Aspects of User Experience in Augmented Reality

Madsen, Jacob Boesen

DOI (link to publication from Publisher):
10.5278/vbn.phd.engsci.00111

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Madsen, J. B. (2016). *Aspects of User Experience in Augmented Reality*. Aalborg Universitetsforlag. https://doi.org/10.5278/vbn.phd.engsci.00111

ASPECTS OF USER EXPERIENCE IN AUGMENTED REALITY

BY
JACOB BOESEN MADSEN

DISSERTATION SUBMITTED 2016

AALBORG UNIVERSITY
DENMARK

Aspects of User Experience in Augmented Reality

Ph.D. Dissertation

Jacob Boesen Madsen

Dissertation submitted May, 2016

Dissertation submitted:	May, 2016
PhD supervisor:	Assoc. Prof. Claus B. Madsen Aalborg University
PhD committee:	Associate Professor, Dr. Matthias Rehm (chairman) Aalborg University
	Professor, PhD Gunnar Liestøl University of Oslo
	Accociate Professor, PhD Dan Witzner Hansen IT University of Copenhagen
PhD Series:	Faculty of Engineering and Science, Aalborg University

ISSN (online): 2246-1248
ISBN (online): 978-87-7112-307-4

Published by:
Aalborg University Press
Skjernvej 4A, 2nd floor
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2016

Author CV

Jacob Boesen Madsen

Jacob B. Madsen received his M.Sc. in Medialogy from Aalborg University. He worked as a research assistant at Aalborg University before starting his PhD study in July 2012 at the Department of Architecture, Design and Media Technology, Aalborg University. While working on his Master's degree, Jacob involved himself with projects related to Augmented Reality and application development. During his PhD studies he had a five month research stay at the Institute for Computer Graphics and Vision, Graz Technical University, Austria. He has been involved in supervision of undergraduate and graduate students in various topics related to the Medialogy study.

His research focuses on developing visualization and interaction experiences for mobile mixed and augmented reality scenarios, allowing users to explore information related to the real world. A major motivation of his work is to overcome the limitations of augmented reality to develop user friendly applications. Currently, his research interests lie in the area of human behaviors, augmented reality, and data analytics and visualization. His ultimate goal is to facilitate interactive visualizations and interfaces for accessing information to enrich our everyday lives.

In his present work, Jacob has conducted several studies centered on user behavior with augmented reality technology, investigating the effects of the technology on the user task performance.

Abstract

In Augmented Reality applications, the real environment is annotated or enhanced with computer-generated graphics. This is a topic that has been researched in the recent decades, but for many people this is a brand new and never heard of topic.

The main focus of this thesis is investigations in human factors related to Augmented Reality. This is investigated partly as how Augmented Reality applications are used in unsupervised settings, and partly in specific evaluations related to user performance in supervised settings. The thesis starts by introducing Augmented Reality to the reader, followed by a presentation of the technical areas related to the field, and different human factor areas. As a contribution to the research area, this thesis presents five separate, but sequential, papers within the area of Augmented Reality. As such, each paper build on results or inspiration gained in the previous paper(s).

Three papers present Augmented Reality installations to the visitors of a danish museum on two projects: *Memories of the Walls* and the *Castle Chapel*. They present the design and development of the installations, and present an evaluation methodology for unsupervised longitudinal studies along with findings of actual in field usage.

Two papers present specific evaluations of areas related to human factors in Augmented Reality, and present specific methodologies for evaluation of user performance: one evaluating user perception of static orientation errors and the other evaluating user task performance related to dynamic view management techniques.

Resumé

Princippet i Augmented Reality er at den virkelige verden bliver suppleret eller udvidet med computer genereret data. I grove træk kan det være at kombinere visuel information fra den virkelige verden med visuel computer data. I forskningsverdenen er dette et emne der har været kendt og forsket i de seneste årtier, men for mange almindelige mennesker er dette et helt nyt emne, de aldrig har hørt om.

Det primære fokus for denne afhandling er at undersøge menneskelige faktorer relateret til Augmented Reality. Dette undersøges delvist i studier omhandlende hvorledes Augmented Reality programmer i et naturligt miljø, uden opsyn, og delvist i specifikke laboratorieforsøg omhandlende brugeres præstationsniveau i forskellige sammenhænge. For at introducere læseren til emnet bliver Augmented Reality introduceret, efterfulgt af relevante emner relateret til teknologiske og menneskelige faktorer.

Afhandlingen præsenterer fem separate, men sammenhængende, artikler indenfor temaet Augmented Reality. Som udgangspunkt bygger hver artikel på resultater eller observationer fra tidligere artikler.

Tre artikler omhandler Augmented Reality projekter lavet i samarbejde med et dansk museum for at introducere Augmented Reality til deres besøgende: *Murenes Minder* og *Slotskapellet*. De præsenterer design og implementering af installationerne, præsenterer en metode for længerevarende studie af brugen, samt resultaterne af brugen på museet.

To artikler præsenterer specifikke evalueringsstudier relateret til menneskers brug af Augmented Reality, samt metoder for at evaluere brugeres præstationsniveau. I et studie undersøges hvor godt mennesker sanser statiske orienteringsfejl i Augmented Reality. I det andet studie evalueres menneskers præstationsniveau i forbindelse med dynamiske teknikker for informationsvisualisering.

Contents

Thesis Details

Thesis Title: Aspects of User Experience in Augmented Reality
Ph.D. Student: Jacob Boesen Madsen
Supervisor: Assoc. Prof. Claus B. Madsen, Aalborg University

The main body of this thesis consist of the following papers.

[A] Madsen, Claus B.; Madsen, Jacob Boesen; Morrison, Ann, "Aspects of What Makes or Breaks a Museum AR Experience" *IEEE International Symposium on Mixed and Augmented Reality (ISMAR-AMH)*, pp. 91-92, 2012.

[B] Madsen, Claus B.; Madsen, Jacob Boesen; Morrison, Ann, "Aspects of What Makes or Breaks a Museum AR Experience" *Unpublished*, The manuscript is presented for completion, and is the full paper submission of paper [A].

[C] Madsen, Jacob Boesen; Madsen, Claus B., "An Interactive Visualization of the Past using a Situated Simulation Approach" *Digital Heritage 2013*, pp. 307–314, 2013.

[D] Madsen, Jacob Boesen; Stenholt, Rasmus, "How Wrong Can You Be: Perception of Static Orientation Errors in Mixed Reality" *IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 83–90, 2014.

[E] Madsen, Jacob Boesen; Madsen, Claus B., "Handheld Visual Representation of a Castle Chapel Ruin" *ACM Journal on Computing and Cultural Heritage (JOCCH)*, 2015, vol. 9, no. 1, pp. 6:1–6:18, 2015.

[F] Madsen, Jacob Boesen; Tatzgern, Markus; Madsen, Claus B.; Schmalstieg, Dieter; Kalkofen, Denis;, "Temporal Coherence Strategies for Augmented Reality Labeling" *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 4, pp. 1415–1423, 2016.

In addition to the main papers, the following publications have also been made.

[1] Larsen, Camilla Horne; Lauritsen, David Skødt; Larsen, Jacob Junker; Pilgaard, Marc; Madsen, Jacob Boesen, "Differences in Human Audio Localization Performance between a HRTF- and a non-HRTF Audio System" *8th Audio Mostly*, no. 5, pp. 5:1–5:8, 2013.

[2] Larsen, Camilla Horne; Lauritsen, David Skødt; Larsen, Jacob Junker; Pilgaard, Marc; Madsen, Jacob Boesen; Stenholt, Rasmus, "Aurally Aided Visual Search Performance Comparing Virtual Audio Systems" *Audio Engineering Society Convention 137*, pp. 9150, 2014.

[3] Larsen, Camilla Horne; Lauritsen, David Skødt; Larsen, Jacob Junker; Pilgaard, Marc; Madsen, Jacob Boesen; Stenholt, Rasmus, "Aurally Aided Visual Search Performance Comparing Virtual Audio Systems" *Spatial Cognition IX*, pp. 70–73, 2014.

This thesis has been submitted for assessment in partial fulfillment of the PhD degree. The thesis is based on the submitted or published scientific papers which are listed above. Parts of the papers are used directly or indirectly in the extended summary of the thesis. As part of the assessment, co-author statements have been made available to the assessment committee and are also available at the Faculty. The thesis is not in its present form acceptable for open publication but only in limited and closed circulation as copyright may not be ensured.

Preface

This thesis is submitted as a collection of papers in partial fulfillment of a PhD study at the Section of Media Technology, Department of Architecture, Design and Media Technology, Aalborg University, Denmark. The thesis is the result of years working with the topic of augmented reality. It presents a series of papers, all in one way or another related to the common topic of Augmented reality, and how Augmented Reality affects user task performance in various ways. The research has presented me with interesting and diversified challenges, from intellectual challenges facing an aspiring researcher to the pragmatic issues of developing software to be used unsupervised in the field by end-users.

The research work for this thesis has been carried out at the Department of Architecture, Design and Media Technology, Aalborg University, Denmark, during 2012–2015, and during a research visit in 2014 at the Institute for Computer Graphics and Vision, Technical University of Graz, Austria.

I thank my supervisor Assoc. Prof. Claus B. Madsen for his advice, and for always having an open door and ear, and my colleagues at the department for inspiring discussions and making this an interesting place to work. Special thanks go to Rasmus Stenholt for many inspiring discussions, and the collaboration in publications included in this thesis. I also wish to thank my colleagues during my stay at the Institute for Computer Graphics and Vision, Technical University of Graz, for warmly welcoming me and for the collaboration and support during my stay.

Finally, I wish to thank friends, family and especially Janni for support through the years, and for reminding me that there is a life outside work.

Jacob Boesen Madsen
Aalborg University, May 12, 2016

Part I

Introduction

Introduction

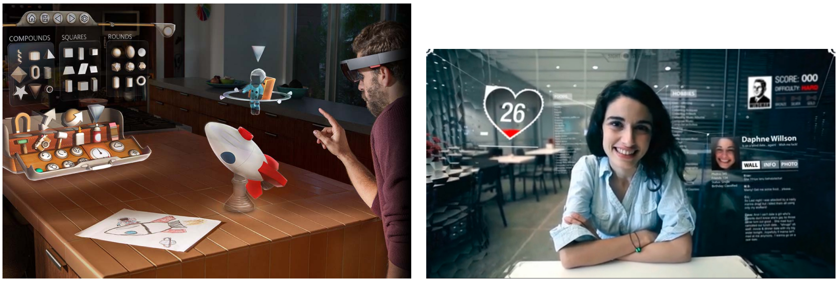
> *"If you wish to make an apple pie from scratch,*
> *you must first invent the universe."*
>
> Carl Sagan, *Cosmos*

Augmented Reality (AR) is the concept of augmenting the user perception of the world with additional computer generated content, which is related to the current view of the world, with either information related to the world or virtual content placed into the world as if it were real objects (Figure 1). AR aims to present an interactive experience, supplementing the real world with additional information. One use case for AR is to give the user a perception which is not available ordinarily.

In his survey of AR [3], Krevelen asks the reader to imagine a world in which you can see more than other people, and hear something others cannot. It is a world where technology enables this imagination, and we get the ability to perceive computational elements augmented into our real world experience. In his world we have creatures and structures that aid us in our daily lives, and interaction with this world is as natural and simple as speaking and gesturing. This is the world many dream of being a reality with AR technology, some day.

A lot of people are working on improving technology in order to get there. And just as many come up with ideas for future implementations, once the technology is ready. As one of the most current examples, Microsoft are presenting the Hololens (Figure 1a), which according to themselves will allow a more natural interface to the technology. An example idea of how the technology might be used as it matures, Sight (Figure 1b), a short futuristic film by Eran May-raz and Daniel Lazo. It presents life with AR technology in contact lenses, and how that might affect everyday life.

But, before we get to that point, there is a series of issues we need to tackle. These issues are mainly in relation to the world, such as getting precise information on location and orientation of the user in the world, and following that, some information about the close proximity of the user, i.e. what he is



(a) Microsoft Hololens marketing example of how their technology might be used

(b) The short film "Sight" by Eran May-raz and Daniel Lazo, as they imagine AR might be used in the future

Fig. 1: Examples of AR interfaces as imagined used.

looking at, is there anything blocking the view or occluding objects? Chapter 5 dives more into some of the technological issues of AR, describing the current state-of-the-art and how far we are from the goal, and how this has motivated my work and this resulting thesis.

Even with all this technology we are not done yet. Having all the technology in place only enables us to move the applications from the laboratory into the real world. The perceptual issues related to the user when seamless blending of computational objects into the real world is the first topic to explore related to the human factors of AR. There is a need for understanding these human factors in order to enable design of applications, which successfully convey information to the user, which is consistent with the users expectations. This is presented in Chapter 6.

Next we need to enable end-users to exploit the technology to the best of their abilities. Learning to use new technology is something many of us take for granted. We set time off to do so, and then we just learn it! How hard is that? Well, for many people it really is a hard task. And with disruptive technology that greatly differs from the technology people are already used to, it becomes even harder. Think about how many people you might know, who cannot use a mobile phone, those who never send emails, or who never use a computer. This is in one end of the spectrum of users, while in the other end of the spectrum we have the technology literate and first movers of the technology world. Getting people to use technology in the real world is a topic which in many ways is out of the scope of this dissertation, and it is not explored in depth in any of the chapters nor publications, but will be shortly covered in Section 6.3.

1 Motivation

The thought-experiment described in the introduction presents us as developers and researchers with a general list of topics to overcome in order to create engaging and convincing AR applications.

In this section, the goal is to motivate the research in relation to the topics mentioned in the introduction, and to motivate why this is necessary to develop engaging AR applications. The topics can roughly be divided into a technology aspect and a human factors aspect, which both must be given adequate attention to create any type of useful AR application.

With continual increase in computational power, increased hardware mobility, and a decrease in size of technology, the barriers of the technology are becoming lower and lower. It is possible to create new technology driven applications which exploit the surroundings of the user to embed context into the world. Augmenting computer generated content into the world through tablets, phones and glasses, wherever the user is in the world, is possible. By doing so we might change the way the user interacts with the surroundings. However, as stated in the introduction, we are still not there yet, and there are lot of technical limitations to create in-field unsupervised AR applications that "just works".

One area, which has been the focal point, motivating the research presented in this thesis, is the idea of getting the technology in the hands of real people, in a real setting, for an extended amount of time. Not a small scale study in the lab for a short duration. By enabling users in a real setting we can better uncover how the technology is being used from their perspective, and how they opt to use it with no supervision or no sense of commitment. However, this presents the problem of not being in control of the environment and the users, and may limit the reproducibility of the experimentation, as well as increase confounding and random variables in the experiment. On the positive side, it allows for getting a better understanding of the application usage of users in a natural setting, where they do not feel obliged to participate as requested by any traditional experimental laboratory study would do.

A way to introduce an audience to AR in a more controlled setting, to limit some of these confounding and random factors, is through museums and historical settings, in which they can be exposed to AR in a limited form, with defined settings and boundaries for the ease of experimentation. Outside the lab, multiple studies have attempted to leverage AR technology for historical sites, such as a situated documentaries about historical events [2, 7, 9], reconstructions of cultural heritage [1, 8, 10], geo-physical Situated Simulations [5, 6], and urban city touring [4]. Within museums AR can be used as a tool with which visitors can extend the information available to them, by adding to the existing information, be it by audio, visualization and/or haptic augmentation. Some recent application examples of AR in museums include museum guides, that attempt to bring added information to the museum visit through augmentation of information into the scene. Similarly, AR exhibitions offer the user the ability to explore AR installations, and AR interfaces can represent artifacts that may otherwise be off limits to the public, for various reasons. Two projects in this thesis deal with applications presented to visitors in a museum context running daily on location, unsupervised. The projects present collaboration with Koldinghus Museum in development and experimentation of two installations: *Koldinghus Augmented: Memories of the Walls* and *Koldinghus Augmented: Castle Chapel*. These projects and the setting will be presented in more detail in Part II of this thesis.

Despite being a large field of research, and with multiple people striving to take AR technology from the research labs and into the world for people to use in their daily lives, we are not there yet. In the research literature, people discuss the presence of "the one app" or "the killer app" to save AR as a technology and research area, or to define AR in the eyes of the world and the media. A lot of aspiring developers and researchers have devised innovative, but sometimes also crazy, AR application hoping to discover this one app.

In the reality of things, I believe this one application does not exist. Nor should AR be treated simply as a technology platform for development. Instead I see AR as an interface to data. It is decoupled of the display technology, and can be implemented on hardware platforms such as a computer, tablet, phone, or a pair of glasses even, and it becomes the interface which displays the data to the user. In that sense, looking for a single app to define AR is sort of like looking for the one app to solve all of peoples problems. The correct thing to do, in my opinion, is to leverage the interface paradigm and not set limit for "AR only", if in reality the application would be better off with a different interface. What is really needed is to focus on the people using this interface technology, and develop applications which are useful to a potential user. Just like with any technology, there are always cases where a specific interface is simply not useful, or there is a more simple way of doing it using another interface. To develop an application for which the AR interface is the most natural thing is the hard part. And we must elevate both the technology and our knowledge of human factors in order to arrive there.

2 Thesis Outline

This thesis consists of three main parts

In this first part of the thesis, Part I, the topic of AR and motivation for the experiments of the thesis was introduced. Following this, the core concepts of

AR is introduced, as well as the history and definitions needed for discussing the topic and sets up this as a common framework for later discussions. It relates the contributed experiments within this thesis to the broader field of AR and presents both the technology and user problem spaces. Furthermore, this part motivates the remainder of the thesis, the contributions.

Part II introduces the research projects and experimentation carried out. It initially motivates the research area of AR applications for unsupervised in-field experimentation. Following the research projects, the main contributions of the thesis are summarized and conclusions are drawn. This part is the summary of the main body of work and generation of new knowledge to the research field.

The final part of the thesis, Part III, contains the papers that make up the core scientific contribution, presented in chronological order.

The AR applications developed during this thesis work, can be categorized into two elements, related to either longitudinal on-site studies or in-lab experiments. For the first element, Papers [A-C, E] address longitudinal user studies on location, in unsupervised settings. These applications demand knowledge of both technical insights in developing AR applications (presented as theory in Chapter 5), as well as insights into how people make use of the technology within the wild (Chapter 6). For the second element, Papers [D, F] presents laboratory work within more perceptual areas of user studies and AR (Chapter 6), where insights gained in the the first part of the research area have been used to instruct focused research.

References

[1] J. Happa, A. Artusi, and P. Dubla, “The Virtual Reconstruction and Daylight Illumination of the Panagia Angeloktisti.” *VAST*, pp. 49–56, 2009.

[2] T. Höllerer, S. Feiner, and J. Pavlik, “Situated documentaries: embedding multimedia presentations in the real world,” in *Digest of Papers. Third International Symposium on Wearable Computers*. IEEE Comput. Soc, 1999, pp. 79–86.

[3] D. W. F. V. Krevelen and R. Poelman, “A Survey of Augmented Reality Technologies, Applications and Limitations,” *The International Journal of Virtual Reality*, vol. 9, no. 2, pp. 1–20, 2010.

[4] G. A. Lee, A. Dünser, S. Kim, and M. Billinghurst, “CityViewAR: A mobile outdoor AR application for city visualization,” in *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*. IEEE, Nov. 2012, pp. 57–64.

[5] G. Liestøl, “Augmented reality and digital genre design — Situated simulations on the iPhone,” in *2009 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media and Humanities*. IEEE, Oct. 2009, pp. 29–34.

[6] G. Liestøl, T. Rasmussen, and T. Stenarson, “Mobile innovation: designing and evaluating situated simulations,” *Digital Creativity*, vol. 22, no. 3, pp. 174–186, Sep. 2011.

[7] D. Pletinckx, D. Callebaut, A. E. Killebrew, and N. A. Silberman, “Virtual-reality heritage presentation at Ename,” *IEEE Multimedia*, vol. 7, no. 2, pp. 45–48, 2000.

[8] M. Russo and G. Guidi, “Reality-based and reconstructive models: digital media for cultural heritage valorization,” *SCIRES-IT*, vol. 1, no. 2, pp. 71–86, 2011.

[9] M. Trapp, A. Semmo, R. Pokorski, C.-d. Herrmann, J. Döllner, M. Eichhorn, and M. Heinzelmann, “Communication of Digital Cultural Heritage in Public Spaces by the Example of Roman Cologne,” in *Proceedings of the Third international conference on Digital heritage*. Springer-Verlag, 2010, pp. 262–276.

[10] V. Vlahakis, N. Ioannidis, J. Karigiannis, M. Tsotros, M. Gounaris, L. Almeida, D. Stricker, T. Gleue, I. T. Christou, and R. Carlucci, “ARCHEOGUIDE: First results of an Augmented Reality, Mobile Computing System in Cultural Heritage Sites,” in *Proceedings of the 2001 conference on Virtual reality, archeology, and cultural heritage - VAST ’01*. New York, New York, USA: ACM Press, 2001, p. 131.

The Augmented Reality Problem Space

> *"If you love what you do and are willing to do what it takes,*
> *it's within your reach."*
>
> Steve Wozniak

This chapter presents the relevant theory of AR and related topics in the human factors area. In recent years, the topic of AR has greatly expanded, and in addition to visual AR, audio, tactile and other senses have been covered by researchers.

As a delimitation for the thesis, this chapter is limited to the topics relevant for the research areas of the work presented, namely visual AR and some human factor topics related to AR. For a more thorough survey of the field as a whole, the reader should consult literature such as [7, 9, 75].

3 Defining the topic of Augmented Reality

The aim of AR is to present an interactive experience, supplementing the real world with additional information. This information is presented to the user as anchored in the real world. The physical world and objects herein acts as the backdrop to the user, on top of which the computer generated content is presented. To the user it can be perceived as real and virtual stimuli co-existing in the world.

In a broad sense, AR can be defined as the artificial, seamless, and dynamic integration of new content into one's current perceptions of reality. Here one's view of a physical environment is supplemented (or augmented) by computer-generated sensory input, including visual, auditory, haptic, olfactory and taste. The augmentation happens in real-time, as well as in the correct semantic context of one's real environment. In the context of this thesis, I delimit this broader definition of AR, and focus mainly on visual AR. Through the papers presented, my research focuses on visual and perceptual issues related to AR [93, 95–98]. In addition to the presented papers, I

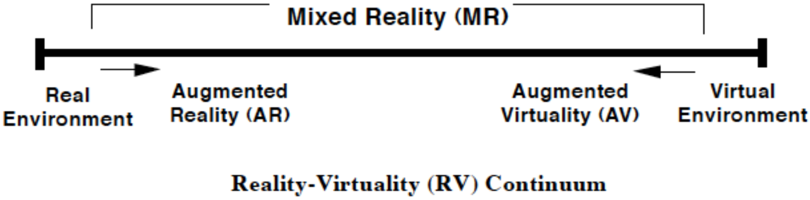


Fig. 2: The Reality-Virtuality Continuum as presented by Milgram et al. [106].

have been part of research projects within Auditory AR and VR and related perceptual issues [76–78].

Most researchers agree on which elements of what constitute AR, but there are differences of opinions on some areas of the research. I follow the definitions for AR by Azuma [9], and at the same time think of AR as a subset of the Virtuality Continuum by Milgram [106]. Building on this, I choose to also include Indirect Augmented Reality [153] and Situated Simulations [85] as being a subset of AR. The reasoning behind this is explained subsequently.

Milgram et al. [106] presents the Reality-Virtuality continuum (Figure 2), and classify it as part of a larger set of Mixed Reality (MR).

Azuma [9], and Azuma et al. [7] define AR as systems which exhibits three characteristics:

1 Combines real and virtual

2 Interactive in real time

3 Registered in 3D

This definition allows the AR application to be decoupled from the display technology, and allows for a broad range of applications to fit the criteria. It can be presented on a 2D display, which is the most common display method (computer screens and mobile displays such as phones and tablets), or it can be rendered to a stereoscopic display in a Head-mounted display (HMD) as just two examples.

Starting from the last characteristic presented by Azuma, being registered in 3D means that the augmented information must be linked to physical point in the world, such that when the user navigates around in the world, the augmented element stays linked to a physical point in real 3D space. Different sensing technologies enable this registration and tracking in 3D to various degrees of precision. It is a topic of discussion how precise tracking is needed for coherent AR, if we don't simply accept the fact that the more precise tracking the better. Depending on the application, the tracker may need more or less precision. Examples of this would be for medical and surgical assistance where high tracking precision is a hard requirement. And in comparison, adding additional information to a poster on your mobile phone might be completely fine with less precise tracking.

Being interactive in real time, the second characteristic, tells us that the application should be a form of interactive media. As such, movies are not part of this definition, and even though most movies today add computer generated content into the scene registered in 3D, it is not an interactive technology, and therefore not AR. As it is not fully defined from the characteristic alone, I define this interactive media as either allowing manipulation of objects in the space or the manipulating the viewpoint looking at the space.

The first criteria, combining real and virtual objects in a real setting, is the idea of augmenting computer generated content into the physical setting. Depending on how wide or narrow this criterion is interpreted may restrict the types of applications fall into the category of being AR. Some initial questions one might ask are:

- *"What is meant when saying combining real and virtual?"*. I.e. does it need to be computer generated visual content onto a live image capture?
- *"Does it have to be in the current setting of the user?"*. I.e. if I am watching a soccer match, and information is drawn on top of the game, registered in 3D on the field, is that AR, even though I am not situated in that setting currently?
- *"Must the coupling of real and virtual happen on the display technology?"*. I.e. must the blend of real and virtual content happen on a screen, or can there be another coupling instead, such as a mental coupling?

For this thesis, I define AR to also provide local virtuality, as presented by Benford et al. [16] in his classification on shared spaces, depicted in Figure 3. Local virtuality happens in situations where the user is presented with synthetic information related to the local physical setting where the user is located. According to this classification, AR provides local virtuality, and applications which seek to present virtual information to the user placed in the correct local physical setting can be considered a subset of AR. For this reason I include Indirect Augmented Reality [153] and Situated Simulations [85–88] as being part of this subset of AR. Both authors present a general idea that it is possible to replace the live video part of the augmentation, and rely on other factors to create AR. In the case of Wither et al. (Indirect Augmented Reality), they use a pre-captured panoramic image of the same location to match the tracking better. And in non-technical terms, there are no "live" elements shown to the user, even though the user is not aware of this. They motivate this by improving the tracking quality while the user maintaining a similar experience to that of a live camera feed. In the case of Liestøl et al. (Situated Simulations), they opts to remove the live image entirely and

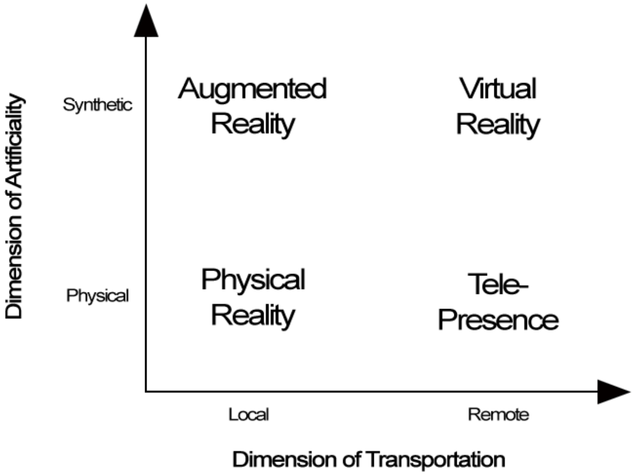


Fig. 3: Benfords classification of shared spaces according to transportation and artificiality, in which AR needs a local presence in space to be considered AR and not simply virtual [16].

replace it with a historical representation of the physical place. I.e. they use the physical presence of the user as the "real" world (from Azuma's first property of "Combines real and virtual"). In case of Situated Simulations, the fact that the simulation happens in the correct physical location makes this AR, similarly as described by Wither et al.

