strategically placed around the mouth, and run a calibration before each experiment.

The robot eyes are instrumented with two cameras allowing a stereoscopic capture of the robot vision. Microphones were inserted at robot ears location, allowing the capture of the sound environment around the robot. A loudspeaker was also inserted in the robot mouth, creating audio source for verbalization reproduction.

A Sony Head Mounted Device (HMD), like to the one from Figure 1, was used to immerse the pilot in the robot space that transmit the views from the two mounted cameras (Figure 2). The HMD earphones receive the sounds captured by the iCub ear microphones. Pilot’s speech is captured by a close microphone and played back via the robot’s mouth loudspeaker. The robot acts as the distant avatar of the pilot. A motion tracking system was used for capturing both global movements of the pilot and mouth articulations.

3.3. Use case selection about mechanism assembly

A scenario that could be played by actors in a reasonable amount of time was expected to make experiences acceptable by subjects. The proposed scenario should be of a bearable difficulty, considering that something too hard to achieve is not useful here. To seek with utility within design and manufacturing applications, a jigsaw system was selected. This system provides a realistic context; on the inside it is a traditional mechanical system, with simple components such as a rotor, a gear, a switch, a rod, etc. and it is a real system for real engineering context whether it is for design, maintenance, production, etc. Additionally, it is easy enough for people to quickly identify its main parts and even if they have never seen such a system before, to be able to handle it without requiring training or whatsoever. This system can be mounted in a reasonable time.

Involved subjects were not familiar with it. The disassembled system is presented to two subjects. The two subjects are expected to collaborate to re-assemble the system (see Figure 3) under the supervision of the pilot. The role of the pilot is limited to supervision (our beaming platform does not have an exoskeleton to support arm control yet). The scenario can be viewed as a training situation where two persons discover and learn a professional task supported by a remote expert. This context is closed to a context of project review where the pilot plays the role of manager and moderator of a discussion between two experts.

The selection of the jigsaw system is a compromise between complexity of task to assemble it, realism of the system and level of details: indeed the pilot guides the robot with a vision replication which has a low resolution and quality; it is due to the resolution of camera installed in the robot eyes (two 640x480 cameras with Bayer filter) and to the HMD which reproduce vision with a better resolution than the camera but creates new sources of uncomfort: weight, vergence for eye accommodation, etc.
3.4. Experimental protocol

The pilot is driving tasks within a predefined sequence of actions. The pilot is expert of the system, knows every details about the mechanism and must have enough possible situations in mind to react promptly as soon as there is a deviation from the predefined sequence of tasks to ensure to make possible comparisons between tests. The sub-tasks are thus organised into a series of short collaborations or parallel actions of the two subjects to ensure that the predefined process is followed.

The position of parts when the subjects enter the room is pre-defined as organised in Figure 4. When entering the room the jigsaw is masked.

**Figure 4: Jigsaw presentation when initializing the process.**

Two configurations of robot behavior are created.

**Speaker configuration : A**: the robot is motionless, it transmits the voice of the pilot trough the high-speaker, but all his body articulations are off.

**Beaming configuration : B**: The robot has full motion of its head, with 3 degrees of freedom, and reproduced the movement of the lips from the pilot according to the model set up before the experiments. In this condition, the pilot has a partly blurred vision in the HMD, except in the center area. This vision mode forces the pilot to use explicit head movements to track or see the focus zone: the interest point he’s gazing at is then clearly identified by the subjects.

Three questionnaires have been prepared which cannot be fully reported here but we provide some question example to fix ideas:

**Memory Questionnaire:** $Q_1$: 10 basic questions to test what the person remind by himself even if no special attention was drawn to these details during the assembly process:

- $Q_{1.1}$ What is the color of the main switch of the jigsaw ? black□ red□ yellow□ green□
- $Q_{1.2}$ How many screw sizes belongs to the system ? 1□ 2□ 3□ 4□
- $Q_{1.3}$ etc.

**Individual perception of collaboration:** $Q_2$: 10 new questions were devoted to measure the perception of a more or less living robot:

- $Q_{2.1}$ How will you qualify the robot from machine to alive ? 1□ 2□ 3□ 4□ 5□
- $Q_{2.2}$ How will you qualify pilot/robot from stupid to clever ? 1□ 2□ 3□ 4□ 5□
- $Q_{2.3}$ etc.

**Individual comparison of configurations:** $Q_3$: 5 questions to compare configurations ‘A’ or ‘B’.

- $Q_{3.1}$ Mark from 1 to 3 the importance of the following modalities ?
  - voice transmission : 1□ 2□ 3□
  - head movement : 1□ 2□ 3□
  - lip movement : 1□ 2□ 3□
- $Q_{3.2}$ In which configuration the pilot specifications were more clear ? A□ or B□ or None□
- $Q_{3.3}$ In which configuration the pilot was the more attentive ? A□ or B□ or None□
- $Q_{3.4}$ In which configuration did you feel more easy when interacting with the pilot ? A□ or B□ or None□
- $Q_{3.5}$ What configuration did you prefer ? A□ or B□ or None□

After filling a participation agreement form, both the avatar and pilot are presented to subjects. The wizard of Oz context is clearly highlighted: they know that the robot is driven by a person. They are informed that they will have to follow the directives of the pilot and that they will assemble twice the same system following different sequences. Each sequence will not overpass 20 minutes: the task is stopped after this delay.

There are two assembly sequences, $S_1$ and $S_2$. For each $S_i$, assembly sub-tasks are executed in different orders and the roles of the two subjects are swapped. Sub-tasks are either individual or collaborative, and can be executed in parallel or sequentially.

For statistics issues configurations $A$ and $B$ are associated to sequence $S_1$ and $S_2$ leading to four conditions that are applied alternatively to the tuples of subjects.

Then, the mask over the mechanism is removed and the pilot drives the first assembly sequence. At the end of this sequence the questionnaire $Q_1$ is filled individually by the two subjects. Then they also answer to the questionnaire $Q_2$ to collect their subjective feedback about the collaboration with the pilot through the robot. While filling the questionnaire the system is disassembled and prepared again for the second sequence. The second sequence is played with the remaining configuration. After the second sequence, the questionnaire $Q_2$ is answered again but in reference to the new configuration. At last, the two subjects answer to the questionnaire $Q_3$.

This protocol was repeated for 9 couples of subjects and indeed 18 different persons were involved. The distribution of participants could be presented as follow: 10 females and 8 males, with 3 teams of women, 2 teams of men and 4 mixed
teams. The average age of the participants in the experiments was 25 years old, and the majority had an engineering background.

4. Analysis of collaboration between subjects and a remote drive

4.1. Analysis of questionnaire \(Q_1\)

The overall objective of the questionnaire \(Q_1\) is to check if subjects have a better performance in configuration \(B\) than in configuration \(A\). The average percentage of correct answers was medium to low in both cases: 45% of success rate for the actors that answered after experiencing the configuration \(A\), and 62% for actors that answered after configuration \(B\). It must be reminded that in both cases they answer after the first sequence of assembly. These values seem to indicate that actors in configuration \(B\) are more engaged in the task because they have a better communication whenever the robot presents a more human-like expression and record more information.

4.2. Analysis of questionnaire \(Q_2\)

The second questionnaire deals with the subject individual impression about the robot. The average mark of each question for all participants are representative numbers to analyse. For every question, the average mark is higher for configuration \(B\) than for configuration \(A\). Figure 5 details results for every question and the configuration \(B\) is clearly appreciated but questions can be separated into two sets depending on a more or less difference between the two configurations. Then it seems that:

- Configuration \(B\) provides a real impact to understand who (between the two subjects) is concerned when the pilot interacts. This configuration also enhances the frequency of interaction and the demand for interaction with the pilot.
- Configuration \(B\) seems to be also better for interaction quality and when assessing the overall pleasure to interact through the avatar but the difference is less obvious.

4.3. Analysis of questionnaire \(Q_3\)

The third questionnaire serves as a summary to confirm the analysis of the first questionnaires. The first question (Figure 7) provides a quite obvious result but which is clearly measured. The voice is the preferred interaction modality but the gaze identified by head movements is a major interaction modality. The mouth articulation are not unnecessary (indeed, they might play an important role for speech intelligibility), but from the conscious self-report task, they are perceived as less important.

The other questions of \(Q_3\) were direct comparisons of the two configurations. The results described by Figure 8 are incontrovertible. The configuration \(B\) is clearly preferred even if 1 or 2 subjects reported that they did not care at all about the robot, which explains a few answers with no preference: video shows that these persons were concentrated on the task while listening the pilot but without watching the robot.
5. Conclusion and perspectives

The current study clearly demonstrates again the importance of gaze and head movements within social interaction. It also demonstrates the capacity of a beaming system to telepresent a remote human coach as soon as the beaming system reproduces the good interaction modalities. It demonstrates also a beaming process within a realistic professional collaborative task. Usually, research about user interfaces focuses on usability demonstration. With the current realistic process, a first step towards utility assessment is passed.

Obviously the statistics sample should be extended but the first results seem incontrovertible. We need to explore which parts of the dialog benefit from gazing at head or lips movements, and how these gaze patterns evolve with the duration of the interaction. Another direction of assessment will be exploited soon to get objective measures: Indeed all the scenarios have been recorded (video, audio, motion capture of the pilot, as well as robot’s sensors streams) providing a rich set of corpus which can be deeply analysed (video labeling of the eye contacts, of the number of head movements from the human partners...)

Another perspective is also to go ahead with this experience using other type of avatars, to check how presence filling may be produced on a distant location. A video of the pilot face may be streamed at the remote location but it can be expected to face the Mona-Lisa cue: a 2D image cannot propagate the gaze direction of the pilot. An image avatar could provide an immersive environment where the importance of gaze could be checked with respect to the quality of the avatar. The image could provide really realistic human face. To avoid the Mona-Lisa cue, stereoscopic displays or holographic displays could create an interesting 3D avatar; for a single remote subject, stereoscopy may be used while for multiple remote subjects holography will be expected. It will be worth exploring the respective properties of the streaming of 3D videos vs. the sensorimotor monitoring of virtual/robotic avatars in terms of users’ experience and interaction efficiency.

**Figure 8**: Results for questionnaire Q3, question 2 to 5.

---

**References**


---

**Aknowledgement**

- This paper was written within the scope of the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01).
- This work has also been partially supported by the VISIONAIR infrastructure project funded by the European Commission under grant agreement 262044.
- The experimental robotic platform has been supported by the ANR (Robotex ANR-10-EQPX-44-0) and SOMBRERO ANR-14-CE27-0014) and the UIF (EMSOC 2012084RECF991).
Beaming for collaboration

Gomez & al. / Beaming for collaboration


Virtual Reality for Collaborative System Engineering supporting ESA experimental vehicles

AGATA MARTA SOCCINI

University of Torino; Torino, Italy; soccini@di.unito.it

Abstract

Benefits of Virtual Reality systems for the sharing of knowledge and results among disciplines during pre-launch phases of Space Vehicles are proved, but the design of easy to use specific interaction paradigms is still difficult and rarely standardized.

This paper presents a solution based on the study of the most frequent cases of need of collaboration in System Engineering of experimental vehicles. We encourage focusing on spotting specific use cases and thinking of single purpose applications, rather than only concentrating on a general purpose wide integrated system.

We finally report the positive qualitative evaluation of this approach based on the feedback of scientists and engineers who tested the applications we developed.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism/Virtual Reality

1. Introduction

The use of Virtual Reality as an interface for scientific tools has became more and more popular in the last few years. The commercialization and diffusion of new low cost devices reintroduced the presence of Virtual Reality in daily habits, and therefore the concept of Virtuality in common technology final users. Among those users are people who belong to the scientific community, who might find VR familiar or at least not completely unseen.

Complex System Engineering, above all other manufacturing related fields, is in the constant need of cutting edge technologies, and the benefits of the use of Virtual Reality in several steps of the collaborative projects have been previously demonstrated. We use Virtual Reality to make System Engineering collaborative [PBM*12].

Space objects are often Complex Systems. In the processes of engineering and designing, the use of VR has been recognized to be an incisive and productive way to perform tests and validate feasibility of products before their actual implementation [WQ10] [Wan11]. Communication anomalies, as well as misinterpretation of results of the disciplines, are dramatically reduced, and big amounts of complex data can be nicely represented and shared [TSO*11].

Not only, but the use of interactive immersive systems, which is what Virtual Reality systems intrinsically are, let users easily perform other disciplines calculations while testing their own, without needing to understand the inner processes and motivation of other research teams.

In this scenario, we performed early studies on how to interface the single disciplines physical analysis software tools with a common Virtual Reality system. We initially spotted the strengths and weaknesses of the approach: while we can take advantage of the several benefits we mentioned, we need to define some design standards, in order to be able to create some guidelines to follow in the future cases.

In general, for what concerns Virtual Reality for Complex System Engineering, we encourage focusing on common specific tasks and developing single-purpose functionalities, one for each spotted case. In particular, for what concerns designing and engineering experimental space vehicles, we suggest to provide the disciplines tools for two main cases.

In the first case, we give users the opportunity to model an environment and test the physical behaviour of the vehicle. In the second case, we provide users a tool that simulates...
the future flight of the vehicle, and can test the physical behavior of the vehicle in some chosen moments of its flight.

Not only scientists and engineers can analyze heterogeneous, complex, mono or multimodal data in a short time and in a visual way, but also they will find specific tools they will need in most cases.

By sharing the results of our research, we hope not only to encourage the use of Virtual Reality in System Engineering processes, but also to share the single-task approach and the specific tools for future experimental vehicles engineering, in order to simplify processes and enhance communication and productivity.

2. Studies and definition of the project

At the COllaborative System Engineering (COSE) Centre [PBR*11] at Thales Alenia Space Italia [web*TAS], as part of the project STEPS2 [PMG*11] [web*ST2], my research team explored the role of Computer Graphics, and in particular of Virtual Reality as ergonomics and collaborative engineering booster in pre-launch phases of space vehicles design and engineering.

“Sistemi e Tecnologie per l’Esplorazione Spaziale – Phase 2” (STEPS 2) is a research project co-founded by EU on the “Misura Piattaforme Innovative” - Phase 2 of POR FESR 2007/2013.

Among Space Exploration current research topics, The European Space Agency (ESA) [web*ESA] for several years has been giving great importance to Re-Entry Technologies, to consolidate a European knowledge base on several disciplines, including Aerothermodynamics. IXV (Intermediate xExperimental Vehicle) [web*IXE] is the main ESA recent project regarding this topic, and Thales Alenia Space Italia is the prime contractor of IXV [web*IXT].

We used VERITAS as a Virtual Reality system designed for space exploration simulation purposes, ATDB Tool as a tool to perform the aero-thermal analysis of IXV.

VERITAS [BMF*08] is a framework that includes several different Virtual Reality applications, designed to run either in a CAVE with up to six screens, or on a desktop computer.

It is property of Thales Alenia Space Italia, and has been developed at COSE since 2004 with the original name of TraVis [MB*08], as it was first designed only to visualize space vehicles and their trajectories. In order to be more flexible and adaptable to meet new needs and support new case studies regarding space missions, the Solar System, the International Space Station (ISS), or detailed terrains of Mars and the Moon, TraVis was extended during the years, and took the name of VERITAS.

ATDB Tool [MFZ*11], acronym of Aero-Thermo-Dynamic Data Base Tool is an aerothermodynamics analysis tool specifically designed and developed for IXV.

It takes as input a model of the environment in terms of Mach, Temperature, Pressure, Density- and an asset in terms of Elevon, Aileron, Sideslip, Attack-, and returns as output a punctual aerothermodynamics analysis.

The two systems, VERITAS and ATDB Tool, can both represent and understand, using dramatically different approaches, the geometry of a spacecraft.

While VERITAS can represent and nicely visualize a wide set of standardized geometries, and is able to simulate their behaviour and interaction with other space objects, ATDB Tool has old style basic low level interfaces and only works with IXV data, but, given an asset and an environment, can calculate some thermodynamic quantities (i.e. pressure, temperature) of the vertices of the represented geometry that VERITAS does not originally consider.

We previously realized a basic generic interface between the two systems and proved the efficiency and ergonomics improvement of the use of Virtual Reality as an interface for Multidisciplinary Physical Analysis of Space Vehicles [SMB*14].

Following this topic, we recently performed a research on internal industrial needs, and collected interviews among scientists and engineers belonging to different disciplines (we will now refer to as users) to spot the most common goals they had while using the prototype we developed to test our early studies.

We worked with the staff of IXV to understand the criticalities of collaboration processes during design and engineering phases. Based on their experiences, we planned to concentrate on these two cases:

Users may need to visualize IXV follow a trajectory, which could be the real one as well as an experimental one, and want to see the Aerothermodynamics behavior of the vehicle in certain critical moments of its flight.

Users may need to visualize the Aerothermodynamics behavior of IXV in specific environmental conditions and with a specific asset. They should be able to visualize the results of the analysis in a required hypothetical case [SMB*15].

As we realized the main punchal goals were common and useful for all disciplines, we decided to suggest and investigate the value –in terms of ergonomics, usability and design efficiency – of developing one single app for every use case instead of a general purpose software, more flexible but more dispersive and less user friendly.

We therefore developed two Virtual Reality applications to provide solutions to the existing issues and test suggested interaction patterns.

3. Implementation and Validation

To validate the two interaction design patterns, we implemented two applications and invited users to test them.

Using VERITAS, based on Open Scene Graph Libraries [web*OSG] we designed and developed two Virtual Reality applications based on the mentioned principles, and according to the spotted needs. We called them ‘IXV-Trajectory’ and ‘IXV-Asset’ [SMB*15].
Given the geometric model of IXV, a trajectory coded in a text file, IXV-Trajectory first visualizes IXV follow the provided trajectory, and then allows users select specific time instants in which they want the aero-thermal analysis to be performed. Once selected the instants, users hit ‘Call ATDB’, and IXV-Trajectory calls ATDB Tool passing as an input the asset of the vehicle, which is coded in the trajectory file, in up to ten specific moments of the flight. Again, ATDB Tool gives as an output the result of the Aerothermodynamics analysis in Tecplot formats [web*TEC]. As shown in Figure 1, The application visualizes one Tecplot at the time: according to the timeline of simulation in VERITAS, only the most recent Tecplot is visualized in every moment.

Given the geometric model of IXV, IXV-Asset lets users model a specific asset and a specific environment they want to investigate. As they hit ‘Call ATDB Tool’, IXV-Asset invokes ATDB Tool, and the result of the call is a visualization of the model of the spacecraft as well as a Tecplot colored according to the output of the aero-thermal analysis performed by ATDB Tool. In Figure 2 we see the representation of IXV as well as the results of two different hypothetical physical analyses [SMB*15].

To validate our approach we invited the staff of the IXV project to test and evaluate the applications we developed in VERITAS. Scientists and Engineers used the apps and were interviewed right after that. We mainly asked to evaluate ergonomics and usability of the apps -compared to the experience given by the use of our early prototypes- and the importance of interfacing physical analysis tools with an immersive stereoscopic interactive visual interface.

Our qualitative evaluation of the apps underlined a common appreciation of the work.

4. Results and Conclusions

These Virtual Reality Applications, as prototypes of interfaces to physical analysis tools, can be considered a core feature for an actual implementation of Collaborative Engineering methodologies.

We first underlined the value provided by spotting single-use functionalities instead of focusing on a general purpose interface during the design of Virtual Reality applications for physical analysis of Complex System Engineering.

We then discovered that the best use of Virtual Reality for physical analysis of Space Vehicles can be designed and implemented in two main applications.

One is centered on the physical behavior of the vehicle during the flight; the other one is made to test hypothetical physical environmental situations.

Scientist and researchers involved in IXV project at Thales Alenia Space were asked to try and test the mentioned applications. The results of the interviews underlined appreciation on:

- Visualizing a mission that has no visual documentation
- Being able to understand other disciplines needs and results
- Being able to perform tests using other disciplines tools
- Being supported in the design and engineering by ergonomic tools specifically made for supporting critical phases and ‘what if’ analysis

We found some criticalities in the design process, and still some quick training before using the apps independently is needed.

Mapping industrial needs in design patterns, finding a compromise between the ease of use and the potentiality of the software has been a circular and difficult process. Involving and motivating users step by step helped in reaching a successful result.

We received enthusiastic feedback by other Working Packages of STEPS2, of IXV, and by Engineers and Scientists at Thales Alenia Space. The described results also confirm the interest and the efforts of Thales Alenia Space Italia -and of the COSE Centre in particular- with the collaboration of the Computer Science Department of University of Torino, towards the advantages of Collaborative Engineering, Model Based System Engineering and Virtual Reality.

References

[BMF08] CHRISTIAN BAR, MANUELA MARELLO, ATtilIO FERRARI:

TraVis – A Virtual Reality Interactive Application For Scientific And Aerospace Visualization And Simulation.

[MB08] M.MARELLO, C.BAR:

TraVis - A Virtual Reality Application For Interactive Simulations In Aerospace And Astronomy.
5th INTUITION International Conference, Turin, Italy, 2008

INTERMEDIATE EXPERIMENTAL VEHICLE, ESA PROGRAM IXV ATDB TOOL AND AEROTHERMODYNAMIC CHARACTERIZATION.


[web*ESA] European Space Agency, website: https://www.esa.int

[web*IXE] IXV at European Space Agency, website: http://www.esa.int/Our_Activities/Launchers/IXV


Figure 1 – IXV Trajectory showing the temperature of the shuttle during the flight

Figure 2 – IXV Asset showing the result of two different analyses
Ergonomic Validation of Manual Processes through Posture Detection Using Optical Sensors

STRATOS ANTONIOU\textsuperscript{1}, LOUKAS RENTZOS\textsuperscript{1}, GEORGE PINTZOS\textsuperscript{2}, KONSTANTINOS SMPAROUNIS\textsuperscript{3}, DIMITRIS MAVRIKIOS\textsuperscript{4}, GEORGE CHRYSSOLOURIS\textsuperscript{5}\textsuperscript{*}

\textsuperscript{1}Laboratory of Manufacturing Systems and Automation (LMS), Department of Mechanical Engineering and Aeronautics, University of Patras; Patras, Greece;

\textsuperscript{*}corr. author e-mail address: xrisol@lms.mech.upatras.gr

Abstract

This paper presents an approach to the ergonomic assessment of manual operations in Virtual Reality simulations as well as to shop floor applications. There is a position posture recognition, based on the relative position of a user’s joints and is coupled with a joint angle-based posture detection algorithm. This allows for conclusions to be drawn regarding ergonomic validity. The output is then communicated to the user through different devices. The approach is applied to a case study.