In both of these cases (most profound in Situated Simulations) there is mainly a mental coupling, where the user must understand how the visual content on the screen fits into the local physical environment. Physical local presence is needed to achieve this mental coupling. If this is not possible it loses the local virtuality aspect, and becomes an interactive visualization not coupled to the real world, and therefore no longer AR. In this thesis, the same concept and definition is used for motivating the visualizations presented in *Koldinghus Augmented: Castle Chapel* (Chapter 9).

4 A Brief History of Augmented Reality

The term *augmented reality* was coined in the early 1990s by Caudell and Mizell while working at developing display technology for the manufacturing process at Boeing [21]. During the last few decades, a multitude of research has developed the technology even further, and investigated use-cases for AR. This section presents a brief introduction to the research area and explains how we got to where we are now.

In the research field, Ivan Sutherland constructed his "Sword of Damocles", the first head-mounted display system for AR, in 1968 [140] (See Figure

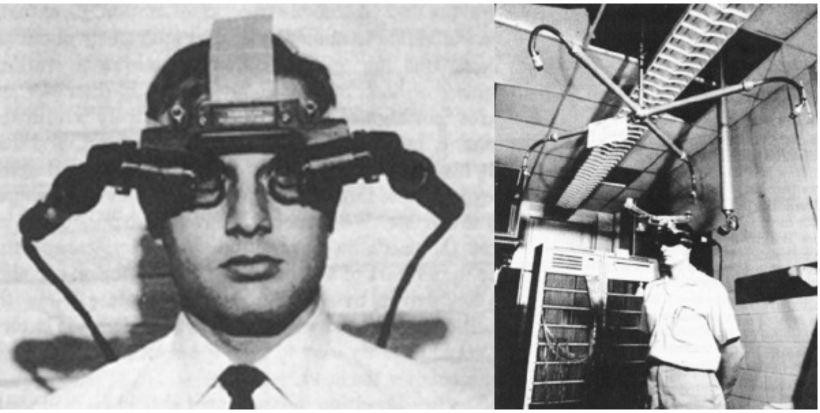


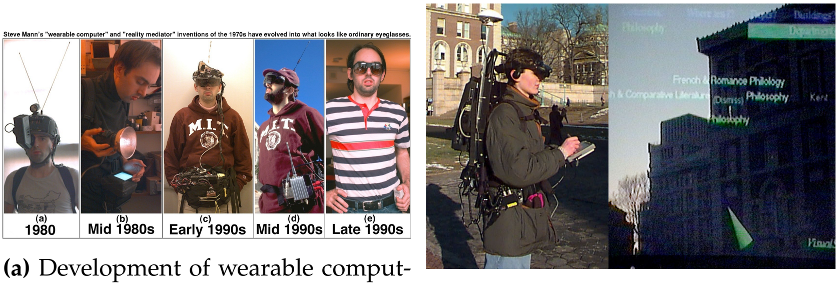
Fig. 4: The Sword of Damocles developed by Ivan Sutherland at the University of Utah [140].

4).

During the 1970s and 1980s as technology matured and computers became powerful and small enough to be worn, industry products such as Sony's Walkman (1979) were introduced, the digital watch and early personal computers led to a boom in the research and industry in the early 1990s, where applications soon started to be published [11, 21, 129]. Mobile computing and AR through wearables and GPS was introduced by Steve Mann [100] and Feiner et al. [38] (Figure 5). Also in the 1990s, scientific conferences started appearing: the International Workshop and Symposium on Augmented Reality, the International Symposium on Mixed Reality, and the Designing Augmented Reality Environments Workshop. From the industry side, technology became smaller and more powerful, and computing devices such as the Apple Newton MessagePad 100 (1993) and the Toshiba Libretto (1996) appeared. In 1997 Ronald Azuma more or less defined the field of AR with his survey on the research field [9], and re-defined it with slightly different wording in 2001 [7]. In the latter part of last century, Kato and Billinghurst introduced ARToolkit [67], a software library making it easier to develop AR applications using camera registration of fiducial markers.

A decade ago computing processors had become even more powerful and energy efficient, and we saw the early camera phone from Sharp (the J-SH04 from 2000), and later the Apple Iphone (2007) and HTC Dream (2008). With this mobile technology researchers had the power to move the technology to a truly handheld and mobile device. With advances in tracking technology, in 2007 Klein and Murray presented their implementation of Parallel Tracing and Mapping using a monocular camera [69].

During the last 10 years, public interest in the topic and related technologies have increased (See Figure 7) and motivated more people to enter



(a) Development of wearable computing from Steve Mann, from 1980s to late 1990s [100].

(b) Prototype mobile AR system from Feiner et al. [38].

Fig. 5: Wearable and GPS computing.

Fig. 6: Tracking and mapping for AR using commodity webcams from Klein and Murray [69].

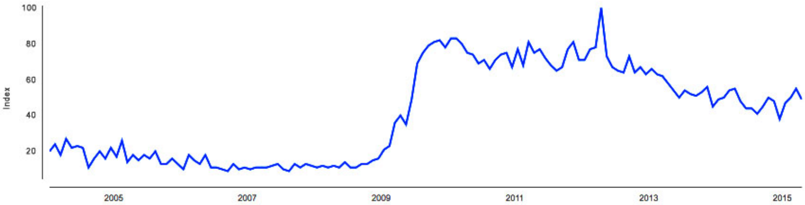


Fig. 7: Google Trends index for popularity of "Augmented Reality" from 2005-2015.

the field. And with new publicly available technology such as the Occulus Rift for virtual reality and the coming Hololens for AR, it might inspire researchers and developers to new use-cases for AR as an interface. Further sparking the interest and dreams (and fears) of the public, Hollywood has set the bar high when imagining AR interfaces in their movies, and what the endless possibilities are.

Over the years, there have been multiple surveys on the field of AR. Anyone interested in the history and technology should refer to the consult liter-

Fig. 8: Examples of visual coherence of AR [83].

ature such as [7, 9, 75] for a more in-depth historical insight to the research area.

5 Technical Challenges in Augmented Reality

The concept of AR, i.e. superimposition of virtual content onto the real world, typically onto a live video sequence of the real world, and to do this in a realistic and convincing manner, is a challenge in many ways.

The theoretical principles behind developing AR applications are well known [123], and have been for more than a decade. However, it is not until recent years AR has started rising in popularity, mainly on mobile devices. To develop applications leveraging on AR technology, a lot of specialist knowledge is needed and little (if any) technology frameworks work out of the box with no hiccups, in a full development cycle.

The general principle is to augment digital computer generated content into the real world. Sort of as the step between the real world and virtual reality. The embedded content is often based on real content in the scene, such as embedding a 3D object on top of a marker or known position, or linking a label to an object or point in the real space.

In order for the user to perceptually identify this augmentation as part of the world, the fidelity, and in the case of this thesis the visual fidelity, i.e. the degree to which it matches the world must be considered in the research and in the application development. Visual fidelity naturally leads to three main technical aspects to consider: tracking, occlusion and illumination.

Figure 8 illustrates examples of these areas. Tracking is the cornerstone of AR, allowing AR applications to be developed according to the three characteristics of Azuma's definition of AR. Tracking alone has the purpose of ensuring that the embedded content is registered correct in 3D. No other chain in the development pipeline has this much responsibility, and even the characteristic of having interactive frame-rates is the responsibility of all components, and if one component use too much processing power or too much time, it will stop being interactive, or the experience will suffer. It is much easier to develop an application if you do not care about occlusion or illumination of the experience, but that does not make those areas unimportant.

And for a good AR application with a good user experience, these topics are a necessity to consider in the development process.

5.1 Tracking

During the years the term *tracking* has been used by the community as a "catch-all" term for detection, registration, estimation and tracking positions and orientations of known or unknown objects in a scene. Or at least for scientific content where the registration or tracking has not been the main purpose of the contribution. For this section, tracking is used in that context, as a catch-all term for the above.

Tracking is the task of observing the state of an object (or multiple objects) in the world over time. This state can be both the position of an object, its orientation and features of the object. It is a topic which has left many developers with many headaches and something practitioners in the field more often than not delegate to specific development libraries or experts rather than doing themselves. In the relation to this thesis, tracking is defined as an adaptation of [14],

- the process of estimating and following the pose of a physical object in the real world over time, relative to the sensing device

and for the application and experiment development presented in the thesis, readily available tracking technology has been employed and the tracking technology itself is not the focus of any reported experiments. This section will therefore serve the purpose of introducing the reader to the topic of tracking for AR, and technology used in the experiments will be presented.

In order to align the virtual computer generated objects to the world, the 6 degrees-of-freedom (DoF) of the object and the pose of the place in the world it is augmented into must be known, and as accurately as possible [6].

For motion tracking for augmented and virtual environments, there are many enabling technologies such as e.g. mechanical, ultrasonic, magnetic, optical and inertial tracking [128, 151]. Welch and Foxlin [151] presents a thorough review of the state of the art within motion tracking techniques, explaining pros and cons of each method in great detail.

Notably from the research field, AR technology has seen much progress with various motion tracking technologies as an enabling technology for the applications. There has been research into tracking from GPS [38], to inertial tracking [45], further to camera vision based techniques, utilizing both 2D fiducial marker based [55, 67] and feature point based techniques [70, 91, 150], i.e. with or without known objects in the camera field of view to track. Researchers have recently attempted combining multiple techniques in sensor fusion tracking solutions [20, 68, 119].

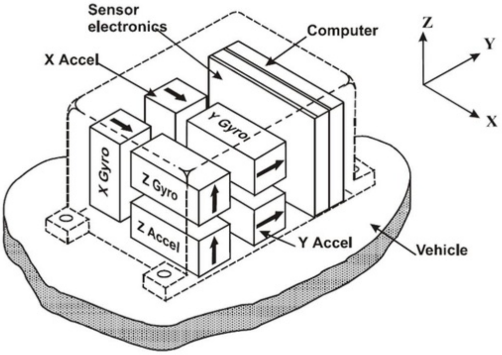


Fig. 9: Example of a strapdown inertial navigation system with 3 axis accelerometer and 3 axis gyroscope. [22]

The systems employed in the experiments presented in this thesis relies on inertial and/or optical tracking, which is specified in the individual experiments. For this reason, this section will only go into detail with these methods for tracking, and the reader is refered to Welch and Foxlin [151] for a more thorough review of the other tracking technologies.

Inertial tracking

An inertial navigation system (INS), is a self-contained device measuring orientation and acceleration using a combination of accelerometers and gyroscopes (see Figure 9 for an illustrated example). Inertial systems have been in use for ships, airplanes, submarines since the 1950s [151], and have since found useful for a wide range of applications from unmanned aerial drones and guided missiles to tracking for AR.

Gyroscopes measure the angular velocity of the system. Accelerometers measure the linear acceleration of the system. From the measurements of angular velocity and linear acceleration, it is possible to determine current orientation and velocity of the system by integrating measurements of the system, and integrating a second time yields position of the system. Figure 10 displays the information flow from the INS to calculate orientation and velocity.

Inertial sensors such as illustrated in the examples in this section are completely self-contained, requiring no setup or installation in the settings to work, and are available in most handheld mobile devices (phones and tablets) being sold today. Other characteristics of inertial sensors are low latency and high sample rate, meaning they can be measured multiple times (up to thousands of samples) per second and high quality units are less prone to jitter effects. However, for inertial tracking components often found in consumer

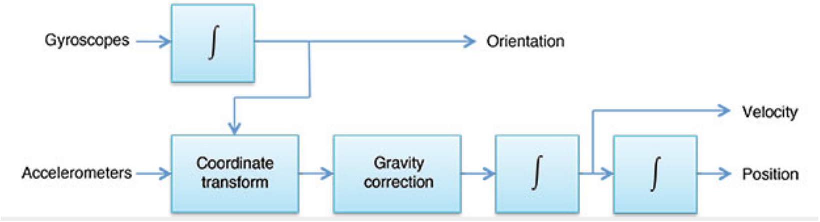


Fig. 10: Funtional flow diagram of a strapdown inertial navigation system with 3 axis accelerometer and 3 axis gyroscope. [125]

hardware, noise is present, and filtering this will introduce latency to the measurements. A negative characteristic for this type of tracking unit is that they are prone to drift over time, due to the nature of integrating measurements. Even a small bias error of the inertial unit measurement can result in large errors after just seconds of usage, and even larger errors for the double integration step in calculating position.

For the experiments described in this thesis, I use and describe inertial based tracking approaches and outcomes in the project *Koldinghus Augmented: Castle Chapel* (Chapter 9). There I explore the usefulness of inertial tracking for commercial applications related to handheld and static museum installations, and present methods to overcome problems of drift within the boundaries of these settings.

Vision-based tracking

Tracking via vision input, i.e. a live video stream, requires extracting a pose of the tracked object(s). This requires the software to make a correspondence between objects identified in the video stream and known locations in the real world. This is not a trivial task, even for a computer. For the purpose of this thesis and how optical tracking has been used in experiments herein, two methods for tracking will be explained in general terms, both based on a monocular camera: marker-based tracking (employing fiducials) and markerless tracking, also called feature based tracking, which uses only the available scene. I realize that there are more approaches to optical tracking, both with monocular camera, stereo-cameras and multiple camera setup in static tracking environments, but it is out of the scope of this thesis to explain these. For a thorough survey of vision based tracking using a monocular camera the reader should refer to Lepetit and Fua [84].

Figures 11 and 12 depicts examples of marker based and markerless vision-based tracking.

From the field of robotics research comes another formulation for this problem, called "visual servoing". A closed-loop system places the camera

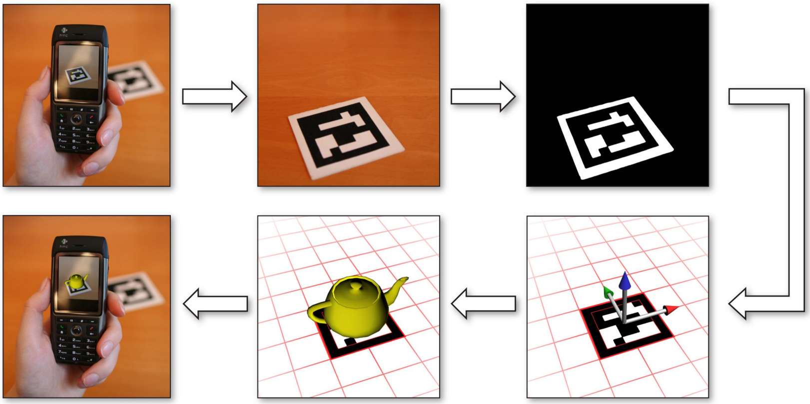


Fig. 11: Example of vision based tracking, based on a known fiducial marker in the scene [46].

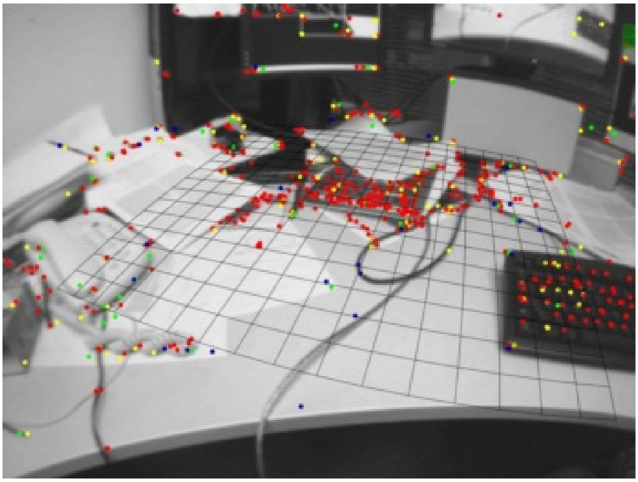


Fig. 12: Example of optical tracking in an unknown scene, using a real-time process of finding scene features [69].

relative to the visual target to match the tracking, inspired by work from Espiau et al [36]. A virtual version of this, "virtual visual servoing" has been used extensively in AR by numerous researchers [15, 101, 137]

Early examples of marker-based tracking, with passive fiducial markers in the scene, are Mellor [104] and Hoff et al [55]. Both authors used small circle markers for tracking the object in the scene, Mellor to investigate registration accuracy for medical x-ray and Hoff et al to display maintenance instructions using an HMD. Here fiducials are detected by first segmenting the image into black and white regions and small patches (noise) are eliminated (as exemplified as step 2 and 3 in Figure 11). Once a minimum of

Fig. 13: Example markers from Kato and Billinghurst's ARToolkit library. A widely used tracking library for AR development [67].

three fiducials have been found, the remainder can be matched from the known pattern and position of the three. The work has since been extended and authors have proposed other fiducial markers for tracking, such as color patches (Neumann and Cho [113]) and black bordered square ficudals (Kato and Billinghurst [67]). An example of markers from Kato and Billinghurst's ARToolkit library is illustrated in Figure 13. The fact that it soared in popularity since its initial release is due to the ease of use, as well as being cross platform and free to use. It has inspired hundreds (if not more) of AR applications in research, where the tracking technology has not been of primary concern.

Mohring et al [109] presented the first AR system running on a mobile device in 2004 using a color marker for tracking, and recently Wagner et al. [150] presented real-time tracking on a modern smartphone using a SIFT tracking approach based on [92, 136] for full 6-DoF tracking.

Branching off from tracking static objects, Pilet et al [121] presents a method for tracking and augmenting deformable objects in real-time, and following on that Moreno-Noguer et al [110] presented a method for recovering 3D shape of a stretchable object.

Early advances in object-based tracking by Harris and Stennett [50] tracks a known object with arbitrary motion from a monocular camera. The estimated pose of the object is found based on measurements of the edges in the video stream compared to the known edges of the object. The pose of the object is then updated to minimize the error between the rendered object and video edges. This base approach from Harris and Stennett, of tracking objects in unknown conditions using an edge-based approach, has been improved over the years by various researchers, who have proposed various techniques for optimizing performance of the tracker [32, 44, 90, 135]. As examples, Gennery [44] and Lowe [90] describes optimizing performance of the tracker through various techniques, such as edge detection on larger area of the input image and rejection of fit if error is to large (Lowe), and using weight for measurements (Gennery).

In recent years tracking systems which integrates point feature tracking into edge-based trackers have emerged [130, 144]. One advantage of feature points over edges is that often a feature can provide a descriptor, which allows correspondence of features over multiple frames, where correspondence of edges usually depends on proximity over multiple frames, rather than descriptors/appearance. The advantages of edges are that they are often invariant to aspect or illumination changes in the scene, where this might change the appearance of the feature point.

Early this century Davison [26] presented a real-time solution for monocular cameras that tracks small 2D image patches and cross-correlation, called simultaneous localisation and mapping (SLAM), which aside from a initial initialization runs entirely as a real-time process. This implementation differs from the previous described, as it does not rely on a known marker or object in the scene, but aims to track the viewpoint (camera) movement within the scene while simultaneously mapping the scene. An adaption of the SLAM approach has since run on mobile devices, as reported by Klein and Murray [70]. An overview of SLAM methods can be found in [33].

Despite the recent advances in markerless tracking for AR, it is even today rarely seen in commercial applications utilizing an AR interface. Typically, applications make use of fiducials in the scene to track and place objects relative to the fiducial, due to their robustness and low computational requirements [71].

In a perfect world, the tracking technology everyone want is one that is small, self-contained, 6 DoF, accurate, fast, robust and cheap. In reality, every tracking technology falls short on some or more of these characteristics. In the end, the best approach is to decide on the tracking technology that achieves the best results for the specific application.

For the experiments described in this thesis, I use vision based tracking technology in all projects except the *Castle Chapel* project, employing using various libraries based on fiducial marker tracking.

5.2 Occlusion

So, what is occlusion and why do we need it? Occlusion is when one object is fully or partially hidden by another object, as seen from the eyes of the observer. When you look around, you may notice that a majority of objects we see are either partially occluded by other objects, or partly occluding other objects. Since so many objects in the environment are occluded, the images they project to your eyes are incomplete. We have spent much of our lives learning the subtle cues telling us one object occludes another. As such, when you notice a partly occluded object, you do not consider this part of an object. Our mental and visual systems allow us to perceive such objects as whole, connected objects.

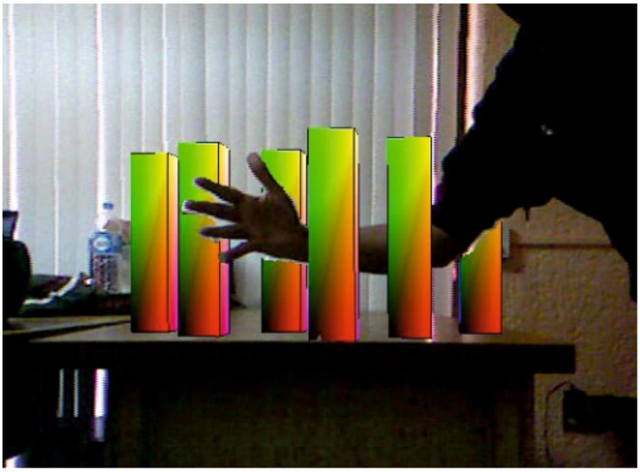


Fig. 14: Example of a real world object (an arm in this case) partially occluding a virtual object. This also demonstrates virtual foreground objects occluding the arm. From [80]

This does not automatically happen in a computer. When virtual content is superimposed onto an image or a video-stream of the real world, the virtual content is unaware of the real world just as the real world is unaware of the augmented content. This means that the application running must be designed to be aware of (or need to figure out) the distance to all objects, real as well as virtual, to present them at the correct distance in the final image. In the context of AR, anything that is rendered by a computer onto the video-stream will by default be in front of the real world, hiding the real objects in the world. In order for objects in the real world to occlude virtual objects, the virtual objects (or part of them) must not be drawn, as illustrated in Figure 14. Alternatively, the occluding part of the real image must be segmented out and drawn on top of the virtual object [40, 82].

Early research in this area use stereo camera setups to estimate a depth map for the real world, used for handling occlusion of virtual object [19, 154]. Kanbara [65] optimizes the process of generating a depth map by only considering the image region created by the bounding box of the virtual object rendered.

For monocular cameras, Berger [17] uses 2D contours observed over time in the scene to estimate depth, and does not generate a depth map.

More recently, Fisher et al [41] and Mulder [112] employ various methods of background subtraction to determine occluders and their position in the scene.

Using stereo vision and contour matching, Li et al. [134] presents a method for calculating depth of an object in the foreground of the video. Similarly, Leal-Melendrez et al. [80] presents a variant of foreground occlusion handling using a Kinect camera.

Using a time-of-flight camera, Dong et al. [31] proposes a method for occlusion handling using information obtained by the camera, and uses the data to suppress background illumination. The authors use a second camera for the actual RGB image, after an alignment step of the two images.

There are many areas which are currently unexplored with regards to occlusion on handheld mobile devices for AR, mainly due to the current constraints of having only monocular cameras in consumer hardware. For the purpose of this thesis, occlusion will not be considered any further, other than the attempts to work around any possible pitfalls during the experiments presented.

For the experiments described in this thesis, I only consider occlusions between virtual objects, and not between real and virtual objects, and design applications and experiments with this constraint in mind.

5.3 Illumination

There are two major problems with illumination worth discussing.

The first problem is related with the two previous topics. I.e. what do we actually know (or can figure out) about the real world surroundings. In relation to the topic of illumination, that is how are the current lighting conditions? We need a robust method of figuring out where the light is coming from and how bright it is. This can be as simple as figuring out where the lights are in the surroundings, or it can be as complex as figuring out where the lights are placed, but in addition to that how bright they are, and all reflections of lights on walls, objects, etc. This is a vastly more complex topic to dive into. Intuitively, this means that not only do we need to know the placement of the lights, but also the material of all objects in the scene AND their placement, to determine the final illumination in the point of the virtual object. Luckily, as this section will describe, we can get by not knowing quite all that. This is the area which this section will mostly focus on.

The second major problem is applying all this information and actually render the virtual objects with illumination consistent to that of the real world surroundings. Think about how all light travels from its source, in a straight line, towards an occluding object, hits the object, loses some of the original power and is bounced off the objects and into new directions. This happens multiple times until the rays finally hits the receivers eye(s) or in our case, the camera lens. To achieve realistic and consistent lighting we basically have to emulate this physical process on the computer and apply the result to the virtual object. In general, for high quality illumination, many algorithms simulate light transport models from optics and light transport in the real world. The rendering equation [58, 62] presents a solution by evaluating the light contribution from the point of view of the camera lens position. Ray tracing [152] and path tracing [62] are examples of implementations of this principle. I will not in this section focus too much on this part, as this in it self is a large subject. I will however introduce this area and suggest literature for the interested reader.

As the field of illumination research is large and complex, it is hard to suggest a single material, such as an article or book, which covers the topic of "illumination". Depending on the readers interest in subjects within illumination, possible references for a more detailed overview are [3, 18, 49, 64]. For in-depth information of realistic rendering of illumination, refer to [34].

Within acquisition of illumination information for AR, it is necessary to differentiate between indoor and outdoor setups. For outdoor environments, the lighting is mostly limited to the direct illumination, as well as the indirect illumination (light bounces) from both the sun and skylight. Unprepared indoor environments can be vastly more complex with not only lighting from the outside coming in, but in addition to that multiple lights of different colors and light intensity as part of the setting.

For AR applications, we cannot always assume to have information of the illumination in the current scene. Therefore, we have to consider to what extend we are interested in modeling (acquiring) the illumination. I consider four types: 1) Accept that illumination is not important and not assume anything about the light. 2) Pre-acquire illumination information and use that to generate lighting conditions for the scene. This resembles how Hollywood collects illumination information for scenes in movies. By collecting the information (by light probes or cameras) beforehand, realistic and fitting light can be created for the scene. It is also possible to use pre-acquired illumination information to model light in a real-time application assuming the light remains static for the application. 3) In a real-time process using a light probe, which from the literature typically is a reflective sphere or a fish-eye lens on a camera capturing the hemisphere (see Figure 15), it is possible to generate a map of the current illumination distribution in the scene and use this information to render augmented objects. 4) Using an adaptive approach which does not rely on a light probe, but instead attempts to estimate illumination conditions based on the currently available information readily available in the camera feed used for the augmentation.

Since we in this section is interested in illumination acquisition for AR, the first step, ignoring the illumination is not an option.

The second option, pre-acquisition of illumination information has typically been done using light probes. Using image information from a light probe allows us to generate a light model covering all possible directions (called an *environment map*, illustrated in Figure 16). Using this information to model how the illumination is distributed over the horizon and use this model to illuminate the virtual objects. Collecting multiple images using different exposures from one probe enables generation of a *high-dynamic range*

(a) Chrome sphere as light probe **(b)** Fish-eye camera as light probe

Fig. 15: Example light probes for illumination acquisition in AR. Both light probe examples can be used for both pre-acquisition and real-time capturing of scene illumination. (a) A reflective light probe, and (b) a fish-eye camera light probe. From [139]

(a) Light probe image example **(b)** Environment map generated from the light probe image

Fig. 16: Image capture of light probe converted to environment map in longitude-latitude format. (Courtesy of Claus B. Madsen)

image (HDRI or HDR image) and use this as illumination information of the scene as an environment map. An HDR image enables us to store more detailed illumination information compared to a single image, due to the nature of camera lenses, which are limited to capture information in 8-12 bit per color channel, and not enough to map the intensity difference light typically produces.

With camera equipment readily available today, generating such an environment map involves taking a series of images of the same scene or probe with different levels of exposure [29]. This process of using an environment map for lighting the virtual objects are also referred to as *Image-based lighting* (IBL) [28]. Multiple of studies have been conducted in obtaining illumination information usable for lighting 3D objects placed in the real world, requiring pre-acquisition steps, including the need for known geometry [5], stereo-camera setup [39], pre-capturing of environment maps, etc. Jensen [61] uses an IBL approach for outdoor environments during changing lighting conditions, in an approach which required both a 3D model and the HDR environment map in an offline step.

The third option describes a real-time process using a light probe, such as applied in studies by Kanbara [66] and Supan [138], where the authors achieve realistic lighting and shading using real-time constructed environment maps, using a reflective sphere in the scene. The main difference here compared to illumination estimation in a pre-capture step is that now the illumination distribution can change dynamically, and this change will reflect onto the virtual objects. This allows for consistently rendered dynamic virtual objects [117]. As a limitation for most methods using a live light probe, is that generating an HDR image becomes harder, as the camera will have to expose the scene differently each frame to acquire the images to combine the final HDR map. In a different study, however related to illumination acquisition, Sato [133] presents a method for recovery of illumination distribution from radiance distribution inside shadows cast by an object of known shape onto a surface of known shape and reflectance. However, a pre-requisite is known scene geometry, making it not fully adaptive, but does remove the need for a light probe or second camera.

For the final (fourth) option, an adaptive approach which does not rely on a light probe, multiple authors have presented novel solutions for estimating illumination of a scene. This might be the "holy grail" of the options. Estimating illumination properties of the scene with no offline step, no light probe or known scene geometry, relying only on the viewpoint camera (either monocular or stereo camera), with information available on the computer (e.g. location and time of day for outdoor usage, to estimate sun position). Currently no method allows for estimation of environment illumination to a degree which allows rendering of virtual objects in a fully photo-realistic manner. Presented methods are currently computationally expensive, and the authors present mainly incremental findings or solves part of the problem. Madsen and Nielsen presents a method for detecting shadows in a scene, and use this information to determine ratio of sky to sun irradiance [94]. Madsen and Lal presents a technique for estimating outdoor illumination conditions from pixel values of of dynamic shadows from stereo video sequences [93]. Using an RGB-D camera, Gruber et al. [47] presents a method using arbitrary geometry as light probe to estimate real-world diffuse illumination, in a real-time method running on modern hardware. Similarly, Jachnik et al. [60] presents a method using a single camera tracked using PTAM, capturing illumination information from general specular surfaces. From this data, they show it is possible to split diffuse and specular data, and use the specular component to generate an environment map.

At this point, it might be clear that acquisition and presentation of realistic illumination for AR is a computationally hard problem. This has led to research into how to implement and optimize computation of illumination rendering, such as Crassin et al. [23] who presents a voxel-based cone tracing for real-time indirect illumination on the GPU, which runs at interactive frame-rates on modern desktop PCs with advanced GPU hardware. For advanced light effects, Franke and Jung demonstrates GPU-based effects such as occlusion and self-shadowing [43]. Similarly Kan and Kaufmann presents a GPU implementation of photon mapping for real-time caustics, refractions, reflections for AR [63]. Recently, Meilland et al. [103] has presented a method for acquiring HDR light fields from several images with different exposures, using images acquired from an RGB-D camera also used for viewpoint rendering.

Immense resources are spent on realistic illumination computation, both research time as well as computational time on devices, and running modern algorithms in real-time is only just possible for not-quite-realistic illumination. But impressive nonetheless. Now the question is whether this is needed at all, or we could do with a more simple approach and it would be "good enough"? In this thesis I will not attempt to quantify the importance of illumination through experiments. However, part of the research does involve illumination and rendering of light on virtual objects. This is the research related to the *Koldinghus Augmented: Castle Chapel*, to be presented in Section 9. I motivate the need for realistic illumination (at least internally consistent illumination) as being about making things believable. Illumination is as much about lighting the scene in a realistic manner as it is about creating a *feel*, aiding the viewer in appreciating the scene. As such, I find illumination an inherent and important aspect of AR applications designed for superimposing virtual content onto a real setting in a realistic manner.

6 People and Augmented Reality

For any system to be used by people, the main objective of that system is (or should be) that users can understand and execute a set of tasks as defined within the system. Following this, the system should provide interactive feedback that indicates whether this task was carried out correctly [131]. The first example coming to mind is the utility of my word processor to indicate to me, whether or not it is responding to me hitting the keyboard, allowing me to ascertain whether the keys reported on the screen match what I typed. This usability of the system, and the ability of me to achieve objectives, is linked with the *user experience* of the system, which is defined as the feelings, assessments and satisfaction in relation to the system [51, 79]. As formulated by Law et al.:

> User experience is seen as *something* desirable, though what exactly *something* means remains open and debatable [79]

Historical mentioning of user experience (i.e. [4, 120]) in the research field had the primary aim of convincing the Human Computer Interaction (HCI) community to view issues besides the task related more seriously [51].

I will not venture too far into a discussion about how to define *user experience*, but merely use it as a distant goal in the experiments presented in this thesis. Systems intended for use by people should be designed with those users in mind, and altered to best fit feelings, assessments and satisfaction for the user in relation to the system. I hold this true for an AR system, as for any other system.

6.1 User centered development of systems

For a majority of the software and products developed, success can be determined by the people using the product. In order for a product to succeed, it must be practically usable. Such systems are often described as being *intuitive*, *enjoyable* or *easy to use*. While these terms in themselves are vague, they do point in the right direction.

Within the last 70 years, research within man and machine interaction has become a field of research within its own right, being presented under terms such as *Ergonomics*, *Human Factors*, *Human Computer Interaction* (HCI), etc.

Ergonomics and *Human Factors* are often used interchangeably. Both disciplines are concerned with user performance in the context of any system, not limited to computer systems. As computer became more widely accessible, an increasing number of researchers specialized in research related to the interaction of humans and computers, which became HCI [30]. *Ergonomics* and *Human Factors* will be explained further in the following section and related to perceptual issues in the thesis.

In the process of developing software for people to use (AR included), human users are key components of the design problem. A component that cannot and should not be ignored because humans are complex to address and design for. In many interactive systems, a majority of the design and development task is related to user interaction.

HCI presents a multidisciplinary field of research, involving elements of computer science, behavioral sciences, and design. It focuses on interfaces between humans and computers. From observing users interaction with computers, as well as research in designing technologies that let users interact with computers.

HCI, according to Dix et al.:

> Involves the design, implementation and evaluation of interactiven systems in the context of the user’s task and work [30]

Comparing ergonomics and HCI with related fields in the literature, Preece, Rogers and Sharp defined Interaction Design (ID) as:

> Designing interactive products to support the way people communicate and interact in their everyday lives [122]

The two quotes seems similar, and the authors of the latter quote goes on to argue that they increasingly see ID as an umbrella term, covering different aspects, including user-centered design, software design, and experience design amongst many others. The focus lies more in an experienced and eclectic way of doing things, and is very much concerned with practice, using a range of methods, techniques and frameworks. [122]

It is about creating a user experience that improve and augment the way people work and interact.

The above definitions have been presented in general, as presenting frameworks, taxonomies or models used within the fields are out of the scope of this thesis. However, relating this thesis to the above definitions, I embrace the general idea from the HCI community that a human user can be seen and analyzed as a human, biological, version of a computing machine [115]. In this model, information enters the mind from various sensing devices, is processed on different levels and a response is produced as a reaction to the presented stimuli. This is of course an oversimplification of real life, but is an example of simplifying human psychology to present theories and design laws, such as Fitts Law [42], the 7 +/- 2 memory limits of the human mind [107], and various visual perception findings related to colors and grouping.

Drawing from both presented fields, designers and developers need to know how design to the eventual users' tasks and how to translate that knowledge into an executable system.

There is far too much expertise here for one person (me) to fully invest in HCI or ID principles, and in practice people tend to take a strong stance on one side or another. In this thesis, I use guidelines and principles from the described fields, but take a strong technical and quantitative focus on both design and evaluations of studies. In my research, I tend to be a pragmatist rather than a theorist.

6.2 Some perceptual issues related to Augmented Reality

In the research literature, *Human Factors/Ergonomics* have been widely used by different groups, to express similar ideas concerning people, their characteristics and abilities, along with their interaction with the world. The Human Factors and Ergonomics Society present multiple definitions for Human Factors [37], from different societies, scientific literature and government agencies, amongst others.

For the purpose of the work presented in this thesis, the definition of Human Factors from NASA [147], considering the term as an umbrella term for several areas of research, provides a good general starting point for discussing what Human Factors are:

> Human factors is an umbrella term for several areas of research that include human performance, technology design, and human-computer interaction. [147]

Using this initial definition, the initial working idea is to aim the presented research towards evaluation of these Human Factors related to the people using the developed AR technology. This currently is in line with the presented definition for User Experience previously presented. In User Experience research, the working definition is the "feelings, assessments and satisfaction in relation to the system". Relating these topics, I see User Experience, in contrast to Human Factors, more as moving towards users expectations and their emotional relationship with a product. Continuing with this definition for Human Factors, it is more fitting to further delimit the working definition of Human Factors (within the context of this thesis) to:

> Human Factors is that field which is involved in conducting research regarding human psychological, social, physical, and biological characteristics, maintaining the information obtained from that research, and working to apply that information with respect to the design, operation, or use of products or systems for optimizing human performance, health, safety, and/or habitability. [48]

Narrowing this even further down, I limit the presented thesis to focus primarily on research relating to the human visual perception, the design and performance of the interactive systems used in the presented work.

Relating this to the previously covered technical aspects needed for the virtual objects to "fit in" with the real world; tracking, occlusion and illumination. These are all relevant aspects to consider when developing AR applications. A natural extension of this, and heavily researched, is that of realistic image synthesis. Within computer graphics, realism is often a primary goal, and we strive to create models which are indistinguishable from an actual scene. Since AR is basically collapsing the real and the virtual into a combined and shared space, it is natural that developers should at least consider the interplay between the real elements, virtual elements, and the user experience.

Within the research field, there is a general interest in the effect of the human visual system (HVS), and how to make use of this knowledge for visual perception. McNamara [102] presents work on defining metrics for measuring visual perception of images based on the HVS, and presents a perception driven rendering approach. Similarly interested in visual perception, Rademacher et al. [124] argues that how observers distinguish between photographic and computer generated images is not well understood. Gaining an understanding of how users choose would make the goal of realistic rendering of images easier. They present an experimental method for measuring the perception of visual realism in images, using an approach from experimental psychology, asking participants for a binomial *realism response* of presented images. In another research direction, though still related to visual perception, Veas et al. [146] uses modulated images to direct user attention to certain parts of images. Using visual saliency modulation, they investigate the degree of modulation needed to attract attention from users. More recently, Knecht et al. [72] has presented a framework for perceptual studies related to photo-realistic AR, based on separation of different visual cues.

Related to visual perception are the issues in computational performance of the AR system, such as the effect of tracking or occlusion issues. Specifically the inherent issues from tracking, such as latency, drift, depth cues, etc. Multiple studies have evaluated and analyzed tracking performance from a technical perspective [12, 45, 54, 57, 89, 99]. As an example, Gilson et al. [45] presents a quantitative analysis of tracking systems and found that tracking performance deteriorates with increase in speed of the object. Also latency of a system has shown to be a problem, with the exception where all tracked objects remain stationary for the duration of the usage. The size of the problem depends on the nature and size of the latency, as explained by multiple authors [2, 35, 99, 141]. A small and constant latency might not be a problem, whereas any non-constant latency or a large latency poses a problem in any setup. Mania et al. [99] presents the results of an experiment of users experiences and perception of latency with varying degrees of scene complexity, and finds that the just noticeable difference (JND) is 15 ms or less. This is in line with earlier investigations, such as Adelstein et al. [2], who measured 8-17 ms and Ellis et al. [35], who found an average of 11.6 ms to be the JND.

These visual issues as well as tracking based issues are all related to human factors, and something developers must be aware of in designing any application relying on AR technology. As argued by Davis [25], there need to be a perceived usefulness and perceived ease of use of the system, in order for people to have an interest in using the system. People will notice if the technology does not behave in line with their expectations, just as they will notice if anything seems out of place. It is about making elements of the experience internally consistent, and believable within this usage context, as well as provide a usability of the system, allowing people to carry out their tasks without the technology hindering them.

6.3 Augmented Reality "in the wild"

In this context, *the wild* is the notion for out of the laboratory, in the real world where we have little to no control over the otherwise very controlled elements of a testing scenario. In this section, I focus on AR in museum contexts, and general Cultural Heritage (CH) settings, as it is a controlled setting for "out-of-the-lab" experiments.

In-situ studies, in the wild, broadly differ from controlled settings in that participants within studies in the wild differ markedly when following and performing instructions, compared to what they do in the lab. Furthermore, researchers are also increasingly developing and experimenting with new technology that changes and disrupts practices instead of developing solutions fitting existing practices [1], with a key concern that people are reacting to, changing and integrating this technology into their lives [126].

In one example of how in-the-wild evaluations are used in research, Bai and Blackwell [10] review usability evaluation papers published in ISMAR between 1993-2007, where they found 71 papers concerned with various categories within usability research. Most papers in the review did not mention using field studies for evaluation. This is despite important factors such as people having increased cognitive load in the field [114, 127]. Evaluating AR in the field is still rather uncommon. Most studies in the field mainly evaluate usability or User Experience, as the success of an AR application to a certain extend depends on the user satisfaction [73]. As a result, User Experience of an AR application must be part of the design strategy [118].

Multiple authors have documented positive effects of evaluating the novelty of AR technology [81, 111, 157], using the technology acceptance model [52, 116], field studies [53, 142, 143], or within human factors related to AR [8, 59, 132, 146, 149]. However, a limitation is that a majority of these studies are of limited duration in the field, or limited to a confined setting, evaluating a specific goal. This, in it self is not a bad thing at all. Limited settings or goals are needed to minimize confounding variables and random effects, enabling actual evaluation of specific goals of the system (or the user). However, while such studies might say a lot about a small area (each being a piece of a larger puzzle you could say), together they provide us with an indication for how people might perceive and use the technology.

In relation to this thesis, doing research in the wild, and in Cultural Heritage institutions such as museums, means that I will most likely get a more unpredictable side of people using the presented technology. They might be interrupted, get distracted, take breaks, etc. just as they would in their everyday lives. This is to be expected when conducting the research [126]. Therefore, it will be difficult to say with any certainty, to what degree their responses are part of the system, and what are part of the environment. However, this will hopefully allow for certain design guidelines to be constructed, related to the usage in a given environment.

There is a vast body of work within the field of AR in museum contexts, ranging from museum guides [24, 108, 142], to building virtual and augmented installations and exhibitions for the museum [27, 74, 105, 145, 155, 156]. Novel technology in museums is an active and vast research area from both a technology point of view, to test new technology in a semi-controlled setting, but also from a CH point of view; considering how to best present information and knowledge to a user. Van Eck [145] considers how to augment paintings in the Van Gogh museum and describes the different information overlays, while [27, 156] present research on information presentation in a way which are normally not readily available to users, such as the universe (S.O.L.A.R. System) and the Interactive Antartica. Woods et al. describes how they have received positive user feedback while using a range of control tools suggesting sufficient freedom to create many interaction scenarios [156]. Wojciechowski et al., describes a system that allows a museum to easily develop and display their digitized content on web or using AR or VR techniques [155]. Another example is Koleva et al. who explore how to best support the creation of hybrid exhibits, which merges the physical and digital elements in a museum environment [74]. A lot of interesting work within museums installations based on Augmented Reality technology has been carried out.

Morrison et al. [111] describes the usage of a map AR application and how this facilitates interaction in a user study. It describes the usage of the application from a user perspective, and describes how mobile AR features need to be developed with a view to the physical environment they will be used within. Lee et al. [81] describes how the user experience differed between participants trying their applicaiton with and without AR. They suggest developing the application around the content, which is the main focus of the user experience. Also Irshad and Rambli [59] describes the importance of evaluating the user experience of Mobile AR technology, by evaluating a mobile AR application through a mix of quantitative and qualitative approaches. They report mobile AR as a novel experience for the user, and that related to the findings, end users want a very efficient and robust experience to be delivered on their mobile devices.

Historically Cultural Heritage institutions have been interesting for researchers for bringing research applications into the real world in order to further document the effects on the user. This provides a confined setting for experimentation, to minimize some of the random effects from evaluations of AR applications. Multiple studies have documented the user of AR technology for historical sites, such as a situated documentaries about historical events [56], reconstructions of cultural heritage [148], geo-physical Situated Simulations [88], and urban city touring [13].

For research in the user acceptance of AR within CH, Haugsvedt et al. [52]

argues that perceived enjoyment and perceived usefulness are important for users intention to use the AR application presented in their study. This is in line with Davis' [25] argument for the acceptance of technology. Consensus in the field is that users appear interested in the novel experience of AR technology. Due to generally short lived studies, it is not known whether this interest is an effect due to presented material or a "wow" effect of AR being novel to a majority of people. Relating this body of research to this thesis, the logical next step is the importance of investigating and understanding how people are using the technology for an extended duration, and not just the short term usage. The experiments presented in this thesis do all, in one form or another, anonymously log all user interactions with the systems in order to describe how the systems have actually been used, and not as users self-reporting their experiences. I use this to get insights into how the applications have been used. It demonstrates an important tool to use in addition to observations and user feedback.

References

[1] "Introduction to the special issue of “the turn to the wild”," *ACM Trans. Comput.-Hum. Interact.*, vol. 20, no. 3, pp. 13:1–13:4, Jul. 2013. [Online]. Available: http://doi.acm.org/10.1145/2491500.2491501

[2] B. D. Adelstein, T. G. Lee, and S. R. Ellis, "Head Tracking Latency in Virtual Environments: Psychophysics and a Model," in *Human Factors and Ergonomics Society Annual Meeting Proceedings (2003)*, 2003, pp. 2083–2087.

[3] T. Akenine-Möller, E. Haines, and N. Hoffman, *Real-Time Rendering 3rd Edition*. Natick, MA, USA: A. K. Peters, Ltd., 2008.

[4] L. Alben, "Quality of experience: defining the criteria for effective interaction design," *The ACM interactions*, vol. 3, no. 3, pp. 11–15, 1996.

[5] M. S. Andersen, T. Jensen, and C. B. Madsen, "Estimation of dynamic light changes in outdoor scenes without the use of calibration objects," in *Proceedings - International Conference on Pattern Recognition*, vol. 4, 2006, pp. 91–94.

[6] R. Azuma, "Tracking requirements for augmented reality," *Communications of the ACM*, vol. 36, no. 7, pp. 50–51, 1993.

[7] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *IEEE Computer Graphics and Applications*, vol. 21, no. 6, pp. 34–47, 2001.

[8] R. Azuma and C. Furmanski, "Evaluating label placement for augmented reality view management," in *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.* IEEE Comput. Soc, 2006, pp. 66–75.

[9] R. T. Azuma, "A survey of augmented reality," *Presence*, vol. 4, no. August, pp. 355–385, 1997.

[10] Z. Bai and A. F. Blackwell, “Analytic review of usability evaluation in ismar,” *Interact. Comput.*, vol. 24, no. 6, pp. 450–460, Nov. 2012. [Online]. Available: http://dx.doi.org/10.1016/j.intcom.2012.07.004

[11] M. Bajura, H. Fuchs, and R. Ohbuchi, “Merging virtual objects with the real world,” *ACM SIGGRAPH Computer Graphics*, vol. 26, no. 2, pp. 203–210, 1992.

[12] M. Bajura and U. Neumann, “Dynamic registration correction in video-based augmented reality systems,” *IEEE Computer Graphics and Applications*, vol. 15, no. 5, pp. 52–60, 1995.

[13] P. J. Bartie and W. a. Mackaness, “Development of a speech-based augmented reality system to support exploration of cityscape,” *Transactions in GIS*, vol. 10, no. 1, pp. 63–86, 2006.

[14] M. Bauer, “Tracking Errors in Augmented Reality,” Ph.D. dissertation, Technische Universität München, 2007.

[15] R. Behringer, J. P. J. Park, and V. Sundareswaran, “Model-based visual tracking for outdoor augmented reality applications,” *Proceedings. International Symposium on Mixed and Augmented Reality*, pp. 1–3, 2002.

[16] S. Benford, C. Greenhalgh, G. Reynard, C. Brown, and B. Koleva, “Understanding and constructing shared spaces with mixed-reality boundaries,” *ACM Transactions on Computer-Human Interaction*, vol. 5, no. 3, pp. 185–223, 1998.

[17] M.-O. Berger, “Resolving occlusion in augmented reality: a contour based approach without 3D reconstruction,” *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 91–96, 1997.

[18] J. Birn, *Digital Lighting and Rendering, 3rd Edition*. New Riders, 2014.

[19] D. E. Breen, R. T. Whitaker, E. Rose, and M. Tuceryan, “Interactive Occlusion and Automatic Object Placement for Augmented Reality,” *Computer Graphics Forum*, vol. 15, no. 3, pp. 11–22, 1996.

[20] J. Caarls, P. Jonker, and S. Persa, “Sensor Fusion for Augmented Reality,” in *Ambient Intelligence*, E. Aarts, R. Collier, E. Loenen, and B. Ruyter, Eds. Springer Berlin Heidelberg, Jul. 2003, no. 1, pp. 160–176.

[21] T. P. Caudell and D. W. Mizell, “Augmented reality: an application of heads-up display technology to manual manufacturing processes,” in *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences*. Ieee, 1992, pp. 659–669 vol.2.

[22] Chris Verplaetse. (1995) Strapdown systems. (online) http://xenia.media.mit.edu/ verp/projects/smartpen/node8.html.

[23] C. Crassin, F. Neyret, M. Sainz, S. Green, and E. Eisemann, “Interactive Indirect Illumination Using Voxel Cone Tracing,” *Computer Graphics Forum (Proc. of Pacific Graphics 2011)*, vol. 30, no. 7, 2011.

[24] A. Damala, P. Cubaud, A. Bationo, P. Houlier, and I. Marchal, “Bridging the gap between the digital and the physical,” in *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts - DIMEA ’08*. New York, New York, USA: ACM Press, 2008, p. 120.

[25] F. D. Davis, "Perceived Usefulness , Perceived Ease of Use , and User Acceptance of lnformation Technology," *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989.

[26] A. Davison, "Real-time simultaneous localisation and mapping with a single camera," *Proceedings Ninth IEEE International Conference on Computer Vision*, vol. 2, pp. 1403–1410, 2003.

[27] C. de Bérigny Wall and X. Wang, "Interactive Antarctica," in *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts - DIMEA '08*. New York, New York, USA: ACM Press, 2008, pp. 319–325.

[28] P. E. Debevec, "Rendering synthetic objects into real scenes," in *ACM SIGGRAPH 2008 classes on SIGGRAPH 08*, 2008, p. 1.

[29] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques - SIGGRAPH '97*. New York, New York, USA: ACM Press, 1997, pp. 369–378.

[30] A. Dix, J. E. Finlay, G. D. Abowd, and R. Beale, *Human-Computer Interaction*, 3rd ed. Harlow, England: Pearson Education Limited, 2004.

[31] S. Dong, C. Feng, and V. R. Kamat, "Occlusion Handling Method for Ubiquitous Augmented Reality Using Reality Capture Technology and GLSL," *Computing in Civil Engineering*, pp. 494–503, 2011.

[32] R. Drummond T. Cipolla, "Real-Time Visual Tracking of Complex Structures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 932–946, 2000.

[33] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part I," *IEEE Robotics and Automation Magazine*, vol. 13, no. 2, pp. 99–108, 2006.

[34] P. Dutré, K. Bala, and P. Bekaert, *Advanced Global Illumination*, first edition ed. A K Peters, 2003.

[35] S. R. Ellis, K. Mania, B. D. Adelstein, and M. I. Hill, "Generalizeability of Latency Detection in a Variety of Virtual Environments," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 48, no. 23, pp. 2632–2636, Sep. 2004.

[36] B. Espiau, F. Chaumette, and P. Rives, "A new approach to visual servoing in robotics," *IEEE Transactions on Robotics and Automation*, vol. 8, no. 3, pp. 313–326, 1992.

[37] H. Factors and E. Society, "(ONLINE) definitions of human factors and ergonomics," http://www.hfes.org/Web/EducationalResources/HFEdefinitionsmain.html, 2005.

[38] S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster, "A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment," in *Proc ISWC '97 (Int. Symp. on Wearable Computing)*, vol. 97, 1997, pp. 74–81.

[39] Y. Feng, "Estimation of light source environment for illumination consistency of augmented reality," *Proceedings - 1st International Congress on Image and Signal Processing, CISP 2008*, vol. 3, pp. 771–775, 2008.

[40] J. Fischer, D. Bartz, and W. Straß er, "Occlusion handling for medical augmented reality using a volumetric phantom model," in *Proceedings of the ACM symposium on Virtual reality software and technology - VRST '04*, 2004, pp. 174–177.

[41] J. Fischer, H. Regenbrecht, and G. Baratoff, "Detecting dynamic occlusion in front of static backgrounds for AR scenes," *Proceedings of the workshop on Virtual environments 2003 - EGVE '03*, pp. 153–161, 2003.

[42] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *Journal of Experimental PSychology*, vol. 74, pp. 381–391, 1954.

[43] T. Franke and Y. Jung, "Real-time mixed reality with GPU techniques," *GRAPP 2008 - Proceedings of the 3rd International Conference on Computer Graphics Theory and Applications*, pp. 249–252, 2008.

[44] D. B. Gennery, "Visual tracking of known three-dimensional objects," *International Journal of Computer Vision*, vol. 7, no. 3, pp. 243–270, 1992.

[45] S. J. Gilson, A. W. Fitzgibbon, and A. Glennerster, "Quantitative analysis of accuracy of an inertial/acoustic 6DOF tracking system in motion." *Journal of neuroscience methods*, vol. 154, no. 1-2, pp. 175–82, Jun. 2006.

[46] Graz Technical University. (2014) Press illustration. from Christian Doppler Laboratory for Handheld Augmented Reality.

[47] L. Gruber, T. Richter-Trummer, and D. Schmalstieg, "Real-time photometric registration from arbitrary geometry," in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Nov. 2012, pp. 119–128.

[48] J. H. Stramler Jr., *The Dictionary for Human Factors/Ergonomics*. CRC Press, 1993.