Keywords: Virtual reality; Optical computing; Computer-aided manufacturing

1. Introduction

Mass customization has become the current trend in which traditional manufacturing systems slowly move towards [Chr_06]. This new paradigm focuses on the specific customer needs, enabling the manufacture of products, in different configurations, by increasing the flexibility of the systems [HKW_11]. However, in most cases this creates the necessity of more complex production stations, which increases the time required for the designing of processes which can cause significant delays in production realization. Virtual Reality (VR) and Augmented Reality (AR) technologies can potentially improve process design and development methods when they are applied to assembly/disassembly (A/D) simulations, support for assembly and maintenance, ergonomic studies, virtual prototyping in the context of conceptual design and product evaluation [MPRC13, RVMC14]. This potential is true, due to the fact that VR offers the flexibility to perform a number of analyses related to the design of processes by incorporating anthropometrics of the human operator using different arrangements of control and signaling devices accordingly. To conduct these analyses, full interaction between the user and the workplace he/she will be working in must be programmed, including collision detection, kinematics of the different devices and possibilities of activating their various functions in relation to other objects in the Virtual Environment (VE). However, on the hardware’s side, VR and AR peripherals are becoming more accessible for small and medium companies due to their use in more mainstream applications [GGZH13].

The setup of VR simulations requires the transformation and composition of many types of data, such as the compilation of 3D VEs, geometries and interactive scenes, which is still considered as a big drawback due the resulting authoring times [PDS_13]. While these data are often pre-existing, they must be transformed, enriched and combined (geometric simplifications, combinations of 2D and 3D) to enable their use in VR/AR applications. In simulations where the goal is the evaluation of ergonomic features of processes or products, the user is immersed in a VE representing the Real Environment (RE) and reproduces a close to realistic metaphor of the real task [PDS_13]. Ergonomic evaluation analyses can then be performed resulting in the redesign of the processes, providing increased ergonomic validity. It has also been shown that the virtual process design can be used to minimize the physical risk factors involved in the musculoskeletal disorders appearance [MCBZ09]. Posture detection is the most used method when it comes to the evaluation of working environments.

Postures can be classified in two main classes; key postures and transitional postures. Key postures uniquely belong to one action so that people can recognize the action from a single posture presentation. Transitional postures are interim between key postures and the action they are part of cannot be recognized. Therefore, human actions cannot be recognized from postures of a single frame that belong to transitional postures [CCCL06].

Apart from the visual representation, related VR studies employ posture tracking mechanisms in order to determine ergonomically acceptable virtual prototypes. The most widely used posture detection method includes using accelerometers placed on the body of the user, in order to monitor changes to their posture during a VR session [JMF_11]. The main advantage of this method of posture recognition is its high resolution and high accuracy of tracking of all posture changes of the user. However, it requires the placement of additional hardware on the user. On the other hand, a wireless solution would allow the ergonomics studies, once val-
2. Proposed methodology

The methodology proposed in this paper is based on an integrated VR-based system for the improvement of ergonomics in shopfloor simulations, with potential for applications on actual industrial environments. The developed idea is targeted for wide applicability by the use of a low-cost setup (low investment risk) and the lack of necessity of a visualization device for the user. In an industrial setup, a worker can perform the routine tasks and at the same time be monitored by the system evaluating the real-time body postures through their comparison to referential ergonomically acceptable postures. Based on the proximity of the two, the system has the capability to calculate the deviation and warn the user in the case of potential risks (big deviation from the ergonomically acceptable posture). In an immersive setup, the user can conduct pre-defined tasks and with the same procedure he or she can receive feedback in the form of changes in the VE.

The system consists of a VR engine that handles the rendering of the virtual scene (only available for observers in case of industrial environments), as well as the handling of interactions between the user and the system. The simulation is run on a computer and gets input from the user through the multi-sensoral tracking device which can be connected either to the same machine or to any other computer connected to the network.

The feedback to the user and observer can be provided through standard visualization devices, HMD or CAVE hardware in case of Immersive VR environments and AR goggles or even sound messages in real environment (Figure 1). The feedback, in the form of audio, text or virtual instructions is selected to be discrete and not interfere with the user’s field of view and perception of the undertaken task.

The core algorithm used, performs position-based posture recognition based on the relative position of the joint sensors to the upper body of the user and is coupled by a joint angle-based posture detection algorithm. One requirement was to provide a widely applicable way of quick training for ergonomically acceptable postures on simulations and the shopfloor and the evaluation of the validity of them by people belonging to different characteristics (ranges of anthropometric values). The tracking device provides live position and orientation data for the user’s body and the developed algorithm compares the data received to predefined ergonomically acceptable postures that correspond to the way that a specific task is designed to be performed. Users can also add new ergonomically acceptable postures of their own that serve as procedures which are stored in the form of arrays in the VR engine. This way, the system is able to detect the kind of task that the user is performing and reach conclusions about whether or not it is performed correctly.
The Kinect sensor has three autofocus cameras: two infra-red cameras optimized for depth detection and one standard visual-spectrum camera used for visual recognition [DZ14]. The selection of Kinect, amongst the low-cost tracker solutions was made based on the free-offered SDK and wide compatibility with Virtual-Reality Peripheral Network (VRPN) that allows its easy integration to VR compositions as a replacement of high-end trackers.

For the connection of Kinect to the VE, a free-licence utility developed by Stanford University called FAAST was used [SLR_11]. FAAST creates a VRPN server [TRH_01] that broadcasts the data of the Kinect-generated virtual markers that correspond to the user joints over a network. The VR engine (3DVIA Virtools) uses a VRPN client that connects to the server of the device and with the correct calibration, transfers the position and orientation data in the VE. In order for the user to be able to interact with the VR engine by pressing buttons during the procedure, Nintendo Wii-mote was selected.

3.2. Application description

The core of the application consists of two proximity calculation algorithms which have been developed for Posture Detection. The first algorithm is based on the 3D position coordinates of the different joints of the user. This allows for the separation of the key-postures from the transitional postures since it activates the posture detection algorithm only when the user is “in position” to perform a predefined task (Figure 2).

Figure 2. System functionality

The second algorithm, is based on the idea of calculating the distance between two joints and generating a vector from them (bone). As an example, for the right elbow - right shoulder joints the algorithm generates a vector A and another one for right elbow - right wrist, vector B. The algorithm can now calculate the bending of the right elbow by calculating 3D angles or Euler angles of the elbow. Based on the deviation of the reference and real-time joint angle difference the system can give suggestions to the user so as to make corrections to their posture in order to get closer to the ergonomically acceptable posture. More specifically, if the distance between the joint position data of the reference posture and the real-time posture is above a certain limit, the posture is considered “bad”. The same applies for the joint angle difference. The VE can be configured to give warnings to the user in case of high-risk postures that deviate a lot from the pre-defined.

Using the above method scalable implementations in different VEs can be performed based on the needs and joint positions or joint angles can be added or removed as criteria in order to increase/decrease accuracy accordingly.

The global coordinates of the head joint of the user are used as the main criterion in order to determine if the operator is in place where the postures for a specific task (set of postures) have previously been recorded (Figure 3).

Once the user’s head gets near one of the predefined positions, the algorithm starts tracking the user’s hands, in order to determine which of the possible tasks the operator is currently undertaking. The positions of both hands for each undertaken task are saved in pairs. Once the live position coordinates of the user’s hands “lock-in” pre-defined task positions, feedback is displayed to the user in the form of on-screen message and the joint angle detection algorithm is initiated. Figure 4 shows an example where the operator would be characterized as “in position” for a task.

Figure 4. Head and hands “locked in position” for a pre-defined task
In order to be able to detect live postures and compare them with ergonomically acceptable ones (stored), the system is “trained” by having ergonomically acceptable postures saved. This happens with the following procedure. A handheld device is used for notifying the system that the user has reached a reference position. Once inside the range of a head reference position and if both the user’s hands do not match a predefined task, another button on the handheld device can be pressed and save a new task. In the corresponding array of the reference tasks of the workstation, a new entry is created containing the reference hand and reference joint-angle data for the new pre-defined task. The handheld device used for this purpose was the Wii-Mote which in some cases allows even the user performing the tasks, to store references.

4. Use cases

Since the developed idea aims to be applied both on VEs and actual shopfloor environments -for training and monitoring respectively- two separate use cases were examined. In both cases the tracking setup consisted of a Kinect sensor positioned to face the user.

The VR use case included the implementation of the algorithm in an immersive virtual scene. The immersive application included the sequential assembly of different components belonging to a laser machine head using only the operator’s right hand. Two workspaces were defined using the corresponding head position coordinates (Figure 5). This task was done using the Wii-mote to save the coordinates and then manually edit the corresponding array entry. The first workspace was the area where the user performed a picking task. The second workspace was the area where the user performs the “place” task. For this prototype, all other postures were considered as transitional postures.

Apart from the equipment described above, the setup included an nVisor SX60 [NVIS] high definition Head Mounted Display (HMD) for the input to the user and an additional 24 inch full HD screen for the observers. A limitation was discovered in this setup; when wearing the device, due to the extra volume of the HMD the sensors assumed a slightly bent head position (Figure 6). This issue can be eliminated with calibration of the virtual head position in the application. However, this limits the rotational range of the user’s head that can be detected.

For the real environment use case (Figure 7), the pick and place task of an object for its storage was performed.

The storage place was on a shelf located over the user’s head and the picking took place below the user’s waist. In this case it was assumed that we had one workspace, but two different tasks. This again meant that two ergonomically acceptable postures were considered as key-postures and the rest were assumed as transitional postures. Figure 8 shows the user picking up the object assuming a wrong posture. In this case the system notified him/her through an on-screen message on the observation screen.

Figure 5. Virtual assembly of laser machine’s header

Figure 6. Kinect joints estimation without HMD (left) and with HMD (right)

Figure 7. Detection of Steps of a Procedure

Figure 8. Pick-up task assuming a wrong posture; (a) live, (b) recorded and (c) stored posture
For the second task of this scenario, the user places the object on the shelf by assuming a correct position (Figure 9).

Figure 9. Place task assuming a correct posture; (a) live, (b) recorded and (c) stored posture

5. Conclusions and future work

The methodology presented in this paper can be used to ergonomically validate user postures in both VEs and real environments through the tracking of specific joints. In the first case it offers an on-screen real-time feedback about the correct posture and it allows the instructors to add or remove conditions to a correct posture, thus resulting in a very flexible training software. Furthermore it can eliminate the need of an observer/instructor during many phases of training. On the other hand it lacks the accuracy that high-end trackers offer and introduces compromises. Although additional joints can be used for identifying more joints of the human body (e.g. waist), additional optical sensors which will include potential sound notifications, it can lead to the decrease of the risk of accidents and improve the long term health conditions for the workers.

The work presented will be followed by a potential extension to include additional optical sensors which will include an additional module for the merging of data coming from different sensor instances (i.e. skeleton merging). Also, the system will be enhanced by additional user interfaces for the shopfloor case using devices such as the ones proposed in section 2.

6. Acknowledgements

This study was partially supported by the project VISIONAIR / INFRA-2010-1.1.29 - 262044, funded by the European Commission.

References


The presentation of procedures in virtual training of assembly tasks: segmented vs. whole task approaches

A. Langley, G. Lawson, S Hermawati, M D’Cruz

Human Factors Research Group, Faculty of Engineering,
University of Nottingham, NG7 2RD, UK

Abstract
Virtual Environments (VEs) are currently being investigated within the automotive industry for training assembly tasks. While previous research has shown that VEs can provide a flexible and adaptive approach to training, most studies have focused on the evaluation of the technology; fewer have investigated the efficiency of the training received. This paper begins to address this deficiency by reporting on a study that adapts principles established within medical research relating to training in the VE, and transferring them to virtual assembly training. The main investigation focussed on comparing training in repeated segments of the task to training by repeating the whole assembly, but also investigated the value of recording and informing the participant of the errors made. The results show that while both groups showed an improved task completion time, the segmented approach resulted in fewer errors in the early stages of training. The small advantage seen in error reduction for the segmented task approach could possibly benefit the automotive industry by providing a more proficient workforce.

Virtual Environments (VEs), Virtual Reality (VR), virtual assembly training, efficiency, training procedures, automotive industry

1. Introduction

The current demand within the automotive industry to provide a broad range of high quality new products, in a shorter time and at minimum cost, is in response to the expansion of consumer needs and competition from global markets [HKW_11]. Product variety is mainly attained in the final assembly stage, where multi-skilled workers perform several assembly tasks. Assembly operations consist of complex tasks, with the quality of the final product relying on the ability of the worker to remember and perform correctly the different assembly operations [MMP_10]. Therefore, a flexible and adaptive training system is required in order to maintain a proficient assembly workforce [GSW_11]. Traditional methods of assembly training are time consuming and expensive [BSG_07] however recent advances in virtual technology have lead to an increasing number of industries introducing virtual training systems to support, or replace these traditional methods [Pat10].

Research has shown that Virtual Reality (VR) has the potential to provide training that is more versatile than the traditional methods. Studies have shown that virtual assembly training can provide workers with guidance and instruction concerning work-related procedures, activities and actions and can improve and accelerate learning when compared to traditional training [BBG12; EETM12; SK07; MSH04]. Most of the research involves the development of high fidelity systems that represent the real-world task under investigation. Although significant time and money has been spent on their development [VFM13; JW12; SD11; SWP09; Joh05; SSV05], many are designed without considering the effectiveness and efficiency of the integrated training strategies, resulting in unstructured programs based on methods that have been directly transferred from the traditional training program [SGRO_02].

Confidence in virtual training can only truly be achieved if the system has a carefully considered and integrated training program [GRCH_05]. Efficient virtual training systems must incorporate the right learning principles to provide the information needed at all levels of investigation [BS10]. However, current research in virtual assembly training systems within industry fails to evaluate the efficiency of the adopted training program or instructional material [Ese02]. The task of establishing virtual training programs for VEs is complex and procedure specific. It requires the deconstruction of the task into its individual components and the definition of optimal performance [GRCH_05].

Within the medical field, research has investigated how to effectively train virtual surgical skills to a sufficient standard that they can be used in the real operating environment [GRCH_05; MKMB_04; GKBB_04; HLRL02; SGR0_02]. This has led to the establishment of a number of principles that have contributed to the successful integration of virtual surgical training into the medical curriculum [GRCH_05]. This current paper reports on a study whose main aim was to compare segmented training to whole task assembly training. This is based on the
principle of distribution of training found within medical education but adopted to suit the practicalities of training in the automotive industry. However, this study also investigates the value of obtaining an error measurement to evaluate the accuracy of the performance and informing the trainee of their mistakes so that corrections can be made. The background to these principles is described below.

1.2. Distribution of Training

Distribution of training refers to the spacing of training materials or sessions over time. A number of recent medical studies have compared training sessions given in one long session (massed session) with training given in multiple, short sessions (interval session) and have shown that the latter is more beneficial [SWMH09; MDMG 06; MMDC_02]. However, this method involves presenting the training over a significant amount of time causing problems relating to the time limitation for medical trainees wishing to perform procedures in the clinical environment [GR08]. Grantcharov & Reznick (2008) proposed a method that overcomes the time limitations by dividing the task into sub-tasks allowing the trainee to identify and follow the procedure in the correct manner and order while performing the procedure. Segmented task (ST) training consisting of splitting a task into sub-tasks is an alternative to more traditional complete task (CT) training and has been used in the training of common surgical procedures such as laparoscopic cholecystectomy, hernia repair, and Nissen’s fundoplication [GSK07; SJ04].

1.3. Error as a performance measure and for feedback

Almost all studies relating to virtual assembly training systems have evaluated performance by measuring task completion time, either during training or during real task performance. However, being able to perform a task quickly does not necessarily mean that it is correct or accurate therefore time may not be the most beneficial measurement [GRCH_05]. For example within the medical field, being able to perform a surgical procedure quickly gives no indication of the quality of the procedure. For critical processes, inaccurate procedures can have severe consequences.

The medical literature has highlighted that error is a more valuable measure than time because it is a key indicator of skill level with senior or experienced surgeons consistently performing well within virtual training environments [GRCH_05]. A number of reports have investigated the efficacy of virtual training to objectively measure technical competence of surgical skills using error as a performance measure [GAPD08; HS06; GLMS 04; MMSD03; AMKD02; SGRO _02; DST99]. The studies have shown that evaluation of error provides the VE with the ability to assess the skills of individuals before they progress to real patients [HS06] and can distinguish between the experienced surgeons and the novice trainees [SGRO02]. Within vehicle manufacture, accuracy of the assembly is often essential for quality control and safety and analysis of error has been used in some studies as a measure to investigate virtual assembly training.

In addition to error being a valid performance measure, the medical literature also highlights the value of error feedback to the trainee. Providing error feedback allows the trainee to learn what they have done wrong and avoid making the same mistakes in the future. Providing the trainee with error information can aid in learning, improve performance, make performance consistent, and reduce errors, and is optimized when the feedback is given as soon as possible to when the error is committed [GRCH_05].

2. Method

The experiment was designed to investigate the effectiveness of virtual assembly training adapting principles that have been established within medical research. Specifically, the experiment compared segmented task (ST) training with complete task (CT) training to determine the optimum method for the retention of information and performance related to the assembly of a LEGO® car. This focus on training strategy has previously been neglected on studies using VEs for assembly operations. While prior research has shown the benefits of interval session training when delivered over several days (Stefanidis, et al 2009; Moulton, et al 2006; Mackay, et al 2002), this would be difficult to implement in the automotive industry as time pressure limits the amount of time spent training (Hermawatia et al 2013). However the method, proposed by Grantcharov & Reznick (2008) that applied segmented task (ST) training in order to overcome the time limitations imposed by extending training programs over an extended period of time was employed in this study. Therefore, the ST training was implemented in one session with tasks presented in segmented cycles which were repeated before progressing to the next stage, and compared to CT training in which the complete assembly task was conducted, and then repeated. This adaptation to the approach aimed to investigate whether any of the benefits of ST training seen in the medical literature could be realised in assembly training when applied in a short-cycle/segmented task approach, albeit delivered within one session. The experiment also adopted principles from the medical literature, such as recording error to determine performance, providing feedback on error to participants and investigating training to a pre-established proficiency criterion.

The method employed in this study and described below received ethical approval by The University of Nottingham Faculty of Engineering Ethics Committee.

2.1. Participant Selection

Power analysis was conducted to calculate the minimum sample size required to accept the outcome of a statistical test with a 0.05 confidence (power = 0.8). A sample size
of 17 per group was deemed acceptable. A total of 34 participants were recruited from The University of Nottingham student and staff population (19 female 15 male; average age 30.24; age range 21-53). All participants were screened to be English speaking adults over the age of 18, have normal or corrected vision and the ability to recognise colour. The participants were randomly allocated into one of the two groups, the Segmented Task (ST) training group and the Complete Task (CT) training groups.

2.2. Apparatus

The study involved two computers, both of which were used during the training session and a model of a LEGO® car, all of which are shown in Figure 1. Language

![Figure 1: The experimental set-up](image1)

One computer displayed the LEGO® Digital Designer (LDD), a computer program produced by the LEGO® Group which allows for the construction of virtual LEGO® models. The second computer displayed the instructions for the assembly given via PowerPoint presentation. Figure 2 shows a screenshot of the LDD displaying the assembled LEGO® car.

![Figure 2: Screenshot of LEGO® Digital Designer and the dismantled pieces of the LEGO® car](image2)

For the purposes of this investigation, a LEGO® car was used because it provided a task that represented skills required in assembly work such as procedure recall, part recognition and part placement (Hermawattia et al 2013). In addition, pilot testing proved it was sufficiently complex to require training, while still being achievable within an appropriate time frame. The LEGO® model of the car is shown in Figure 3.

![Figure 3; Screenshot of the LEGO® Digital Designer with the assembled LEGO® car](image3)

2.3. Performance Criteria

The performance criterion was established during a pilot experiment involving the mean values obtained from three participants. The task completion time was set at less than five minutes and the error rate was set at three or less. The criteria would have to be attained on two occasions to avoid the measure being obtained by chance. However, due to experimental restrictions on time, if the criteria had not been attained after three training cycles the experiment ended, but the measures were still included in the analysis.

2.4. Experimental Procedure

Before the experiment commenced, informed consent and demographic data was obtained and the participant was assigned a random participant number. Verbal instructions on the purpose and procedure of the experiment were given. Both experimental groups followed the same protocol. The participants were trained in the assembly of the LEGO® car using the LDD followed by the trial involving the assembly of the real car where performance measures were taken. The participant completed two cycles of training (either ST or CT training) followed each time by the assembly of the real car. The training stage and the trial stage were then repeated until the proficiency criteria had been reached on two consecutive occasions. A summary of the experimental procedure is shown in Figure 4.
Initial stage

The initial stage involved all participants undertaking a pre-training session which aimed to give instruction in the use of LDD program. The researcher demonstrated the assembly of a simple LEGO® model and how to rotate the parts, zoom in and out and move the completed model. The participant was then allowed to practise these actions until they felt comfortable with the use of the software. This took no longer than a couple of minutes for all participants.

Training stage

Each group trained on the LDD in the virtual assembly of the LEGO® car but the PowerPoint assembly instructions, and therefore the order of their practice sessions, were presented to each group differently. The ST group was presented with the instructions in five separate sections and they worked through each section twice before moving on to the next section. The CT group worked their way through the full set of instructions, twice.

Trial stage

At the end of the training, all participants undertook the trial involving the assembly of the real LEGO® car. The participants were encouraged to remember the assembly but if required the PowerPoint instructions were made available. After each trial performance measures were recorded and the participants were informed of the errors that they made.

2.5. Data analysis

Performance measures were collected during the final stage in the form of error, instructional reference and task completion time. Errors were defined as when a part was placed in the wrong position, when the wrong colour part was used or when there were parts left over at the end. Instructional reference was established by recording the number of times the participant referred to the PowerPoint instructions during the trial stage. Task completion time was defined as the time taken to perform the real assembly task.

In addition, once the study had been completed, each participant was asked to rate the following statements on a seven-point Likert scale ranging from strongly disagree to strongly agree:

- I am satisfied with the way the training was presented
- The training program made it simple to learn
- The training program made it fun to learn
- The training program made it easy to remember how to perform the task
- The LEGO® Digital Designer was easy to use
- The LEGO® Digital Designer was fun to use
- It was easy to recognise the parts within the LEGO® Digital Designer
- It was easy to position the parts within the LEGO® Digital Designer

3. Results

3.1 Objective data

A one-way between subjects ANOVA was conducted to compare the different measures obtained from each group during the three real trial stages. Figure 5 shows the mean and standard error of the task completion times between the two groups at the three trial stages. There was a main effect of trial (F (2, 64) = 75.58, p = < 0.001, Eta²=0.79) which showed that task completion time decreased over the three trial stages. Pairwise comparisons with Bonferroni corrections (only accepting a significant result if p < 0.0117) showed significant differences in task completion time between all three trials (p < 0.001).