[49] J. Happa, M. Mudge, K. Debattista, A. Artusi, A. Gonçalves, and A. Chalmers, "Illuminating the past: state of the art," *Virtual Reality*, vol. 14, no. 3, pp. 155–182, Feb. 2010.

[50] C. Harris and C. Stennett, "RAPID - a video rate object tracker," *Procedings of the British Machine Vision Conference 1990*, pp. 15.1–15.6, 1990.

[51] M. Hassenzahl and N. Tractinsky, "User experience - a research agenda," *Behaviour & Information Technology*, vol. 25, no. 2, pp. 91–97, 2006.

[52] A.-C. Haugstvedt and J. Krogstie, "Mobile augmented reality for cultural heritage: A technology acceptance study," in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Nov. 2012, pp. 247–255.

[53] S. Henderson and S. Feiner, "Exploring the benefits of augmented reality documentation for maintenance and repair." *IEEE transactions on visualization and computer graphics*, vol. 17, no. 10, pp. 1355–68, Oct. 2011.

[54] W. Hoff and T. Vincent, "Analysis of head pose accuracy in augmented reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 6, no. 4, pp. 319–334, 2000.

[55] W. A. Hoff, K. Nguyen, and T. Lyon, "Computer vision-based registration techniques for augmented reality," in *Proceedings of Intelligent Robots and Computer Vision XV, SPIE*, vol. 2904, 1996, pp. 1–10.

[56] T. Höllerer, S. Feiner, and J. Pavlik, "Situated documentaries: embedding multimedia presentations in the real world," in *Digest of Papers. Third International Symposium on Wearable Computers*. IEEE Comput. Soc, 1999, pp. 79–86.

[57] R. L. Holloway, "Registration error analysis for augmented reality," *Presence: Teleoperators and Virtual Environments*, pp. 1–25, 1997.

[58] D. S. Immel, M. F. Cohen, and D. P. Greenberg, "A radiosity method for non-diffuse environments," *ACM SIGGRAPH Computer Graphics*, vol. 20, no. 4, pp. 133–142, 1986.

[59] S. Irshad and D. R. A. Rambli, "User Experience Evaluation of Mobile AR services," in *Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia - MoMM '14*, 2014, pp. 119–126.

[60] J. Jachnik, R. a. Newcombe, and A. J. Davison, "Real-time surface light-field capture for augmentation of planar specular surfaces," in *ISMAR 2012 - 11th IEEE International Symposium on Mixed and Augmented Reality 2012, Science and Technology Papers*, 2012, pp. 91–97.

[61] T. Jensen, M. S. Andersen, and C. B. Madsen, "Real-Time Image-Based Lighting for Outdoor Augmented Reality under Dynamically Changing Illumination Conditions," in *Proceedings: International Conference on Graphics Theory and Applications, Setubal, Portugal*, 2006, pp. 364–371.

[62] J. T. Kajiya, "The Rendering Equation," *ACM SIGGRAPH Computer Graphics*, vol. 20, no. 4, pp. 143–150, 1986.

[63] P. Kan and H. Kaufmann, "High-quality reflections, refractions, and caustics in Augmented Reality and their contribution to visual coherence," in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Nov. 2012, pp. 99–108.

[64] P. Kán, "High-Quality Real-Time Global Illumination in Augmented Reality," Ph.D. dissertation, T.U. Wien, 2014.

[65] M. Kanbara, T. Okuma, H. Takemura, and N. Yokoya, "A stereoscopic video see-through augmented reality system based on real-time vision-based registration," *Proceedings IEEE Virtual Reality 2000 (Cat. No.00CB37048)*, 2000.

[66] M. Kanbara and N. Yokoya, "Real-time estimation of light source environment for photorealistic augmented reality," *Proceedings - International Conference on Pattern Recognition*, vol. 2, pp. 911–914, 2004.

[67] H. Kato and M. Billinghurst, "Marker tracking and HMD calibration for a video-based augmented reality conferencing system," in *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*. IEEE Comput. Soc, 1999, pp. 85–94.

[68] G. Klein and T. Drummond, "Tightly integrated sensor fusion for robust visual tracking," in *Image and Vision Computing*, vol. 22, no. 10 SPEC. ISS., 2004, pp. 769–776.

[69] G. Klein and D. Murray, “Parallel Tracking and Mapping for Small AR Workspaces,” in *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE, Nov. 2007, pp. 1–10.

[70] ——, “Parallel Tracking and Mapping on a camera phone,” in *2009 8th IEEE International Symposium on Mixed and Augmented Reality*. IEEE, Oct. 2009, pp. 83–86.

[71] M. Klopschitz, G. Schall, D. Schmalstieg, and G. Reitmayr, “Visual tracking for Augmented Reality,” *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, no. September, pp. 1–4, 2010.

[72] M. Knecht, A. Dünser, C. Traxler, M. Wimmer, and R. Grasset, “A Framework for Perceptual Studies in Photorealistic Augmented Reality,” in *Proceedings of the 3rd IEEE VR 2011 Workshop on Perceptual Illusions in Virtual Environments*, F. Steinicke and P. Willemsen, Eds., 2011, pp. 27–32.

[73] M. Koelle, P. Lindemann, T. Stockinger, and M. Kranz, “Human-computer interaction with augmented reality,” Embedded Interactive Systems Laboratory, Passau University, Tech. Rep., 2014.

[74] B. Koleva, S. R. Egglestone, H. Schnädelbach, K. Glover, C. Greenhalgh, T. Rodden, and M. Dade-Robertson, “Supporting the creation of hybrid museum experiences,” in *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*. New York, New York, USA: ACM Press, 2009, pp. 1973–1982.

[75] D. W. F. V. Krevelen and R. Poelman, “A Survey of Augmented Reality Technologies, Applications and Limitations,” *The International Journal of Virtual Reality*, vol. 9, no. 2, pp. 1–20, 2010.

[76] C. H. Larsen, D. S. Lauritsen, J. J. Larsen, M. Pilgaard, and J. B. Madsen, “Differences in human audio localization performance between a hrtf- and a non-hrtf audio system,” in *Proceedings of the 8th Audio Mostly Conference*, ser. AM ’13. New York, NY, USA: ACM, 2013, pp. 5:1–5:8. [Online]. Available: http://doi.acm.org/10.1145/2544114.2544118

[77] C. H. Larsen, D. S. Lauritsen, J. J. Larsen, M. Pilgaard, J. B. Madsen, and R. Stenholt, “Aurally aided visual search performance comparing virtual audio systems,” in *Audio Engineering Society Convention 137*, Oct 2014. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=17473

[78] C. Larsen, D. Lauritsen, J. Larsen, M. Pilgaard, J. Madsen, and R. Stenholt, *Aurally Aided Visual Search Performance Comparing Virtual Audio Systems*. SFB/TR 8, 2014, vol. 8684, pp. 70–73.

[79] E. Law, V. Roto, A. P. Vermeeren, J. Kort, and M. Hassenzahl, “Towards a ahared definition of User Experience,” in *CHI 2008 Proceedings - Special Interest | Groups*, 2008, pp. 2395–2398.

[80] J. A. Leal-Meléndrez, L. Altamirano-Robles, and J. a. Gonzalez, “Occlusion handling in video-based augmented reality using the kinect sensor for indoor registration,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8259 LNCS, no. PART 2, 2013, pp. 447–454.

[81] G. A. Lee, A. Dünser, S. Kim, and M. Billinghurst, "CityViewAR: A mobile outdoor AR application for city visualization," in *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*. IEEE, Nov. 2012, pp. 57–64.

[82] V. Lepetit and M.-O. Berger, "Handling occlusion in augmented reality systems: a semi-automatic method," *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, pp. 137–146, 2000.

[83] V. Lepetit. (2014) Augmented reality (710.008). University Lecture in Augmented Reality, Graz Technical University.

[84] V. Lepetit and P. Fua, "Monocular Model-Based 3D Tracking of Rigid Objects," *Foundations and Trends® in Computer Graphics and Vision*, vol. 1, no. 1, pp. 1–89, 2005.

[85] G. Liestøl, "Situated Simulations Between Virtual Reality andMobile Augmented Reality: Designing a Narrative Space," in *Handbook of Augmented Reality*, B. Furht, Ed. New York, NY: Springer New York, ch. 14, pp. 309–319.

[86] ——, "Augmented reality and digital genre design — Situated simulations on the iPhone," in *2009 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media and Humanities*. IEEE, Oct. 2009, pp. 29–34.

[87] G. Liestøl and T. Rasmussen, "In the Presence of the Past. A field trial evaluation of a situated simulation design reconstructing a viking burial scene," in *Media Inspirations for Learning. Proceedings of EDEN 2010*, A. Szücs and A. W. Tait, Eds., Budapest, 2010.

[88] G. Liestøl, T. Rasmussen, and T. Stenarson, "Mobile innovation: designing and evaluating situated simulations," *Digital Creativity*, vol. 22, no. 3, pp. 174–186, Sep. 2011.

[89] M. A. Livingston and Z. Ai, "The effect of registration error on tracking distant augmented objects," *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 77–86, Sep. 2008.

[90] D. G. Lowe, "Robust model-based motion tracking through the integration of search and estimation," *International Journal of Computer Vision*, vol. 8, no. 2, pp. 113–122, 1992.

[91] ——, "Object recognition from local scale-invariant features," *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, 1999.

[92] ——, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[93] C. B. Madsen, J. B. Madsen, and A. Morrison, "Aspects of what makes or breaks a museum AR experience," in *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*, no. November. IEEE, Nov. 2012, pp. 91–92.

[94] C. B. Madsen and M. Nielsen, "Towards Probe-Less Augmented Reality - A Position Paper," *International Conference on Graphics Theory and Applications (GRAPP)*, pp. 255–261, 2008.

[95] J. B. Madsen and C. B. Madsen, "An interactive visualization of the past using a situated simulation approach," in *2013 Digital Heritage International Congress (DigitalHeritage)*. Ieee, Oct. 2013, pp. 307–314.

[96] ——, "Handheld visual representation of a castle chapel ruin," *J. Comput. Cult. Herit.*, vol. 9, no. 1, pp. 6:1–6:18, Nov. 2015. [Online]. Available: http://doi.acm.org/10.1145/2822899

[97] J. B. Madsen and R. Stenholt, "How wrong can you be: Perception of static orientation errors in mixed reality," in *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, Mar. 2014, pp. 83–90.

[98] J. B. Madsen, M. Tatzgern, C. B. Madsen, D. Schmalstieg, and D. Kalkofen, "Temporal coherence strategies for augmented reality labeling," *IEEE Transactions on Visualization and Computer Graphics (Accepted)*, 2016.

[99] K. Mania, B. D. Adelstein, S. R. Ellis, and M. I. Hill, "Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity," in *Proceedings of the 1st Symposium on Applied perception in graphics and visualization - APGV '04*, no. study 1. New York, New York, USA: ACM Press, 2004, pp. 39–47.

[100] S. Mann, ""Smart clothing"," in *Proceedings of the fourth ACM international conference on Multimedia - MULTIMEDIA '96*, 1996, pp. 163–174.

[101] E. Marchand and F. Chaumette, "Virtual visual servoing: A framework for real-time augmented reality," in *Computer Graphics Forum*, vol. 21, no. 3, 2002, pp. 289–297.

[102] A. McNamara, "Visual Perception in Realistic Image Synthesis," *Computer Graphics Forum*, vol. 20, no. 4, pp. 211–224, Dec. 2001.

[103] M. Meilland, C. Barat, and A. Comport, "3D High Dynamic Range dense visual SLAM and its application to real-time object re-lighting," in *2013 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2013*, 2013, pp. 143–152.

[104] J. Mellor, "Enhanced Reality Visualization in a Surgical Environment," Massachusetts Institute of Technology, Tech. Rep. 1544, 1995.

[105] D. Michael, N. Pelekanos, I. Chrysanthou, and P. Zaharias, "Comparative Study of Interactive Systems in a Museum," in *Proceedings of the Third international conference on Digital heritage*. Springer-Verlag Berlin, 2010, pp. 250–261.

[106] P. Milgram and F. Kishino, "A Taxonomy of Mixed Reality Visual Displays," *IEICE Transactions on Information Systems*, vol. E77-D, no. 12, 1994.

[107] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *The Psychological Review*, vol. 63, no. 2, pp. 81–97, March 1956. [Online]. Available: http://www.musanim.com/miller1956/

[108] T. Miyashita, P. Meier, T. Tachikawa, S. Orlic, T. Eble, V. Scholz, A. Gapel, O. Gerl, S. Arnaudov, and S. Lieberknecht, "An Augmented Reality museum guide," *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 103–106, Sep. 2008.

[109] M. Mohring, C. Lessig, and O. Bimber, "Video see-through AR on consumer cell-phones," *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 3–4, 2004.

[110] F. Moreno-Noguer, M. Salzmann, V. Lepetit, and P. Fua, "Capturing 3D stretchable surfaces from single images in closed form," in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, 2009, pp. 1842–1849.

[111] A. Morrison, A. Oulasvirta, P. Peltonen, S. Lemmela, G. Jacucci, G. Reitmayr, J. Näsänen, and A. Juustila, "Like Bees Around the Hive: A Comparative Study of a Mobile Augmented Reality Map," in *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*. New York, New York, USA: ACM Press, 2009, pp. 1889–1898.

[112] J. Mulder, "Realistic occlusion effects in mirror-based co-located augmented reality systems," *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, vol. 2005, pp. 203–209, 2005.

[113] U. Neumann and Y. Cho, "A self-tracking augmented reality system," *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, pp. 1–7, 1996.

[114] C. M. Nielsen, M. Overgaard, M. B. Pedersen, J. Stage, and S. Stenild, "It's worth the hassle!: The added value of evaluating the usability of mobile systems in the field," in *Proceedings of the 4th Nordic Conference on Human-computer Interaction: Changing Roles*, ser. NordiCHI '06. New York, NY, USA: ACM, 2006, pp. 272–280. [Online]. Available: http://doi.acm.org/10.1145/1182475.1182504

[115] S. Nilsson, *Augmentation in the Wild User Centered Development and Evaluation of Augmented Reality Applications*. Linköping: Linköping University Electronic Press, 2010.

[116] S. Nilsson and B. Johansson, "Acceptance of augmented reality instructions in a real work setting," *Proceeding of the twenty-sixth annual CHI conference extended abstracts on Human factors in computing systems - CHI '08*, pp. 20–25, 2008.

[117] D. Nowrouzezahrai, S. Geiger, K. Mitchell, R. Sumner, W. Jarosz, and M. Gross, "Light factorization for mixed-frequency shadows in augmented reality," in *2011 10th IEEE International Symposium on Mixed and Augmented Reality*. IEEE, Oct. 2011, pp. 173–179.

[118] T. Olsson and M. Salo, "Narratives of satisfying and unsatisfying experiences of current mobile augmented reality applications," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '12. New York, NY, USA: ACM, 2012, pp. 2779–2788. [Online]. Available: http://doi.acm.org/10.1145/2207676.2208677

[119] T. Oskiper, S. Samarasekera, and R. Kumar, "Multi-sensor navigation algorithm using monocular camera, IMU and GPS for large scale augmented reality," in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Nov. 2012, pp. 71–80.

[120] K. Overbeeke, T. Djajadiningrat, C. Hummels, and S. Wensveen, “Beauty In Usability: Forget About Ease Of Use!” in *In: W. Green and P. Jordan (Eds.), Pleasure with Products, Beyond Usability*. Taylor & Francis, 2000, pp. 9—-18.

[121] J. Pilet, V. Lepetit, and P. Fua, “Augmenting deformable objects in real-time,” in *Proceedings - Fourth IEEE and ACM International Symposium on Symposium on Mixed and Augmented Reality, ISMAR 2005*, vol. 2005, 2005, pp. 134–137.

[122] J. Preece, Y. Rogers, and H. Sharp, *Interaction Design*, 1st ed. New York, NY, USA: John Wiley & Sons, Inc., 2002.

[123] D. Prochazka and T. Koubek, “Augmented Reality Implementation Methods in Mainstream Applications,” *Acta Universitatis agriculturae et silviculturae Mendelianae Brunensis*, vol. LIX, no. 4, pp. 1–22, Jun. 2011.

[124] P. Rademacher, J. Lengyel, E. Cutrell, and T. Whitted, “Measuring the Perception of Visual Realism in Images,” in *Proceeding Proceedings of the 12th Eurographics Workshop on Rendering Techniques*, 2001, pp. 235–248.

[125] Ramsey Faragher and Robert Harle. (2013) Innovation: Getting closer to everywhere. (online) http://gpsworld.com/innovation-getting-closer-to-everywhere-accurately-tracking-smartphones-indoors/.

[126] Y. Rogers, “Interaction design gone wild: Striving for wild theory,” *interactions*, vol. 18, no. 4, pp. 58–62, Jul. 2011. [Online]. Available: http://doi.acm.org/10.1145/1978822.1978834

[127] Y. Rogers, K. Connelly, L. Tedesco, W. Hazlewood, A. Kurtz, R. E. Hall, J. Hursey, and T. Toscos, *UbiComp 2007: Ubiquitous Computing: 9th International Conference, UbiComp 2007, Innsbruck, Austria, September 16-19, 2007. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, ch. Why It’s Worth the Hassle: The Value of In-Situ Studies When Designing Ubicomp, pp. 336–353. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-74853-3_20

[128] J. P. Rolland, L. Davis, and Y. Baillot, “A survey of tracking technology for virtual environments,” *Fundamentals of wearable computers and augmented reality*, vol. 8, no. September, pp. 1–48, 2001.

[129] L. Rosenberg, “Virtual fixtures: Perceptual tools for telerobotic manipulation,” *Proceedings of IEEE Virtual Reality Annual International Symposium*, pp. 76–82, 1993.

[130] E. Rosten and T. Drummond, “Fusing points and lines for high performance tracking,” in *Proceedings of the IEEE International Conference on Computer Vision*, vol. II, 2005, pp. 1508–1515.

[131] J. L. G. Sánchez, N. P. Zea, and F. L. Gutiérrez, “From Usability to Playability : Introduction to Player- Centred Video Game Development Process,” in *First International Conference, HCD 2009, Held as Part of HCI International 2009*, M. Kurosu, Ed. Springer Berlin Heidelberg, 2009, pp. 65–74.

[132] M. E. C. Santos, A. Chen, T. Taketomi, G. Yamamoto, J. Miyazaki, and H. Kato, “Augmented Reality Learning Experiences: Survey of Prototype Design and Evaluation,” *IEEE Transactions on Learning Technologies*, vol. 7, no. 1, pp. 38–56, 2014.

[133] I. Sato, Y. Sato, and K. Ikeuchi, “Illumination distribution from shadows,” in *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, vol. 00, no. c. IEEE Comput. Soc, 1999, pp. 306–312.

[134] A. Shinohara, A. Shinohara, J. Tomita, J. Tomita, T. Kihara, T. Kihara, S. Nakajima, S. Nakajima, K. Ogawa, and K. Ogawa, “Human-Computer Interaction. Interaction Platforms and Techniques.” in *Human-Computer Interaction. Interaction Platforms and Techniques*, 2007, vol. 4551, pp. 936–945.

[135] G. Simon and M.-O. Berger, “A two-stage robust statistical method for temporal registration from features of various type,” *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, 1998.

[136] I. Skrypnyk and D. G. Lowe, “Scene modelling, recognition and tracking with invariant image features,” in *ISMAR 2004: Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, no. Ismar, 2004, pp. 110–119.

[137] V. Sundareswaran and R. Behringer, “Visual Servoing-based Augmented Reality,” *Proceedings of the First International Workshop on Augmented Reality*, pp. 1–11, 1998.

[138] P. Supan and I. Stuppacher, “Interactive Image Based Lighting in Augmented Reality,” in *Central European Seminar on Computer Graphics*. Citeseer, 2006.

[139] P. Supan, I. Stuppacher, and M. Haller, “Image Based Shadowing in Real-Time Augmented Reality,” *The International Journal of Virtual Reality*, vol. 5, no. 3, pp. 1–7, 2006.

[140] I. E. Sutherland, “A head-mounted three dimensional display,” in *Proceedings of the December 9-11, 1968, fall joint computer conference, part I on - AFIPS ’68 (Fall, part I)*. New York, New York, USA: ACM Press, 1968, pp. 757–764.

[141] R. J. Teather, A. Pavlovych, W. Stuerzlinger, and I. S. MacKenzie, “Effects of tracking technology, latency, and spatial jitter on object movement,” *2009 IEEE Symposium on 3D User Interfaces*, pp. 43–50, 2009.

[142] A. B. Tillon, I. Marchal, and P. Houlier, “Mobile augmented reality in the museum: Can a lace-like technology take you closer to works of art?” in *2011 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities*, no. Figure 1. IEEE, Oct. 2011, pp. 41–47.

[143] J. Torpus and B. Tobler, “lifeClipper3 — An augmented walking experience field evaluation of an experience design approach for immersive outdoor augmented reality,” in *2011 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities*. IEEE, Oct. 2011, pp. 73–82.

[144] L. Vacchetti, V. Lepetit, and P. Fua, “Combining edge and texture information for real-time accurate 3D camera tracking,” in *ISMAR 2004: Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, no. Ismar, 2004, pp. 48–57.

[145] W. van Eck and Y. Kolstee, "The augmented painting: Playful interaction with multi-spectral images," in *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*. IEEE, Nov. 2012, pp. 65–69.

[146] E. Veas, E. Mendez, S. Feiner, and D. Schmalstieg, "Directing attention and influencing memory with visual saliency modulation," *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, p. 1471, 2011.

[147] A. Vera and P. So, "(ONLINE) nasa - human factors 101," http://human-factors.arc.nasa.gov/web/humanfactors101/index.html, 2010.

[148] V. Vlahakis, N. Ioannidis, J. Karigiannis, M. Tsotros, M. Gounaris, L. Almeida, D. Stricker, T. Gleue, I. T. Christou, and R. Carlucci, "ARCHEOGUIDE: First results of an Augmented Reality, Mobile Computing System in Cultural Heritage Sites," in *Proceedings of the 2001 conference on Virtual reality, archeology, and cultural heritage - VAST '01*. New York, New York, USA: ACM Press, 2001, p. 131.

[149] D. Wagner, M. Billinghurst, and D. Schmalstieg, "How real should virtual characters be?" in *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology - ACE '06*. New York, New York, USA: ACM Press, 2006, p. 57.

[150] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond, and D. Schmalstieg, "Pose tracking from natural features on mobile phones," *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 125–134, Sep. 2008.

[151] G. Welch and E. Foxlin, "Motion tracking: no silver bullet, but a respectable arsenal," *IEEE Computer Graphics and Applications*, vol. 22, no. 6, pp. 24–38, Nov. 2002.

[152] T. Whitted, "An improved illumination model for shaded display," *ACM SIGGRAPH Computer Graphics*, vol. 13, no. 2, pp. 343–349, 1979.

[153] J. Wither, Y.-t. Tsai, and R. Azuma, "Indirect augmented reality," *Computers and Graphics*, vol. 35, no. 4, pp. 810–822, 2011.

[154] M. M. Wloka and B. G. Anderson, "Resolving occlusion in augmented reality," *Proceedings of the 1995 symposium on Interactive 3D graphics - SI3D '95*, pp. 5–12, 1995.

[155] R. Wojciechowski, K. Walczak, M. White, and W. Cellary, "Building Virtual and Augmented Reality museum exhibitions," in *Proceedings of the ninth international conference on 3D Web technology - Web3D '04*, vol. 1, no. 212. New York, New York, USA: ACM Press, 2004, pp. 135–144.

[156] E. Woods, M. Billinghurst, J. Looser, G. Aldridge, D. Brown, B. Garrie, and C. Nelles, "Augmenting the science centre and museum experience," in *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Austalasia and Southe East Asia - GRAPHITE '04*, no. Figure 3. New York, New York, USA: ACM Press, 2004, pp. 230–236.

[157] R. C. M. Yusoff, R. Ibrahim, H. B. Zaman, A. Ahmad, and S. Suhaifi, "Users Acceptance on Mixed Reality Technology," *iacis.org*, vol. XII, no. 1, pp. 194–205, 2011.

Part II

Contributed Research

Contributed Research

> *"I may not have gone where I intended to go,*
> *but I think I have ended up where I needed to be."*
>
> Douglas Adams, *The Long Dark Tea-Time of the Soul*

This thesis presents its main contributions in the form of five papers. Each paper is based on investigating AR applications in relation to a target audience, a set of users who have experienced the application and provided feedback, either directly or indirectly. With exception of the first project described, papers [A-B], subsequent papers use findings from the previous papers as a driving motivation for the research.

In Chapter 1 I shortly introduced the collaboration with Koldinghus Museum, with a promise that we would return to it later.

Initially the research started as a collaboration with Koldinghus Museum to present novel technology to visitors at the museum, aimed at attracting visitors who are not interested in the "old and dusty" image many museums today have. The castle of Koldinghus, where the museum resides, was built during the 13th century, and in the following centuries the castle was one of Denmark's most important royal castles. In 1808 a fire left the castle in ruins. It has since been restored, and today the castle houses a museum which is a centre for cultural activities and exhibitions. As a result of the collaboration with the museum, two AR applications have been developed for, and used extensively by, visitors at the museum. This has allowed us first hand insights and knowledge of AR applications *in the wild*. For the museum the collaboration has been a way to introduce and promote novel technology to their visitors, in an attempt to attract new visitors. The first two projects presented are a result of this collaboration, presented in papers [A-B, D, E]. They focus mainly on application development, data collection and presentation from unsupervised longitudinal studies.

The remaining two projects are inspired by insights resulting from the projects related to the collaboration with Koldinghus Museum. These focus mainly on human factors related to AR, inspired directly from findings in the

Koldinghus collaboration projects. One relates to human visual perception (paper [C]) while the other relates to user task performance for a novel view management system (paper [F]).

7 Research motivation

Part of the experimentation is taking place in unsupervised natural settings, in a *semi-controlled environment*. Another part of the research is conducted in the laboratory, in a *fully-controlled environment*. The initial vision for the research topic was investigation of AR applications as they are being used by people *in the wild*, i.e. they should feel free to use the applications or not, however they want, and how they feel is most appropriate, for the duration they want to use them. Not in the closed environments of a laboratory and not for a limited amount of time.

It is relevant to look into how AR as a technology can benefit the user outside of the lab, in the correct physical setting, unsupervised and for an extended duration of time. Most prior research and development within the field of AR has a main focus on enabling or improving AR applications from a technology perspective. Only recently has research begun investigating the potential of mobile AR with focus on the user (often through HCI principles) in natural settings. There is still not a clear and final answer for what AR has to offer as a tool the user is interested in using on a daily basis. The research presented here aims to investigate in which ways we can narrow the gap between the technology and the user.

In general, I have worked with the following research statement in my experiments:

> "There is an unlocked potential within the field of AR, some use-case and design guidelines, which are not understood yet, and are not used in development of AR applications."

7.1 Research methodology

In the traditional model of scientific research (or science in general), there are three main elements: theory, operationalization and observations [1]. This, in science, often takes the form of *deductive reasoning*. Here a general understanding or theory of a subject allows deducing an expectation and a testable hypothesis. In the opposite end, *inductive reasoning* starts from a set of observations and proceeds to search for a pattern within the data.