There was no main effect of group (F (1, 32) = 0.19, p=0.665, Eta²=0.006) or interaction between trial and group (F= (2,64)= 0.146, p = 0.865, Eta²=0.006) or interaction between trial and

![Figure 5; Mean task completion time for both groups at the three trials](image)

Figure 6 shows the mean and standard error of the number of instructional references used by the two groups at the three trial stages. There was no main effect of group (F (1,32) = 1.398, p = 0.246, Eta²=0.042) interaction (F (2,64) = 2.710, p = 0.074, Eta²=0.078). However there was a significant main effect for trial (F= (2,64)= 120.061, p < 0.001, Eta²=0.79). Pairwise comparisons with Bonferroni corrections (only accepting a significant result if p < 0.0117) showed significant differences in task completion time between all three trials (p < 0.001).
Figure 6; Mean number of instructional references for both groups at the three trials

Figure 7 shows the mean and standard error of the number of errors by the two groups at the three trial stages. There was no overall effect of group (F (1,32) = 3.422, p = 0.074, Eta2=0.097). However, there was significant effect of trial (F (2,64) = 45.586, p < 0.001, Eta2 = 0.588) and an interaction between trial and group (F= (2,64)= 4.120, p = 0.021, Eta2 = 0.114). Pairwise comparisons showed a significant difference between the groups at trial 1 (p < 0.001).

Figure 7; Mean error results for both groups at the three real trials

Figure 8 shows the amount of participants that reached the proficiency criteria (task completion time ≤ 5 minutes and ≤ 3 errors) in each group during three trials. A chi2 test was undertaken to investigate if there were any significant differences between the groups in the three trials. The results show that there was no significant difference (chi2 = 2.60, df = 2, p = < 5.99).

Figure 8; Number of participants reaching proficiency during each of the real assembly trials

3.2 Subjective data

Individual t-tests with significance set at 0.00625 (p<0.05 with Bonferroni correction) or less were performed on each of the questions proposed in the Likert scale to investigate if there were any differences between the groups and the results are shown in Table 1.

Table 1; The questions posed in the Likert scale and the corresponding results.

<table>
<thead>
<tr>
<th>Question</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am satisfied with the way the training was presented</td>
<td>t (1,32) = -0.254, p = 0.801</td>
</tr>
<tr>
<td>The training program made it simple to learn</td>
<td>t (1,32) = 0.841, p = 0.407</td>
</tr>
<tr>
<td>The training program made it fun to learn</td>
<td>t (1,32) = -0.182, p = 0.857</td>
</tr>
<tr>
<td>The training program made it easy to remember how to perform the task</td>
<td>t (1,32) = 1.279, p = 0.210</td>
</tr>
<tr>
<td>The LEGO® Digital Designer was easy to use</td>
<td>t (1,32) = 0.702, p = 0.488</td>
</tr>
<tr>
<td>The LEGO® Digital Designer was fun to use</td>
<td>t (1,32) = 1.667, p = 0.105</td>
</tr>
<tr>
<td>It was easy to recognise the parts within the LEGO® Digital Designer</td>
<td>t (1,32) = -0.356, p = 0.724</td>
</tr>
<tr>
<td>It was easy to position the parts within the LEGO® Digital Designer</td>
<td>t (1,32) = 0.614, p = 0.544</td>
</tr>
</tbody>
</table>
The results show that there was no difference between the groups on the answers given for each question.

4. Discussion

This paper reports on a study that aimed at employing principles established within medical research into virtual surgical training and adapting them for a study into virtual assembly training. The study compared the efficiency of segmented task (ST) training to complete task (CT) training in the assembly of a LEGO® car. In accordance with best practise from the medical domain, the experiment also obtained an error measurement to evaluate the accuracy of the performance and effectiveness of the training, informed all participants of the errors made during the real trial so that they could avoid the mistake in successive trials and trained to a pre-established proficiency criterion.

All participants in both groups showed an improvement in performance in each of the successive trials. Analysis showed no statistical significance between the groups in task completion time and instructional reference during any of the real trials. A factor that allowed all the participants to learn the assembly, in a relatively short space of time, could be the LEGO® model used in the study. Although, the model of the car was thought to provide enough cognitive load, the speed of which all the participants in both groups learnt the assembly without the aid of instructional reference suggests that it may have been too simple.

A significant difference between the groups was demonstrated during the first real trial, with the ST group incurring less errors when compared to the CT group although significance was not seen at any of the other trials. These results suggest training given in segments in the early stages of the training program can improve the accuracy of the performance. The potential reasons for the lack of significance for error in the other trials could relate to the principle of providing the participants with error information. From observations made by the researcher at the time of the real trials, once the participant was aware of a mistake made, most did not make the same mistake again.

From observations made during the study it appeared that the participants in the ST group followed the procedure that was presented to them during training in the real trial more closely when compared to participants in the CT group who did not necessarily assemble the car following the procedure presented during training. This could be the explanation for the ST group committing fewer errors and suggests that for tasks that are procedure specific ST training is beneficial.

Although the results showed that there was no difference between the groups on the amount of participants that reached the proficiency level within the three training cycles, the study did show that it was possible to define the skills level for the assembly task and chart the performance of a participant. The proficiency measure may have been more sensitive to the differences between the training approaches for a more complex task (see Future Studies below). The setting of a performance criterion enables trainers to be confident in the ability of the trainee prior to performing a real assembly operation and ensures a much more homogeneous skill set within a group of workers.

The results of this study indicate a slight advantage to a ST training approach in non-complex tasks. When considering the thousands of parts which comprise a modern car, as well as the thousands of assembly line operators, this small advantage could multiply to a significant cost saving (through higher quality and therefore reduced returns) when implemented in the automotive assembly training. The following section describes suggested further work to fully understand the differences between ST training and CT training.

5. Future Studies

VEs have the potential to be a powerful training and assessment tool when integrated into a well-structured curriculum (Gallagher et al 2005). However, the initial lack of research within the industrial literature regarding how to effectively apply a virtual training program could be one reason for the overdue acceptance of this technology within manufacturing. More research is needed in the area of curriculum development for virtual assembly training systems before confidence in the results of an evaluation of a virtual training system can be achieved.

To further assess the benefits of segmented task training on complex assembly operations, the above study will have to be repeated using a more difficult assembly model. This could produce results in which the differences in the groups appears more pronounced and significant. It would also be interesting to investigate the contribution of informing the participants of the errors made to the speed of learning, by comparing a group that received error feedback with a group that did not receive feedback. More studies are needed to assess the effect of training on a proficiency based VR curriculum versus traditional methods of assembly training. Informing the participant of the proficiency criteria before training commences might be an element that motivates learning and results in the acquisition of proficiency at an earlier stage and would therefore be a valuable investigation. Analysing errors made during the real trial was found to be a valid measurement for the assessment of accuracy of performance but to be confident in the data further research would have to involve an increased number of participants.

Future studies could also incorporate other principles established within virtual surgical training including the value of pre-training in virtual reality, virtual instructions and how to present them, determining learning curves and using training strategies such shaping and fading (Gallager et al 2005). Shaping involves starting the training with a relatively simple task that gradually become more difficult and fading involves providing clues and guides which are
gradually faded out as the task becomes more difficult. Both shading and fading would involve different learning levels and trainees would not progress until they consistently met defined performance criterion (Gallager et al. 2005).

All these principles could be investigated using the methods and equipment presented in this study including the use of different LEGO® models that increase in difficulty of assembly.

6. Conclusion

In this study, a slight advantage was recorded for segmented task training over massed session training shown by the reduction in error over time. The results validate the benefits of introducing principles established in virtual surgical training. Virtual training is more likely to be successful if it is integrated into a well-thought-out training program with an appropriate training schedule and objectively determined proficiency criterion and valid performance measures relevant to the task being trained.

Acknowledgments

This study is one part of the VISTRA (Virtual Simulation and Training of Assembly and Service Processes in Digital Factories) project (FP7-ICT-285176) which aims to develop a consistent and complete digital toolset for the virtual training of assembly processes.

7. References


Towards a Low Cost High Immersive Virtual Environments
for Training Heavy Machinery Operators

F. J. Lacueva1, M. A. Gracia1, J. Gracia1, D. Romero1, A. Sanchez de Olavarrieta2

1Instituto Tecnológico de Aragón - ITAINNOVA, Huesca, Spain
2TAIM WESER S.A., Zaragoza, Spain

Abstract
This paper presents the work of ITAINNOVA Multimedia Division (IMD) in creating low cost training Immersive Virtual Environments (IVE) for our heavy machinery customers. First, we present the background of industrial expectations in the general application of IVEs. Then, we complement these expectations with a state of the art (SoA) of IVEs supporting technologies, existing research projects and available commercial products in order to detect opportunities that improve our results. In particular we present the projects we are developing together with TAIM WESER (TW) and with the Aragon Government’s Education, Culture and Sports (AGDECS) support which are related with the use of IVEs for pre-selling tasks and operator training. Finally we present the preliminary results of our work in the creation of a methodology to evaluate existing Commercial-of-the-Shelf (COTS) driver technologies to be used in real IVEs. This methodology will evaluate those technologies based on the user Quality of Experience (QoE) KPIs for determining its usability and feasibility but also as a way to determine their applicability in other scenarios through the entire virtual continuum.

Categories and Subject Descriptors (according to ACM CCS):

1. Introduction
This paper presents IMD’s work in creating low cost IVE for adding value to our industrial customers at different points of their value chain. We centre this paper in the civil engineering, construction and machinery-manufacturing (CECMM) sector due our close relationship with the Spanish National Association of this kind of enterprises (AN-MOPYC). More specifically, it is centred in big cranes and forklifts. In projects such as TAIM WESER VR Crane (TWVRC) or the Forklift Simulator (FS), we cover different product life cycle stages as the design and fast virtual prototyping of new crane products, their visualization embedded in the destination environment (i.e. a crane in an actual dock) or their operator training.

The paper is structured as follow: first, in point 2, we analyse the industrial background looking for enterprise motivations to use IVE (point 2.1) within the CECMM sector. Then we review the state of the art (SoA) of Serious Games (SG) for training (point 2.2) as well as their current application for training crane operators (point 2.3). The conclusions of these SoAs are the basis (point 2.3) for projects that are presented in point 3. We present here two of them: our technical developments for improve and create IVE (points 3.1, 3.2 and 3.4) as well as our work in the evaluation of the user’s QoE (point 3.2). This QoE approach is a very interesting point to consider for assuring success of low cost IVEs as training tools. It will provide an objective framework to evaluate COTS devices’ usability and feasibility in real training tasks. The presented QoE study is part of the work we’re developing in the creation of an evaluation framework that will allow the evaluation of existing technologies to support different use cases and requirements.

2. Background

2.1. Industrial Background and Reasons of Interest

The conclusions of the CECE 2014 congress [CEC14] show CECMM as a sector in a deep restructuring and recovery after the economic crisis. However, it has showed its resistance by improving its competitiveness.

CECMM enterprises, and in particular SMEs, should create more innovative and high quality products and services while they should reduce their operational costs and time-to-market of its products in order to improve their competi-
tiveness. They need to apply technologies to face affordable, efficient and precise customization of their products and services to their customer demands.

IVEs already contribute to achieve these goals. IVEs support working in a virtual collaborative environment simplifying the visualization of complex technical information. IVEs are applied through all the value chain of the enterprises: validating design prototypes; visualizing complex simulations of the different subsystems; providing information to workers during the manufacturing processes; training final users; remotely guiding operators or maintainers. Immersive technologies support user deeper understanding of the product and/or service he is buying or using. They reduce the time-to-market, in particular with big machines such as TAIM WESER cranes.

IVEs deployed using COTS devices will create competitive opportunities for enterprises as they would provide low cost solutions for interacting with virtual prototypes of the products working in its intended environment. They would also reduce physical risks and save time during training phases because low cost IVE would allow training sessions to be performed just after signing the contract and, in the case of big machinery, before its final assembly would have even been started [BRU13].

2.2. Simulation, VR and SG For Training

The term Serious Games [SB07][MC05] includes several technologies for developing and executing videogames to improve users capabilities by confronting them to goals in a virtual environment simulating the real world. SGs were first developed for military training and then they were adopted in other fields [RHM05]. The consequence is that military demands have guided the development of the SGs and of the Game Engines (GE) supporting them.

First, for reducing costs of the military SGs, it became necessary to standardize the way in which different 3D models, RT simulation results, etc. should be integrated in a simulation execution. High Level Architecture (HLA) [HLA10] was created as the standard set of tools to integrate, to execute, to supervise and to reproduce the different components of a SG simulation. HLA has become the de facto standard for distributed RT simulations within the military field but also in the industrial one [HG03]. The consequence is that numerous GEIs (VB52 [VHP15], Unity3D [UHP15], etc.) have developed interfaces to participate as federated nodes within HLA simulations (VR-Link for Unity of VT MAK [VBS15], etc.).

The demanded realism for military SGs requires a higher dynamism of every element, to increase user immersion within the virtual tasks the user/operator is performing. In recent years, these dynamic IVEs are becoming a reality [LH13]. There are simulators where all the elements (terrains, objects, sound, etc.) are designed to react to user(s) interactions (modifying their shape, position, etc.). They optimize training results, reducing the time and the cost versus the real training [CON12][LIN15]. The more realism is demanded, the more computational power is required and might exceeds the actual capacity of centralised systems. HLA capabilities for distributing computational requirements between computers would support the integration of specialized physics simulators (fluids, synthetic 3D sound, complex mechanical models, etc.).

Military SGs have also benefited of Games Engines (GE) development. They have taken advantage of their capacity to incorporate new COTS interaction devices. This has contributed to obtain low cost high IVE that replaces very expensive VR Rooms (aka caves) with high resolution projectors, movement’s sensors, etc. Head Mounted Displays (HMD), Gesture Recognition Devices (GRD) and haptic devices create a much more immersive user experience (UX) with a fraction of the ownership cost.

Figure 1 summarizes the maturity level (ML) of the technologies involved in the creation of low cost IVEs [PG13][LAU13]. These technologies are evaluated from the user point of view, its interactivity and their responsiveness to the stimulus from the virtual world (VW).

![Figure 1. Interaction Technologies Maturity.](image)

The maturity of the synthesizers for visual and auditory channels reflects the high percentage of information human beings gather through them. It is also represented by the number of existing COTS devices for these specific communicative channels: visual (Oculus Rift [OHP15], SONY[SGA15], etc.) and auditory (3D positional audio, bi-aural algorithms, …).

On the other side, reacting to user stimulus developments are at a lower development level. Sometimes IVEs have taken advantage from the real world adoption of solutions created for virtual worlds: joysticks are an example of devices highly used by heavy machinery. However, there are many simple interactions (pressing a button or operating a lever), which are not well implemented in VWs. Gesture recognition devices are able to recognize hands movements and they can contribute to improve interaction. COSTs devices (Leap Motion [LHP15], MS Kinect [KHP15], etc.) allow capturing natural user actions, but they do not allow reproducing the tactile feedback from the system.

HLA frameworks and GeEs present other capabilities interesting for training people. They support the recording of the events generated during a simulation execution so they can be reproduced later to detect and to correct errors. As they are able to distribute computational power, they can support multiplayer execution and visualization to allow more realistic, immersive and collaborative interaction of the operators. Moreover, multiplayer supports Command
and Control (C2) modes to allow RT supervision of the training task.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Research Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-Screen</td>
<td>[ZO14] [REZ11]</td>
</tr>
<tr>
<td>Cave</td>
<td>✓</td>
</tr>
<tr>
<td>Interoperability</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>Physic Interaction</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>Component Physic</td>
<td>✓ ✓ ✓</td>
</tr>
<tr>
<td>Load Physic</td>
<td>✓ ✓</td>
</tr>
<tr>
<td>Environment Physic</td>
<td>✓ ✓</td>
</tr>
<tr>
<td>Training Evaluation</td>
<td>✓</td>
</tr>
<tr>
<td>Multi-exercise</td>
<td>✓ ✓</td>
</tr>
<tr>
<td>RT Supervision</td>
<td>✓ ✓</td>
</tr>
</tbody>
</table>

Table 1. Crane IVE research projects.

HLA, GE and COTS devices contribute to reduce the cost of developing and executing training simulations on IVEs. This, together with the reduction in size and cost of the required equipment, has increased their mobility. That is, SGs developed on COTS devices can be applied for training of heavy machine operators with a reduce cost and risks almost everywhere.

2.3. Training SG for machine operators

With this background, SGs have been naturally adopted by educational [DBB14] [ARN12], medical [GSS12], corporative [ARH12], manufacturing [MAV13], logistic [KÖN11], architecture [XZ02] or construction sectors [LCS12]. In this paper we focus on the application of SGs for training in the use of heavy machinery and, more specifically, in the use of machinery for loading and unloading containers (cranes) and pallets (forklifts).

Tables 1 and 2 summarize a SoA of existing crane simulators, both results of research projects and commercial products. It evaluates different criteria such as: the HLA compliance; the capabilities of visualization (as a way to determine its mobility, cost, etc.); the kind of physics they can reproduce; the kind of human interaction; the possibility of define different scenarios based on multiple machines models and/or different exercises; the capabilities for evaluate and supervise the training exercise.

Table 2. Crane IVE commercial products.

2.4. Detected Opportunities to Improve

Having the objective to create low cost, portable and high quality IVEs able to be used in different training scenarios customized by supervisor users., we use previous SoA to detect opportunities which allow us to increase the value we give to our clients by providing IVEs.

A first opportunity is the generation of a framework to support the selective reduction of the complexity of machine’s CAD 3D models in order to adapt that model to the simulation realism required. This framework will also advance in the automation of importing models in the GEs that will contribute to the reduction of the SGs development costs but also to their flexibility.

HMDs can contribute both to the reduction of the cost and to the increase of the realism of the simulation. Moreover they can contribute to the mobility of the complete solution.

One of the VR HMDs drawbacks is that they don’t allow user to see its hands. Using Gesture Recognition (GR) could solve this problem. Once it would be solved a second step would be integrate the interaction with complex consoles. Here is where Augmented Reality (AR) technologies would help to increase interaction realism. Basic AR markers can be use to locate and to show console elements in the virtual world. Moreover a “fake” console would be able to use to create a basic haptic feeling.

Finally we see that an additional opportunity to improve is to extend the use of IVE to other practices within the heavy machinery such as enterprise’s workers guiding in building the machine or to be used for remote assistance and even for remote machine operation.
3. IMD’s works in low cost IVE development.

Next sections shows IMD’s work in creating low cost IVEs. The projects represent innovative tools for designing, virtual prototyping, pre-selling, or training services as well as, internal projects to advance in the SoA: to increase user interaction and immersivity. We are also creating a methodology to evaluate user’s Quality of Experience (QoE) when using HMDs.

![Figure 2. Commercial IVE types.](image)

3.1. Automation of 3D Model Importation.

Automate importing of 3D Model is a common task for the use of existing CAD/CAM/CAE 3D models in either IVEs or AR applications and it will have a great impact in their success.

Creating a framework able to simplify the importation of existing models is a necessity because of several reasons: first, because of the information these files contain; second, because of the detail level which does not usually correspond to SGs needs; finally there are different formats for exporting, importing and exchange models.

Examples of the results of our work are the TWVRC project or the FS we present next sections. At the moment, we are working on a semiautomatic framework which will allow the user to interact with elements from the original model.

In the TWVRC project we are using the grab semiportal crane 3D CAD model provided by the company in order to test our developments. The 3D model is composed of 15878 parts. Managing these numbers in an automatic and reliable way will contribute to project success: reduction of the computational power; better immersive UX; better detail level of the relevant parts; more autonomy for TW users.

3.2. Low cost forklift driver training simulator

The purpose of the Forklift Simulator (FS) is to enable users to acquire basic skills for handling a forklift before starting to drive a real one, and thus reducing possible dangers or damages caused by inexperience or insecurity. The simulator is part of an official course of the AGDES for obtaining the forklift driver license. It consists of several training practices: from the most basic (such as the forklift power-on protocol) to the more advanced ones (e.g. carry loads from compact shelves into a truck).

One of the key factors of the FS development has been to increase the sense of realism allowing students to become familiar with every aspect related to the forklift. We have worked in the different aspects: visual fidelity, motion fidelity and the interaction with realistic controls.

To achieve visual fidelity, we prioritized to model the virtual world by mimicking as many details as possible of the real forklift and its usual environment: instrument panels, cabin, forks, etc. Physics, materials and textures were also carefully included.

![Figure 3. Forklift Simulator during User Evaluation.](image)

Several studies support that motion fidelity is a factor even more important than visual fidelity [BSH 05], thus especial attention has been paid to develop the physics of the system so the movement behaviour and reactions of the real forklift were closely simulated in terms of speed, rotation axis, response delays or acceleration. In addition, an important effort has been made to mimic the interaction between the user and the virtual forklift. For this purpose, COTSs have been used (a gaming steering wheel with control buttons and pedals) to mimic the interaction between the user and the virtual forklift.

The FS is currently compatible and working in two different environments: a PC with a conventional 2D/3D computer screen and it also has a version for the Oculus Rift™ HMD.

As the HMD completely covers the user’s eyes, he/she is not able to watch the real surrounding environment. On the one hand, this fact allows increasing the sense of immersion but, on the other hand, the user loses any kind of link with the real world (he/she cannot see his/her real hands nor the “physical” interaction devices, etc.). To decrease the latter effect, self-awareness techniques are used. An avatar accurately maps each of his/her movements to the virtual world: the position of the hands over the wheel, as well as the feet and the head pose when the user looks around.
performed a study to evaluate how the Oculus Rift from a learning (task-oriented) point of view. Then we focused on the realism that this device provides: the immersive level they provide. Lacks HMDs have are the lost of visual connection with the real world because of the immersive level they provide. This blocks the possibilities of “virtual tactile interaction” within the virtual world.

3.3. Quality of Experience of IVE.

After finishing the FS, it was evaluated by final users from a learning (task-oriented) point of view. Then we focused on the realism that this device provides: the immersive level they provide. This blocks the possibilities of “virtual tactile interaction” within the virtual world.

Figure 4 shows a proof of concept we developed in order to get a virtual hands image of user’s real hands that will allow him to interact in the virtual world by using natural gestures [SG15]. We use Leap Motion™ as the sensor to gather information about hand and fingers positions and their movements.

Although we do not incorporate it to the FS yet, our test shows how the realism would improve the learning capabilities of the final simulator.

3.4. Gesture Recognition for Improving Interaction in IVE

As we introduced when we describe the FS, one of the lacks HMDs have are the lost of visual connection with the real world because of the immersive level they provide. This blocks the possibilities of “virtual tactile interaction” within the virtual world.

Figure 4 shows a proof of concept we developed in order to get a virtual hands image of user’s real hands that will allow him to interact in the virtual world by using natural gestures [SG15]. We use Leap Motion™ as the sensor to gather information about hand and fingers positions and their movements.

4. Conclusions and Future Work

Low cost IVEs seems to be good tools for training heavy machine operators. With COTS characteristics and computational power improving fast, the realism of obtained simulations should improve at same pace and it would take advantage of distributed environments such as HLA.

However, answering to industrial requirements will require the combination of several technologies in order to improve SGs cost and flexibility. The combination of AR, VR, CAD, etc. will allow it. An example of this advance is the combination of Gesture Recognition and HMD for advancing in the blurring of the frontiers within the virtual continuum. These combinations would also allow the extension of VR technologies to other applications: remote control, collaborative works, etc.

In order to determine which technologies are applicable to which scenarios and under which conditions, we are working in a methodology to evaluate technologies against necessities. This methodology does not only consider the technological dimension but also UX one by providing ways to evaluate its QoE when using a specific implementation in order to avoid potential hazards.

Finally, for increase flexibility, reduce efforts in IVEs creation, operational costs (maintenance and updating) and protect industrial know-how, automatic importing frameworks have to be developed. This will allow that projects, such as the TWVRC project, can evolve from the preselling visualization of one crane model and the simulation of the crane behaviour for training purposes to the visualization and simulation of potentially all the cranes models TW has in its portfolio.

Acknowledgment

This work has been partly funded by the Aragon Government Department of Innovation, Research and University.

References


Virtual reality environment for disassembly sequences generation and evaluation

Peter MITROUCHEV¹, Chen-gang WANG¹, Jing-tao CHEN¹

¹Univ. Grenoble Alpes, G-SCOP, F-38000, Grenoble, France, CNRS, G-SCOP, F-38000 Grenoble, France, Peter.Mitrouchev@es-cop.enspg.fr

Abstract

In virtual reality environment (VRE) a human model is often involved in a digital mock-up (DMU) model for assembly/disassembly (A/D) operation simulation. However, it has limited application areas because of its high cost investment. In this context, a mixed Virtual Reality Disassembly Environment (VRDE) is developed based on Python programming language using mixed VTK (Visualization Toolkit) and ODE (Open Dynamics Engine) libraries. First it allows the generation of the possible disassembly sequences for selective disassembly operations. Based on the lowest levels of a disassembly product graph the proposed method considers the geometric contact and collision relationships among the components in order to generate the proposed Disassembly Geometry Contacting Graph (DGCG). The latter is then used for disassembly sequence generation thus allowing decreasing the number of possible disassembly sequences. The method is applied for automatically generating the selective disassembly sequences thus allowing reducing the computation resources and is illustrated through an example. Secondly, focusing on the evaluation of disassembly operations in a VRE, a method based on five new criteria is proposed. The criteria are presented by dimensionless coefficients automatically calculated thus allowing evaluating disassembly sequences complexity in a VRE. The proposed method is tested and an example for disassembly operations evaluation of a mechanical assembly is presented. The results of the analysis and findings demonstrate the feasibility of the proposed approaches for disassembly sequences generations and their evaluation thus providing significant improvement of the Product Development Process (PDP).

Applications in manufacturing and engineering

1. Introduction

Nowadays, Virtual Reality environments (VRE) have significantly evolved towards assembly/disassembly (A/D) simulation, highlighting new needs for their simulation preparation, evaluation and integration in Product Development Process (PDP). A/D simulations address different objectives such as: sequencing modelling, path planning, collision detection, operational time etc., which are often complementary to each other [AC11, IML11]. They play a very important role during the initial design phase of an industrial product.

VRE and Augmented Reality environment (ARE) for disassembly sequences generating are based on two disassembly approaches which are: interactive and automated (or semi-automated) [LT07]. However, both of these have some limits. Interactive approaches, for instance, require extensive user input usually in the form of answering questions, whereas automated approaches are usually used to generate disassembly processes for products with relatively simple component configuration and geometry. In VRE where interactive simulation is critical, fast and accurate collision detection among moving objects is also a challenging problem. Most of the recent work on A/D related with Virtual Reality (VR) technology focuses on the simulation itself. They try to build an environment to assemble or disassemble products and to compare the simulation results with the results of real A/D processes. From this perspective, Aleotti and Caselli [AC11] proposed a physics-based VRE for task learning and intelligent disassembly planning. Ladevere et al. [LFP10] proposed an interactive path planning method for haptic assistance in assembly task. Mohd et al. [MWS12] proposed some criteria for A/D sequences evaluation, mainly used for complexity analysis of assembly in order to maximize the flexibility of A/D sequencing, or to minimize the assembly time through parallel execution of some assembly tasks. The purpose of these studies was to obtain approximate disassembly time for a product using formulas derived from the information pertaining the connexion of the parts instead of disassem-
bling the product in reality. However, none of them considered the ergonomic evaluation related to product disassembly operation. Thus, Ergonomics evaluation considering the disassembly operation should be involved in the future design phase. Moreover, the disassembly time is much different according to each disassembly sequence and the operators involved. Some commercial software and tools were also proposed to perform ergonomic evaluation during assembly [SVO11]. However, this evaluation is relatively expensive. Moreover, VR-based applications use real-time interactions and immersive techniques allowing enlarging the user perception of digital models.

After reviewing some current approaches, which have been partly presented here above, it can be stated that the existing VR approaches still have limitations in the generation and evaluation of disassembly sequences. Consequently, there is a strong need to evaluate disassembly operation for immersive simulations with a larger set of possible movements and to get more realistic results.

In this context, a Virtual Reality Disassembly Environment (VRDE) is presented here. First it allows generating the possible disassembly sequences based on contact identification method (performed by ODE libraries) and collision detection (performed by VTK libraries). Secondly, it allows disassembly operation evaluation based on five criteria (scores). Instead of the ergonomics simulation with a human model, it introduces some new sources in performing disassembling task in a VRE. The Environment is more efficient than those of Enomoto [EYS13] which only involves one ergonomic score evaluation. Here we propose five criteria divided in two categories for: ergonomic evaluation (visiblity score) and traditional processing evaluation (disassembly angle, number of tools' changes, path orientation changing and sub-assembly stability). The Environment is also more efficient than the simulation system using human model in a digital mock-up (DMU) [WZW_12, LFP10] whose main areas of application are limited to big mass production lines because of its high cost.

The results of this study may be useful for designers and industrials, allowing them to: take into account of the constraints of disassembly operations by automatically generating the selective disassembly sequences and their evaluation in a VRE during the initial phase of product’s design.

2. Virtual Reality Disassembly Environment (VRDE)

The general structure of the VRDE with the relationships between VTK and ODE libraries, which include the two main aspects of this work namely: the method for selective disassembly sequences generation and their evaluation, is shown in Fig. 1. The two loops in the flow chart for collision detection (Fig. 2) are simultaneously executed. Loop1 is the interaction with the model performed by VTK. Loop2 is the collision detection performed by ODE. In the disassembly process, these two loops affect the position of the parts’ models created by VTK.

2.1 Collision detection

In order to detect collisions the Kinematic criteria of ODE mass method was used. For this purpose a mass is associated to each ODE body. If the mass of the ODE Kinematic is active, the associated model is too heavy to be moved by the collision detection force feedback. Thus, the method can only be used to simulate the unmovable characters of the components because of the friction’s influence. The developed software, integrated in VRDE can support WRL and STL format files. In this application, the models created by SolidWorks were imported in the application in STL format [MWL-15].

![Figure 1. Structure of the mixed VTK and ODE VR Disassembly environment (VRDE)](image)

![Figure 2. Flow Chart for collision detection.](image)

2.2 Pixel calculation

In order to calculate the visibility score for the target part (here a bolt, Fig. 3), the camera in VRDE should be in the position of the human eyes and in the direction of the bolt. The target’s colour (here in red) should be different from the other components. For this purpose the other components are becoming black coloured in shades of grey as shown in Figure 4. The pixels’ counting is based on the

134
OpenCV library (http://opencv.org/). In the proposed method, two images are taken by the camera: the bolt itself (labelled in red, Picture 1 of Fig. 3) and the bolt in the assembly surroundings (Picture 2 of Fig. 3). The operator is turning the camera around the target component in order to obtain maximum visibility (see Section 3.1).

Thus, the proposed visibility score $v$ is defined as the ratio between the number of red colored pixels in the current image $v_a$ of the target part (Picture 2 of Fig. 3) and the number of red colored pixels of the whole image $v_b$ (Picture 1 of Fig. 3) captured by the camera:

$$v = \frac{v_a}{v_b}$$

with $v_b \neq 0$. If there is no obstacle part to hide the target, the visibility score is 1. If the target part is completely hidden by other parts, the visibility score is 0. Thus, the average visibility for a disassembly sequence is:

$$V = \frac{1}{m} \sum_{i=1}^{m} v_i$$

where $m$ is the number of components in the assembly.

3. Implementation

An application for disassembly simulation was developed running on the proposed virtual reality disassembly environment (VRDE). It is illustrated by an example of a five-part mechanism (mechanical assembly) disassembling (Fig. 5). The experiment consists in moving all the parts from the mechanical assembly to the destination vertical surface as shown in Fig. 5a. As previously said, the collision detection is performed with ODE. If a collision happens, the collision force changes the moving direction of the VTK model [WML_15].

3.1 Operation process

It consists in manipulating the camera. As previously said, the operator may move/rotate the camera to a convenient position for observation. The environment coordinates for the camera position and the object (target) position related to the human height (175cm) are first built. When the operator decides to disassemble the target, its pixels and position are automatically recorded for later analysis.
3.2 Calculation process

During the disassembly simulation, the realized application calculates the values of the proposed five criteria for disassembly operation evaluation. Those criteria are chosen because they cover two aspects of the disassembly operation evaluation. The first one is the ergonomic evaluation represented by the visibility score. The latter includes two movements: lateral and forward rotation angles of the neck, and trunk bending movement. The second aspect is the traditional processing evaluation represented by four criteria whose principal functions are: disassembly angle which defines the set of directions of removal (SDR); number of tools’ changes which characterizes the number of tools necessary to carry out the disassembly operation; path orientation changing which characterizes the execution time and finally sub-assembly stability which guarantees the safety of the operator. The main functions of these criteria were the reasons to choose them.

i). Visibility score (see Section 2.2): During the disassembly simulation, the realized application calculates the values of the proposed five criteria for disassembly operation evaluation. The first one is the ergonomic evaluation represented by the visibility score. The latter includes two movements: lateral and forward rotation angles of the neck, and trunk bending movement. The second aspect is the traditional processing evaluation represented by four criteria whose principal functions are: disassembly angle which defines the set of directions of removal (SDR); number of tools’ changes which characterizes the number of tools necessary to carry out the disassembly operation; path orientation changing which characterizes the execution time and finally sub-assembly stability which guarantees the safety of the operator. The main functions of these criteria were the reasons to choose them.

\[ C = \frac{1}{4\pi} \int_{0}^{\beta} \int_{0}^{\pi} \sin(\theta) d\theta d\phi \]  

where \( \theta \) and \( \phi \) are polar and azimuthal angles respectively.

The value of \( C \) ranges from 0 to 1. The best situation is when \( C=1 \) (all the possible movements are feasible) and the worst one when \( C=0 \) (there are no possible movements).

ii). Disassembly angle: Prior to disassembly operation simulation, the set of direction for removal (SDR) is calculated. It consists in detecting polar and azimuthal angles according to the assembly relationship amongst the components. The relative score for the disassembly angle \( C \) for a component (calculated in real time) is the ratio between the disassembly surface angle and the whole surface of the sphere (Fig. 6):

\[ C = 1 - \frac{\sum_{i=1}^{n} f_i}{m} \]  

where \( f \) is the number of the components, falling down in the gravitational field, calculated by the developed software. The value of \( Sta \) ranges from 0 to 1. For \( f=0 \), the stability is maximum, consequently \( Sta=1 \). The worst situation is for \( f=m \), when \( Sta=0 \). Note that component 5, (Fig. 5) being the base component, is not concerned by falling down under the effect of gravity. After disassembling the components 1 and 2, if component 3 is the next to be disassembled, components 3 and 4 will be in an unstable state thus threatening the operator’s safety. In this case, to continue the simulation, additional fixtures for components 3 and 4 have to be added in order to ensure the stability of the sub-assembly. If it is necessary to add a fixture to a component, the assembly time will increase. For that reason, in the performed VRE for disassembly sequences’ evaluation, a punishing time for this component is allocated.

iii). Stability of the sub-assembly: The stability score \( Sta \) of a sub-assembly is defined by:

\[ Sta = 1 - \frac{f}{m} \]  

In this case, the local vector (here in C) becomes the referent vector. The value of \( P \) ranges from 0 to 1. The ideal path is when \( t=0 \), \( \alpha_i = 0 \).
For $i$, $P=0$, which is the worst situation.

4. Results and discussions

4.1 Disassembly sequences generation

In order to compare the proposed method with other works an example for disassembly sequences generation of an electrical motor with sixteen components, similar to the work of Popescu [PI13] is presented here below (Fig. 8).

The first step of the proposed method for disassembly sequences generation consists in building the Disassembly Geometry Contacting Graph (DGCG) [MWL_15]. The assembly relationships amongst the components (parts) are automatically generated from the 3D CAD models (SolidWorks) of the products (assemblies). According to these relations a computer application was realized allowing building the associated DGCGs. Thus, the five-levels DGCG for Cover 5 of the electrical motor is built as shown in Fig.9.

![Figure 9. DGCG of the electrical motor (target Cover 5).](image)

The second step consists in disassembly sequence generation performed by VRDE (based on Python programming). The input 3D assembly models are based on VTK (Visualization Toolkit) library and acquired through VRML files coming from CAD software (SolidWorks). The contact identification is based on ODE Geom (Open Dynamics Engine) libraries. The twenty four results for the feasible disassembly sequences for Cover 5 of the electrical motor are shown in Fig.10.

![Figure 10. Results for possible disassembly sequences, Cover 5](image)

4.2 Disassembly sequences evaluation

In order to compare the trajectories of the different components in the assembly of Fig. 5, during the disassembly sequence evaluation, the path lines (trajectories) for each component in $O,x,y$ plane are recorded (Fig. 11). There are four possible disassembly sequences for this assembly, namely: $\{1,2,3,4,5\}$, $\{1,2,3,5,4\}$, $\{1,2,5,3,4\}$ and $\{1,2,5,4,3\}$. Note that parts 1 and 2 have the same order in all these sequences. Their trajectories are the same and consequently it is useless to compare them. The trajectories of parts 3, 4 and 5 are shown in Fig.11 for sequences $\{1,2,5,4,3\}$ and $\{1,2,3,4,5\}$ respectively. Concerning criterion paths orientation change, the best one, for part 4, belongs to sequences $\{1,2,5,4,3\}$ (Fig.11a) as its path is a nearly straight horizontal line. However, its worst path change, belongs to sequences $\{1,2,3,4,5\}$ (Fig.11b), because it requires some steering to reach the destination surface (see Fig. 5.a).
After performing the four disassembly sequences, the scores for the five proposed criteria are calculated (Table 1). The last column presents the sum of the five criteria (sum scores). The higher the value, the better the sequence. Thus, the best one is sequence \{1,2,3,5,4\} with \textit{SUM}=3.12236.

<table>
<thead>
<tr>
<th>Sequenc.</th>
<th>Visib. score</th>
<th>Disassem. angle</th>
<th>Num. tools’ chang.</th>
<th>Path orient. change</th>
<th>Stability</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2,3,4</td>
<td>0.6985</td>
<td>0.7333</td>
<td>0</td>
<td>0.5276</td>
<td>0.6</td>
<td>2.5595</td>
</tr>
<tr>
<td>1,2,5,4</td>
<td>0.6868</td>
<td>0.6849</td>
<td>0.0625</td>
<td>0.5267</td>
<td>0.6</td>
<td>2.5610</td>
</tr>
<tr>
<td>1,2,3,5</td>
<td>0.6632</td>
<td>0.7146</td>
<td>0.1667</td>
<td>0.7777</td>
<td>0.8</td>
<td>3.1223</td>
</tr>
<tr>
<td>1,2,3,4</td>
<td>0.6023</td>
<td>0.6333</td>
<td>0.1667</td>
<td>0.5212</td>
<td>1</td>
<td>2.9235</td>
</tr>
</tbody>
</table>

Let us note that the values for visibility score and path changing depend on the way that the operator is handling the components in the VRDE. However, the values of disassembly angles, the number of tools’ changes and the stability are not related to the operator’s abilities and consequently only depend on the mechanical assembly and the disassembly sequence itself. Two subjects were involved in the experiment for disassembly simulations. In order to improve the reliability of the proposed method, the average duration of the disassembly time for these two subjects were recorded as well. The results for average disassembly time for each sequence are shown in Table 2. The shortest time is 39.0890002 sec for sequence \{1,2,3,5,4\} performed by Subject 2, which is consistent with the previous evaluation, thus showing that this sequence is the best evaluated one according to the five criteria here proposed.

<table>
<thead>
<tr>
<th>Sequences</th>
<th>Subject 1 time (sec.)</th>
<th>Subject 2 time (sec.)</th>
<th>Average time (sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2,3,4</td>
<td>48.45850001</td>
<td>46.2220001</td>
<td>47.3535001</td>
</tr>
<tr>
<td>1,2,5,4</td>
<td>46.2139999</td>
<td>44.1860001</td>
<td>45.2000000</td>
</tr>
<tr>
<td>1,2,3,5</td>
<td>40.40400000</td>
<td>39.0890002</td>
<td>39.7465001</td>
</tr>
<tr>
<td>1,2,3,4</td>
<td>44.39800000</td>
<td>44.601111</td>
<td>44.4995557</td>
</tr>
</tbody>
</table>

5. Conclusion and future work

Some limitations of the available techniques for disassembly operation simulations stimulated this research on disassembly sequences generation and evaluation. Based on the proposed method for disassembly sequences generation and evaluation, an application integrated in a Virtual reality disassembly environment (VRDE) based on Python programming language associated with VTK and ODE libraries is developed. A set of five criteria is proposed, namely: visibility of the sub-assembly part, disassembly angles, stability of the subassembly, number of tools’ changes and path direction change. It allows the evaluation of the disassembly operation complexity during the initial stage of product design or during the Product Life Cycle (PLC) in production processes, product maintenance and at the end of PLC. The case studies demonstrated the efficiency of the proposed method. The SUM score result of the five criteria allows the selection of the best disassembly sequence. It was confirmed by experimental tests, thus allowing validating the proposed method.

However, at this stage, the work is not considering the ranking of the proposed criteria. Future work will focus on ranking the criteria according to their importance. For this purpose different weights will be allocated to each of them, thus allowing a more comprehensive evaluating method.

Acknowledgments The research leading to these results has received funding from the European Community's Research Infrastructure Action - grant agreement VISIONAIR 262044 - under FP7/2007.-2013 (http://www.infra-visionair.eu) and has been partly supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025) (http://www.persyval-lab.org/index.html)

References


AR in support to helicopter maintenance activities, experimentation results

ROBERTO SANGUINI¹, MICHELE SESANA²

¹ AgustaWestland Training and Helicopter Support System; Sesto Calende, Varese, Italy; roberto.sanguini@agustawestland.com
² TXT e-solutions; Via Frigia 27, Milano, Italy; michele.sesana@txtgroup.com

Abstract
The aeronautical industry is a very promising domain for Augmented Reality (AR) applications because complex assembly processes and maintenance tasks are experienced and, at the same time, the high standardisation simplifies the standardised access to the material. This paper discusses AgustaWestland, worldwide leader in rotary wing productions, experience in using AR to support helicopter maintenance activities. In particular, after the analysis of the previous research activities, results from the TELL ME project are discussed. The ultimate goal of the experimentation is to verify the adoptability of current solutions, devices and procedure in such a regulated environment as well as the business sustainability of the solution. Results come from an extensive experimentation that involved several people in a simulated but realistic environment.

K.3.1 [Computer and Education]: Human Factors – Aeronautical Maintenance J.2 [Computer Applications]: Support at the workplace

1. Introduction
Customers in the helicopter industry have historically pushed OEMs towards technological excellence while they have been relatively less demanding in terms of service performance over the last decades. However, this trend has rapidly changed and service performance has become a more and more critical buying factor for fleet operators with a clear target to minimize the time aircraft spend on the ground and to maximize lifecycle productivity of their helicopters.

For these reasons AgustaWestland is committed to test new solutions in order to provide additional value for our customers, drastically reducing maintenance time per flight hour and simplifying the maintenance tasks as much as possible.

In particular, the aeronautical industry is a very promising domain for Augmented Reality (AR) applications because complex assembly processes and maintenance tasks are experienced and, at the same time, the high standardisation simplifies the standardised access to the material.

The need of AR is pervasive but especially related to maintenance activities. During maintenance is important to provide to technicians all the support useful to carry out the work in the best way; especially important when the procedures are performed infrequently or are particularly complex.

Another important aspect to be considered is that the technicians involved in an aircraft maintenance task are charged with a great responsibility and their work can be considered as very stressful because a maintenance error can have catastrophic consequences. Human errors are one a major cause that can compromise a maintenance activity, from which attention to all systems that can reduce errors due to human factors.

The present paper reports the experience of AW in using of the AR application developed as a part of R&D project TELL ME. The results of the trial are going to evaluate in terms of effectiveness, cost and any constraints by the aviation regulations.

2. Previous AgustaWestland experimentation
In 2009 AW decided to invest in developing a prototype of AR system with the aim of:

- experience the state of the state of the art of the technology available at the time (hardware and software) and verify their reliability in a real workplace;
- define what kind of information to be superimposed to real-world (2D/3D graphics, annotations, etc.) to achieve the most effective result for the technician;
- define a man/machine interface;
- evaluate the effort necessary to implement AR inside AW Technical Publication;
- collect feedback from internal/external stakeholders by organizing demo live on the field.

The initial AW requirements for the prototype were affected by applications already existing in aeronautical and especially in automotive sectors in which the graphic aspects was emphasized; so AW decided to develop a solution based on the alignment of 3D virtual word and the associated real world. The deci-
sion was facilitated by the fact that the 3D models of components and tools were already available.

The developed solution included a marker-based tracking system using a fixed video camera to detect marker (predetermined black and white pattern) positioned in a known location in the environment (marker-less technology was not precise enough for our scope). On the basis of an open source software, a framework developed in C++ was developed for information processing video, the analysis of the frame and then the identification and the tracking of the marker [HF07].

Software engineering developed the solution in close collaboration with AW Subject Matter Expert technical department for contents and instructor for methodology. Due to the fact that wearable devices were not available, real world and superposed virtual information were shown in a laptop PC put in front of the technician.

![Figure 1: First prototype of AR application in AW.](image)

The prototype developed was strongly application oriented to overcome the scepticism and show the potential of AR thought a practical use case: the removing of left main wheel of AW139 helicopter.

![Figure 2: Marker-based tracking with AR Tag.](image)

After some trials on the field major strengths and weaknesses highlighted by the users were the followings:

- the system is very useful to avoid human errors especially for activities performed infrequently;
- information provided by AW are clear and complete also by suggestion about use of tool;
- use of markers is not feasible on an aircraft;
- superimposition of 3D, is very impressive but sometimes it hides the working area;
- at the moment the system is not applicable for activity outside the hangar;
- effort to develop AR for a single maintenance task was too high and the it implies the involvement of software developers to prepare and tune the application.

In the final analysis this first feasibility study confirmed the effectiveness of the use of AR to support aeronautical maintenance technicians but, at the same time, demonstrated that the technology was not yet so mature to be accepted by workers. Has to be highlighted that users are aeronautical technicians that are deeply specialized and they need very accurate information and so they have very high expectations. Furthermore the process to develop AR was not really sustainable.

Nevertheless AW decided to continue monitoring technology and applications waiting for a solution closer to expectation.