Designing experiments which involve people in the field often have a high level of uncertainty. Often prior expectation of users does not match their behaviour in the experimentation. For the projects described in this thesis, both inductive and deductive reasoning processes are involved at some stage in the project, but differ in whether data exploration or hypothesis testing is the main focus of the investigation. Thus, I have worked under the assumption that one of two possible scenarios are present in the experimentation:

> With only a slight assumption of the system behaviour and how people might interact with the system, what is the most likely explanation for the observations gathered? (inductive reasoning)
>
> With a specific expectation and hypothesis for how the system should behave and how people might interact with the system, we observe a system and analyse data related to the hypothesis (deductive reasoning)

Qualitative and quantitative research approaches differ in their purpose, where a quantitative approach aims at generalisation, prediction and causal explanation, and qualitative research tend to be contextualising, interpreting and understanding perspective of users involved.

In this thesis I attempt a quantitative approach to data exploration in the projects, letting the logged data tell the story, and attempt to construct the most likely explanation based on the data. This differ from other research where the objective performance measures, related to the study, are often collected in the laboratory. As an example of this approach, in the collaborations with Koldinghus Museum, which will be presented in the following chapters, the main driving motivation is exploring how users interact with installations in an unsupervised settings. From that I draw insights which inform about the usage patterns and suggest directions for further research in information facilitation and interactions methods relevant for this or related settings.

The described approach for the collaboration projects with Koldinghus Museum takes the form of a more inductive approach, where an assumption of user interaction is set initially in the project. Based on this assumption, quantitative data relevant to user interaction is logged, and based on this data likely explanations for this data are drawn, and insights into user interaction are presented.

Observational anecdotes presented in the projects have been of a qualitative nature, and used as inspiration for hypotheses for sequential projects, or as suggestions for potential explanation to outcomes of quantitative analysis. An example of hypothesis testing inspired by observational anecdotes are the user studies in the projects *How Wrong Can You Be?* and *Experiments in Viewpoint Management*, which both are presented later. Common for these projects is an inductive reasoning approach, in contrast to the Koldinghus Museum collaborations presented. Here, the approach is flipped. Using the insights into user interaction previously gained, specific hypothesis for user interaction are presented and evaluated quantitatively in the laboratory. A system is systematically developed to observe and analyse data related to the hypothesis and conclusions are drawn based on these observations.

The methodology for this thesis is a mixed method approach, mixing quantitative data and qualitative anecdotes, aiming to use the languages of quantitative data, descriptive and inferential statistics to present insights gained. And to use qualitative anecdotes and narratives to further clarify and support significant contributions. As such, the observational anecdotes are not used as a means to draw conclusions.

Data quality

To be able to conduct any scientific investigation, the data collected must adhere to certain rules to be used, be it for exploration, heuristic analysis [3, 4] or statistical analysis [2].

Specifically, for statistical tests, choices about measurements in the experiments must be considered.

Overall two types of data can be gathered from an experiment: qualitative and quantitative data. Shortly explained, qualitative data is often complex data such as opinions and experiences, not easily represented as numbers or put into categories without losing some of the original meaning. With quantitative data, measurements are represented as a number, or represent categorical data such as hair color, or ordinal data such as user preference (e.g. on a scale of 1-10). Qualitative data allows for gaining complex insight into a subject or situation, to draw conclusion by understanding how or why users have a specific opinion. Quantitative data on the other hand allows for statistical analysis, to express experimental data in terms of probabilities, error rates and uncertainties. Most of the experimentation in this thesis deals with quantitative data, or a quantification of qualitative data to derive insights from observations.

Another dimension for measurements are the nature of whether the data is of objective or subjective nature, i.e. is the data based on the subjective opinion of a user or is the data not reported by the user at all, in which for most cases we assume the data is objective (such as application logging). Quantitative data can be both subjective and objective, whereas qualitative data is of subjective nature. Through experimentation in this thesis where users are directly involved both subjective and objective measurements are used.

Depending on the approach used in the studies, exploratory or hypothesis testing, we must find relationship hidden the data. One way this can be done, is by ordering the chaos in the data (for exploratory studies) using *descriptive statistics*, to summarizing sample data in order to describe and communicate the important characteristics. Alternatively, we must find whether sample data accurately represent a particular relationship in the population using *inferential statistics*, i.e. decide whether we believe that the sample data is adequate to describe the relationship to the population.

For the initial projects revolving around the Koldinghus Museum collaboration, the main driver is *descriptive statistics*, aiming to inform about the visitors and their usage of the applications.

For the remaining projects, *How Wrong Can You Be?* and *Experiments in Viewpoint Management*, the nature of the experimentation where specific hypotheses are tested, *inferential statistics* are used to deduce whether the data follows the expectations of the hypotheses.

8 Koldinghus Augmented: Memories of the Walls

Paper [A] (and the corresponding original full manuscript [B]) is entitled *Aspects of What Makes or Breaks a Museum AR Experience*. This paper presents the contributions related to the collaboration and research project of Memories of the Walls at Koldinghus Museum.

Motivation

In collaboration with Koldinghus Museum and other partners, this project is aiming at bringing AR inspired applications to their visitors. Specifically, the AR application presented in this project is designed for children aged 8 to 12 and mixes AR and mini-game elements to convey dramatized historical events to users. This means that the considerations for development presented in Section 6, and specifically Subsection 6.3 are relevant material to consider.

The goal of the paper is to identify and critically evaluate central aspects of the design and development process of a handheld AR application designed for children, and used in a museum context.

Project description

Visitors, children and their families, at the museum get to play through one of a series of wall memories. One memory is fully designed and implemented: the memory of Kirsten Munk, who from 1615 to 1628 was married to King Christian IV and bore him 12 children. Kirsten Munk was nobility but not royal, and could not be Christian IVs formal queen. The story presented, the *memory* of Kirsten Munk, is that the King's sisters are plotting to poison Kirsten in an attempt to thwart her marriage to Christian IV due to her not being of royal birth.

In this research project we demonstrate an interactive AR application, allowing visitors to experience a memory of the history at the museum in an unsupervised setting, traveling through the castle museum in their quest to solve the game. A set of collectible trading cards are handed to the children at the information desk along with a tablet. A subset of the collectible trading

(a) Subset of the trading cards **(b)** Example of an AR frame on the stand

Fig. 17: Illustration of trading cards used in the memory of Kirsten Munk [17a], and an example of an AR frame stand the visitors must find around the castle [17b].

cards are illustrated in Figure 17a. The objective is then to find the matching stand, as illustrated in Figure 17b, for each card in sequential order to solve the game. Initially, when the correct card is placed onto the frame in the chapel part of the museum, and filmed with the tablet, the *Wall* character tells the story of Kirsten Munk, the King's evil sisters and their plot to poison her, and invites the child to assist in preventing the assassination. To prevent the assassination, the child must seek out the assistance of the wine glass as well as the decanter, and a set of nice clothes to wear at the ball where the poisoning is to take place.

The paper presents findings related to the initial pilot study as well as usage of the application in a longitudinal study.

The collaboration enabled a foundation for user study experiments in the real world, in a limited form, through data collection on devices. This non-intrusive data collection framework was designed for this project, and later extended for subsequent projects.

Lessons learned

The project present insights into the children's usage of the applications and how they respond to various phases of the gameplay, and in interplay with the physical setting and collections at Koldinghus.

Based on usage patterns logged of the visitors, a set of suggestions for alterations/improvements are presented, believed to alleviate problematic areas of the experience. The most important findings in the paper are: 1) the application has a long, non-interactive narrative introduction resulting in many users dropping out early, 2) the application does not feature an AR "training level" for users to become accustomed to using visual tracking and be exposed to the AR "wow"-effect, and 3) the application does not encourage visual exploration during the AR application elements. This is further explained in the paper.

Resulting from the project was the development of an activity logging framework for extensive on-device logging of application activity, in a non-intrusive and anonymous manner, for data mining and analysis. It demonstrates a first step in moving from laboratory work and into the world in an unsupervised setting, using a framework for logging activity for anonymous and non-intrusive data gathering in the field.

9 Koldinghus Augmented: Castle Chapel

Papers [C] and [E] are entitled *An Interactive Visualization of the Past using a Situated Simulation Approach* and *Applying Handheld Augmented Reality to Visually Resurrect a Ruin*, respectively. These papers jointly present the contributions related to the collaboration and research project of the Castle Chapel at Koldinghus Museum.

Motivation

Building on the external collaboration with the museum presented in the first paper, this project focuses on the development of an interactive visualization of a ruin. This means that the considerations for the visualizations as presented in Sections 5.1 to 5.3 must all be accounted for.

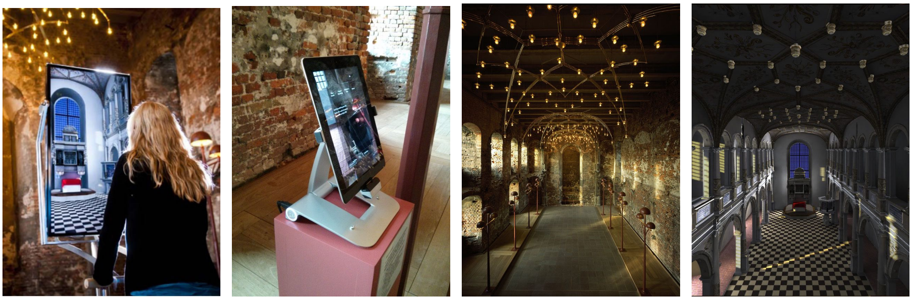
The goal of the project is the virtual resurrection and visualization of the church chapel at Koldinghus castle, and possible interaction and exploration of the chapel itself through presented installations on location.

With a framework developed for logging natural interactions with the device in the previous paper, we sought to investigate usage pattern in the natural unsupervised setting of the castle chapel.

Project description

The Koldinghus castle dates back to the mid 13th century, and has historically been a place for both protection from south, a residency for kings and the royal family, and has played a central role in Danish history. The visualization presented in the project is a virtual resurrection of the chapel as it most likely appeared in 1604, after the restoration commissioned by King Christian IV. During the Napoleon wars in 1808, the castle burned to the ground in a fire lasting two days. The castle was left as a ruin until a basic restoration in the 1970s (See Figure 18c), with the current appearance being no more than a ruin of the former chapel.

For this research project we demonstrate two practical applications for the museum that allow visitors to experience the past grandiosity of the castle



(a) Static **(b)** Handheld **(c)** Physical chapel **(d)** 3D visualization

Fig. 18: Presenting the static and handheld installations for the visualization of the physical chapel

chapel in an unsupervised setting in both a static installation and a handheld version of the same visualization (See Figures 18a and 18b). Due to complications of the gyroscopes' unreliable drifting we developed methods to overcome this issue for short-term usage of both the static and handheld versions of the visualization. Furthermore, we present a method for realistic daylight illumination at interactive frame rates on the device for a single point in time.

The papers present results of an almost one year long unsupervised longitudinal study and explore gathered data to uncover usage patterns and limitations of the visualization and interactions.

Lessons learned

The projects presents three inherent aspects of mobile AR and solutions to overcome them for this particular setting: 1) a procedure for combating gyroscope drift, 2) achieving realistic illumination computation and 3) an approach for re-localization within the setting.

For combating drift, we present a novel approach relying on the return desk of the handheld device to act as a calibration device for the application. For the static installation we use a method for using microphone access on an tablet device to reduce drift by attaching custom hardware to the installation sensing orientation and relaying it as audio input.

Our approach to realistic rendering of the chapel adds in creating the right atmosphere for the chapel. We developed a method for rendering all view-independent effects in an pre-computation step, and only compute view-dependent effects on the device. In a final step, we combine the information computed in the light shader with a lookup in multiple precomputed light maps.

The findings from the unsupervised usage of the interactive visualization, from both a static installation and multiple handheld versions, indicate that the detailed visualization in itself is not enough to keep visitors interested for long. More information must be available to visitors to engage them. The data clearly shows that visitors do not appear to be too interested in looking around, actively and extensively exploring the visualization.

An observational anecdote form on-location observation, inspiring the next project, suggests that part of the visitors at the museum does not recognize when they experience the application with a considerable amount of drift.

10 How Wrong Can You Be?

Paper [D] is entitled *How Wrong Can You Be: Perception of Static Orientation Errors in Mixed Reality*. Related to qualitative findings from the previous project (The Castle Chapel, Chapter 9), this paper focuses on user perception of static orientation errors for AR (called classical AR in the paper) and indirect AR (called indirect AR in the paper) scenarios. These are illustrated in Figure 19.

Motivation

Qualitative findings from paper [C] indicates that users are not aware (or do not care) that the orientation of the visualization does not match the orientation of the physical environments. It is given that we as developers pay attention to detail, and we should not expect the general public to hold the same standards. However, even in situations with large drift (more than 30° offset yaw rotation) this was observed. Based on this, this project aims to find the minimal static orientation offset where 1) people no longer notice an offset for both classical AR and indirect AR, and 2) investigate to what degree the rotation axis (yaw, pitch, roll) had an influence on this.

We present a study of human perception of static orientation tracking errors in video see-through classical AR and indirect AR, with an experiment designed to uncover the lower boundaries for human registration of static orientation errors.

Project description

Using experimental software developed for tablet devices, the experiments are carried out using an iPad3 for tracking the laboratory setting. The main feature of the laboratory is an aluminium cage used for mounting lab equipment (Figure 19a). This cage serves as the main feature of the setting, and was chosen to ease the the task of correcting the offset, to find the minimum limit for perceptual errors.

(a) Laboratory setting showing the main features of the setting **(b)** Screen-capture of the AR test application **(c)** Screen-capture of the IAR test application

Fig. 19: Illustrations of the setting and applications used for the experiments

The first experiment is focusing on both classical AR and indirect AR. It investigates the lower boundaries for human perception of static orientation errors, and simultaneously investigates the effects of the error on the 3 local rotational axes. The purpose is finding the absolute minimum boundary for human error perception when manually calibrating orientation offset.

In the second experiment, we attempted to eliminate the possibility of relying on memory and learning when solving indirect AR tasks. Furthermore, we sought to eliminate solution tricks involving fitting the position of physical tablet features or GUI elements to virtual features on the screen. This was achieved by using three different physical platforms for the tablet during the experiment on which only one set of yaw, pitch, roll trials would be performed, and by removing the classical AR scenarios from the experiment.

Lessons learned

The presented method contributes three interesting findings: 1) people are more perceptive of static orientation errors in classical AR than in indirect AR scenarios, 2) the ability to detect static orientation errors is dependent on the rotation axis affected by the error, and 3) roll error perception seems consistent in indirect and classical AR, but yaw and pitch errors are perceived differently in indirect and classical AR.

Investigating user sensitivity of orientation errors, results showed that people are more than 50 times more sensitive to static orientations errors in classical AR, compared to indirect AR. Furthermore, the results show that users have different sensitivity to the three axes (yaw, pitch and roll). In indirect AR, roll has smaller errors than yaw, which in turn has smaller errors than pitch. This means that users are more sensitive to roll errors, and least sensitive to pitch errors in indirect AR.

There are several other interesting findings in the experiments. One worth mentioning here is that participants reported using various means to attempt to correct the error, such as memory of prior trials within the experiment, aliasing effects in the rendering (they assumed that least *jagged* lines indicated most correct), amongst other. Based on this we believe that the reported numbers in the paper are lower bounds, and that the error tolerance in a more realistic, dynamic situation might be larger.

11 Experiments in Viewpoint Management

Paper [F] is entitled *Temporal Coherence Strategies for Augmented Reality Labeling*. This project focuses on view management techniques for AR, as a natural future extension of the visualizations presented in papers [C] and [E].

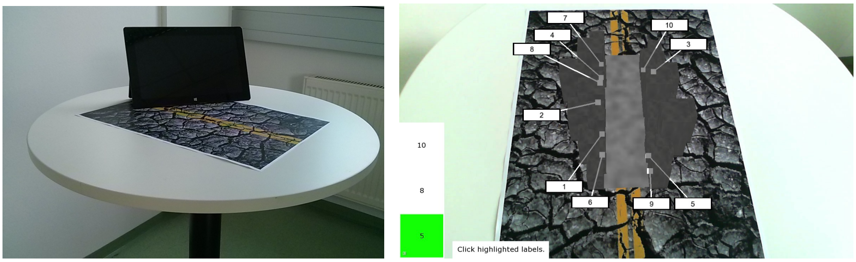
Motivation

Interactive visualizations, such as the Castle Chapel studies presented earlier, ideally should include additional information to present to the user, in addition to the visualization itself. We, and the museum at Koldinghus, are interested in how to present additional information to visitors at the museum who experience the installations. This information could as an example be related to objects in the visualizations, places in the chapel or other points of interest. One way to do so is by annotating interesting objects in the scene by the means of labels in the scene presenting the information to the visitor. However, in a dynamic setting such as AR, the system must be able to adapt to ever changing viewpoints of the interface as the user moves around, and the labels should not obstruct the objects of interest in the current viewpoint rendering.

Project description

To investigate how such a system should be constructed, we evaluate implementations in user studies related to task performance and subjective opinions of users. We have developed four view management techniques for label management, and evaluated user task performance and user opinion in two user studies. The implementations enforce constraints from the literature by placing annotations in the vicinity of the object, avoid overlaps between annotations and the annotated objects, and avoiding crossing leader lines. Examples of the implementations are illustrated in Figure 20.

Using experimental software developed for windows tablets, the experiments are carried out using a Surface 2 Pro tablet.



(a) Minimal experimental setup for the studies

(b) Showcasing the application running for the first study

Fig. 20: Illustration of the experimental setup and application running. The purpose of the user was to identify specific labels, select them and then move closer to the anchor point.

In the first experiments, we evaluate user task performance of the four implementations in a manner similar to a novice user in a learning scenario. We are interested in how different factors interact on task performance. The factors include *rendering space* (2 levels: 2D and 3D), *update rate* (2 levels: continuous and discrete) and *grouping* (2 levels: balanced and unbalanced) of labels. Thus the experiments became a factorial mixed model design, with *update rate* and *rendering space* as within subjects factors, and *grouping* as a between subjects factor.

In the second experiment, based on the results of the first, we limit the experiment to subjective preferences of only one factor, *update rate* with the two levels from experiment one. For this experiment *rendering space* is set to 3D only, based on previous results, and *grouping* set to an unbalanced grouping, as this is deemed "worst case" for both implementations.

Lessons learned

We present findings of two user studies, and present suggestions for usage of viewpoint management technologies for practitioners in the field. The project describes the implementation and evaluation process using four different view management techniques for dynamic handheld AR.

The contributions of the project are four relevant findings: 1) participants perform tasks faster with a management system using object space rendering compared to one with screen space rendering, 2) discrete updating in object space outperforms any screen space implementation presented, 3) from the constraints presented in the literature it appears only re-arranging of labels have an effect on performance, and 4) participants prefer a discrete updating system when presented with an object space rendering technique. In general terms, this means that users were fastest using the object space rendering method, and when presented with various update methods for this particular technique, they preferred that the system was discretely updating.

To explain the constraints from the literature, we limited the data to a subset of the original experiment data, relevant for investigating this. This showed that the *re-arranging labels* constraint had an impact on performance.

The insight gained leads to new interesting venues for further investigations in this area. One example of this is selective updating of only labels breaking constraints as an alternative to updating the entire system as it is currently. This might increase user task performance as well as computational performance.

12 Contributions

To summarize, the contributions made in this thesis are:

- **User evaluation for visual perception of static orientation errors for direct and indirect Augmented Reality**
  The method presented in paper [D] demonstrates a user acceptance threshold for static tracking errors, revealing (not surprisingly) that the threshold is lower for classic AR than indirect AR. Additionally, detection of errors is highly dependent on orientation axis.

- **User evaluation applied to Augmented Reality technology for view management**
  The methods presented in paper [F] demonstrates increased task performance for object space rendering techniques in comparison to screen space rendering techniques. Further analysis reveals re-arranging of labels might have an impact on performance.

- **Demonstration of practical framework for realistic daylight simulation in applications for augmented realty for low-powered mobile devices**
  The illumination framework presented in paper [C] includes a precomputation of light maps in combination with on-device rendering of viewpoint specific illumination effects such as specular highlights on low-powered mobile devices.

- **Method for a practical solution approach for tracking related issues**
  The methods for combating gyroscope related issues presented in papers [C] and [E] describes novels methods to overcome drift-issues using a practical solution. The relocation method presented in the paper [C] describes a practical solution for relocation in the virtual setting, allowing the user to reposition himself accordingly in the physical space following the virtual viewpoint.

- **Demonstration of practical applications using Augmented Reality technology**
  The methods presented in this thesis have been applied and demonstrated in several unsupervised real world scenarios, presented in papers [A], [C], and [E].

- **Demonstration of practical logging framework for non-intrusive anonymous logging**
  The descriptions presented in this thesis have been applied to facilitate data gathering and analysis of in-field usage study, presented in papers [A], [C], and [E].

13 Conclusion

This thesis has presented methods and applications for investigations of in-field usage of *Augmented Reality* (AR) technology. With a main focus on usage by people, the work has been organized around two major themes: explorative analysis from in field usage of commercial applications, and laboratory experimentation resulting from insights gained from the previous explorative analysis.

Two applications, presented in papers [A-C, E], study usage patterns in unsupervised longitudinal field testing. The first one, papers [A-B] is presented as an application targeted for children as a game in a museum context, while the second, papers [C, E] present a static as well as handheld installation of a visualization of a historical ruin. Based on data collected and analysed in both settings, usage patterns have been uncovered and further work has been proposed. These projects demonstrate practical solutions for using AR technology within the field, as well as usage in unsupervised real world scenarios. Furthermore, papers [C, E] present a novel and practical solution for combating gyroscope related issues. Paper [C] presents an illumination framework for pre-computation of light maps in combination with on-device rendering of view-point dependent illumination effects, such as specular highlights. This presents a framework for time of day specific illumination for low-powered handheld devices.

Based on proposed future work and observations on locations, two studies in the laboratory, papers [D, F], deal with user performance in a series of experiments related to human factors; visual perception and task performance in selection, respectively. The first project [D] is based on a developed method for participants aligning offset orientation to the best of their abilities for a single axis, in a repeated experiment. The paper demonstrates a user acceptance threshold for static tracking errors, and illustrates that the threshold is lower for classic AR than that of indirect AR, unsurprisingly. Lastly, it reveals that detection of errors is dependent on orientation axis. The second project [F] in the laboratory environment is based on a method investigating user task performance in a series of consecutive tasks related to identification and selection of annotations for dynamic handheld AR. The paper demonstrates an increase in task performance for rendering in object space, in comparison with screen rendering techniques. Additional analysis within the paper suggests that re-arranging of labels might affect task performance.

During the thesis, all methods and applications have been tested on real world data, showing promise for AR technology running unsupervised in longitudinal field trials. The collected data demonstrate potential for usage in real world application for information collection. Focusing on how the applications are received and used over an extended time, there is a lot we can learn from such usage, such as how and what interests people, and where to apply more effort in understanding the user. It might just be that this is the way to unlock the potential for AR and uncover use-cases and design guidelines, which are not yet understood.

The thesis present ways and methods for bringing AR into the world and systematically collect data to explore how it is used. I believe this brings us a step closer to understanding the relationship between the interface and the user.

Chapter 12 listed six main contributions of this thesis. The research presented has resulted three possible avenues for further investigations which will be described subsequently.

13.1 Future work

From the papers [A-C, E], my thesis contributed insights into users' interest in exploration of the presented mixed-reality world. It was found that users does not appear to be interested in visually exploring the AR areas. In mixed reality settings, it is unclear to what extend users are interested in experiencing the presented mixed-reality world, and how the curiosity of the user can be stimulated to further explore an area in a mixed reality setting sufficiently. Specifically, papers [A-C, E] found that users' does not appear to be interested in visually exploring the AR areas, neither in the game (paper A) nor the visualization of historical ruin (paper C). For future work, it would be interesting to explore the relationship between the AR technology and users' interest in visually exploring the area, to uncover the potential of tracked AR and user immersion within the environment. Another possible future work in this area is investigation into the mental connection the user has, between the real environment and the virtual setting presented. In Cultural Heritage (CH) environments, visualization of historical settings or artifacts within the correct physical setting might elevate the overall experience. The considerations for user exploration is interesting as this is a hard subject to gather information on. To the best of my knowledge, this is an unexplored research area, however relevant for user acceptance of the enabling technology.

From the paper [D], knowledge of perception of static orientation errors was uncovered in a laboratory experiment. Specifically a lower threshold for perceptual user acceptance of static tracking errors were uncovered. A next logical step in research of perception of static orientation errors, related to paper D, is to evaluate the effect of allowing the tablet and the participant to move around in a realistic manner and the effect of taking the study out of a controlled lab context and into a more realistic setting. Other interesting areas of future study might be incorporation of head tracking, a higher level of realism in the virtual model, and similar studies of static position errors.

Paper [F] gave recommendations for designing a view management system related to task performance of an AR system. Related to that research, future work should isolate which factor influences the performance of the participants. Object-space labeling may have performed better due to stable label placement or due to the additional spatial cues from registering the labels as 3D objects in the scene. Additionally, to achieve better coherence between layouts, a better solution would take into account the layout of the previous viewpoint when calculating the layout of the new viewpoint.

13.2 A final outlook and perspectives

Despite the ever increasing quality of AR technology and related applications, there are a series of issues which make mass adoption problematic at this point in time.

First of all, for normal people, for whom AR is something they haven't heard of yet, getting started with mobile applications presenting this feature in unprepared environment, will be a hard task. If the technology is not robust or prepared for any given environment, developers will be tasked with explaining the user in a simple way why the applications fail to live up to any expectations the user might have. I.e. why the tracker does not work, which might be related to issues from lack of scene features, to illumination or even unstable handling by the user. This is as much a technology issue as it is a presentation issue. And if the default expectation of the user becomes that it does not work part of the time, it might not be worth it to the user to actually use the system.

When technology has matured and we are able to track these "hard to track" environments in multiple conditions, there are still issues related to occlusion and illumination presented in the theoretical parts of the thesis. With Hollywood presenting their visions for the future of AR, people will be expecting things to "just work". Too many hiccups or not really understanding what the limiting factors are might drive potential users away.

This relates to a need for additional user studies of both user adoptions of natural AR applications in the field, but also of perceptual studies related to the main technologies (tracking, occlusion, illumination). New knowledge in these areas can lead to improvement of technology and applications to be more natural, user friendly and useful to the individual.

References

[1] E. R. Babbie, *The Practice of Social Research*, 2010, vol. 12th edition.

[2] G. W. Heiman, *Basic Statistics for the Behavioral Sciences*, 6th ed. Wadsworth, Cengage Learning, 2011.

[3] J. Kjeldskov, M. B. Skov, and J. Stage, "Instant data analysis," in *Proceedings of the third Nordic conference on Human-computer interaction - NordiCHI '04*. New York, New York, USA: ACM Press, 2004, pp. 233–240.

[4] J. Nielsen and R. Molich, "Heuristic evaluation of user interfaces," in *Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI '90*, no. April. New York, New York, USA: ACM Press, 1990, pp. 249–256.

Part III

Contributed Papers

Paper A

Aspects of what makes or breaks a museum AR experience

Claus B. Madsen
Jacob B. Madsen
Ann Morrison

The paper has been published in the
Proceedings of 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-AMH), pp. 91–92, 2012.