3. The TELL ME experience

Since 2012 AgustaWestland is involved as an industrial partner in the TELL ME (Technology Enhanced Learning Living Lab for Manufacturing Environments) R&D project which is co-funded by EC under the Seven Framework Programme lead by TXT e-solutions.

Inside the project AgustaWestland is responsible to test the system in aeronautical industry by providing the maintenance technicians with the hardware/software aids provided by TELL ME:

- capable to refresh the knowledge of the procedure before executing (e.g. display video and/or interacting virtual reality with the procedure to be performed).
- ready to go where the maintenance has to be performed (e.g. wearable hand-free visual/audio I/O systems such as glasses with Augmented Reality).
- able to guide and control step by step the technician while performing its activities (e.g. with a voice guide and/or Augmented Reality signalling things to be done and possible errors).
- providing a mechanism to collect and deliver feedbacks related to lesson learned.

The goal is that technicians will continuously be trained directly at their usual workplace, enriched with new devices supporting the mobility of the worker traditional, mobile and augmented reality devices through which they are offered learning content and learning experiences that better meet their needs and preferences.

By the TELL ME project AgustaWestland had the opportunity to test again the application of AR taking advantages from the new emerging technologies now available. The aim is to develop a sustainable solution to be integrated in the existing tools such as the IETP (Interactive Electronic Technical Publication) repositories [PdML02].
Taking into account the previous experience, AW was able to define more precise requirements, in particular:

- application of a marker-less tracking system really efficient;
- solution deployable on different mobile devices such as Smartphone, Tablet, Google Glasses;
- technology available for maintenance in hangar (line maintenance) and on the field (ramp maintenance);
- use of standard superimposed symbology.

As testers have been involved three certified Airframe Maintenance Technicians which are also instructors in Practical Element courses, and ground instructors for maintenance theoretical course at the AWTA.

The two users’ roles allowed to evaluate the application in terms of both operation and training effectiveness point of views.

The AR applications have been integrated into maintenance workcard loaded from IETP repository in the standard format AECMA S1000D. The workcard (see next picture) is a step-by-step procedure that provides all the instructions to carry out the task.

For each step a AR call button is available to the technician; clicking on it the object recognition is started and, when the item detected, the AR information are showed to the user.

Instruction can be very disparate, from arrows showing which component to interact with, to text providing additional information like the right sequence for detaching object, to the over imposition of a 3D element (see next picture).

At the end of the trial, technicians have been extensively interviewed to gather their feedbacks.

From a general perspective the impression of the AR solution is very positive; the solution can be really useful to support technicians providing information directly at work place when necessary. The AR support on top of certified IETP procedures allows the technician to work as usual and, when needed, gets the help from the system for example refreshing location of components, panel, etc. The optional usage of the AR allows to not loose time when
operations are well known and where the fluency on the job is high, but concentrating on complex operations saving the time to look at printed material.

When looking to the information conveyed the reaction speed of the system has been rated as good; augmented information are immediately available on the AR layer; the efficiency into the understandability of the meaning of the data has been achieved by including simple elements, like arrows, in the overlay. The visualisation of complex 3D objects or dynamic elements (e.g.: moving wrench) didn’t provide any value added to the worker and, in some cases proved to be counterproductive hiding important elements in the real world.

Compared to other applications of AR for maintenance, the one developed in TELL ME is characterized by an essential augmentation (standard icons, simple 3D graphics, no animations) provided only if requested by the technician. For this reason the solution of a continuous tracking to keep aligned overlaid graphics with the real image has not been adopted. The application developed provides for taking a photo from the best user point of view using the mobile device; then the system augments the image with graphic information contextualized to the step of maintenance activity in progress. In this way the calculation power required is smaller for the benefit of autonomy (for example when using Google Glass).

The discussion about mobile/wearable devices used provided several results, even unexpected. The best rated solution are Google Glass (that have been abandoned by Google right after the experimentation); the main reason is that they allow to carry a monitor mounted on the head leaving the hands completely free as well as the sight that is 100% as without glasses. Interaction with voice commands and touch device was sufficient, while the audio is very good, even in case of background noise, thanks to the bon conduction headphones. Anyway the low maturity of the technology (battery and stick heat mostly) prevents their usage in the real environment but the feeling is that something similar will be the right device for the future.

Moverio Smart Glasses has been used to test see through displays. Their appealing for worker was not satisfactory due to the opacity of the screen. The opacity prevents the worker to be sure to have correctly operated and can cause accidents. Also the interaction modality by the hand device to be sure to have correctly operated and can cause due to the opacity of the screen. The opacity prevents the displays. Their appealing for worker was not satisfactory for their usage in the real environment but the feeling is that something similar will be the right device for the future.

Moverio Smart Glasses has been used to test see through displays. Their appealing for worker was not satisfactory due to the opacity of the screen. The opacity prevents the worker to be sure to have correctly operated and can cause accidents. Also the interaction modality by the hand device to be sure to have correctly operated and can cause due to the opacity of the screen. The opacity prevents the displays. Their appealing for worker was not satisfactory for their usage in the real environment but the feeling is that something similar will be the right device for the future.

4. Conclusions and next steps

The on the field trial confirmed the strengths already highlighted about the adoption of AR solution in a highly specialised environment as the aeronautic maintenance.

The technological advancements in terms of devices, software libraries and human-computer interaction techniques are making the full adoption closer.

AR application can really represents an added value for the IETP (Interactive Electronic Technical Publication) that are now available electronically, providing technicians with an additional technical support at the workplace. Of course it is mandatory to cope and maintains the conformance with aeronautic standards, regulation and legislation.

Another important aspect that must be considered is the cost. An IETP (one for each helicopter model) covers several hundreds of maintenance tasks each of which might require an AR application. From this point of view a huge step forward is the availability in the near future of sustain- able authoring systems for the development of AR easy to use by non-experts in IT, but experts in maintenance.

As a potential evolution of the system, it is envisioned the combination of a remote assistance system by which a remote expert can support the technician by usage of AR to solve unforeseen faults or behaviours of the systems [A97].

References


[dcCFpdISAA11] F. DE CRESCENZIO, M. FANTINI, F. PERSIANI, L. DI STEFANO, P. AZZARI, S. SALTI:


[LJ02] DIETER LANGER, ANDREAS JAROSCH: Augmented Reality Maintenance Assistant for the NATO Helicopter NH90. Paper for Ismar 2002. 1


AR/VR-enabled Architecture for Technology Enhanced Learning in Manufacturing Environments

STEFANO BIANCHI1, STEFANO GENOLINI2, MAURIZIO MEGLIOLA3, KAJ HELIN4, MICHELE SESANA2

1Softeco Sismat, Italy, stefano.bianchi@softeco.it; 2TXT e-solutions, Italy, {stefano.genolini, michele.sesana}@txtgroup.com; 3PIKSEL, United Kingdom, maurizio.megliola@piksel.com; 4VTT, Finland, kaj.helin@vtt.fi

Abstract
This paper describes the AR/VR enabled Architecture developed to support life-long vocational training as well as answering ad-hoc specific request of support of Blue Collar Workers in small/medium manufacturing environment. The ultimate goal is to increase work performances in the workplace combining training methodologies and latest technologies to the daily work of technicians. This paper first of all will discuss the technical architecture developed, and subsequently the specific application and impact of AR/VR technologies in the training and support of technicians. Applications of the AR/VR architecture are spanning all over the different training phases of Blue Collar workers, but especially i) during the briefing in which VR technology is used to illustrate specific manual processes and operations and ii) on-the-job training where AR technology aims to improve learning-by-doing experiences by enriching the scenes with contextual information generated according to the workplace and the user profile.


1. Introduction

Small and Medium Enterprises (SME) are key assets to Europe’s economic future, in particular for the manufacturing sector. The fast reaction to new challenges and the fast adaptation to new standards, processes and the arising of technologies is essential element for manufacturing SMEs to survive and compete in the new world. Those high levels goals should be addressed globally at companies by very different perspectives including business, organisational, and financial.

Specifically related to the training there is a growing need for managed training approaches and appropriate tools for learning and skills/knowledge development. Faster mastering of workers skills, faster reskilling, refreshing of competencies and support during working activities are just some of them. Solutions should be at the same time efficient and effective as well as tailored on SMEs need of having low cost and reliable solutions with low technological barriers.

The needs introduced before has to be crossed with the ones arising from the workers. In fact, looking from their perspective, BCWs while working in complex environment or on complex machineries, have several needs to be supported in order to accomplish their work faster, with less stress, and finally with better performances, in particular they need fast access to information, better instructions, and support when in troubles.

This paper illustrates an AR/VR-enabled architecture (and included components) [BSM_15] developed to implement an operational methodological framework tailored to improve the learning of Blue Collar Workers (BCW) in manufacturing environments. The environment concept is heavily stressed because the architecture is thought to target technicians directly at the workplace following the working by doing paradigm.
Technological solutions are differently exploited to improve the way training is carried out during two complementary interlaced phases: briefing time – when the know how is recapped before the execution of manual procedures, eventually with the support of Virtual Reality (VR) applications, and on-the-job, when learning-by-doing physical activities are better supported by Augmented Reality (AR) applications.

The architecture can be customised easily and scales up nicely according to the needs of the applicable scenario, including and customizing components and tools in compliance with the technical and operational requirements defined by the working environment context.

The methodological steps (enquire, mix, experience, match, and optimise) are supported by three distinct architectural layers: a presentation layer (holding multimodal user interfaces, including AR/VR applications), a middleware layer (including the core business logic components that implements the methodological framework), and a repository layer (including distributed repositories and legacy systems).

Fig. 1. AR at the workplace

2. Architecture

The AR/VR-enabled service-oriented architecture was conceived to improve vocational training in small manufacturing environments by introducing advanced learning technologies and innovative training methods to SMEs, with particular emphasis on AR and VR applications in different training phases.

A new pedagogical framework was designed by the TELL ME project to be supported by state-of-the-art technologies and tested during two piloting cycles in three different application scenarios. This framework [WSD_2013] [WDS_2013] [SNM_2014] encompasses several steps (enquire, mix, experience, match, and optimise) and is tailored to the needs of Blue Collar Workers’ (BCW) in manufacturing environments, in order to rapidly increase their work performance, with training taking place directly at the workplace and being supported by interactive AR/VR experiences.

Fig. 2. Architecture (logical view)

2.1 Presentation layer

The presentation layer includes tools for the Training Manager, the Process Manager, and the BCW, to support configuration, content preparation (upload, tagging, editing etc.), job/activity scheduling and monitoring, and learning content consumption while the briefing and the on-the-job training phases.

The Training Manager can include available learning contents in the managed assets by uploading (or simply referencing) contents and tagging them with proper domain taxonomy tags for later indexing and retrieval.

The Process Manager can assign jobs to BCWs: the system acts as a Decision Support System (DSS) identifying the training needed by the selected BCWs to accomplish the assigned task, according to previous experiences (in same or similar tasks) and to personalisation criteria (e.g. the expertise in the specific sector).

The BCW uses the learning consumption User Interface (UI), where notifications on assigned jobs and associated learning mixes are available. The UI includes a set of players to run the learning contents (which include e.g.
video tutorials, interactive Computer Based Training contents, Precision Teaching lessons, extracts of manuals, electronic jobcards, AR/VR engines etc.) plus tools to monitor the consumption and test results where applicable (e.g. in Precision Teaching modules). The adaptive UI and the included engines allow blending learning contents and manual activities in order to correctly accomplish the assigned operations.

2.2 Middleware layer
The middleware layer provides 1) communication and integration functionalities (e.g. Enterprise Service Bus, for the registration and orchestration of services, Messaging Server, for the communication between core components and the IoT infrastructure) and 2) business core functionalities (services that implement the methodological framework to provide personalised and contextualised learning mixes to the BCW).

2.3 Repository layer
The repository layer includes repositories for data and meta-data (e.g. Learning Contents and Mixes and the associated meta-data, Domain Taxonomies, User Profiles) as well as the Repository Data Access functionality, providing a unified API for accessing the stored data. All training material is homogeneously described by using tags extracted from domain taxonomies modelling the workplace in terms of relationships associating categories such as Subject, Object, Mediator, Activities, plus technical information such as Format and contextual information such Rule (i.e. motivation for a specific learning need) and Target (i.e. objective of training).

3. General Use Cases
The architecture allows training and process manager to define, optimize and monitor the on-the-job training activities of BCW: 1) at configuration time, rules and models are defined to fine tune the smart enquiry/match/optimize functionalities for the application domain, and learning contents are described and classified to ease retrieval and creation of learning mixes; 2) at monitoring time, learning mixes are assigned to trainees (blue collar workers) in order to fill specific knowledge gaps, so that results (e.g. performances) can be tracked and evaluated afterwards; 3) at consumption time, blended learning activities can be experienced on-the-job and off-the-job by several content delivery engines and technologies.

A generic yet significant use case can be the following: after a recap of general safety instructions on a tablet and a VR-enabled interactive tutorial illustrating the workplace and the features of tools and devices to be used therein, the trainee explores the workplace with the AR app to familiarize with the new manufacturing process and the associated constraints, then carries out a guided step-by-step jobcard in hands-free modality and finally checks what has been learnt by means of dedicated training exercises and interactive questionnaire-based assessment.

All smart functionalities – those which associates the correct learning contents to the specified learning need by proper matching and optimization – operates on metadata, which are specified at configuration time when feeding the TELL ME system with learning contents for the given application domain: all contents are tagged with domain-specific taxonomies (subject, object, mediator, activity, target) defined according to the applicable workplace model. Such an approach allows to decouple the value added services provided by TELL ME and the content repositories: by means of available API, plugins, or simply by access over the internet (e.g. external websites/resources, YouTube videos etc.) any relevant content can be managed and seamlessly proposed to trainees.

4. Learning contents
Learning contents can be consumed in several – even blended – ways, according to the objectives, the contexts and the trainees they have been designed for. In case a classic briefing session is necessary to coordinate process manager and blue collar workers, learning contents can be consumed on a pc or tablet via the responsive web UI, as well as on modern smartphones, thanks to a complete set of dedicated embedded (for videos, recordings, pictures, documents, etc.) and external players (Precision Teaching, SCORM). For complex AR enhanced manual procedures, VR applications can project the user into the recreated scene to illustrate better the operations that will be later experienced.

Alternatively, if specific training sessions with detailed traceable instructions must be completed by the blue collar workers on the job, jobcards can be downloaded from the web UI on a smartphone and consumed with multimodal interactions – text-to-speech and speech-recognition – via the dedicated app;

Finally, when immersive and interactive training-on-the-job sessions are necessary (e.g. to make blue collar workers familiar with new/updated manual processes, or simply to introduce new employees in the workplace), learning-by-doing can be supported by a cutting-edge Augmented Reality (AR) app where the workplace is automatically enriched with information (e.g. safety signage overlays on materials, devices, machines, etc.) and contents according to the trainee’s profile and contextual needs.

There are conceptual differences in how learning is best supported in different application contexts: in particular, continuous learning while doing vs. interrupting work for learning (real on-the-job learning vs. teaching factory).

Traditional web-based learning content delivery (long text paragraphs, multimedia files presented in players, SCORM players etc.) is still a valuable mean, but should be supported and complemented by more immersive and interactive approaches (AR/VR-enabled), in particular for those contexts where manual operations and interactions with tools and machinery are expected.
Traditional AR has three main characteristics [AZU_1997]: it combines real with virtual elements; it is interactive in real time; and it provides registering in 3D space. AR enables the contextualisation of learning content, changing its consumption from a mono-dimensional approach to a fully multidimensional interactive mode. Personalization axes, such as user’s expertise or profile or simply user’s preferences, allow for making the AR/VR experience even more adaptive and effective, delivering the useful information in the correct place at the required moment.

The following sections illustrates the different AR/VR-enabled approaches developed by the TELL ME system and provides some insights on the technical solutions adopted and the results obtained.

5. Selection and delivery of Virtual Reality contents (briefing)

During the briefing phase, the BCW gets prepared for the upcoming work: it can happen right before the execution of the assigned job, or even hours/days before, depending on the specific time constraints. Especially for complex maintenance procedures, such as maintenance on complex vehicles as helicopters are, quite sophisticated methods can apply. Virtual Reality (VR) is one of the most sophisticated techniques to provide an immersive training experience. It allows the worker to apply maintenance procedures on a complete synthetic environment that reproduces, exactly, not only the appearance but also the behaviour of the modelled equipment.

Complex environments are usually not available in the workplace to the BCW before the operation. It is easy to think to a work in a nuclear plant in which the timing for operation should be very short, or to a maintenance of an airplane to be done in a few minutes to be accomplished during the loading/unloading of passengers between two flights. In this case, the usage of VR allow the BCW to know about the target place/object even without having it available on premises, lowering the working time during operation and reducing possible errors.

Another important usage of VR is dedicated to the operations in which the worker cannot see the target space because the view fied is occupied by other stuff like panels etc.. In this case, VR allows to have a look at the hidden components before the operation while getting prepared for its proper execution.

Depending on the user contents, the system will apply different rules in order to select which is the best VR object to be used as learning content to get the worker practicing with the work to do. The method by which it is provided is different: it can be an exploratory VR content to navigate and see the item before to have it into reality, interact with it by means of virtual harnesses located in the scenario.

Fig. 4. VR to support maintenance activities training

It can be a quiz in which the component to act on is identified, extracted from the major objects and isolated. VR can help the worker to recognize which are the different parts of the object and how to act on them.

6. Selection and Composition of Augmented Reality contents (on-the-job)

Augmented Reality provides new possibilities to improve proficiency and safety on the workplace, while at the same time reducing training costs in particular in manufacturing industries. Among the applications developed to support workers on the shop floor, the possibility to perform training-on-the-job using an AR app and approach for showing contextualised information, helps and such novel learning contents represent a valuable and innovative approach. The TELL ME project developed several AR prototypes, described in the following sections: these include AR players launched on tablets as well as native applications for smart glasses, providing complete training support as well as on-demand-only specific helps.

The following two chapters reports what has been implemented within TELL ME about AR solutions; the former is based on a full flavoured tablet based AR environment guiding the user to the goals, the player is named ARgh!. The latter is based on a step by step jobcard execution support both by tablet and wearable devices (glasses) triggering AR only when needed.

7. Augmented Reality Player

The Augmented Reality Player focuses on an activity model (managed as a special learning content to be consumed on a tablet or smart glasses), which prompts contextual information (e.g. safety signs) according to environment (scene) recognised and objects detected. The tablet UI enables the presentation of both informative HTML5 content and simple/contextualised AR content.

Fig. 5. ARgh! internal architecture
The tablets are technically very suitable for AR. Since the user looks at the environment through the live video produced by the tablet camera, the overlays can be placed accurately without the need of any offset correction. Being the AR layer added using this hand-held window, the user can simply put the AR layer aside, whenever it is not needed.

Fig. 6. ARgh! prototype on a tablet

Usability problems with tablets arise in cases where it is necessary to use both hands for manual activities and get information through the AR layer as well. Since many on-the-job learning cases require hands-free operation, the player running on tablets cannot be the only available option for training delivery, although the tablet might offer the most reliable AR experience: TELL ME provides such an alternative in terms of e.g. hands-free smartphone apps and smart glasses apps.

8. Augmented Reality recall from jobcard

In this solution, the main access and navigation point is the jobcard, a self-consistent description of the different steps that must be mandatorily completed to accomplish a specific operation. Here AR support is recalled by the BCW only when strictly needed: the goal is to maintain a low usage of resources and let the user to decide when it is necessary to launch the AR application to get additional support.

At each jobcard step the worker may be interested to ask to specific questions: how to perform the step? Where is the item I need to work on? How to use the harness required in this step? Many kinds of learning contents: textual, annotated pictures and video pills can be used for this purpose. AR seems to be the proper tool to provide help in detecting the right element (localisation), suggest possible actions (rotate, pull, push…) and suggest the correct harness to be adopted. The right learning content shall be provided at each step of the jobcard, that means, in particular cases, AR is not meaningful and a simple picture may be enough to provide the right support. This simplify enough the construction of the jobcard player. A specific editor shall support the construction of the jobcard, combining learning contents at each step. The editor allows to collect the content interactively during the jobcard execution. Of course the complete procedure shall be validated before to become ready to use in the plant.

9. Augmented Reality wearable devices in manufacturing workplaces

The piloting activities carried out in the TELL ME project allowed to test AR solutions on different devices, collecting valuable feedbacks on the usability of the prototypes released and on the acceptance of edge technologies in manufacturing working environments. In particular, AR applications were tested on smart glasses (Google Glasses and EPSON Moverio) and on more classic mobile devices such as tablets and smartphone.

Notwithstanding the technological trends in the consumer market, the BCW generally considered the smart glasses as an interesting and promising device for AR applications and learning content delivery, but generally preferred tablets and smartphones to get the support provided by AR only when considered useful and necessary.

Mobile devices are impressive, for example Google Glass solution, but there are some doubts on its maturity (autonomy, fragility and cost). Although Google Glasses cannot be considered the right device for AR (it is not possible to impose images on the current view but only on images captured by the internal camera) they seems to be quite interesting because not really intrusive in the maintenance operations, can be driven by voice commands allowing a good hands free behavior.

Fig. 7. AR on tablet

Fig. 8. Google Glasses

Moverio Glasses supports better the field view with the AR contents over-imposed, even if they are heavy and require additional cable connected device.

Fig. 9. EPSON Moverio

As a conclusion, these consumer devices are not really ready to be used in the industrial environment. More so-
phisticated and robust solutions are available, but they are considered too intrusive for a light maintenance environments as considered within TELL ME. For this reason smartphone and tablet are still considered as the most appropriated choice, at least when combined with voice recognition.

10. Conclusions

At the time of writing, the project Consortium is preparing an EU-wide campaign of live demonstrations of the final TELL ME prototype. The campaign is expected to involve many companies (mainly SMEs) from the manufacturing sector and a large number of Blue Collar Workers and operation managers, with engaging hands-on/multimodal experiences for all participants.

Furthermore, during the final piloting, validation and evaluation phase, the prototype will be tested by end users in three different application scenarios: supplies and maintenance for the helicopter industry; fitting out luxury yachts (furniture making); and in the textile industry (quality inspection). The results will allow to start the industrialization phase for the tools and reduce the time-to-market for SME-tailored learning systems based on the proposed SOA.

As for the feedback collected during the first piloting phase, the application of AR/VR technologies in the manufacturing sector seems a valuable approach to support maintenance phase, overall if combined with electronic jobcard management.

ACKNOWLEDGEMENTS. The results illustrated have been achieved by the research project TELL ME ("Technology-Enhanced Learning Livinglab for Manufacturing Environments", FP7-ICT-2011-8, G.A. N. 318329, www.tellme-ip.eu), a EU-co-funded initiative held by industry, universities, research labs, technology firms and innovation companies from all over Europe. The authors thank especially all partners participating in this initiative (in alphabetical order): AgustaWestland Spa (IT), Alfamicro - Sistemas de computadores, Lda (PT), Asociacion de Investigacion y Desarrollo en la Industria del Mueble y Afines AIDIMA (ES), Association for Advancement of Radical Behavior Analysis (IT), Clear Communication Associates Ltd (UK), Deutsches Institute für Textil und Faserforschung Denkendorf (DE), Kungliga Tekniska Hoegskolan (SE), Laboratoire Virtuel Europeen dans le Domaine de l'Interoperabilite des Entreprises AISBL (BE), Profitex Textil Service Helmbrechts GmbH & co (DE).

References

Augmented Reality authoring by blue collar workers as key enabler for AR adoption by manufacturing companies

MICHELE SESANA¹ PAOLO CIVARDI¹, JACOPO MARULLO¹

¹TXT e-solutions, TXT; Milano, Italy; {michele.sesana, paolo.civardi, jacopo.marullo} @txtgroup.com

Abstract
Augmented Reality in manufacturing (e.g.: production, maintenance and overhaul) is a research topic addressed since several years and presented in several papers. This paper approaches the usage of AR from a different perspective than usual discussing the importance of the authoring of AR by Blue Collar Workers on the field in manufacturing companies. The paper aims to highlight the importance of the efficacy and efficiency of the authoring as one important barrier to the adoption of AR in manufacturing environment, as well as presenting the prototype implemented in the TELLME research project focusing on authoring of AR instrumented job/workcards by Blue Collar Workers on manufacturing environments. The software solution is based on a tablet with embedded a job/workcard editor that can be used by the technician, on the field to create the AR instructions in an efficient way, test them immediately and collaborate with the engineering departments that should approve them. The application is thought for any engineering work with a special focus on the aeronautic being compatible with the A1000D standard.

K.3.1 [Computer and Education]: Computer Uses in Education - Computer-assisted instruction (CAI); J.2 [Computer Applications]: Physical sciences and engineering - Engineering

1. Introduction
The topic of the adoption of Augmented Reality into manufacturing operation is widely addressed especially in production, maintenance and overhaul. Augmented Reality is a promising technology for such kind of operations especially when those that are targeting complex objects and/or in complex and difficult environments.

Despite the great advancements carried out in the last year both from devices and software technologies the silver bullet for the wide Augmented/Virtual reality adoption in manufacturing companies is yet to come [GAR]. The reason for this scenario for sure has to be investigated not only into the TRL of devices and players software, but as well as, into barriers preventing the creation of AR solutions into the manufacturing environments.

In fact AR is still felt as an IT engineering topic and most of the company are scared about costs of development of the solution and the need to keep the system up to date. Despite players and demos of AR adoption are improving monthly the topic of the AR generation is less addressed but more than critical for the AR adoption. It is easy to understand why a company having 1.000 procedures to instruct and maintains a conservative approach to the AR adoption and should be demonstrated the economic feasibility of the operation as important as the technical feasibility.

2. Barriers to adoption
Here below a set of barriers for the smooth and wide adoption of AR into manufacturing environments is reported and discussed. The focus is on barriers preventing a smooth adoption and economical sustainability at design time during the creation of the AR contents and procedures. A specific attention is devoted to complex manufacturing environments where most of the operations are human driven.

A first block to the adoption is linked to structural processes barrier that are focused on the development of the product and the development of services in two separated silos. The reason for that is essentially historical and originated by the fact that manufacturing companies started with
products and few supporting services; only on the last year the servitisation trend is highlighting the importance of the co-evolution of the services [MSEE] [PSY]. The approach prevents the flowing of product design material becoming supporting material for services like assembly, maintenance and overhaul. An sustainable AR application needs that processes for product and services are as interconnected as possible providing at design time useful design phase material for building AR services, as well as allowing to access to them during the execution of operations.

A second element is an organisational barrier; companies departments or organisations that create supporting procedures and/or material are different from the ones that conceived the product. The first are usually customer support and/or training departments/companies or even marketing in some cases, while the second category are generally engineering departments/companies. Usually these two entities are totally separated and are speaking very different languages. This scenario leads to the fact that who has the deep knowledge about the product is different from who creates the supporting material that knows more about how to conceive the message. This scenario originates a lack of communication and an inefficiency of the process: who created the basic material (instructions, images) has few access to engineering material and if access is permitted the organisation of engineering material follows the PLM approach that is not what is required by the Service Lifecycle Management. For having a better support for AR it is necessary that representatives of the different departments/companies work in strict cooperation and material is thought and achieved since the beginning thinking to the services that will be developed on top of it.

Another aspect of the organisational barrier is the lack of communication between the operators on the field and the engineers. Feedbacks coming from the field are sometimes reflected into new versions documentation but is very difficult that can impact the material created by the engineering department.

Another aspect to consider is the physical object availability barrier. Expensive complex products (eg: production machineries, aerospace objects, elements of nuclear plants, etc.) are manufactured by “make to order” strategy and are not available at the time of the creation of the documentation. 3D CAD and prototypes can help into creating the “usual” documentation but are not enough for the AR creation where to create assembly/maintain instructions the physical object should be available. Moreover the expensive complex products are not stored in the hands of “documentation” people. For having a take-up of AR it is needed that the complex object has to be made available to people having the knowledge of the item as to create the documentation.

A technological barrier is another aspect to take into consideration; the maturity of modellers and players is quite low and contents and procedures are created nowadays by ICT technicians with, usually, a computer science degree, that don’t have any glue about the product. This provides to double time and costs in order to create material and procedures due to the fact that two persons are needed increasing the possibility of misunderstanding. To simplify the creation of AR contents the technology should be as simple as possible with a short learning curve to be utilized by the people having the knowledge of the product and procedures, which usually don’t have high IT skills.

In several organisation a limit to AR adoption is the mental barrier in reluctance into changing the status quo of the documentation. The customer support department should, instead, ask for new objects to be used in documentation like 3D design, tips from the field operators, etc. The “invasion” of digital natives in the companies can be a breakthrough of the current scenario accelerating the process of changes creating a new digital thinking approach.

Last but not least is the precondition economic sustainability barrier. The time (and corresponding costs) to create the AR is quite high and not affordable for manufacturing companies having several procedures to implement and maintain over the time. The reason is originated by both the low maturity of the technology and the less research on modellers in respect to players.

3. Proposed approach to overcome barriers

The six before mentioned barriers can be put into two different clusters to be addressed and tried to be solved.

The first cluster has to deal with company processes; the need is to improve the information flow to correlate elements that are currently separated: product vs services (structural), engineering dept. vs customer support dept. (organisational) and old vs new approach (mental). In general the companies willing to take-up AR should change the mindset thinking since the beginning to create product design material in the way that can be used by people that will build the documentation and secondly to allow to these information and material to flow along the supply chain in order to be really utilized.

The second cluster of barriers focuses on the efficient availability of the enabling elements for the AR contents creation. In fact it is necessary to have the availability, in the same period of time and space, of both tools to create AR material, namely the object (real world layer), and the enabling technological solution to create the virtual world layer and correlate the two worlds. Moreover the expert of the tool (non-IT people) should be put in the condition to easily use the technological solution obtaining results in a short period of time in order to be economically sustainable.

4. The technical solution

TXTeye is a software designed to support Blue Collar Workers during their activities in the workplace in order to improve the work performances reducing time, error and increasing the safety of workers into the workplace. The application is composed by a player TXTeyePlayer and an authoring tool TXTeyeEditor.

4.1. The TXTeyeEditor solution

TXTeyeEditor is the software solution presented in this paper aiming at overcoming the previously mentioned barriers realising the approach described in the previous chapter. The business goal is to let the user easily and quickly create
multimedia AR enabled operational (assembly, maintenance, overhaul) procedure to support specialized technicians during complex manual activities on the field.

In order to instantiate the former approach mentioned in the previous chapter the solution aims at gathering information from both the engineering department (3D objects), and from the customer support (IETP), spreading the need to have more integration between product and services.

The second objective is to enable the blue collar workers working in the shopfloor to create the AR contents. Those peoples are the actors that really have the physical object available in their hands, the knowledge about procedures to be done and should be enabled by a software solution easy to use, fast in creating procedures to allow them to create AR contents to be streamed and consumed during activities by the other blue collar workers.

Main features of the solution is to add to reality, in a simple way, up-to-date information in the form of 3D standard symbol, text, images, videos allowing low cost content development and high performance and stability of the recognition process. The use of interactive 3D objects is discouraged and limited only for critical procedures, this due to the high cost of implementation. The created procedures are stored in a remote database and can be accessible by from the player.

AR enabled procedures are taken in charge by the TXTEyePlayer; the complementary application which is able to retrieve and play the procedures at runtime on smartphones, tablets or wearable devices like smart glasses.

4.2 Blue collar workers

In order to overcome the barriers the user of the software are Blue Collar Workers that, having knowledge of the object and the procedures, having the physical object at its own disposal as well as a the software for the authoring (TXTeyeEditor) will be able to create the AR enabled procedures. In particular different kinds of blue collar workers will take part to the process of AR enabled procedures. In particular, we identified three roles of BCWs:

1) Content creator user: this user is a subject matter expert and will create learning contents that are used in a second time to create the body of the procedures steps. A learning content is composed by a name, a description, a group (or maintenance domain), optional tags and a number of different multimedia learning contents (photos, videos, audios and Augmented Reality scenes);

2) Procedures creator user: the task of this user is to import or create procedures, using as procedure steps the contents created by the content creator user or other procedures with additional information like a name and a description. One procedure is formed by a name, a description, a group (or maintenance domain), optional tags and a number of steps which have to be properly linked together, with the possibility also to create different branches and flows;

3) Approver user: the task of this user is to approve the created contents and the procedures. Every content or procedure has a status that the approver user changes. As soon as a BCW realized a learning content or a procedure, the status is equal to “created”; the approver can change this status to “approved” after he has performed a quality check on the new element or reject if not satisfied. When the status of an element is equal to “approved” it can be used by the player. The approver can be an external person from the quality department and so not a BCW.

4.3 Major Functionalities and procedures creation workflow

Profile editing. As introduced before, when the user follows the registration mechanism he has to insert some personal information. These data are going to form the user profile. The user can access his profile and modify the displayed data selecting the profile button on the main menu.

Content creation. The content creation is performed selecting the contents button on the main menu. This will take the user in the contents page. In this page we have the list of all the contents that are stored in the remote database. The user can select an existing content and edit it, or create a new one selecting the “add content” icon on the right bar. On the content page the user has to insert all the data that identifies the content, like name, description, group, tags and one or more multimedia learning contents of different kind.

Multimedia contents creation. The multimedia contents can be photos, videos, audios and Augmented Reality scenes. For the first three elements the user can choose the data files from the device gallery and sounds (in case of tablet/smartphone application) or from a defined path in the file system (in case of desktop deployment) or capture the contents one-shot using device camera and microphone (obviously in case of they are connected). The photos that are chosen for the current content can be edited at runtime clicking the edit button. After this, one simple photo-editing program is launched. The user can insert on the photo some symbols of a chosen colour (like arrows, geometric forms and labels) that are hopefully used to describe all the different parts of the represented mechanic environment. The available symbols are defined in a given well-defined library.

Augmented Reality scenes creation. The multimedia content that is the most effective and advanced in terms of tech-
nological efforts is without any doubts the Augmented Reality scene. For doing this the TXTeyeEditor application will launch a small 3D editor program, that helps the user to set up the AR scene that has to be reconstructed by the player during the recognition and tracking/registration phases.

The user can create an AR scene doing two important things:

- Set the 2D or 3D target
- Insert the virtual objects

The first thing to do is to choose the target that will be detected by the device camera of the TXTeyePlayer application during the recognition phase. The user can use a photo chosen from the remote database or a 3D object.

After the target is set in the 3D scene, the user can now insert virtual objects in the scene. These 3D objects are available from a given well-defined standard library, shared with the models used by the TXTeyePlayer. The position, rotation, scale and colour of every object can be adjusted using mouse/ touch controls.

Finally, the AR scene can be saved giving it a name. All the information that identifies the scene (target and virtual objects relative position, rotation, scale and colour) are saved in the remote database. The player will use these data to recreate the scene during the recognition and tracking/registration phases.

![Figure 1: AR Scene Creation](image)

4.4 Procedure creation

The procedure creation is performed selecting the procedures button on the main menu. This will take the user in the procedures page. In this page there is the list of all the procedures stored in the remote database. The user can search and filter the procedures that are shown. By clicking on a single procedure the user can edit it, by clicking on the “add procedure” icon on the right bar the user can create a new one. On the procedure page the user has to insert all parameters that identifies the procedure, like a name, a description, a group and one or more tags.

Step creation. By clicking on the “steps” icon the user can watch the procedure steps, edit them, or add a new one. Parameters identify the step such as a name, a description, a photo, if the step is the first and if the step is a procedure or a content. The step photo can be selected from the existing content photos stored on the remote database, form the device gallery or file system data path or can be captured using the device, or desktop camera. After doing this the photo can be also edited, as described before. A step can contain a procedure or a content. In this way, the reuse of existing procedure is promoted decreasing time and costs.

Steps linking. After the procedure steps are created, they have to be linked in a next/previous order, in order to be played properly when the TXTeyePlayer runs the procedure. In this application the user can set the next, or the follower, step from the current step page. By clicking on the “links” button on the right bar the user can choose one or more followers of the current step from the pool that contains all the other steps of the procedure (obviously except the first).

Branching creation. If the user select two or more followers for the current step, a branching is created. The branch manager page is loaded in order to allow the user to set the branching conditions. The branching conditions define the conditions that the player has to follow to decide what the next step in the execution flow is. From the perspective of the player, the branching happens when a question is asked to the mechanical maintenance technician and he/she has to reply. The next step to load is chosen depending on the answer given by the technician. For example, if the current step explains the way to unscrew a particular screw from a complex mechanical part, at the end of the step a question could be asked: “Is the screw rusty?” Depending on the answer (“yes” or “no”) the procedure continue in a different branch.

4.2. Conclusions

The application presented in this paper is under experimentation into six companies from three different domains. The first feedbacks gathered from users (blue collar workers for the practical usage and managers about business sustainability) are positive. In particular the BCWs have appreciated the facility in procedure creation that was expected to be more technical and managers appreciated the easy procedure for the control of the process and the final quality check. The access to Original Equipment Manufacturer elements is for sure one of the most important and challenging topic; a future challenge for the application is to make as easier as possible the access and the navigation (including the selection of the material) of remote databases that can contains tons of procedures and materials. Next developments are focusing, starting form user feedbacks, in improving more and more the interfaces of the application making as easier as possible the user interaction.

References

[GAR] Gartner technology hype cycle 2014
[MSEE] Manufacturing SErvice Ecosystem research project, co-funded by the European Commission under the FP7 framework - www.msee-ip
Exploiting ARgh! on-job-learning tool in Mixed reality training environment

Helin Kaj¹, Karjalainen Jaakko¹, Kuula Timo¹

¹VTT Technical Research Centre of Finland Ltd; Espoo, Finland; e-mail:firstname.lastname@vtt.fi

Abstract
This contribution reports on design decisions and evaluation procedures as well as findings studying an on-the-job Augmented Reality (AR) learning system, namely the ‘ARgh!’ system, which has been exploited in Mixed Reality (MR) environment. The objective of this paper is to develop the on-the-job learning system and evaluate its usefulness as learning environment. This study contributed to the research programme of the EC-funded ICT project TELLME (‘Technology Enhanced Learning Livinglab for Manufacturing Environments’) which aims to develop and trial in authentic contexts (SME-driven human-centric and service-oriented manufacturing workplaces) an innovative cross-enterprise methodology and IT platforms for continuous education and training in heterogeneous business ecosystems, blending lifelong learning and participative co-creation aspects in ways that can address more business needs than traditional training[TELL_15].

Categories and Subject Descriptors (according to ACM CCS): H.5.1[Information Interfaces And Presentation]: Artificial, augmented, and virtual realities, I.2.6 [Learning]

1. Introduction
Augmented reality (AR) and Virtual Reality (VR) systems are regarded as promising training platforms for industrial maintenance tasks and for visualising complex operations [GGE_13], [WKH_14], [WBE_11]. The Internet of Things (IoT), is the network of physical objects/things embedded with software, sensors, and connectivity to enable objects to collect and exchange data. IoT allows objects to be sensed and controlled remotely across existing network which gives opportunities for more direct integration between the physical world and computer-based systems[WW_10].

This paper presents the on-job-learning system is developed and tested in this study. The test case is maintenance work on heavy machinery, i.e. a rock crusher. There were 7 test subjects, all working as researchers at VTT. In the test case setup, all manually manipulated objects such as the IoT control box have been made available as hardware, while the rock crusher, however, is visualised only virtually on the surrounding video walls, even the rock crusher was controlled via the IoT control box. The test subjects are not familiar with the work task, thus resembling situations where people face new, previously encountered challenges in their work. Existing skills and the use of new approaches such as opportunistic collaboration are needed to respond and to improve attitudes towards learning and change. Subjects are deploying the ARgh! AR-system using a tablet see-through video-overlay (see Figure 5).

2. Methods
2.1. ARgh! AR-system
ARgh! system is AR-system which main idea is to be easy to use system, what non-expert can exploit. ARgh! system components are illustrated on Figure 1. The whole system is configured based on the Activity ML and the Workplace ML file. Activity ML describes all the action steps and what content should be active in each of these steps. Activity ML is parsed by the HTML5 layer, which controls also the AR presentation in addition to HTML5 presentation. Workplace ML describes all the devices and other workplace related components [WSK_15].

AR presentation is controlled by the Presentation Engine component. This component stores a set of C# functions that have bindings with a matching set of Javascript functions on the HTML5 layer. This means that the HTML5
layer can call a C# function that e.g. places an action symbol for pointing on an AR enabled physical hammer object.

Figure 1. ARgh! system components

2.2. Mixed reality environment

Technical set up of Mixed Reality environment in VTT Tampere (see Figure 2) [HKV_15]. System includes (1) three-screen system with active stereo (3 x Barco RLM-W12 stereo projectors) [BAR_15]; (2) optical tracking system / mocap (Vicon T20) [VIC_15]; (3) several goggles or head mounted display; (4) game controls, (5) Motion Platform: MeVEA 3 Degree of Freedom (DOF) [MEV_15] and (6) real chair with real controls. Visualization software is Unity3D [UNI_15].

Figure 2. Mixed Reality environment in VTT Tampere

The calculation of the visualization and device management were distributed over three computers. Communication between subsystems was handled throughout by the Virtual Reality Peripheral Network (VRPN), except IoT models (see next section). The fundamental idea behind the established MR system was that it was relatively low cost, easily re-configurable, and the immersion level was reasonable enough for designing and learning purposes.

2.3. Internet of Things (IoT) control box

MR environment model was connected by IoT enabled operational mock-up of a control box assembly quite usually found on various types of machines (See Figure 3). The assembly consisted of removable modules with potentiometers for controlling the module output voltage. There was also a third module, which had been made faulty by adding a simple resistor to the output channel. The logic was built with an Arduino Uno [ARD_15] and it monitored the voltage coming to box. If the output voltage was below 3.5V, the module with the low voltage was considered to be in “low”-state. If the voltage was above 4.5V, the module was considered to be in “high”-state. A voltage between 3.5V and 4.5V made the module to reach the targeted “ok”-state. If the input voltage was above 4.5V and the output voltage was less than 3.75V, the module was considered to be in “faulty”-state. The overall state of the whole assembly was defined in three stages: green (both modules in “ok”-state), yellow (one of the modules in “ok”-state) and red (neither of the modules in “ok”-state).

Figure 3. IoT enabled control box assembly

3. Results

The main result is ARgh! based learning system which exploits MR system. The results presented in this chapter are divided in two sections including ARgh! on-job-learning tool in Mixed reality training environment and learning scenario.

3.1. ARgh! on-job-learning tool in Mixed reality training environment

Test subject was acting in middle of Mixed Reality lab in VTT Tampere (see Figure 4). Virtual rock crusher was visualized in three-screen system without stereo and its sound was generated with 3D system. Main physical working area was located to front of rock crusher. Lab layout supports observation and there were two persons observing the task from human factor and AR-system point of view.
Manually manipulated objects were realized as IoT control box and devices. An iPad tablet was deployed for running the ARgh! system, with the application connected to a server for fetching the data about work task sequences. All the augmented IoT information and symbols were visualized on the tablet (see Figure 5).

The VR model was based on a model of a rock-excavating site where a model of a Metso Minerals [MET_15] rock-crushing machine was placed. The rock-crushing machine had 4 simulated condition stages: machine off, green (machine operating without any faults, Figure 6), yellow (incorrect voltage only on one module, Figure 7) and red (incorrect voltage on both modules, Figure 8). In the green condition, crusher sounded like a normal crusher and emitted crushed rocks in a stable stream from the conveyer. In the yellow condition, some sounds imitating mechanical sounds were added together with additional emitter that shot rocks from the crusher jaws, which should not happen on normal operation. Also the crushed rock stream was made more random. The red condition added excessive smoke emitter, even more mechanical fault sounds and the crushed rock emitter shot rocks all over.

Users received instructions on a iPad tablet device running ARgh!. ARgh! was able to react to the changes in the IoT control box and also display the real-time voltage values of the modules, Figure 9.
Controller box was tracked in ARgh! with image based tracking. IoT logic was used as a trigger for advancing to next steps e.g. by detecting if the user turned off the machine as instructed. This meant that the ARgh! application was able to move forward automatically in most of the steps. Automatic state changes were indicated by sound feedback. The removal of the module from the control box was not automatically detected by the IoT logic so the user had to confirm the action by clicking a text box in ARgh!. All the elements that required this type of interaction were clearly indicated with a big icon in the text box, Figure 10.

Figure 10. ARgh! indicating a requirement for clicking

3.2. Learning scenario

Two types of introduction

Before the actual test task, the test participants were given an introduction to the basic functions of the system. Two types of introduction were created for this purpose: video and "free ride". In the video (01:38 min.) the basic ideas of utilising the tablet view with AR-symbols and the physical controllers were shown without revealing the idea of the test task. Guidance for instructional video was explained as follows: “First we would like you to watch a video, where the basic functions of the system used in the test are explained. The video lasts about 2 min and you can watch it freely for about 5 minutes.”

The alternate introduction, free ride, allowed the test participant to try hands-on the same basic functions of the system as shown in the video. Guidance for free ride was explained as follows: “This is the system used in the test. It includes the big screens and the helmet, tablet, and the equipment on the desk. The view of the virtual environment changes when you move in the environment. There are controllers in the cabinet and on the desk, which you can also view through the tablet. You can now freely try out the different functions of the rock rusher for about 5 minutes.”

Test task

The job description for the actual test task was as follows: “You are a maintenance engineer and have this system in use. Now you need to check the status of the engine and do the necessary procedures. The system tells you how to proceed.”

The test task consisted of the following phases / orders:

- Start the machine (rock crusher)
- Adjust the modules: set voltages to range of 3.5-4.5 V
- Fault detected in module [1 or 2]
- Turn off the machine
- Disconnect cable from module [1 or 2]
- Remove module [1 or 2] from the chassis
- Pick up new module and place it the chassis
- Connect cable to module [1 or 2]
- Start the machine
- Adjust the modules to correct voltage of 3.5-4.5 V

In the test, it was randomly decided by the researchers, which one of the modules 1 or 2 (IoT boxes) was malfunctioning.