The layout has been revised.

Abstract

The paper critically evaluates central aspects of an iPad AR application developed for a museum context. The application is designed for children aged 8 to 12 and mixes AR and mini-game elements to convey dramatized historical events. The game has been deployed for roughly 3 months and the findings in the paper are supported by extensive in-application activity logging. Actual usage of the application at the museum proved far less extensive than envisaged. Hypotheses for this finding are presented and discussed, with support from the logging data.

1 Introduction

We describe an Augmented Reality (AR) based interactive experience designed for a museum, but it is not a typical success story. A number of design decisions were made, which turned out to prevent the designed experience from living up to its potential. We expose and discuss these flaws, and provide guidelines for improvements that can be useful for practitioners and researchers in related contexts. Koldinghus Museum in Kolding, Denmark, is a castle dating back to the 13th century and has played a central historical role as part-time residence for a row of kings. The AR game/experience is called Memories of The Walls, and the story-driving element in the game is the walls (personified as a character) of the ruin remembering and letting the user experience dramatized historical events that occurred at Koldinghus.

The application is developed for children and pre-teens, runs iPads, and uses the AR concept as its main technological focus. The partners on the project have been Koldinghus Museum (contractor), the Board of Tourism for Southern Denmark (project management and funding), Aalborg University (technical development and implementation), No Parking and Baaring Stories (concept, manuscript, asset production).

The museum has been handing out iPads with the Memories of the Walls app to visiting children and their families since the launch on February 9th, 2012, (Figure A.1). In this paper we perform a longitudinal study of the use of the app, based on in-app activity logging, in addition to conducted interviews, and qualitative observations of groups interacting with the app.

2 Story and the game

The experience is centered around a “if only the walls could speak” concept. The Wall provides narration and quests by addressing the user directly, and users get to play through one of a series of wall memories. One memory is fully designed and implemented: the memory of Kirsten Munk, who from

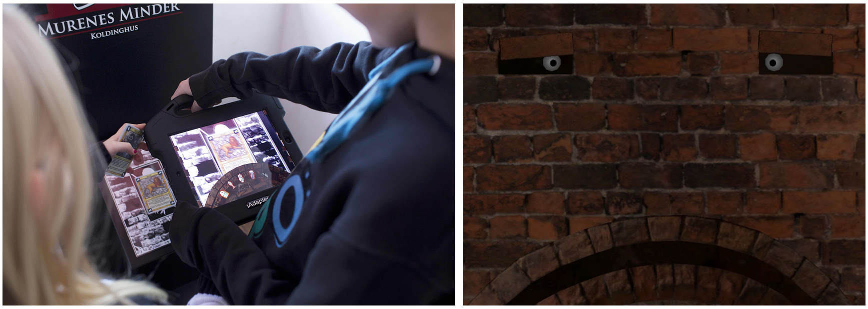


Fig. A.1: Left, two children playing Memories of the Walls with iPad in rugged casing. Right, screen shot from the developed app showing the wall character with eyes and mouth. The wall characters presents back story and quests to the user.

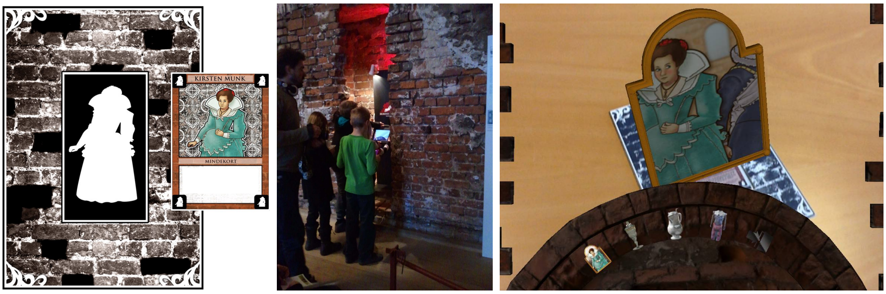


Fig. A.2: Left to right: 1) a static frame placed on a stand in the Castle Chapel, and the memory card of Kirsten Munk, 2) a stand in a recess in the Chapel wall, and 3) animated augmentations are displayed when the user places a proper card on static frames around the castle.

1615 to 1628 was married to King Christian IV and bore him 12 children. Kirsten Munk was nobility but not royal, and could not be Christian IVs formal queen. The story implemented in the memory of Kirsten Munk is that the King's sisters are plotting to poison Kirsten Munk in an attempt to thwart her marriage to Christian IV. The quest of the child(ren) playing Memories of the Walls is to prevent the murder. The actual game experience takes the user through a large part of the castle while playing mini-games and unlocking items needed to complete the quest.

Memories of the Walls is an application for iOS 5.0 iPad2 tablets. The application is developed in the Unity game engine using Qualcomm's augmented reality SDK (QCAR) for tracking. In-app logging collects time of completion for individual parts of the game, total playtime, and progress in mini-games. During game stages entailing augmenting the position and orientation of the tracked marker relative to the iPad camera is logged at 2 Hz.

3 Longitudinal study

Usage data of the application has been logged on the 6 museum iPads from March 8th 2012 to May 9th 2012. During this time the application has only been used 54 times. One third (18) played alone, while the remaining (36) played in a group of two or more. Only 33 (61%) ventured further into the game than the introduction. The average playing time for all participants is 38 minutes, whereas average playing time for those that complete the game is 65 minutes. Subsequently, we address three main problem areas we have identified from the longitudinal study.

3.1 Making it past the introduction

Log data shows that, if users made it past the introduction 88% (29 out of 33) complete the game. Apparently, the introduction represents a hurdle to users. The introductory stage of the game is not really completed before users make it to the chapel and launch the presentation of the back story on Kirsten Munk. The average time to accomplish this for the users that complete the entire game is roughly 12 minutes, so on average it takes 12 minutes before gaming, interaction and the AR “wow” effect kick in.

3.2 Reality not augmented

The goal of AR technology is to augment the physical world, but we noted that most groups spent more time looking at the screen and playing the games and only noticing the museum itself as if in passing, in a quest to find the static frames and locations of mini-games. The game is an orienteering exercise too focused on playing mini-games and getting from station to station. This is substantiated by the logging data. Average playing time for those that complete the game is 65 minutes, roughly 45 minutes of which is spent on moving from station to station across the entire castle. People who know their way around the castle are estimated to be able to walk that tour in under 10 minutes. Thus, on average roughly 35 minutes are spent on figuring out where you are and where you need to go.

3.3 Visual exploration in AR sequences

In Memories of the Walls there are 5 stages where animations are augmented onto the video feed. The in-app logging allows us to data-mine the positions of the iPad relative to the markers at the stages. Figure A.3 shows a heat map of viewing positions (viewing directions projected onto unit circle) for the station where Kirsten Munk is introduced. The plot clearly shows that users do not exploit the possibility to visually explore the augmentations

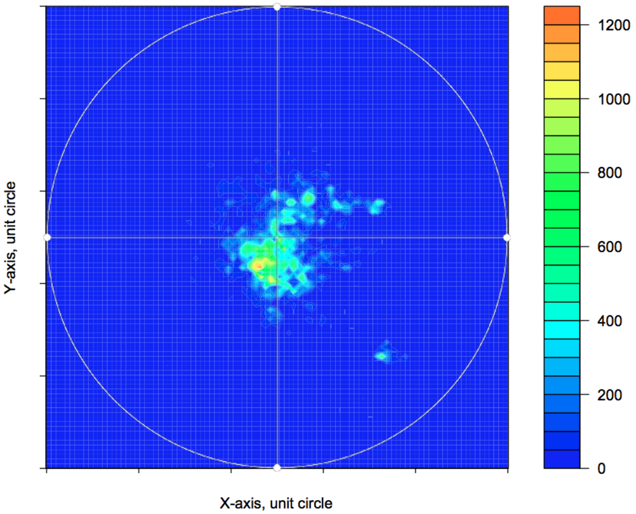


Fig. A.3: Heat map of viewpoint positions relative to the Kirsten Munk augmentation.

from arbitrary viewing directions. Had they looked more evenly from all viewing directions the plot would not have shown such a clear hot spot. This is clear evidence that users do not get the "wow" effect of tracked AR; they are passively watching an animation that might as well have been a regular 2D animation.

4 Discussion

We subsequently present some alterations/improvement we believe would alleviate the problematic issues.

Abbreviate the introduction sequence. The design team behind Memories of the Walls championed a cinematic style presentation of the setting for the experience. Based on the longitudinal experiment we suggest that the experience be altered to a more game-oriented, explorative style, and design the introductory stages such that an interactive AR "moment", or "wow" effect occurs immediately upon starting the game.

Make the game more grounded in surroundings. Nothing in the gaming experience forces the user to take advantage of, or to "need", the surroundings while playing. We recommend changing the gameplay towards actively involving the surroundings, e.g., getting codes or names from paintings, counting the number of doors, locating the best place to hide and pointing it out on displayed floor map, anything that forces the users to take notice of the space and perceiving it, not just sensing it.

Motivate visual exploration during AR sequences. Data showed that users do not visually explore the augmentations. We suggest that this is addressed by 1) move stands from wall recesses into open floor space to allow users to walk around animations, 2) design augmentations to provide occlusions, for example multiple objects near each other, forcing the user to move in order to create motion parallax, 3) design mini game elements that force users to look behind/around objects to find, e.g., a year of manufacture or a code, 4) have static marker frames placed on the floor to enable body-sized augmentation which require actual body movement to walk around, not just a swaying from side to side. We conjecture the embodiment of the visual experience to be really important for the perception of the augmentations being grounded in the real environment.

5 Conclusion

During design and development there was a clash of cultures between the design people voting for a cinematic style, and the implementation people voting for an interaction-oriented style. The longitudinal study clearly indicated three issues with the current design: 1) the introduction sequence is too long and not interactive, 3) the physical location is not sufficiently integrated into the experience, and 2) augmentations are only experienced as cinematic cut-scenes, not visually explored.

We believe these experiences indicate that there is still a lot to be learned about how to design good applications around AR and how to optimally exploit the affordances of AR for interactive edutainment applications.

Acknowledgements

European Union Regional Fond programme funding for "The Walls Remember" and the "Play and Learning, Kids'n'tweens Lifestyle: ERDFD-08-0051" projects is gratefully acknowledged.

Paper B

Aspects of what makes or breaks a museum AR experience

Claus B. Madsen
Jacob B. Madsen
Ann Morrison

The manuscript is presented for information completeness, and is the original full paper submission of paper [A]
Unpublished.

The layout has been revised.

Abstract

In this paper we critically evaluate central aspects of an iPad AR application developed for a museum context. The AR application is designed for children aged 8 to 12 and mixes AR and mini-game elements to convey dramatized historical events to users. The game has been deployed for roughly 3 months and the findings in the paper are supported by extensive in-application activity logging. Actual usage of the application at the museum proved far less extensive than envisaged. Hypotheses for this finding are presented and discussed, with support from the logging data. The most important findings are: 1) the application has a long, non-interactive narrative introduction resulting in many users dropping out early, 2) the application does not feature an AR "training level" for users to become accustomed to using visual tracking and be exposed to the AR "wow"-effect, and 3) the application does not encourage visual exploration during the AR application elements. These relatively easily remedied technical/design issues, combined with problems with the narrative of the application have resulted in the museum not promoting the application heavily to its visitors, leading to low usage. We contribute impacting cross-effects useful for the future design and deployment of AR in educational and recreational settings.

1 Introduction

There are many, many excellent examples of experiences and applications based on new technologies designed for museums. Museums in many ways provide a good setting for communicating with visitors through the use of technology. In this paper we describe an Augmented Reality (AR) based interactive experience designed for a museum, but it is not a typical success story. In fact, despite all the best intentions, a number of design decisions were made, which turned out to prevent the designed experience from becoming a successful one, or at least living up to its potential. The paper exposes these flaws, discusses them and provides guidelines for improvements that can be useful for practitioners and researchers in related contexts.

Koldinghus Museum in Kolding, Denmark, is a castle dating back to the 13^th century and has played a central role in the history of Denmark, primarily as part-time residence for a row of kings. In 1808 the castle burnt to the ground leaving only the bare walls, and in the 1970s the castle was restored into a museum, with absolute minimum alteration to the ruin (see Figure B.1).

The museum statutes obligate the museum to convey the history of Koldinghus also to children and pre-teens. In order to address this age group funding was secured for a cross-disciplinary team to develop a location-specific game-like experience for the museum, to complement regular museum exhibitions with a technology-driven element.

Fig. B.1: Photo of interior of reconstructed ruin of Christian IVs caste chapel at Koldinghus Castle Museum. Here the chapel is being used for a temporary exhibition of bronze statues.

Fig. B.2: Two children playing *Memories of The Walls* with iPad in rugged casing.

The application is developed for a mobile platform (iPads) and uses the AR concept as its main technological focus.

The developed AR game/experience is called *Memories of The Walls*, in reference to the fact that the story-driving element in the game is the walls (personified as a character in the game) of the ruin remembering and letting the user experience dramatized historical events that have occurred at Koldinghus.

The museum has been handing out iPads with the *Memories of The Walls* app to visiting children and their families since the launch on February 9^th, 2012, (Figure B.2). In this paper we perform a longitudinal study of the use of the app, based on in-app activity logging, in addition to conducted interviews, and qualitative observations of groups interacting with the app.

The partners on the project have been Koldinghus Museum (contractor and provider of historical facts), Syddansk Turisme (board of tourism for Southern Denmark, project management and funding), Aalborg University (technical development and implementation), No Parking and Baaring Stories (concept, manuscript, content production).

2 Related Background

Museums and public gallery spaces have often proven to be useful environments to start testing real use of emergent technologies, see for example recent work with multi-touch and AR systems, [1, 4, 10]. There are multiple reasons for this. These types of environments generally have information and clear instructions beside each work. The purpose is two-fold: the information states clearly what the work is, piquing interest and instructions on how to use entice the potential user. These factors are particularly important for novice users. In addition, there are staff on hand to advise and instruct, even at times guided tours or individual museum guides that demonstrate use. Further, technical staff are available to maintain the work, restart the work and make sure the work is functioning correctly. All of these factors ensure that for an audience member using a new technology there are less obstacles to overcome to produce a successful experience.

2.1 AR in museums

Within museums AR can be used as a tool with which visitors can extend the information available to them by adding to the existing information, be it by audio, visualization and/or haptic augmentation. Some recent usage examples of AR in museums include museum guides [1, 5], that attempt to bring added information to the museum visit through augmentation of information into the scene, AR exhibitions [9, 10], offering the user the ability to explore the installation, and attempts at multimodal AR experience [1], or more natural AR interfaces [3], as a representation for artifacts that may otherwise be off limit to the public for various reasons.

2.2 AR with Posters and Maps

There are many examples of AR technology working with paper surfaces, particularly with paper maps and posters. As an example of hybrid technology in action, studies indicate the use of paper surfaces with AR technology has much potential. A projection-based system by Reitmayr et al. [8] augmented a paper map directly with dynamic, geo-referenced information. Reilly et al. [7] used RFID tags to associate locations on the map with digital information. There are many instances of AR working well with paper surfaces to dynamically update awareness of friends' locations or activities [6].

Fig. B.3: Screen shot from the developed app showing the wall character with eyes and mouth. The wall characters presents back story and quests to the user.

More common applications may include updating altered bus locations on bus stop maps or updating information for concerts or festivals on poster adverts closer to the event when more information is at hand.

In this project paper posters are added to the museum to indicate information and next clues, and are used as markers in an AR application for marker-based tracking in order to entice a young audience in a museum setting.

3 Story and the game

The Castle of Koldinghus has been at the center of many important historical events in Denmark during a 500 year period. This, combined with the fact that all that remained after the fire in 1809 are the castle walls, inspired the development team to center the experience around a "if only the walls could speak" concept. The *Wall* (Figure B.3) is an animated character in the game, providing narration and quests by addressing the user directly.

The basic idea of the concept is that users (children) get to play through one of a series of wall memories. Currently, one memory is fully designed and implemented: the memory of Kirsten Munk, who from 1615 to 1628 was married to King Christian IV and bore him 12 children. Kirsten Munk was nobility but not royal, and could not be Christian IVs formal queen.

The story implemented in the *memory* of Kirsten Munk is that the King's sisters are plotting to poison Kirsten Munk in an attempt to thwart her marriage to Christian IV due to her not being of royal birth. This plot is not historically substantiated and represents "artistic freedom" for dramaturgi-

Fig. B.4: On the left, a static frame placed on a stand in the Castle Chapel, and the memory card of Kirsten Munk. On the right, the stand in a recess in the Chapel wall.

cal purposes. The quest of the child(ren) playing *Memories of The Walls* is to prevent the murder.

After having paid admission the child, and its family if present, is issued an iPad and informally informed about the game. The app is already launched and prompting the user to press *START*. The custodian instructs the family to go to an adjoining room and commence the game. In addition to the iPad the child is given a set of "trading cards". The first card is a memory card, depicting Kirsten Munk. Three cards depict objects; a wine glass, a wine decanter, and a set of clothes, respectively. Finally, two cards represent traits: wisdom and courage, respectively.

Upon pressing *START* the *Wall* is awakened from its sleep and in an intro sequence recounts of the Big Fire, of the old days and of the memories. Then the child is instructed to find his/her way to the Castle Chapel and place the *memory* (Kirsten Munk) in a **static** frame to be found in the Chapel and then "The adventure begins". The duration of the intro sequence is 2:47 minutes.

When the child places the memory card in the frame in the Chapel, and films it with the iPad (Figure B.4), the Wall character tells the story of Kirsten Munk, the King's evil sisters and their plot to poision her, and invites the child to assist in preventing the assassination. While this is being told a magic mirror is augmented onto the iPad camera image, see Figure B.5.

To prevent the assassination, the child must seek out the assistance of the wine glass as well as the decanter, and a set of nice clothes to wear at the ball where the poisoning is to take place. The task of the child being to make sure Kirsten Munk's poisoned wine is replaced with the non-poisoned wine.

The remainder of the game now to move around in the castle and, in

Fig. B.5: The Kirsten Munk memory card is tracked and a "magic mirror" augmentation visually underlines the back story for the Kirsten Munk memory and the quest to prevent the assassination.

Fig. B.6: Left to right screen shots from the wine glass, decanter and clothes mini games, respectively.

succession, find the locations of the glass, the decanter and the clothes, respectively, and place the appropriate trading cards in the frame. When the app recognizes a correct combination of a static frame and a trading card an animated character, e.g., the wine glass, is augmented onto the iPad camera view and presents the child with a challenge (mini-game) to be completed before the decanter will assist the child in the overall quest.

The wine glass mini-game is to correctly replay (by touching various glasses on the screen) a 9 note phrase of a piece of renaissance music. The decanter mini-game is a motor-skill game where the child must pour wine into the glass. The clothes mini-game is to arrange the colors of a set of clothes to match a real set of renaissance clothes on a dress form, see Figure B.6.

When all mini-games are completed the child has everything needed to proceed to the big hall and actually ward off the assassination by placing either the *courage* or the *wisdom* trait card on a static frame. A corresponding ending is then presented as an animation augmented on the screen.

4 Implementation

Memories of the Walls is an application for iOS 5.0 iPad2 tablets containing camera and gyroscope. The application is developed in the Unity game engine using Qualcomm's augmented reality SDK (QCAR) for tracking.

When a single marker (static frame or trading card), or combinations of markers (static frame and trading card), are viewed through the tablet camera, the system identifies the marker or markers. Depending on progress in the application, a scripted augmented narrative or interactive mini-games will be revealed for the user. Naturally, the user must keep the frame/trading card combination in the camera's field of view for the tracking to function. If tracking is lost for a short while, audio will continue but animation is not shown. If tracking is re-obtained within 10 seconds, animation continues from where it would have been without tracking loss so as to be in sync with audio. If tracking is lost for more than 10 seconds, audio is also terminated, and both audio and animation are restarted from the beginning when tracking is re-obtained.

The implemented app was developed specifically for the iPad2, although both Unity and Qualcomm SDK makes it possible to extend this to Android tablets, as well as mobile phones. The application operates at between 20-30 FPS depending on the progress in the game and number of simultaneously tracked markers.

4.1 Guiding the users around

Assisting static markers are spread throughout the castle in order to aid the users on their journey. When identified with the iPad camera, a map of the castle with location of every station (locations of static frames) appears, along with the user's current position in the castle. This has been added to aid the users in navigating the large castle across several floors when moving between posts in response to the progression of the quest.

When a mini-game associated with an object has been completed the corresponding object icon on the graphical user interface (GUI) is highlighted. The GUI for the application is illustrated in Figure B.7. If, while en-route to the next post, the child forgets what to go and look for, it is possible to touch an icon on the GUI upon which a small audio file is played, explaining what to look for and in which part of the castle to look for it.

5 Early pilot study

A trial was held at the museum in early February 2012, with 16 school children of mixed gender from 10-11 years of age (5 girls and 11 boys) com-

Fig. B.7: The GUI for the application.

posed into a set of six teams. The children were accompanied and observed throughout and then interviewed in teams after working through an AR experience of the museum. There were several configurations of numbers of children to numbers of iPads: 1) 2 groups of 2 children with iPad per group. 2) 3 singular children, each with their own iPad. 3) 3 adults and 3 children with their own iPad. The purpose for this initial pilot test was to get constructive feedback from the children about the game. The evaluation team, led by Syddansk Turisme, wanted to gain an understanding of several elements of the AR experience. They wanted to firstly, better understand how the game components worked and whether the children learnt the story and understood how mini-games worked. Secondly, ascertain what were the best team composition(s) so that the children could achieve the best possible gaming experience. Third, understand from how the children described their experience what worked and what did not. Fourth, know how well the children could use the individual components: the cards, the frames and the iPad interfaces. Fifth, determine how the children would manage navigation of this large castle. Sixth, see if the experience impacted their learning. Did the children recall the historical facts the game required they learn in order to move to the next place? And seventh, if so, could the children distinguish between fact and fiction in the game. Lastly, the evaluators wished to uncover if the children could or needed to distinguish between reality and augmented reality (prior work showed children moved seamlessly between these two states) and finally, to find out how the children found the overall visual form and expression of the game.

The priority was to uncover any obvious anomalies with the work, to manage improvements and undertake any necessary modifications while implementing a user centered design approach, before releasing the game to a wider general public.

The users were observed while playing the game and then were interviewed in their teams by a small panel. We report here findings from interviews and observation.

5.1 Findings from interviews

There are limitations in the interviews where the children reported face to face with a group of adults who had been involved in the making of the game (including a nationally famous television role playing game master).

The children stated that they preferred to play with a friend (not a parent or grandparent) and the best combination was identified as two children playing with one shared iPad. All groups found the maps useful for navigation.

All groups with exception of one (discussed later), rated the experience high on games that they have enjoyed most. They reported that the aspects they liked the best were the mini games (identifying particularly the glass and decanter game) and that they were in a big old castle they had to roam all over. All groups reported that they would like to play more history games, for example, the "game of Vikings... second world war... games on the other king and queen". All groups had a good understanding of the story itself determined from a "before and after" test on historical knowledge. The groups could identify what was fact and what was fiction, except for in the ending of the game, where there was a choice of how to prevent the assassination.

The one group that responded differently was composed of children with more gaming experience. They had been talking and missed the information in the long introduction and subsequently boredom set in early. The game system did not allow them to go back and access that information again. They asked if they could stop the game but a teacher went with them and made them do it again.

5.2 Findings from observations

From observing the groups, we could see that all teams caught on quickly how to use the AR markers, the map and the iPad in tandem. Several of the children forgot the cards on the way so there was some learning of necessary sequence of events required. The beginning introduction was long and hard for some players to hear, we noted some children lost focus. Introductory narratives to the mini-games were sometimes skipped, in favour of going directly to the games, meaning there needed to be a way to repeat mini games introduction later in the sequence. In addition, the children needed to be able to see the map on the iPad after each mini-game. With three-person teams there was often a third person more in the background, not able to access the AR system.

We could see that the groups were very social when they bumped into other groups, but the teachers then structured it so this did not occur. We noted that the carry all-version of the iPad works well in that the casing seemed sturdy enough, was easily passed from student to student and obscured any buttons not needed for interaction. Generally we found that the groups were most often either walking fast through the museum or they were confined and working together solving mini games in smaller spaces. This meant the focus was very honed in, which appeared somewhat strange given the grandness and largess of the surroundings.

5.3 Changes in relation to the pilot study

During the month following the pilot study the prototype was altered to the final version used for the longitudinal study. Small changes were made to the application; specifically a map of Koldinghus Castle was now to appear after completion of a narrative or mini-game. This was added in order to aid the children in their tour around the castle. Additionally, the narrative introductions to the mini-games were made mandatory, as all users skipped those without listening, and subsequently did not know what to do following the mini games.

It was noticed that users often would lose track of the marker during the narratives, an issue which caused some frustration amongst the users. A short delay before stopping the narrative was introduced, so users did not have to start over due to lost tracking, in order to complete the narrative (as described in Section 4).

Logging of data was implemented in the application for an longitudinal study. The aim of the logging is to have a look into the usage of the application and how the users play through the experience as well as the experimentation they do with the iPad.

6 Longitudinal study

It was specified by the collaborators that the iPad should not be able to access the internet or communicate with other devices, meaning the logging of data had to happen locally on the iPad and stay there until an authorized person could connect to the network and upload the usage data. As a result of this, the data is saved continuously on the iPad itself, collecting data through logging continuous observations of the users' behavior with the application.

With this restriction in mind, the data collection is designed to not flood the iPad when playing or the server when uploading. In order to get the total playtime, every event was logged. It is however possible to suspend the application and stop it without sending an event in Unity. In these cases the total playtime might be a few seconds or minutes more than portrayed. This is not seen as a problem as early testing has shown users usually take a substantial amount of time to complete the entire game.

The logging collects data concerning time of completion for individual parts of the game, and the total playtime, along with progress in the games. During the game stages that entailed augmenting an animation on to the video feed the position and orientation of the tracked marker relative to the iPad camera was logged a 2 Hz. This data was logged to investigate how the users exploited the 6 degree of freedom tracking during the AR sequences.

Subsequently, we address three main problem areas we have identified from the longitudinal study.

6.1 Characteristics of actual usage

Usage data of the application has been logged on the iPad and gathered in a period covering two months, from March $8^{t}h$ 2012 to May $9^{t}h$ 2012. The first noticeable information from the data is that the application has only been used 54 times during the two months period of testing. This is less than once per day, and as the museum is open all week and has six iPads for handing out, the limiting factor does not lie with hardware or accessibility to the museum. Clearly, this indicates that the museum has not actively promoted the game to its visitors. Equally clearly, there is a good reason for this which we shall discuss in Section 7.

For those who played the game, one third (18) played alone, while the remaining (36) played in a group of two or more. Of the 54 plays, only 33 (61%) ventured further into the game than the introduction, i.e., the 2 minute 47 second narrative section where the Wall character wakes up and introduces the whole thing, see Section 3. 29 users (54%) completed the entire game. Conversely, 21 users (39%), or groups, returned the iPad to the custodians without even getting to the stage where the memory card with Kirsten Munk is placed on the frame and the back story of the plot to assassinate her is presented. This data shows that, if users made it past the introduction 88% (29 out of 33) complete the game. Apparently, the introduction represents a hurdle to users. And the introduction is not just the narrative presentation by the Wall character (which takes 2:47 minutes). The actual introductory stage of the game is not really completed before users make it to the chapel and successfully launch the presentation of the back story on Kirsten Munk. The average time to accomplish this for the users that complete the entire game is roughly 12 minutes, so on average it takes 12 minutes before gaming, interaction and the AR "wow" effect kick in.