4. Discussion

Augmented reality based ARgh! on-job-learning tool and IoT components are effective tool for showing normally non-visible information to learner. Especially real time feedback from system state allows user to understand system status and find solution with ARgh! system support. Most of the test subjects were able the finish the task even no additional support wasn’t given by researcher. ARgh! system was effective enough to train subject to finalize the task.

Trainees were feeling less uncertain because they were acting with virtual machine, which is safe for trainees and they couldn’t break the real machine. Mixed reality environment gives real feeling of the situation because trainees were acting.

This develop set-up with IoT allows to transfer the training situation to real environment. Real machine / control box could use also IoT to transfer information to ARgh! system which allows same situation as MR environment.

Generally, learning to operate the system was seen as easy. The test participants were especially satisfied with the instructions in terms of accomplishing the test task. The use of terms was seen as very consistent as well. Furthermore, the participants were happy with the organization of information and position of messages on tablet screen. Thus, the visual implementation of on-screen information was successful.
Acknowledgements

These studies have been funded by the European Commission’s Seventh Framework Programme FP7 under grant agreement 318329 ‘TellMe-project’.

References


A New Approach for Haptic-relevant Simulation of Soft Bodies

Julien Piera¹, Jerome Perret², Arnaud Geneslay³

¹EISTI, Master 2 Research SIC, University of Cergy-Pontoise - ENSEA - CNRS, Cergy-Pontoise, France, julien.piera@haption.com
²Haption GmbH, Aachen, Germany, jerome.perret@haption.com
³Haption S.A., Soulgé-sur-Ouette, France, arnaud.geneslay@haption.com

Abstract
The interactive simulation of soft bodies at haptic rates is necessary for many applications, ranging from assembly simulation of deformable objects to Virtual-Reality based training of surgical operations. Departing from the classical approaches, we propose a new method which focuses on the haptic perception of the user. While the visual feedback is based on a traditional computation of the body deformations, the haptic rendering uses a non-physical model, based on a force profile and on the penetration volume or depth between the non-deformed geometries. In the article, we present the two models and how they work together, and we demonstrate an initial implementation. Finally, we discuss the limitations of our approach, and how they can be overcome in some cases.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [Information Interfaces and Modeling (e.g., HCI)]: Haptic I/O; H.5.1 [Information Interfaces and Modeling (e.g., HCI)]: Artificial, augmented, and virtual realities.

1. Introduction
The interactive simulation of soft bodies at haptic rates (close to 1 kHz) is still a challenge after more than twenty years of research. Many applications depend on such a simulation, foremost Virtual Reality-based surgical training, but also such activities as assembly and maintenance validation in the industry, where cables, hoses or rubber parts are present. From a computational point of view, soft bodies raise two major difficulties as opposed to rigid bodies. Firstly, because of the deformation of their geometry, many optimization methods for collision detection cannot be applied, so that the time needed for identifying contact points is much larger. Secondly, the calculation of the deformations in real-time is a challenge itself: the classical finite elements method (FEM) is not applicable without modification, because the time needed to compute each time-step is too long and not constant. Many methods have been proposed to overcome those difficulties, which are discussed shortly here after.

In this article, we address a further difficulty, often overlooked in the literature. Indeed, all methods proposed for the interactive simulation of soft bodies rely on a mathematical model of the physical behavior of the bodies and its simplification for the purpose of its integration in real-time. The methods of simplification are generally focused on computational efficiency, while enforcing macro-level properties such as volume and energy conservation. However, they don’t guarantee that the haptic feedback received by the user is in any way realistic. Indeed, it is often necessary to add damping factors in order to conserve energy in the model, which degrades the haptic feedback significantly.

We propose a new approach to the interactive simulation of soft bodies, which departs from the classical methods. Our approach combines two different models of the bodies, which don’t share the same representation. The first model is classical, based on a mass-spring system or on simplified FEM, but it is used only for the visual rendering of the bodies. Therefore, the time constraint on the first model can be lowered from haptic frame-rate (1 kHz) down to viewing frame-rate (30 Hz), thus allowing for a higher fidelity of the visual behaviour. The second model is dedicated to the haptic rendering, and does not have to be physics-based at all; it can even be omitted and replaced by an open-loop haptic signal. The parameters of that model are chosen so as to optimize the quality of the haptic feedback, with no immediate consideration of the underlying behaviour of the soft body.

We describe hereafter different possible implementations of the two models, and how they are linked together so as to build a working interactive simulation, and explain how the decomposition solves the problems of stability. We also discuss the domain of applicability of our method, for which the two models can be expected to deliver compatible results, and what would result if they happen to diverge significantly.

2. State-of-the-art
As presented in the articles [Pet09, CMOL12, KPY09], we will need both haptic and visual realistic rendering with a frequency of 30Hz for graphics and 1 kHz for haptics. For the visual part, it was shown in the articles [Hin96, AAML12] that it plays a very important role even in the haptic feedback because it "improves" it. For two similar haptic feedback, the one with the best visual rendering will be felt to be better. That’s due to the fact that Haptic
sensation is mainly caused by ‘“LOOK’ and feel” [Hin96]. Based on the article [GZC07] we can clearly see why this type of method worth it. In manufacturing, the assembly process is the most costly part. So it is of the most importance to well design that part in haptic using system. In order to achieve that, the authors decided to use a kind of ‘Look and Feel’ process, dividing the haptic and graphical parts in two distinct major parts, using a spring-damper model to link them together, so even if the haptic model was not totally realistic, it was still giving an interesting perception with a graphical part nicely designed. It has also been shown in [AAML12] that the visualization of shadows increase the perception of pressure without making the deformation more expensive.

We reviewed different deformation models [Her13, Pet09, DDKA06, AAML12, BJ08, BC97, KPY09]. The most realistic ones are those using non-linear laws for deformable tissues. The FEM (Finite Element Method) defined in the article [Bro97] is one of the most used in haptics because it represents even complex objects with simplified equations. In general, the behaviour laws don’t need to be applied globally on the whole scene, but can be computed locally, on each individual deformable object. Each object is represented as a surface triangular mesh and a volume mesh, which coincide at the surface nodes, and objects don’t overlap. ‘SOFA’ which is an open-source, free and quite recent library for medical simulation was presented in several articles [FDD*12, Her13, PND*11] all saying that it is quite powerful.

In the case of surgery simulation, as specified in the article [Bro97], due to the fact that two human beings can never be identical, a highly precise simulator is not required: no matter how accurate we want to make the model, somehow it would still be false, and only the realism really matters in that type of simulator. So it would be beneficial to base the simulator on measurements carried out on a real human body ([KFN06] demonstrates that a non-physical model based on measurement can be very realistic) and/or on a description thereof given by senior surgeons. As discussed in [VPP*14], what matters in the end is the Transfer Of Training (TOT) validity of the simulation, and most studies agree that the quality of the haptic feedback is paramount.

In the domain of high-precision manufacturing, [SLR*14] demonstrates that a haptic rendering based on experimental data significantly improves the performance of the user of the simulation. In the article, the authors base the haptic rendering on a database of pre-computed haptic behaviours, depending on the dimensions and parameters of the micro- and nanostructures.

3. Visual model

In our approach, the visual model is responsible for computing the deformations of the soft bodies and displaying them in a realistic way. We discuss hereafter the two most used techniques, i.e. Mass-Spring models and FEM.

In a Mass-Spring model, a body is represented as a set of nodes associated to a mass and ideal weightless elastic springs between nodes; it is important to note that their stiffness and damping coefficients can be different from one to another. The number of springs for each node can also be parameterized. Mass-spring models can be computed very efficiently in real-time, and are easy to implement. However, they cannot display a realistic behavior of a real soft body like a rubber hose or a living organ. Typical issues are energy and volume conservation.

The FEM is a well-known method for integrating differential equations numerically. The quality of the results depend on the type of finite elements, their density or size, the choice of equations and their parameters. The integration scheme plays also an important role, especially for the computing time. One solution for reducing computing time is to simplify the underlying equations, increase the size of elements and choose a fixed time-stepping integration scheme. As it was presented before, the keywords for a good haptic perception are ‘“LOOK’ and feel’. So the visual model, influencing our feeling for the haptic model, cannot be overlooked. For a good visual perception a frequency between 25Hz and 60Hz is enough.

Yet, we cannot fully base the model on the true properties of the object, because we just do not know them especially in medical simulation, where real measurements on living organs are problematic. So there are a lot of assumptions made behind this model.

The visual model is composed of one or more soft bodies, and a number of rigid bodies such as tools or mechanical parts. The rendering scheme consists of computing collision points between the objects of the scene, and converting them into constraints. In the case of a soft body, the collision will be transformed into the displacement of a number of nodes. In the case of a FEM, it needs to be turned into “boundary conditions”, expressed as a position for a node, or a force, or even a force/position relationship.

Up to now, we have experimented with position constraints, i.e. the haptic model enforces the position of the objects in the scene, and the visual model updates the deformations of the soft bodies. We intend to test other approaches later on.

4. Haptic model

In our approach, this model is voluntary much simpler and not based on the physical properties of the soft body. Our intention is to recreate a realistic feeling of the interaction, using a perception-based model.

As shown in [KFN06], it is possible to measure the real perception of an interaction with an object and reproduce it through a haptic device. However, such a scheme is valid only for transitory events. In the case of a soft body, the interaction is continuous, and there is a need for an action-reaction loop. Therefore, we need to design a model for the haptic rendering.

Our proposition for measuring the action of the user is to compute the volume of penetration between the non-deformed geometry of the objects. That volume can be computed very efficiently, and is usually proposed by the major collision detection libraries. If not available, then the penetration distance can be used as a replacement.
Alternatively, the proxy method could be used and then the distance to the proxy becomes the penetration distance.

This model, totally freed from mechanical equations can exert the force we wish, which is allowing us to fully focus on the realism of the haptic feeling for the user. That feeling can be defined subjectively by asking experts (e.g. manufacturing specialist or senior surgeon, depending on the application), or it can be measured on a real sample of the soft body to be simulated. It is usually expressed as a “force profile”. In case a precise FEM of the body is available, the profile could be generated through an off-line simulation done in advance. Fig 1 shows the force profile measured on a wet vegetal sponge.

The easiest way to implement the haptic rendering is to decompose the force profile into a series of linear or quadratic segments, which can be computed at a very high frame-rate so as to optimize the quality of the haptic feeling while guaranteeing stability. As discussed in [Per12], the profile can include some form of hysteresis, as the behaviours when pushing into the object and moving out of it might differ. Such hysteresis can be seen on fig 2, where the probe was moved in and out several times. It can be reproduced easily in the haptic rendering by adding a simple state automaton to keep track of the current profile.

Fig 1 : Force profile of a Vegetal Sponge

The figures 4 and 5 show the deformation of a hose modeled in Bullet Physics (mass-spring model) combined with haptic rendering.

5. Combining the two models

In the final simulator, the two models are running in parallel, each at its own update rate (1 kHz for the haptic model, typically 30 Hz for the visual model). They have to coincide as well as possible, so that in the user’s brain the visual model would be influencing the haptic domain due to the ‘LOOK and feel’ sensitiveness. But actually, only the haptic model is affecting the visual model, and no information is fed back from the visual model to the haptic one. All objects in the scene are displayed according to their positions and deformations as computed in the visual model (fig 3).

Fig 3: System architecture

As a consequence, there might be a discrepancy between the current position of the user’s hand in reality and that of the tool he’s holding in the visual feedback, due to the relaxation factors in the computation of the deformations. However, the user won’t notice the discrepancy unless it is very large or it results in a latency in the motion of the tool.

6. Implementation

As a physics engine for our first implementation, we have chosen Bullet Physics. For the visual deformation, we used the soft body already implemented in Bullet, using the mass-spring model. We were able to display an interesting soft body quite easily, with only a few parameters being set:
- Global Mass & individual vertex mass;
- Number of slices and stacks;
- Internal pressure (only for closed objects);
- Number of iteration for velocities, positions and drifts happening during deformation;
- Number of static vertices;
- Stiffness and Damping coefficients of elastic springs.
In a second implementation, we used SOFA deformation for the visual rendering with FEM deformation (fig.6). We were able to validate our approach on simple geometries like spheres and cubes, but didn’t pursue the tests any further because we lacked the expertise needed to optimize the FEM computation of complex objects in SOFA.

Up to now, we have only used simple models in order to check the validity of our method. The next step is to use more complex shapes like a body organ (e.g. liver, tongue) or a deformable manufactured object (e.g. hose, rubber seal). During the conference, we will show a demonstration of our approach on the Haption exhibition booth, featuring complex objects like the Suzanne model in Blender (fig 7), so that visitors can appreciate the quality of the results.

7. Discussion and future work

The advantages of our method are that we can simulate complex objects like organs or deformable manufactured objects with a haptic realistic rendering and that we can implement it in a real simulator. Further advantages are:

- The haptic rendering is easy to implement, even for a non-specialist, and can be very realistic if based on an actual measurement of the force profile.
- The deformation model does not need to be designed specifically for haptic rendering.
- The stability of the haptic feedback can be guaranteed by limiting the slope of the force profile to the maximum stiffness of the haptic device.

There are several limitations with our current implementation, the most critical one being that we cannot cut or tear the soft bodies, nor stretch them out of their initial geometry. Indeed, because our haptic rendering is based on the computation of the penetration volume between the non-deformed objects, the method might be considered as too limited. Nevertheless, there are many practical applications where the deformation is either small and within the bounds of the geometry at rest (for example the insertion of a car window through the rubber seals), or where a larger deformation is considered as a failure on the part of the user, and can be handled as a “game over” event (e.g. in eye surgery). We believe that for such applications, the benefits of having a realistic haptic feedback overcome the inconvenients by a large measure.

Still, we are planning to investigate different ways to simulate large deformations of objects. For this, we plan to segment the soft bodies into sets of overlapping rigid objects linked together with kinematic constraints. The role of the overlapping is to prevent the appearance of holes when the rigid objects move with respect to each other. The haptic rendering is to be based on the largest penetration volume or depth, and the motion of the kinematic chain will be computed by applying the repulsion force back on the object in collision.
We also believe that some changes of topology can be taken into account, as a modification of the force profile. For example, in the case of a puncture with a needle, there could be one force profile before the puncture and one after, with the switch between the two being triggered by a discrete event coming from the visual model.

Finally, we are also planning to investigate the handling of a soft body with multiple contact points, such as in the case of bimanual operations.

Fig 8: First author testing the SOFA deformable sphere with a Virtuose 6D haptic device

8. Conclusion

We have presented here a new approach to the simulation of soft bodies in the context of haptic rendering, which resolutely departs from former contributions. Our method combines two models, one for the haptic rendering and one for the visual feedback, but contrary to the classical approaches, in our case the haptic model is not a simplification of the visual one, and is not based on the physical properties of the soft body. Instead, it focuses on the haptic perception of the user, and is based on an actual measurement of the force profile, and/or on the expertise of an expert user.

We are aware of the limitations of our method, due to our depending on the non-deformed geometry of the objects for collision detection and penetration volume computation. However, we have pointed out possible extensions of our current implementation, which should widen the applicability of our approach.

References


A Virtual Environment for mission planning and supervision of Remotely Piloted Aircraft Systems

G.P. Viganò, N. Colosio, E. Romano and M.G. Todeschini

1 CNR-ITIA Istituto di Tecnologie Industriali e Automazione, via Bassini 15, 20133 Milano, Italy, tel: +39 02 2369 9602, fax: +39 02 2369 9616, e-mail: {giovannipaolo.vigano, nicola.colosio, enrica.romano, mauro.todeschini}@itia.cnr.it

Abstract

Situation Awareness has been studied since the late 1980s as one of the most interesting area within human factors research and has become a central concept in human decision making models. Since its origins it is related to the aviation domain, where providing a high level of SA is still a major challenge. Moreover, the increasing spread of Remotely Piloted Aircraft Systems, also known as Unmanned Aircraft Systems, needs to be matched with the development of tools that support pilots in crucial procedures. This paper presents a software solution to improve the Situation Awareness of a Remotely Piloted Aircraft operator carrying out the tasks that belong to mission planning and flight supervision in particular. The tool displays a 3D Virtual Environment that shows the operational scenario and contextualizes relevant information about it: exploiting Virtual Reality the software enables the operator to visualize a representation of the environment and then to perceive the possible obstacles and issues for the flight.

Categories and Subject Descriptors (according to ACM CCS): H.5.1 [INFORMATION INTERFACES AND PRESENTATION (e.g., HCI)]: Multimedia Information Systems - Artificial, augmented, and virtual realities, I.3.7 [COMPUTER GRAPHICS]: Three-Dimensional Graphics and Realism - Virtual reality, D.2.2 [SOFTWARE ENGINEERING]: Design Tools and Techniques - User interfaces

1. Introduction

A Remotely Piloted Aircraft System (RPAS) is the complex system that includes the aircraft, the navigation and the flight control tools, the payload, the sensors, the landing tools, the Ground Control Station (GCS) and the communication channel between the aircraft and GCS [WAH12]. Due to the physical separation of the operators from the aircraft the human-vehicle interface has an essential role, allowing the communication and interaction between them. Several software for Remotely Piloted Aircrafts (RPAs) mission planning and flight control already exist, which also provide the pilot with additional information about the system, e.g. the energy level, mission timers, RPA position, etc. [TLDP13]. Common users often experience difficulties controlling the aircraft and interpreting the data collected because most of these software are oriented to expert pilots, who are skilled and trained in flight issues. For this reason human-interface designs for RPASs need to be improved in order to provide a higher level of Situation Awareness (SA) [DJRG06].

SA was formally defined as "the perception of the elements in the environment within a volume of time and space" (level 1 of SA), "the comprehension of their meaning" (level 2 of SA) "and the projection of their status in the near future" (level 3 of SA) [End88]. A software implementing a human-interface for RPASs should allow a fast comprehension of the vehicle status, in relation to different classes of elements needed for SA [End99]:

- Geographical SA: location of the RPA, waypoints, geographical objects as terrains, cities, etc.
- Spatial/Temporal SA: RPA’s coordinates, velocity, path, mission progress, etc.
- System SA: status of the system, fuel/battery, etc.

SA is context dependent thus, in order to increase its level, it is important to understand which information is essential to
the intended users: the specific requirements depend on the goals of the particular RPA mission.

Achieving a good level of SA has a positive feedback in hazard awareness: a big issue in RPA domain. The risk of losing the control of the aircraft and of its fall down due to a lack of reliability of the system or a human mistake reduces the applicability of these vehicles. During the flight information is constantly changing and the environment needs to be frequently monitored to update the knowledge about it, in this context it is critical to make decisions in a strict time limit [Cas93].

In order to increase safety the civil aviation institutions and authorities of many countries have therefore defined regulations. In Italy the sector is regulated by the "Regolamento Mezzi Aerei a Pilotaggio Remoto" issued by the Ente Nazionale per l’Aviazione Civile (Italian Civil Aviation Authority) on December the 16th, 2013 [Ena13] and deferred until April the 30th, 2014 and currently in draft status [Ena15].

We have implemented a software tool (DroneAGE, Drone Advanced Graphic Environment) for mission planning and flight control of RPAs, which should provide a higher level of SA and, as a consequence, a good hazard awareness also to users not particularly expert or trained in flight related themes. Moreover the aim of DroneAGE is to present relevant information in an intuitive way, to omit distracting information or representations and to highlight important circumstances or aspects. To achieve this goals DroneAGE use Virtual Reality (VR), displaying a 3D Virtual Environment (VE) in addition to a more classical 2D Graphical User Interface (2DGUI) [RTVS14].

DroneAGE development has been conducted in the frame of Space4Agri project supported by the bilateral agreement between Regione Lombardia and Consiglio Nazionale delle Ricerche (CNR, National Research Council of Italy), and the preliminary requirements have been defined in collaboration with the Italian company Aermatica S.p.A., operating in the market of RPASs design, production and commercialization.

2. DroneAGE

The requirements for DroneAGE have been defined starting from the possible improvements to the flight control software of Aermatica ANTEOS RPAS [Aer14]. This software shows a GUI employing a 2D map both to plan the UAV mission and to manage the flight.

The first step of DroneAGE development has been the implementation of these stages in the VE, providing a 3D interface in order to improve the SA (based on osgEarth [Pe115], OpenSceneGraph [Osf15] and Qt [Dig15]).

The principal elements managed by DroneAGE, both in the VE and in the 2DGUI, are:

- Flight Space (FS): composed of Flight Zones (FZ)/No Flight Zones (NFZ), i.e. areas where it’s allowed/denied the flight, and Landing Zones (LZ)/No Landing Zones (NLZ), i.e. areas where it’s allowed/denied the landing;
- Flight Path (FP): composed of flight and mission waypoints.

In mission planning and during RPA flight it is important to have an effective perception of these two classes of elements: thanks to the visual depth perception the VE raises their awareness in comparison to interfaces displaying only a 2D scenario which are widespread in RPAS sector.

Moreover, during the ANTEOS flight, DroneAGE allows:

- the connection to the GCS;
- the download of telemetry data.

DroneAGE supports two different operating modes: mission planning (scenario editing) and flight supervision.

2.1. Mission planning

DroneAGE supports a special operating mode that can be enabled to edit the scenario and to plan the mission (Figure 1). Once in editing mode the application shows a different layout, including panels and menu items to allow various tasks:

- integrate GIS data: it is possible to load data from disk (e.g. GeoTIFF, JPEG+TFW raster data and ESRI Shapefile, GeoJSON vector data) or link to Internet data services (e.g. OpenStreetMap, MapQuest, Ready-Map);
- build and modify the FS: the base perimeter can be defined with a sequence of clicks with the mouse pointer on the terrain or edited in the VE or in the 2DGUI; the perimeter of the selected FZ is also projected on the terrain (Figure 2);
- define the FP: waypoints can be added clicking with the mouse pointer on the terrain and edited in VE or in the GUI; the FP is also projected on the terrain.

Once a FP is defined a basic animation represents the flight of the RPA along the defined FP. Both the FS and the FP can be imported or exported from or to files (currently only in the format defined by Aermatica).
2.2. Flight supervision

It is important to notice that DroneAGE does not aim to replace the GCS software, because the one provided by the RPAS vendors is granted to work with their systems and often implements low-level safety mechanisms. Furthermore, replacing a software of a certified RPAS (e.g. Aermatica ANTEOS) with a third party software (e.g. DroneAGE), can lead to a system that does not meet anymore the normative requirements to operate [Aer14].

DroneAGE can link to a GCS (currently only the one bundled with Aermatica ANTEOS) thanks to an auxiliary module that provides a read-only access to the RPAS telemetry (see paragraph 3). Connecting to the GCS DroneAGE can use the telemetry data to show flight status details in a proper panel and to animate the drone 3D model according to the received data (Figure 3).

A simple dead-reckoning function can be activated to compensate low data rates in order to provide a more fluid animation [LCL07]. Energy levels received from the GCS are used to update the battery status indicators, both in the overlaid HUD and in the GUI.

2.3. Improving the situation awareness

DroneAGE provides some features that are useful to aid the user when planning a mission and to enhance the SA while supervising the flight of the RPA (Figure 4).

The flight scenario is represented on a virtual globe that can be fully customized, by means of data retrieved from web services or from files loaded from the local file system. Before and during the mission planning several types of information can be loaded into the VE and saved with the scenario, providing useful information about the environment:

- raster maps coming from a web service, satellite or aerial maps: digital elevation models (DEMs), ortophotos and scientific data can be composed and mixed together;
- vector data loaded from a local topological database or raster maps coming from a web service can be used to
Figure 4: The virtual RPA in the VE

Figure 5: Automatically generated woodlands and date/time dependent lighting

Figure 6: The RPA drops shadow on the terrain with a generated detail map

display roads, rivers, parks, building areas, etc. (Figure 5);
• elements with vertical extension can be loaded as 3D models: buildings, trees, streetlights, etc.;
• real-time illumination reproduces the realistic lighting condition according to the current date and time (these settings can be changed by the users in real-time);
• a model of the flying RPA is displayed in the VE: some indicators are displayed on the screen to show its position and direction even when it is too far or out of sight;
• shadows can be enabled to aid the perception of the RPA altitude when flying: a detail map can be automatically generated to better evaluate the speed of the RPA shadow on the terrain (Figure 6).

These features can aid the mission planner in the design of FS and FP and provide an immediate feedback about the elements surrounding the RPA during its flight; its current status can be evaluated from any point of view with respect to the surrounding environment, thus augmenting both the feedback provided by the GCS and the direct visual observation.

3. Integration with an RPAS

Since DroneAGE is not a tool that completely replaces usual GCS features, and requires to access data coming from the RPA for its operation, a significant importance lies in information exchange and system components interaction.

• DroneAGE and ANTEOS as a whole can be considered highly heterogeneous system: most of the development of ANTEOS modules has been accomplished in the Java programming language;
• Aermatica GCS (AEGCS) runs on a conventional Windows PC; for reliability reasons a rugged laptop is generally used;
• RPA on board software runs on an embedded platform that can fulfil higher restrictions in terms of energy consumption and reliability if compared to a Windows PC;
• ANTEOS adopts a proprietary communication protocol for RPA command & control that has been developed on top of commercial telecommunications hardware raising reliability levels and failure detection and response;
• DroneAGE has been developed mostly as a C++ Windows PC application; this way is has been possible to exploit the features and availability of open-source production grade tools and frameworks for VEs creation and management.

The main considerations that have been done regarding DroneAGE development can be summarized as follows:

• DroneAGE must integrate with ANTEOS, an already production ready system; therefore modifications on ANTEOS side should be stripped down not to reduce reliability or require extensive effort both on development and/or validation;
• any choice regarding DroneAGE development should have come after the evaluation and selection of suitable interaction and coordination tools/techniques between ANTEOS and DroneAGE and should have been compliant with them; in particular a data exchange should have been possible between the two platforms while preserving data semantics.
Both file exchange and network communication have been considered appropriate; files may be used for FS and FP since no frequent exchange is necessary, while a network based communication approach has been estimated to be mandatory in the case of telemetry data exchange that requires high frequency and near real-time interaction. Therefore, for telemetry only, Aermatica implemented specific access capabilities in AEGCS and a Java API (AETAPI) to access it over the network through a proprietary network protocol. We then integrated the API in TIMDA (Telemetry Interconnection Module in DroneAGE), a custom build JVM application (see Figure 7), that:

- collects the telemetry data from AEGCS; TIMDA periodically requests updated data from AEGCS by using a polling process;
- serializes the data: a multi-platform and multi-language framework is used to preserve data semantics among the Java/JVM platform and the C++/PC one;
- transfers the data to the DroneAGE components that need it using a network protocol.

The whole process is depicted in Figure 8 in particular:

- AEGCS and RPA modules are, more or less, a black box from the perspective of DroneAGE and TIMDA: most of the inner workings are covered by industrial secrets and implemented through proprietary solutions. As an example it is unknown how much postprocessing is done on telemetry data inside AEGCS or if they are transferred almost untouched from RPA to TIMDA; this is irrelevant for DroneAGE implementation of operation;
- telemetry data is consumed both in AEGCS (e.g. displaying numerical values or updating the position on the 2D map) and in DroneAGE;
- TIMDA offers mainly intermediate services for telemetry data producers and consumers; e.g. data retrieval, data distribution/transfer, data conversion/serialization. No significant data consumption is performed: the user interface is used mostly to configure its runtime behaviour.

4. Conclusions

Since the early stages of development (as depicted in \[RTVS14\]) a lot of improvements and new features have been introduced in DroneAGE; in particular pertaining two areas:

- editing mode: in the beginning DroneAGE was basically a "read-only" tool able to display FS and FP related data that was previously defined in AEGCS. Now DroneAGE can export such data and support the user to define it from scratch. This achievement entailed the definition of suitable user interaction, the evaluation of the capabilities and the limits of the exchange file format and its handling by AEGCS, and the development of export features;
- real-time telemetry connection: by means of TIMDA module is now possible to retrieve live telemetry data
from the RPA or its simulator. This is essential for flight evaluation and overall SA about the mission.

DroneAGE novel approach lies in the integration of the features offered by RPASs software with enhancements that don’t subvert the workflow and behaviour the user has been used to and trained for. This is obtained through additional modules (DroneAGE, TIMDA) that are highly decoupled from existing and well established ones, but collaborate actively with them to offer the enhanced features.

The ongoing development of DroneAGE focuses on the improvement of its interface and the human-machine interaction:

- input devices: support for 3D mice, joysticks and game controllers is under development;
- voice synthesis: a text to speech module will be integrated into the next prototype to provide warnings and additional information during the flight supervision, without distracting the user from the VE;
- audio feedback: sounds will be played to emphasize alerts or dangerous conditions;
- acquired georeferenced data display: the data already captured by the RPAS will be contextualized in the VE to aid the mission operator to detect areas not yet covered by the RPA sensors.

Shortly we plan to apply the tool in missions in the area of Precision Agriculture to further evaluate the benefits and shortcomings of the tool.

5. Acknowledgements

The authors would like to thank Aeromatica for having shared its experience in design, manufacture and marketing of RPASs, for having contributed to the definition of the requirements and to actively cooperate on interaction/interface issues between DroneAGE and ANTEOS.

The authors also thank Regione Lombardia: this work has been conducted in the frame of Progetto Space4Agri supported by AQ-Regione Lombardia-CNR, Convenzione Operativa n. 18091/RCC, 05/08/2013 [TRVS14].

References


Augmented Reality using Point Clouds for Survey and Design in Outdoor Public Construction

EMERIC BALDISSER$^1$

SIG-IMAGE SARL
BIDART, France
e.baldisser@sig-image.fr

ROMAIN BRONSZEWSKY$^2$

ESTIA RECHERCHE
BIDART, France
romain.bronszewski@etu.u-bordeaux.fr

JEAN-MARC CIEUTAT$^2$

ESTIA RECHERCHE
BIDART, France
j.cieutat@estia.fr

PASCAL GUITTON$^3$

Université de Bordeaux
INRIA
BORDEAUX, France
pascal.guitton@inria.fr

Abstract
Worksite surveying in public construction requires highly skilled technicians who are able to accurately use laser stations, prisms and RTK GNSS. Our research work argues that augmented reality is capable of shortening this process by assisting field operators. We are designing an AR framework allowing geo-registered CAD 3D and BIM models, as well as GIS information to be directly superimposed onto reality. On the one hand, initial experiments coupling Laser and Vision on a tablet do not handle the occlusion of sketches or provide sufficient accuracy regarding French regulations. On the other hand, we know that colored point clouds will be increasingly available for any urban location thanks to recent trends, such as SFM with huge public databases of photos and thanks to 3D acquisition (eg. iTowns). This paper describes a method allowing reality on a tablet to be augmented with decimeter accuracy geo-registered CAD and GIS sketches. It merges – in real time on the field – vision of a surveyed terrain and its pre-processed point cloud from a scanner. The user places 3D BIM models by touching the scene on the screen. Augmentations are regulated by the geolocation of models regarding their position and the vision device’s pose in the point cloud. The main challenge consists of registering the image on the point cloud. We compare state-of-the-art methods and present our approach, which consists of aligning 2D images of scenes with 3D colored point clouds. Solving this issue, designing interaction methods and building the AR framework will considerably change survey workflows.


1. Introduction
In public construction, the first step in any project is surveying the terrain. It consists of recognizing fields, sketching and planning building equipment. In this paper we focus on outdoor materials and buildings, such as: electricity, gas and telecommunication networks, urban furniture and road facilities. During these steps high accuracy Laser and GNSS technologies are used for geo-registering locations of buried or off-ground facilities with an error margin of one decimeter. However, they require highly skilled technicians and time costly methods to provide accurate geolocation. In [BCG14] we presented how these works were split between real and virtual spaces.

[LLTK14] and [SZR13] respectively defend augmented reality and help workers in public construction by simulating future or invisible assets and by simplifying decision-making processes before and during work. [GQZX08]
argues that AR leads to safer work conditions in the field and speeds up maintenance tasks.

**Figure 1:** (Left) Surveyor measuring balanced laser station height. (Center) Pointing a prism with laser for geolocating the area on the floor. (Right) Surveyor geolocating floor points using GPS stick.

**Figure 2:** MR-UP couples monocular AR with a 200 meter capable laser range finder to survey the field. Sky view map helps user to check he/she exactly aims at desired points.

After having explored existing applications of AR in public construction in [BCG13], we proposed an augmented reality framework, named MR-UP, coupling a Laser and monocular camera on a tablet device, for sketching CAD & GIS geo-registered maps directly on a real time view of the field [BCG14]. Experiments gave rise to two particular observations: firstly, unhandled occlusion phenomena cause the unrealistic rendering of sketches in AR that can lead to erroneous on-field decisions. It is quite obvious as MR-UP displays 3D lines and polygons onto two dimensional representations of landscapes. Second, MR-UP does not efficiently reduce time costs as users still have to deal with laser, prism and GNSS initialization.

The first issue can be solved by introducing computer vision methods that would provide MR-UP respectively with a sparse model or with a dense 3D representation of the field in real time. The second issue requires GNSS and Laser Robot Stations that were designed on purpose for this kind of application to be removed. Investigations on topographic tools lead us to consider point clouds. In our case, it solves both issues. Moreover, point cloud representations of urban places will become increasingly available.

We aim to prove that using point clouds within our AR framework allows workers to draw technical sketches on a live tangible representation of the field in-situ, without using GNSS or Laser. We present an original method using computer vision and photogrammetry algorithms. This consists of aligning point clouds with augmented reality views of the field. The point cloud is anchored on the scene. Users only see a 2D representation of the field provided by a monocular camera on the handheld tablet device, but they interact with aligned invisible point clouds. This solves the second issue as every pointed zone is geolocated with an accuracy of less than a decimeter. It also solves the first issue because 3D line representations will be based on point cloud structures, but displayed on 2D video.

Among the technical issues to be tackled, point cloud alignment on live field views and continuous registering are discussed in this article. It is organized in four parts as follows: 1) related works and state of the art 2) global algorithm presentation 3) further works 4) conclusion.

**Figure 3:** Objective of our work: the foreground image is the tablet’s view of the field and the background image is the pre-processed RGB point cloud with high accuracy geolocation information. The user points to the field’s areas on the screen and places an electrical sheath (red). It is rendered as a live 2D view but occlusion and geolocation are handled by the invisible point cloud running in the background.

2. Related Works

2.1. AR frameworks for public construction

Augmented Reality seems to be a promising tool for use in the field of Construction. [DBFK13] mostly discusses office work for construction and proposes a framework called ARVITA based on markers to collaboratively design and plan constructions. [GPSR09] presents D’AR that considers the specifications of a construction project – such as schedule, 3D models of facilities, building chronology – to display on a photograph of the field the elements ahead of schedule or late, by automatically recognizing already built and missing structures.
On field augmented reality offers pertinent applications thanks to recent tablet/smartphone and network performances. [BAE13] presents HDAR + DAR’s extension — that allows workers on the field to take photos of buildings and to automatically annotate certain recognized elements, like windows, for example. These real elements then become clickable and link to information. [JSS08] describes the VIDENTE mobile application that shows buried facilities on site in the tablet camera preview, using global coordinates and GNSS, plus IMU real time information from the tablet device. [YHF12] describes VIDENTE’s new features allowing users not only to visualize, but also annotate zones on the AR view of the terrain and [SZR13] comments on surveying functionalities. It uses coarse GNSS plus IMU enhanced with laser range finder and pose estimation to collect global coordinates of zones pointed by the user on the tablet’s touch screen. Errors of geolocation on each axis vary from 10 to 20 centimeters. Similarly [BCG14] introduces MR-UP coupling the monocular camera and laser range finder on a tablet device accurately geolocated on the field as it is fixed on a prism stick, tracked in-situ by a total station, with one centimeter accuracy. Results slightly improve with an error of less than one decimeter in the survey. [SU13] shows results on an AR framework providing excavators with uncertain areas overlaid on a view of the field, highlighting potential buried and invisible facilities.

2.2. Point cloud alignment on images

In order to obtain exact GPS information for any point viewed by our AR framework, we need to match the monocular vision to the pre-processed colored point cloud in real time. This is a registration problem that occurs in many applications, such as LiDAR point cloud texturing for aerial [DLZ08] [KSSS09] [MRW*12] or urban scenes [Ding 2009] [ZNH05] [MK09], location recognition [IZFB09] [SLK11], augmented reality [BAE13] 3D model generation and in-situ visualization [YHF12].

Several methods are used depending on available equipment and information. Concerning monocular camera vision only, all the methods try to find correspondences between 2D live images and 3D point clouds. The differences lie in the intermediate steps:

- 2D-2D alignment considers monocular images as references and 2D projections of point clouds as inputs. [KSSS09] measures edge similarities between references and 2D projections through an edge cost function. The algorithm works, but results show duration of over one minute. [DLZ08] extracts structural corners and aligns them with 2D corners of live views with a success rate of 61% for urban locations. [BBD*11] extracts a planar surface from a raw cloud, performs edge detection on the matching plan in 2D scenes, and computes edge costs in a loop trying different poses. It is not automatic, since it requires the user’s action to select corresponding surfaces in the point cloud. [MKF09] registers LiDAR uncolored the point cloud on oblique aerial imagery by finding mutual information maxima, looping on point cloud projection and 2D image similarities. Results show between 10 and 15 seconds with a success rate of 95% to 100% in registration. 

- 3D-3D alignment generally includes a processing step requiring the user holding the vision device to move it. It allows SFM to compute point cloud construction of the work environment. Hence, it uses two point clouds for the terrain: the first one was previously acquired by a LiDAR or scanner, the second one comes from SFM. Their structures are compared using ICP, for example, in [ZNH05] that results in successful but only meter accuracy video registration on the point cloud (every 9 frame). [WWS13] also uses ICP to match dense point cloud with point cloud computed by SFM applied on several Google Street View photos of the scene.

- 2D-3D registration consists of finding matching correspondences directly between the image and point cloud, as described in [SLK11]. In this algorithm, SIFT descriptors used for SFM point cloud constructions are computed. Every 3D point in the cloud saves the mean value of its SIFT descriptors and is stored in an indexed visual vocabulary that allows filtering and speeds match searches up. Registration time varies from 1 to 1.5 seconds in the test results with a median error around 150 cm. [LSHF12] provides similar results, but presents a more robust method, estimating poses anywhere in the world by comparing 2D images to 3D SFM point clouds computed from 2 million images.

Methods featuring a loop over point cloud crops generally include RANSAC and M-Estimator. Existing techniques also differ regarding initializations:

- GPS coupled with compass and Inertial Measurement Unit data is conventional. It provides a more or less approximate estimation of the vision device’s position, heading and roll angles. This prior information speeds registration up in [YHF12] [KSSS09] [MRW*12] [DLZ08].

- Image features, such as corners, Hough lines and vanishing points [DLZ08].

- User initialization as in [YHF12] that allows users to specify pairs of matching points between LiDAR point cloud and real views of buildings.

3. Global Algorithm Presentation

The main originality of our work lies in the available information and in our goal. Related works previously quoted used point clouds from LiDAR or from motion stereo processes. These datasets do not provide centimeter accuracy geolocation of realistic colored points. Besides they focus on the rendering and visualization of buried or non-existent facilities. It is worth highlighting the fact that
terrain surveys are our main purpose. This includes recognition of the terrain and high accuracy geolocation of assets or Points Of Interest.

3.1. Hypothesis on data

MR-UP works on fields where point cloud data acquisition was previously carried out with a scanner (FARO Focus). The point cloud provides a centimeter resolution. The scanning process includes panoramic snapshots directly overlaid on the points. Hence, we have a 3D colored, discrete and geolocated representation of the field. We also have the possibility of obtaining global GPS coordinates for every point in panoramic snapshots. We cannot use SIFT-based approaches similar to [SLK11] and [LSHF12] as point clouds are not computed from SFM.

Figure 5: Point cloud sky view. Circles stand for scanner locations during the survey. They are also centers of panoramic photos.

Furthermore, we assume that the user has a RTK GNSS in the back, providing MR-UP with latitude, longitude and altitude every second with an accuracy of one meter. External IMU measures roll with 0.1 degree accuracy and the tablet’s compass gives heading without perturbations, except when working between metal and close surrounding buildings.

The handheld device is a Panasonic FZ-G1 featuring SSD drive and 4 giga-octets of RAM. The Intel I5 processor drives it fast enough to browse a 55 million geolocated point dataset. The embedded 8 mega-pixel camera has a digital zoom that does not affect focal length. We want the user to come and work directly on the field. 3D-3D alignment is out of context as it requires SFM computation of the point cloud to match the terrain’s 2D view with the pre-processed 3D point cloud. Therefore, our research initially focused on 2D-2D registration.

3.2. Matching scene view and pre-processed point cloud

The case study is located in an area of the Izarbel Science Park (Bidart, France) that we scanned with a FARO Focus placed on geolocated points. Therefore, any point of the cloud has centimeter accuracy global GPS coordinates and RGB values.

- We input prior information, such as RTK GNSS geolocation, embedded compass heading values and roll angles from external IMU. Initialization gives an approximate pose for the table’s camera regarding the point cloud.
- When the user points to an area, we compute the estimated position to return the matrix T that defines the affine transformation [ROV10] between the real scene and point cloud.
- P is the known projective matrix which projects the 3D point cloud onto the 2D viewport based on the approximate camera position and similar eye position in the virtual world.
- The user points to a pixel on the screen. Our algorithm then computes the coordinates of matching points in the cloud representation of the field using $T^{-1}$ and $P^{-1}$.
- Then, we request GPS accurate global coordinates for the point that is saved in the survey as a part of the red sheath.

Figure 4: Proposed global architecture for on-field survey using an AR framework

Computing transformation matrix $T$ from pose estimation between the live view screenshot and the point cloud screenshot is problematic. We are currently working on smoothing robust image processing screenshots of colored point clouds to match their descriptors with live view
screenshots. This will result in the pose estimation and in matrix $T$.

Another possibility for matching pointed zones with corresponding zones in the cloud is to exploit topographic tool specs. A scanner uses calibration to register panoramic photographs on point clouds. So, there is a known function that links a pixel of the panoramic photo - centered on the scanner’s acquisition point – to geolocation triangulated data computed from the point cloud. In this case, we could use the [DSH00] technique to match images from a monocular camera and the panoramic snapshot since their resolutions differ. However, we propose a general method illustrated in figure 4 that is convenient for any RGB geo-registered point cloud.

The first experiments on image-cloud matching used ASIFT and SIFT descriptors with unsuccessful results. Harris corner matching resulted in too many false correspondences to compute pose estimation. We are now working on image reconstruction combining Moving Least Squares with edge and corner detection on point cloud representations of the field. This way, we aim to match point clouds with images.

3.3. Interaction with real objects using a pre-processed point cloud

**Figure 6:** In any case, the operator points at real objects by touching its representation in the 2D live video

We used Qt, C++ and OpenGL ES to develop our GUI on a handheld device Panasonic FZ-G1 with Window 8 OS. Without any additional user-interface device or ubiquitous computing tool, we privileged the touch screen of our tablet to support user interactions. When a point cloud is aligned to the video stream of the camera, the operator can select a specific area. The Moving Least Squares method enables the smoothing of surfaces from a point cloud to be reconstructed. The callback associated with a touch event is similar to a ray shooting which intersects the point cloud. Shooting a ray in OpenGL requires the screen pointing position from the viewport space to be translated to the eye space first, then for a ray to be propagated in the direction of the far plan in order to determine which polygon is intersected. Then, the next steps depend on the tasks:

- **Simulation:** the operator tries to visualize on the field the facilities to be set. Objects are laid on the floor using a point cloud, but the user can specify the height (off the ground) or depth (buried equipment).
- **Survey:** collecting information on existing facilities and field details. 3D CAD georeferenced models are segmented, extracted and geolocated using a pre-processed point cloud.
- **Maintenance:** modifying previously collected information on surveyed or simulated facilities: their 3D CAD georeferenced models exist in our database.

4. Conclusion and Future Work

The originality of our work firstly lies in the fact that our goal, is to sketch geolocated CAD and GIS models on a live 2D representation of a scene by making the background interact with the dense colored point cloud of the scene. Research also differs with regard to input data as image-cloud state-of-the-art registration methods have generally used uncolored and not geo-registered point clouds, mainly from SFM and LiDAR respectively. Our current experiments compare the state-of-the-art with our 2D-2D pose estimation methods with images of a scene and colored dense point clouds of the same scene. Research aims to find the best combination of features and descriptors to look for so that matching is robust in outdoor urban locations.

iTowns from IGN (www.itowns.fr) enables users to navigate in a Google Street View-like environment where we can annotate buildings, sketch shapes, measure distances, surfaces, volumes and extract cropped point clouds of selected areas. It navigates in a road centered anachronistic 360° photograph. Our technology will allow us to do the same with the live scene of any location of the terrain and will not be restricted to the centers of photographs.
This will dramatically change survey processes and methods. Point clouds will be soon available for almost every urban location. When this is not the case, computing with a scanner lasts around 30 minutes for a 40 000 m² zone. Therefore, surveyors do not need any specific training to achieve decimeter accurate geolocation.

Future work will concern AR occlusion rules for BIM models in scenes regulated by model geolocation and scene point clouds. The next step will tackle interaction paradigms to draw on reality through tangible 2D representations of the field, in-situ.

5. References


[BBT*11] BENNIS A., BOMBARDIER., THRIET P., BRIE D. : Recalage d’un nuage de points de scanner laser terrestre avec une image de bâtiment. XXIIIe Colloque
Gretsi Traitement du Signal & des Images, Gretsi 2011, Bordeaux, France. CDROM, (Sep. 2011). <hal-00605130>