Fig. B.8: The children often became so immersed in the task that they became oblivious to the magnificent surroundings.

The average playing time for all participants is 38 minutes, whereas average playing time for those that complete the game is 65 minutes.

6.2 Reality not augmented

While the goal of AR technology is to augment the physical world, in the pilot study we noted that most groups spent more time looking at the screen and the posters or playing the games and only noticing the museum itself as if in passing, in a quest to find the AR posters. Of course, this is an element of treasure-hunt type game play, where there is a perceived level of urgency from the participants to finish fast (despite there being no penalties for taking more time). There are numerous instances where the immersion in the tasks strongly outweighs any awareness of the surrounding environment (Figure B.8). This is made even more obvious because of the vastness of the castle in contrast to the small recesses the static frames, or game *stations* as it were, are placed in.

Other location-based game/experiences have attempted to manage this by offering prizes for different kinds of activities, e.g. prizes for best photo of the environment, designing the best new task in addition to the most commonly found game strategy of fastest and most correct task completion. The players need a reason built into the game (and the place itself) to explore the environment beyond finding "tokens". As it stands the *Memories of the Walls* experience is one of orienteering, where players achieve a task and hurry on to find the next station. They become confined by the narrative tasks, rather than having these expand their enjoyment of the whole environment. This is substantiated by the logging data. Average playing time for those that complete the game is 65 minutes, roughly 45 minutes of which is spent on moving from station to station across the entire castle. People who know their way around the castle are estimated to be able to walk that tour in under 10 minutes. So, roughly 35 minutes on average is spent on basically figuring out where you are and where you need to go.

6.3 Visual exploration in AR sequences

In *Memories of the Walls* there are 5 major sequences where animations and/or static objects are augmented onto the video feed, constituting the AR "gimmick" around which the experience is designed. Marker-based tracking is typically considered to provide 6 Degrees-of-Freedom (DoF) movement information, 3 positional and three orientational. In reality, though, since marker-based tracking requires the marker to remain in the field of view of the camera, only 4 DoF can be utilized by the user, as the camera has to point in the direction of the marker while the user for example is walking around the marker. Additionally, one of those 4 DoF is rotation around the camera's optical axis, which is futile as it does not provide any changes in the visual appearance of the augmented object(s). Thus, from an actual usage point of view, marker-based AR only provides 3 DoF for the user to exploit while visually exploring the augmentations. Essentially, the user can move the iPad around on a sphere centered at the marker, while keeping the iPad camera pointed towards the marker.

This kind of movement obviously is sufficient to visually explore the augmentations from all angles, move close for detail and move back for overview. *Memories of the Walls* offers two types of exploration and it is relevant to discuss this in relation to the eye-in-hand versus world-in-hand metaphors from the Virtual Reality and Computer Graphics fields. The two exploration modes are:

World-In-Hand
If a trading card, e.g., the card depicting the decanter is held in the hand in front of the camera, the card is augmented with the 3D model of the decanter. This is simply a mode for the users to have fun with, and not an intricate part of the completing the experience.

Eye-In-Hand
Trading cards placed on a static frame result in an animation being played as an augmentation. Users are free to move the iPad around to visually explore the augmentation. This is the main exploration mode in the experience.

Logging the movement of users while they interact with a 3D world can provide interesting information. In [2] logging of head- tracking was used

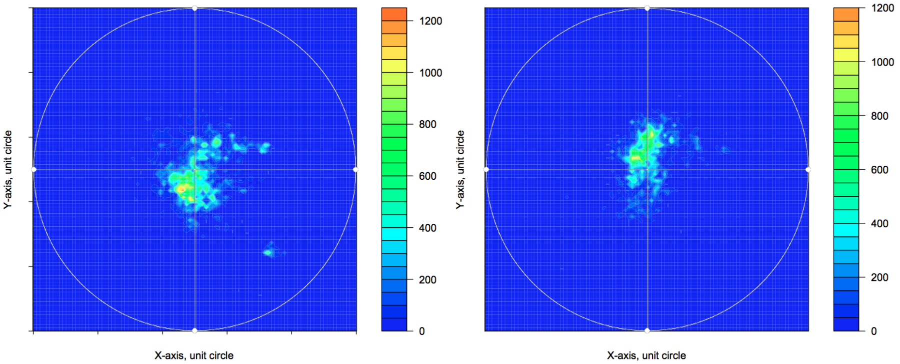


Fig. B.9: The left plot shows a heat map of viewpoint positions relative to the Kirsten Munk augmentation; the right plot shows positions for the assassination prevention augmentation.

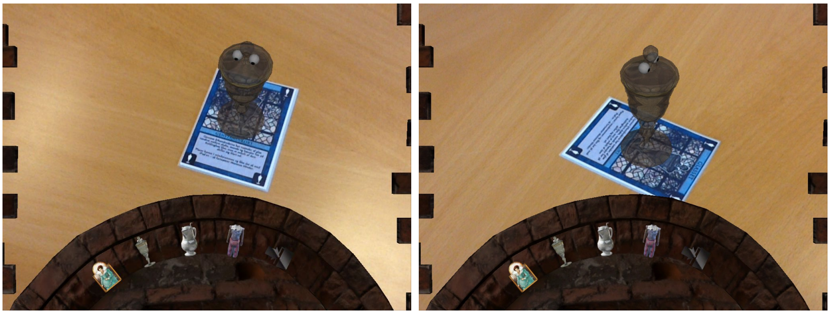


Fig. B.10: Two views of the wine glass representing outer extremes of the viewpoint hot spots.

to explore how augmentations influence people's movement patterns and revealed that people are reluctant to "walk through" objects they *know* are just visual augmentations and not real, physical objects.

In the context of the present paper we logged the positions of the iPad relative to the markers at the various stations of the game. Figure B.9 shows heat maps of tracked viewing positions (viewing directions projected onto unit circle) for the station where Kirsten Munk is introduced and the station where the assassination is prevented, (please refer to Section 3 for more detail).

This is clear evidence that users do not actually get the "wow" effect of tracked AR, in effect they are passively watching an animation that might as well have been presented as a regular 2D animation sequence, i.e. a film/cut sequence. Figure B.10 illustrates just how little movement users are performing while watching an augmentation. The two views shown represent outer extremities of the hot spots in the heat maps.

7 Discussion

Having obtained only 54 plays with 29 completions over a 2 month trial period at a busy museum can only be consider dismal. Obviously, the museum is not promoting the product heavily enough to visitors, and there can be only one explanation for that: the museum is not impressed with it. The museum invested itself in the development process, providing much input and also took great care to design and construct the stands used for the static marker frames, the lighting design at those locations, etc., see Figure B.4. The main reason the museum is not proud enough of *Memories of the Walls* is the low rate of users completing the game, and secondarily it is the daily experience that users have problems navigating the castle, and lastly, a segment of the users have some initial problems figuring out how to use the marker-based tracking.

We subsequently present some alterations and improvement we believe would alleviate most, if not all, of these issues, in addition to addressing the issues presented in Sections 5 and 6. These suggestions also serve as recommendations and inspiration for practitioners and researchers in related projects.

7.1 Abbreviate introduction sequence

On average users spend 12 minutes from pressing ”start game” to getting to the first element of interactivity, namely tracked AR. Already during design discussions arose concerning the length of the intro sequence. The design people were championed a dramaturgic presentation of the setting for the experience, whereas the technical and interaction design people advocated a more game oriented, explorative approach to the opening stages. It became a sort of clash of cultures with the design/storytelling people voting for a more cinematic style, which ended up being the implemented design. The actual introduction sequence is 2:47 minutes long (for the narrative part only) and could perhaps have been acceptable, but in the design phase we did not take into account that in reality it takes users 12 minutes to get through the introductory narrative and get to the first station where the AR presentation of the memory could take place, and another roughly 10 minutes for users to get to the first mini game.

In view of the experiences with the current version we recommend that a future revision, and other similar products, design the introductory stages such that an interactive AR “moment”, or “wow” effect occurs immediately upon starting the game, for example with showing a 3D model of the castle augmented on a marker in the reception area where iPads are issued. This would provide two advantages: 1) immediate stimulation of users’ interest in the technological aspect, and 2) make sure users learnt how to use the marker-based tracking while a custodian is present to assist if needed.

7.2 Motivate visual exploration during AR sequences

Data mining on logged marker tracking data clearly showed that users do not utilize the option to visually explore the presented 3D animations. This essentially reduces the experience to one comparable to a static film sequence, or cut scene. The potential of tracked AR is not at all exploited.

We recommend that this is addressed through various relatively straight forward approaches: 1) move the stands with the static tracking frames from recesses in the walls into open floor space to allow users to walk around animations and look at them from varying angles, 2) design augmentations so as to provide occlusions, for example multiple objects near each other, forcing the user to move in order to see everything and create visually stimulation motion parallax to heighten the experience of watching a 3D object, 3) change stand design from having a slanted top surface to having a horizontal top surface on top of which it is more natural to have a augmentation “standing”, 4) design mini game elements that force users to look behind/around objects to find, for example, a year of manufacture, a code, a pattern, a name, anything, and 4) have some static marker frames placed on the floor so as to enable body-sized augmentation which require actual body movement to walk around, not just a swaying from side to side of the upper body. We conjecture the embodiment of the visual experience to be really important for the perception of the augmentations being grounded in the real environment and for creating the illusion of them being physically present in reality.

7.3 Make the game more grounded in surroundings

Memories of the Walls is a location specific gaming experience, but is it not sufficiently grounded in the museum surroundings. The locations of the stands with static marker frames are placed inside the museum, but they might as well all have been in the same room. There is nothing in the gaming experience that forces the user to take advantage of, or to “need”, the surroundings while playing. This detaches the game from the museum and renders it an orienteering exercise too focused on playing the mini games and getting from station to station. The average playing time spent on locomotion during play is 45 minutes out of an average total playing time of 65 minutes.

It is recommended that a revision takes this into account by substantially changing gameplay towards actively involving the surroundings, e.g., getting codes or names from paintings, counting the number of doors, locating the best place to hide and pointing it out on displayed floor map, anything that forces the users to take notice of the space and perceiving it, not just sensing it.

8 Conclusion

We presented the design of an AR based gaming experience for a castle museum. The castle constitutes a fantastic location for such a product and the main metaphor and idea behind the design, *Memories of the Walls*, has a very high potential and extendibility to many different games focusing on various historical periods, figures and events. Yet, the actual product is not successful.

A two month longitudinal study of the application in daily, unsupervised deployment, combined with observations and interviews revealed that there are serious flaws in the design of the experience. These flaws were pointed out, and substantiated by significant in-app usage data logging.

The main flaws center around the design of the application not being sufficiently focused on an experience and exploration oriented style. There is not enough focus on the users' exploration of the augmentations, nor of the museum setting. We presented a range of guidelines for future revision, serving as inspiration to other similar projects.

Acknowledgements

European Union Regional Fond programme funding for "The Walls Remember" and the "Play and Learning, Kids'n'tweens Lifestyle: ERDFD-08-0051" projects is gratefully acknowledged.

References

[1] A. Damala, P. Cubaud, A. Bationo, P. Houlier, and I. Marchal, "Bridging the gap between the digital and the physical," in *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts - DIMEA '08*. New York, New York, USA: ACM Press, 2008, p. 120.

[2] C. B. Madsen and L. W. Lorentzen, "Using Visual Augmentations to Influence Spatial Perception in Virtual Representations of Real Scenes," in *The 8th Annual International Workshop on Presence, PRESENCE 2005*. London, United Kingdom: Department of Computer Science, Aalborg University, 2005, pp. 351–352.

[3] F. Mata, C. Claramunt, and A. Juarez, "An experimental virtual museum based on augmented reality and navigation," *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '11*, p. 497, 2011.

[4] D. Michael, N. Pelekanos, I. Chrysanthou, and P. Zaharias, "Comparative Study of Interactive Systems in a Museum," in *Proceedings of the Third international conference on Digital heritage*. Springer-Verlag Berlin, 2010, pp. 250–261.

[5] T. Miyashita, P. Meier, T. Tachikawa, S. Orlic, T. Eble, V. Scholz, A. Gapel, O. Gerl, S. Arnaudov, and S. Lieberknecht, "An Augmented Reality museum guide," *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 103–106, Sep. 2008.

[6] A. Morrison, A. Oulasvirta, P. Peltonen, S. Lemmela, G. Jacucci, G. Reitmayr, J. Näsänen, and A. Juustila, "Like Bees Around the Hive: A Comparative Study of a Mobile Augmented Reality Map," in *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*. New York, New York, USA: ACM Press, 2009, pp. 1889–1898.

[7] D. Reilly, M. Rodgers, R. Argue, M. Nunes, and K. Inkpen, "Marked-up maps: Combining paper maps and electronic information resources," *Personal and Ubiquitous Computing*, vol. 10, no. 4, pp. 215–226, 2006.

[8] G. Reitmayr, E. Eade, and T. Drummond, "Localisation and interaction for augmented maps," in *Proceedings - Fourth IEEE and ACM International Symposium on Symposium on Mixed and Augmented Reality, ISMAR 2005*, vol. 2005, 2005, pp. 120–129.

[9] R. Wojciechowski, K. Walczak, M. White, and W. Cellary, "Building Virtual and Augmented Reality museum exhibitions," in *Proceedings of the ninth international conference on 3D Web technology - Web3D '04*, vol. 1, no. 212. New York, New York, USA: ACM Press, 2004, pp. 135–144.

[10] E. Woods, M. Billinghurst, J. Looser, G. Aldridge, D. Brown, B. Garrie, and C. Nelles, "Augmenting the science centre and museum experience," in *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Austalasia and Southe East Asia - GRAPHITE '04*, no. Figure 3. New York, New York, USA: ACM Press, 2004, pp. 230–236.

Paper C

An Interactive Visualization of the Past using a Situated Simulation Approach

Jacob B. Madsen
Claus B. Madsen

The paper has been published in the
Digital Heritage 2013 International Congress (DigitalHeritage) Vol. 1,
pp. 307–314, 2013.

The layout has been revised.

Abstract

This paper describes aspects of the development of an interactive installation for visualizing a 3D reconstruction of a historical church chapel in Kolding, Denmark. We focus on three aspects inherent to a mobile Augmented Reality development context; 1) A procedure for combating gyroscope drift on handheld devices, 2) achieving realistic lighting computation on a mobile platform at interactive frame-rates and 3) an approach to re-location within this applications situated location without position tracking. We present a solution to each of these three aspects. The development is targeted a specific application, but the presented solutions should be relevant to researchers and developers facing similar issues in other contexts. We furthermore present initial findings from everyday usage by visitors at the museum, and explore how these findings can be useful in connection with novel technology for facilitating information transfer to a museum audience. The installation is in active commercial use and is currently logging further user interactions via in-application logging for future investigations in line with this project.

1 Introduction

Museums provide a great opportunity for introducing new technology to a user-base in a semi-controlled environment, in order to investigate user behavior and user acceptance out of the lab and in a contextual setting. The setting for this project is Koldinghus Museum, a historical castle dating back to the mid 1200s, placed in Kolding, Denmark. In 1808 the castle burned to the ground, leaving only the bare walls standing. The castle was restored to a museum in the 1970s.

This project is part of a currently ongoing series of Cultural Heritage (CH) projects in collaboration with Koldinghus Museum, aiming at creating new technology driven installations for the museum to facilitate information and learning of the history at the museum in a novel way. The aim of this project has been to conceive a visualization of the chapel as it appeared when built, using off-the-shelf hard- and software. The installation has to operate robustly 10 hours a day, 7 days a week in a room with no staff. This meant, that in addition to an interactive visualization running in real-time and facilitating information transfer to guests, there was an added constraint of an autonomous installation which requires no supervision. Furthermore, the location of the installation has multiple purposes, which means the installation must be easily transportable to other locations within the chapel.

The purpose of the installation is to deliver an interactive visualization of Koldinghus Chapel as it appeared in 1604 after a large renovation. This has been facilitated through Augmented Reality (AR) technology, to display the visualization of the past chapel room through a window (tablet) into the

Fig. C.1: Usage of the installation on the launch day. A child experiences the systems 3 degrees of orientation freedom of the visualization.

past, placed at the physical correct location in the present setting.

Visitors enter the chapel which is a 10 by 20 meter open space. Along the wall is a podium with docking stations holding two iPads, each running the application. As depicted in Figure C.1, users can grab an iPad, hold it and use as a viewfinder exploring the space. Interface options are given on the iPad, allowing the user via touch to translate the viewfinder position to predefined positions in the chapel. The user should then position himself accordingly in the physical world to achieve a coherent experience of the two spaces.

This contribution describes an approach for the development of the interactive installation using AR technology. We focus on three aspects that were considered crucial for development of the project; 1) A procedure for combating gyroscope drift, 2) achieving realistic lighting computation on a mobile platform at interactive frame-rates and 3) an approach to re-location within this applications situated location without position tracking. We focus this paper on development challenges that are inherent to a mobile AR development context.

The paper is organized as follows: In Section 2 the work presented in this paper is positioned in relation to previous work. Section 3 gives an overview and summery of the location and setting. Section 4 describes the system, in which the three crucial aspects for development is explained in detail.Section 5 presents an initial evaluation of the system on visitors at the museum, based on user data autonomously collected on the device, before summing up with conclusions, and directions for future work in Sections 6 and 7.

The installation has been actively running in multiple iterations since November 11th 2012, and in its current iteration since April 24th 2013, which is the version described in this paper. During this time the application has logged usage data to assist in uncovering usage patterns.

2 Related works

There is a vast body of work within the field of AR in museum contexts, ranging from museum guides [4, 15, 19], to building virtual and augmented installations and exhibitions for the museum [5, 8, 14, 21, 23, 24]. Novel technology in museums is an active and vast research area from both a technology point of view, to test new technology in a semi-controlled setting, but also from a CH point of view; considering how to best present information and knowledge to a user. Van Eck [21] considers how to augment paintings in the Van Gogh museum and describes the different information overlays, while [5, 24] presents research on information presentation in a way which are normally not readily available to users, such as the universe (S.O.L.A.R. System) and the Interactive Antartica.

Within CH, there is a lot of work being done with Virtual Reality [22], ranging from specialized work within 3D model reconstruction to user interaction and acceptance evaluations. Guidi [18] describes two approaches to 3D modelling in CH. One is the representation of the moment "as is" through different approaches and technology, and the other is the previous hypothetical state through a scientific reconstruction process, and presents two examples of work in relation to this. Kersten [7] presents work on modelling a city based on a 3D scanning approach. Trapp [20] describes how the user is now able to explore CH artifacts in real-time, and presents the technical concept for implementing this. Another example is work done by Zöllner et al. [25] for a museum setting allowing in one case a single degree of freedom for interaction by horizontal rotation of an installation stand (MovableScreen) and in another case a handheld device (UMPC) with other affordances. The aim here was to present information from remote CH sites at museums. Another example of work close to the described application in this paper is TimeScope 1 for Ename 974 [17], a church visualization for an archaeological park as a static installation with no interaction aspect, aiming at presenting an early medieval abbey on video shot of the foundation.

It is discussable whether the work presented in this paper falls under the umbrella "Virtual Reality" or "Augmented Reality". One could argue that the entire visualization is, in terms of technology, not connected to the real world, and thus is a virtual reality enabling technology. Another argument is that since it is in fact linked to the geo-physical setting, it becomes an augmentation of the setting, thus it is an AR technology in this setting, and this setting only. This point of view has been discussed also by Liestøl et al. [9], who stress that according Azuma's discussion on AR it is explicit that the term should be general and not based on technology [1, 2]. Thus, it can be inferred, that instead of merging real and virtual realities on the device, the users mentally connect the geo-physical setting and the virtual presentation, based on real and virtual landmarks in combination. This discussion will be elaborated further in further in Section 6 for further research within this area.

Liestøl et al. defines this augmentation in a defined setting as Situated Simulations, here defined as a virtual reality enabling technology for augmenting the geo-physical correct setting where it belongs [9–11].

The ideas of visualization and presentation of information on site is closely related to the work presented in this paper. We aim to further expand the freedom of exploration to allow for 3DOF orientation by hardware sensors in the system, and allowing for semi-freedom in translation by facilitating translation changes in the application, and let the user adapt to the virtual position in the geo-physical space. This builds on the previous work, which presents and adds more freedom in the user exploration.

While the presented work all evaluate their efforts in AR as being generally accepted by users and present findings that users are very interested in this novel technology, it can be speculated whether this is a "wow" effect of novel technology, or whether the effect will last. We consider it relevant to look into the realism of the presented objects, to create a closer connection to the physical world. In order to achieve added realism in the virtual representation of the chapel, we consider how to use pre-rendered and on-device rendering of illumination information to achieve a high degree of realism in the final visualization which is robust for position changes in the physical surroundings.

We also describe in this paper some of the aspects needing considerations for most AR development projects, which has not been discussed in the related works, such as combating gyroscope drift and achieving translation changes in a novel manner. We also enter the area of visual realism in the visualization and considers how to facilitate this at interactive frame-rates on a mobile device. Lastly we discuss methodology for usage of an application in a geo-physical correct environment to investigate how the user experiences the link between the real and virtual worlds when the link is not on the device itself.

3 Location overview

The Koldinghus castle dates back to the mid 1200s, and has through history been a place for both protection of the Danish borders, residency for kings and the royal family, and has played a central role in the history of Denmark.

The construction of the current chapel and tower of Koldinghus was started in 1597 due to a fire in that part of the building. The simple chapel of the time was not as grandiose as envisioned a church chapel should be,

Fig. C.2: Overview of the physical castle chapel (left) and the 3D visualization as represented in the installation (right).

according to the king. He wanted a new and bigger church, which was to be the base for the tower to be built. This new chapel was to be a reflection of the king as God's representative in both ecclesiastical and secular affairs. The new chapel was finished in 1604. [16]

During the Napoleon wars in 1808 the castle burned to the ground due to a fire started from a chimney in the guardroom. This fire destroyed the castle completely over a period of two days. However, the chapel and tower are not restored until the 1970s. The chapel was restored to a bare minimum with little alterations made to the standing ruin. The bare walls are displayed, and the chapel received a new floor and ceiling. The present chapel can be seen in Figure C.2 in combination with our rendered visualization.

4 System description

In order to develop a usable product for both the visitors and the museum staff, some requirements are considered for the design of the system. These are listed in Table C.1. First, as the task of the system is to facilitate information to the guests, who can range from families with children to elderly couples, the system in itself should be self-explanatory and easy to use for the average unskilled visitor with limited experience in technology. The system should be robust enough that it will not end up behaving in a way that the user does not expect, and thus ending up confusing or frustrating the user. Second, the museum personnel require a system with low cost and low daily maintenance as well as a highly transportable system. The chapel area, which

Museum guests	Museum personnel	Technical
Light weight	Flexible placement	Low power consumption
Easy to handle	Low daily maintenance	Long running time
Visualization true to world	Easy to setup and use	High frame-rate
Self-explanatory	Low cost	Limited polygon count
		2-DOF or 3-DOF orientation
		Translation of user perspective
		Robust

Table C.1: Overview of system requirements

the system will augment, is regularly a forum for exhibitions and events that require the floor space of the chapel. In these events, the system might need to be set aside for a small period of time, such as an evening or for a couple days, and then brought in again.

Apart from the user and staff requirements, Table C.1 also describes technical requirements to be met for a successful and functional product. The system itself should be able to process the visualization of the chapel ruin at interactive frame-rates, and be able to run during opening hours without being charged. The processing power of the system should be sufficient to handle these requests while the requirements of the visualization. The polygon count and the shader performance should be optimized to fulfill these requests.

The following sections will elaborate on some of the problems encountered during the development and how to overcome these problems in the areas within hardware, software, reconstruction and realistic daylight simulation.

4.1 Hardware

Using the gyroscope as the only sensor for estimating orientation is not feasible due to accumulated drift over time. Figure C.3 illustrates an early test of the gyroscope drift measurement over the course of one full day. It revealed more than 10° drift over a period of 28 hours. As the application is required to function unsupervised for a full day, this amount of drifting over time suggests that additional information of the orientation is required.

Initial considerations for the hardware were to use a combination of internal compass and gyroscope sensors in combination to estimate rotation and set calibration offset from north, both to combat drift from the gyroscope over time, but also to calibrate the orientation to the geo-physical room itself. However, experiments with the iPad showed that the compass was far too inaccurate to provide any decent magnetic orientation estimation for the chapel setting. A lab experiment was conducted, emulating the conditions of the chapel (electronic devices nearby, indoor) with exception of the granite

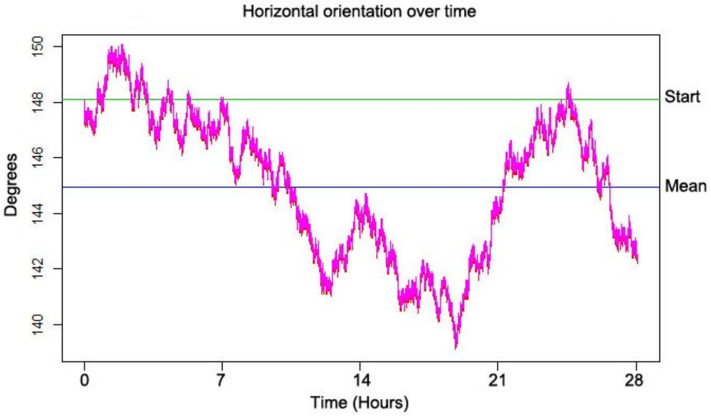


Fig. C.3: Horizontal drift of the gyroscope over the course of 29 hours in a static placement.

Fig. C.4: A version of the handheld installation, as it was displayed during testing in the Koldinghus Chapel.

structure of the chapel. The measurements gained from the iPad compass varied from -14° to +56° from north. With these inaccuracies of the compass for this particular setting, we opted to find a novel solution to calibrate the iPad to the chapel setting and limit the drift from the gyroscope over time and not rely on the compass in addition to the gyroscope.

This solution was implemented in the next iteration of the prototype. As a substitute for the compass, the docking station was used to calibrate the iPad to the orientation in the geo-physical world (Figure C.4). The use-case dictates that the application, and the iPad, can be in two different modes of operation: 1) the iPad is in the docking station, is totally static and charging, and 2) the iPad is held in the hands of a visitor and will move around for some length of time. As software detection of the charging state is simple, the static docking state is easy to detect in the application. The application behaviour in the two states is thus:

Charging:

Reset model orientation to the calibrated orientation

Not charging:

Rely on gyroscope readings to track the orientation

An added benefit to using the docking station is the ability to estimate the total number of uses by assuming that each usage is occurring when one user takes an iPad from the charging station (start) until it is returned to the charging station (end). It furthermore allows us to only log interaction data during this time period.

In the final version of the installation, currently at display and in active use on location, the application is able to function autonomously for an entire day after being setup by the museum staff. This setup and calibration procedure has three steps: 1) Start the application if it is not already running (most of the time it will be running continuously for days). 2) In handheld mode, touch-drag on the screen to calibrate the horizontal orientation by orienting the visualization to the desired orientation to match the physical space. 3) Disable the touch-dragging via a hidden in-application menu. Following this, the iPad will work autonomously during the day, and should there be any problem due to drift, the museum staff can repeat the procedure again, or restart the application to reset everything, if desired.

4.2 Software

The museum, and we, wanted to provide users with the freedom to walk around and explore the chapel. This of course requires position tracking, which we deemed technically and economically unrealistic. The compromise was to re-locate the viewpoint to predefined circles marked on the floor in the visualization, allowing the users to move and experience the model from multiple locations.

To handle in-application translation changes we implemented a method for jumping between pre-defined locations in the virtual chapel via the user interface. These predefined locations are depicted in Figure C.5. The user must place himself accordingly in the geo-physical world to experience the link between the real and virtual environment. With this interaction method,

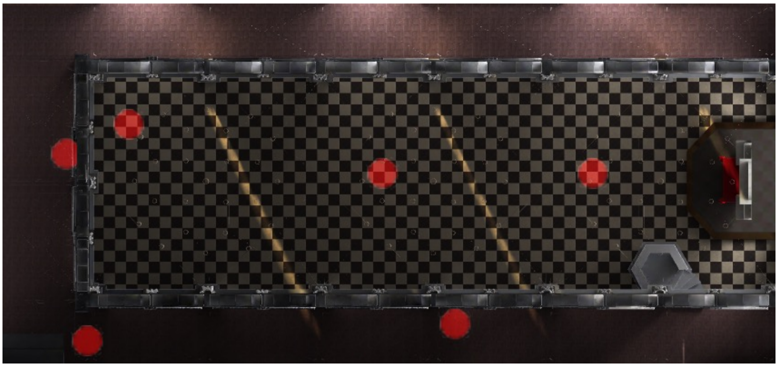


Fig. C.5: Depiction of the floor plan of the chapel, with red marking visualizing areas the user can translate to and experience the visualization from.

the user is able to experience the chapel from multiple locations using a low-tech translation approach.

For this project, we develop the software using the Unity game engine[1], allowing for efficient development for multiple platforms: iPad tablets initially, with option to easily deploy to other platforms later.

In order to process the visualization on the iPad at interactive frame-rates, we had to set strict limits on polygon-count for the visualization, limit the amount of draw calls and static objects to be rendered for each frame, and pre-compute most of the lighting information to light maps split into direct and indirect lighting. The latter will be elaborated later in the paper, when discussing the shading of the model. To a large extend objects should be combined to reduce individual calls to draw objects. As Unity supports a maximum texture size of 2048x2048 for mobile devices, this set a natural restriction in the size of objects, without having to use multiple texture maps per object. In a trade-off between detail and real-time visualization of the model, the polygon count for the model was reduced to an acceptable level, which allowed for a high amount of details from the possible viewport position. This estimate was determined subjectively be the developers on a per-object basis. The overall polygon count for everything in the viewport never exceeds 150.000 polygons at any one time for any position and view direction on this hardware[2].

---

[1] http://www.unity3d.com/

[2] We used Apple's 3rd generation iPad featuring an Apple A5X chip (Dual-core 1 GHz Cortex-A9 processor with a PowerVR SGX543MP4 GPU), and a 2048x1536 (264 ppi) resolution display

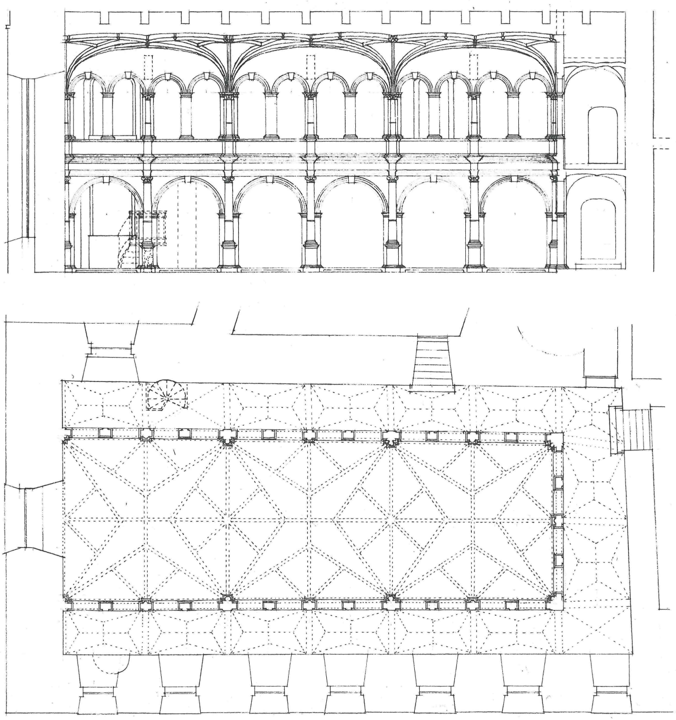


Fig. C.6: Example scanning of illustrative literature on Koldinghus Chapel interior. Here top and side view of the chapel.

4.3 3D model generation

For the 3D visualization of the chapel, information was collected in three ways. 1) research in literature, 2) scanning and modelling from artifacts available at Koldinghus and 3) informed guesses to fill in the gaps.

For the first part, research in literature, information was collected from available books about the chapel, informational posters and paintings in general from the castle which could aid in the generation of the virtual model (Figure C.6). There are no paintings or detailed informational drawings of the chapel from prior to the castle burning. The majority of the information available on the castle and the chapel is from research and reconstruction drawings.

Secondly, the scanning and modelling process was separated in three parts, 1) the core dimensions of the chapel and window placements were acquired by laser distance meter, and the room manually modelled in these dimensions in Maya[3], 2) details manually modelled from existing historical

---

[3] http://www.autodesk.com/products/autodesk-maya/overview

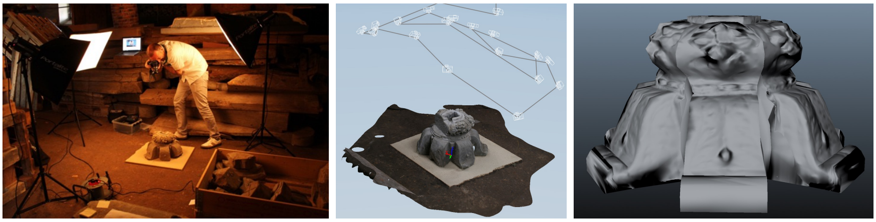


Fig. C.7: Process of generating the virtual content from image capture using 123D Catch software. The generated high-res model is then manually remodelled into a low-res object plus a texture and normal map.

sources and drawings, and 3) whenever possible, details of stone ornaments were used by 3D reconstruction from images (Figure C.7) using 123D Catch[4] from Autodesk.

Third, as there unfortunately are no paintings or detailed drawings of the chapel from the period before the castle burned, a lot of information has been lost. In order to compensate for this lost information, in the visualization we have filled in the gaps of missing information with "best guesses" of how it probably might have appeared in 1604. History informs us that there are more useful data than what is present at Koldinghus. Frederiksborg Castle Chapel was built shortly after Koldinghus Chapel, ordered by the same king. In an attempt to further expand the knowledge of the interior of Koldinghus Chapel at the time, Frederiksborg Chapel was used as inspiration to the generation of the virtual model for areas in which there were limited or no information available of the true decor at Koldinghus Chapel. Additionally, the alter placed at Vor Frue church in Aalborg, was sculpted by the same sculptor as the original alter in Koldinghus. This alter at Vor Frue church was destroyed in a fire in 1902. There exists pictures of this alter from that time, which was heavily used as inspiration for the visualization of the Koldinghus Chapel alter.

The final visualized model is a results of a compromise between what is factually known about Koldinghus Chapel prior to the fire in 1808 and what is by experts considered to be a very plausible appearance considering what was ordered built by the king in other areas of the country during the same period as building the chapel at Koldinghus. It is presented to the visitors as this compromise, with a supporting physical note stating that this is a "best guess" representation of the chapel. Giving visitors the option to choose between "known" and "best guess" options in the application is considered a future implementation.

---

[4] http://www.123dapp.com/catch

4.4 Shading

As stated above much energy was put into creating a realistic 3D model of the chapel, but all this effort is in vain if the model is not rendered with a high degree of realism. An essential part of the aesthetics of architectural visualization is in how the light travels through the space. In a case such as this, the illumination is fundamental in creating the right atmosphere, i.e., a lush renaissance chapel. In Figure C.8 we illustrate how our illumination rendering adds in creating the right atmosphere for the chapel. Obviously, real-time full global illumination rendering is not computationally realistic, especially on a mobile device. Luckily, there are some constraints that can be utilized: 1) the scene is static, and 2) the museum is satisfied with a visualization based on a fixed time of day, i.e., the direction vector to the Sun can be treated as a constant. With these two constraints/assumptions it would be natural to opt to pre-compute the entire illumination. Nevertheless, since the user can re-locate the viewpoint to various locations within the chapel, the viewpoint-dependent effects (specular reflection) must be rendered at run-time. We therefore propose to pre-compute all view-independent lighting effects, and only render specular reflection in real-time.

A formal description of our approach to achieving this takes a starting point in the rendering equation, [6], describing the reflected radiance, $L(x, \vec{\omega}_o)$, at a point x in the scene in a certain observation direction, $\vec{\omega}_o$:

$$L(x, \vec{\omega}_o) = \int_{\Omega} f_r(x, \vec{\omega}_i, \vec{\omega}_o) L_i(x, \vec{\omega}_i)(\vec{\omega}_i \cdot \vec{n}(x)) d\vec{\omega}_i \tag{C.1}$$

Where Ω is the hemisphere defined by the surface normal at the location, f_r is the Bidirectional Reflectance Distribution Function (BRDF), and L_i is the radiance from an incidence direction given by $\vec{\omega}_i$.

We pre-compute the direct and the indirect illumination parts of the view-independent illumination in Maya and store it in separate light maps, for reasons that will be explained shortly. The light maps are rendered with a standard daylight model, and the chosen date and time is November 3rd, 2012 at 14:00. This choice gave an aesthetically pleasing fall of light through the main windows in the wall facing West. Given the pre-computed light maps the rendering equation can be re-written as:

$$\begin{aligned} L(x, \vec{\omega}_o) = {} & \frac{\rho(x)}{\pi}(LM_i(x) + LM_d(x)) + \\ & k_s(x) L_s(x)(\vec{\omega}_o \cdot \vec{R}_s)^{\alpha} \end{aligned} \tag{C.2}$$

Where $LM_i(x)$ and $LM_d(x)$ are the indirect and the direct illumination light maps, respectively, which in radiometric terms store irradiance information. The diffuse part of the BRDF is represented with the albedo, $\rho(x)/\pi$.

The specular reflection is simplified as there is only one light source, namely the Sun. The specular reflection is modeled from the Phong reflection model with a specular reflection coefficient, k_s, the radiance of the Sun, L_s, a geometry term with the dot product of the observer direction and the reflection direction for the Sun, R_s. Given the almost infinite distance to the Sun this direction is not position dependent.

The specular contribution in eq. C.2 is computed in real-time in a fragment shader (Figure C.9 is an example of this). The specular reflection coefficient is manually tuned to get a desired glossy appearance of especially the floor tiles. The position (fragment) dependent incident sun radiance, L_s, is also manually tuned, but the real challenge is that obviously in most positions inside the chapel, the Sun is not directly visible, i.e., many points are not illuminated directly by the Sun. We handle this in the shader implementation by thresholding the value read from the fragment's direct light map (hence the need for having separate direct and indirect maps). If the values are above a certain low threshold, the fragment is in direct light and the specular contribution is computed. If below the threshold, no specular contribution is added. In the shader the albedo is read from the texture map. Normal maps are used in conjunction with the geometry when rendering the light maps, but not in the shader, as the normal information is already taken into account in the light maps.

5 Evaluation

Usage data of the application has been logged on the deployed devices during the period from March 3rd to April 24th. The purpose of the logged data is to investigate which areas within the chapel are of interest to the users. This allows for further development of the application, with focus on which areas should be augmented with additional information of relevance to chapel, in order to inform or inspire visitors to learn historical facts from the chapel.

Orientation data from the iPad has been logged with a resolution of 0.1° at 0.5 second intervals, throughout the day. This gives us an idea of the areas of interest with sufficiently high accuracy. Figure C.10 shows a long-lat map of the visualization of the iPads as well as a plotting of the iPad orientation in relation to the virtual model from a specific position in the virtual space. It is relevant to consider to what extend the users' interest in specific areas is representative for what they are experiencing in the virtual space, or do they aim to connect it to the physical space. The data presented in the figure is only of value if we can assume the users are positioned physically within a small radius of the virtual position in the chapel. This question is a case for further studies following this project.

The degree of exploration of the virtual space is an interesting observation

Fig. C.8: Direct lighting, indirect lighting, and a combination of direct and indirect light in an example scene of the visualization of the chapel.

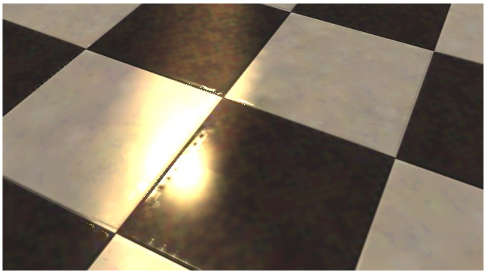


Fig. C.9: Example of real-time computed specular reflections in the visualization.

from the data, for one (or more) of three possible reasons:

1. It would appear that visitors either are satisfied with the interaction and information from the horizontal plane, or that they simple do not give much attention to the ceiling and floor.

2. Perhaps vertical motions with handheld devices are unfamiliar for most

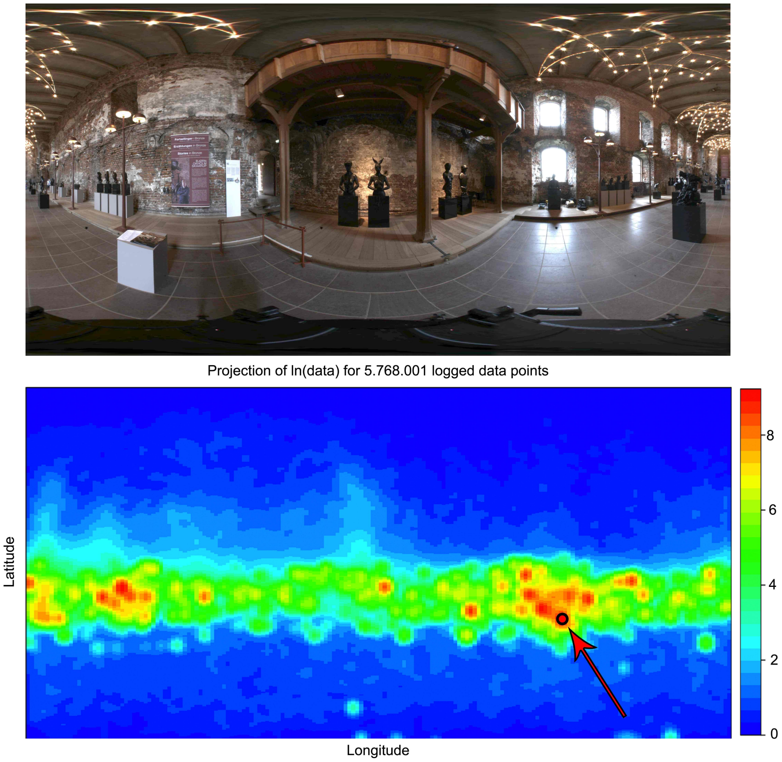


Fig. C.10: The virtual model of the chapel at Koldinghus expressed in longitude-latitude format (top). The sculptures depicted were part of a temporary exhibitions, and not permanently in the chapel. Additional orientation data logged from one specific logged location, plotted on a correlating map (buttom). The red dot illustrates the location of the charging station.

visitors, making it seem out of place for people to doing so in a public space.

3. Maybe the visitors are simply not curious for exploring the area.

An observation mentioned by the staff at Koldinghus, is that they noticed visitors appearing to be very interested in the application, but did not interact with or touch the tablet, which in stationary mode is displaying a bare wall. Next to the stand there is a clear short written guide stating the purpose of the installation which makes it clear that the main purpose of the installation is for the device to be actively used. We can only speculate the reasons for this. One argument is that visitors are used to artifacts on a museum is to be seen and not touched, and thus they, based on prior experience, observe objects from this mental notion, either in a conscious or subconscious way. Another argument is that these visitors simply focus on the installation and are not paying attention to any writings near the installation. This could be interesting to consider for further investigations in facilitating information using AR technology.

6 Further studies

In the current version of the application, the user has the ability to move around between 6 different locations in the chapel and have the application follow along to this position. This gives the user options for a more wide and free use of the application and allow for the possibility to come in close to objects of interest, or to inspect the chapel from the gallery at the first floor balcony. Both the virtual position and orientation are being logged continuously, in an attempt to uncover in more detail what visitors areas are interested in. In a future study, this data is coupled with user data from staff and visitors, to investigate to what degree the visitor link the virtual and physical world using a mental connection. One outcome of such a study is to give an idea what the situated simulation adds to the experience, or whether the experience would have sufficed with the same application in another setting, not linked to the visualization? And is a Situated Simulation part of AR if this is the case?

In mixed reality settings, it is unclear to what extend users are interested in experiencing the world, and how the curiosity of the user can be stimulated to further explore an area in a mixed reality setting. In order to test user exploration in virtual scenes, Madsen and Lorentzen investigated the use of visual augmentations to influence user movement within a small region of exploration [12]. User exploration was also one defining factor from Madsen et al. [13] in lessons learned from a previous project in collaboration with Koldinghus Museum to facilitate knowledge transfer through novel technology. The considerations for user exploration is interesting as this is a hard subject to gather information on. To the best of the authors' knowledge, this is an unexplored research area, however relevant for the user acceptance of the enabling technology. This work could bring attention to areas of the visualization that could benefit from additional information by adding active data, text or images as part of the application to convey information to the visitor.

Future work in the area of rendering for this application includes relieving the constraint of a fixed time and date for the Sun position. We are currently further developing techniques presented in [3], enabling us to render the chapel at any time of day, such that the user gets to experience how the illumination of the chapel changes over the course of a day. Exploring the increased visual realism and how users perceive this in a CH context is another interesting continuance to this line of development.

7 Conclusion

In this article, we have described the implementation process of a novel application leveraging on AR technology. The implementation process has been described in detail from gathering of relevant information to the construction of the virtual model in 3D.

The contributions of this paper are in detailing the visualization process of the castle chapel in the following areas: 1) 3D reconstruction from images is a mature technology, being used in this project to enable highly detailed models of objects, 2) a novel approach to combat drift in mobile applications relying on available hardware, 3) an approach to realistic rendering of global illumination for a single point in time on a mobile device, while maintaining freedom in translation and orientation.

A preliminary investigation of the application usage has been completed with interesting results of the users interest area within the frame of this situated simulation. It appears that visitors are mostly interested in looking at the virtual scene horizontally despite efforts to create a full implementation of the chapel itself. The findings point out a couple of obvious considerations and opportunities for further studies in the area of user exploration of virtual scenes, such as how to design an interaction model to direct the users focus to specific interesting artifacts or other points of interest.

Acknowledgment

This project has been partly funded by Koldinghus Museum. The authors would like to thank Curator and head of exhibitions Axel Johnson, Ph.D., and Koldinghus Museum for their collaboration and efforts in undertaking and realizing this project, and Peter Skotte for his invaluable efforts and collaboration in creating the 3D models and textures used in the project.

References

[1] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, “Recent advances in augmented reality,” *IEEE Computer Graphics and Applications*, vol. 21, no. 6, pp. 34–47, 2001.

[2] R. T. Azuma, “A survey of augmented reality,” *Presence*, vol. 4, no. August, pp. 355–385, 1997.

[3] C. Bahnsen, A. Martin dit Neuville, C. Pedersen, G. Danielle Monique Tranchet, and C. Madsen, “Realtime global illumination using com-

pressed pre-computed indirect illumination textures," in *WSCG 2012 Conference Proceedings*. Union Agency, 2012, pp. 231–238.

[4] A. Damala, P. Cubaud, A. Bationo, P. Houlier, and I. Marchal, "Bridging the gap between the digital and the physical," in *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts - DIMEA '08*. New York, New York, USA: ACM Press, 2008, p. 120.

[5] C. de Bérigny Wall and X. Wang, "Interactive Antarctica," in *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts - DIMEA '08*. New York, New York, USA: ACM Press, 2008, pp. 319–325.

[6] J. T. Kajiya, "The Rendering Equation," *ACM SIGGRAPH Computer Graphics*, vol. 20, no. 4, pp. 143–150, 1986.

[7] T. P. Kersten, "3D Scanning and Modelling of the Bismarck Monument by Terrestrial Laser Scanning for Integration into a 3D City Model of Hamburg," in *Third International Conference, EuroMed 2010, Lemessos, Cyprus, November 8-13, 2010. Proceedings*. Springer Berlin Heidelberg, 2010, pp. 179–192.

[8] B. Koleva, S. R. Egglestone, H. Schnädelbach, K. Glover, C. Greenhalgh, T. Rodden, and M. Dade-Robertson, "Supporting the creation of hybrid museum experiences," in *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*. New York, New York, USA: ACM Press, 2009, pp. 1973–1982.

[9] G. Liestøl, "Situated Simulations Between Virtual Reality andMobile Augmented Reality: Designing a Narrative Space," in *Handbook of Augmented Reality*, B. Furht, Ed. New York, NY: Springer New York, ch. 14, pp. 309–319.

[10] ——, "Augmented reality and digital genre design — Situated simulations on the iPhone," in *2009 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media and Humanities*. IEEE, Oct. 2009, pp. 29–34.

[11] G. Liestøl, T. Rasmussen, and T. Stenarson, "Mobile innovation: designing and evaluating situated simulations," *Digital Creativity*, vol. 22, no. 3, pp. 174–186, Sep. 2011.

[12] C. B. Madsen and L. W. Lorentzen, "Using Visual Augmentations to Influence Spatial Perception in Virtual Representations of Real Scenes," in *The 8th Annual International Workshop on Presence, PRESENCE 2005*.

London, United Kingdom: Department of Computer Science, Aalborg University, 2005, pp. 351–352.

[13] C. B. Madsen, J. B. Madsen, and A. Morrison, "Aspects of what makes or breaks a museum AR experience," in *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*, no. November. IEEE, Nov. 2012, pp. 91–92.

[14] D. Michael, N. Pelekanos, I. Chrysanthou, and P. Zaharias, "Comparative Study of Interactive Systems in a Museum," in *Proceedings of the Third international conference on Digital heritage*. Springer-Verlag Berlin, 2010, pp. 250–261.

[15] T. Miyashita, P. Meier, T. Tachikawa, S. Orlic, T. Eble, V. Scholz, A. Gapel, O. Gerl, S. Arnaudov, and S. Lieberknecht, "An Augmented Reality museum guide," *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 103–106, Sep. 2008.

[16] Nationalmuseet, *Kirkerne i Kolding: Slotskirker på Koldinghus*, ser. Danmarks Kirker. Frederiksholms Kanal 12, 1220 København K: Nationalmuseet, 2009, vol. 11, (Book; In Danish).

[17] D. Pletinckx, N. Silberman, and D. Callebaut, "Presenting a monument in restoration," in *Proceedings of the 2001 conference on Virtual reality, archeology, and cultural heritage - VAST '01*. New York, New York, USA: ACM Press, 2001, p. 197.

[18] M. Russo and G. Guidi, "Reality-based and reconstructive models: digital media for cultural heritage valorization," *SCIRES-IT*, vol. 1, no. 2, pp. 71–86, 2011.

[19] A. B. Tillon, I. Marchal, and P. Houlier, "Mobile augmented reality in the museum: Can a lace-like technology take you closer to works of art?" in *2011 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities*, no. Figure 1. IEEE, Oct. 2011, pp. 41–47.

[20] M. Trapp, A. Semmo, R. Pokorski, C.-d. Herrmann, J. Döllner, M. Eichhorn, and M. Heinzelmann, "Communication of Digital Cultural Heritage in Public Spaces by the Example of Roman Cologne," in *Proceedings of the Third international conference on Digital heritage*. Springer-Verlag, 2010, pp. 262–276.

[21] W. van Eck and Y. Kolstee, "The augmented painting: Playful interaction with multi-spectral images," in *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*. IEEE, Nov. 2012, pp. 65–69.

[22] D. M. White and K. Walczak, “Cultural heritage applications of virtual reality,” *Proceeding of the Eighth International Conference on 3D Web Technology, Web3D*, pp. 182–183, 2003, (Workshop overview).

[23] R. Wojciechowski, K. Walczak, M. White, and W. Cellary, “Building Virtual and Augmented Reality museum exhibitions,” in *Proceedings of the ninth international conference on 3D Web technology - Web3D '04*, vol. 1, no. 212. New York, New York, USA: ACM Press, 2004, pp. 135–144.

[24] E. Woods, M. Billinghurst, J. Looser, G. Aldridge, D. Brown, B. Garrie, and C. Nelles, “Augmenting the science centre and museum experience,” in *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Austalasia and Southe East Asia - GRAPHITE '04*, no. Figure 3. New York, New York, USA: ACM Press, 2004, pp. 230–236.

[25] M. Zöllner, J. Keil, H. Wüst, and D. Pletinckx, “An Augmented Reality Presentation System for Remote Cultural Heritage Sites,” in *10th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, 2009, pp. 112–116.

Paper D

How Wrong Can You Be: Perception of Static Orientation Errors in Mixed Reality

Jacob B. Madsen
Rasmus Stenholt

The paper has been published in the
Proceedings of 2014 IEEE Symposium on 3D User Interfaces (3DUI), pp. 83–90,
2014.

The layout has been revised.

Abstract

Tracking technologies are becoming an affordable commodity due to the wide use in mobile devices today. However, all tracking technologies available in commodity hardware is error prone due to problems such as drift, latency and jitter. The current understanding of human perception of static tracking errors is limited. This information about human perception might be useful in designing tracking systems for the display of AR and VR scenarios on commodity hardware. In this paper we present the findings of a study on the human perception of static orientation errors in a tracking system, using two different setups leveraging a handheld viewfinder: a classical augmented scenario and an indirect augmented one. By categorizing static orientation errors by scenario and local orientation axis, new insights into the users' ability to register orientational errors in the system are found. Our results show that users are much more aware of errors in classical AR scenarios in comparison to indirect AR scenarios. For both scenarios, the users registered roll orientation errors differently from both pitch and yaw orientation errors, and pitch and yaw perception is highly dependent on the scenario. However, the users performance ranking for orientational errors in AR scenarios was unexpected.

1 Introduction

As smartphones and tablets are becoming commodity hardware, equipped with numerous sensors, such devices have rapidly become an attractive platform for applications augmenting our everyday lives, ranging from GPS navigation to mixed reality (MR) gaming and location based browsers. One of the principal goals of most mixed reality applications is to deliver a convincing experience to the user in merging the real and virtual worlds. Errors in the tracking system are often one of the major factors in diminishing the overall perception or sense of presence of the experience for the user. These tracking performance errors can be split into two categories, dynamic errors such as measurement noise and jitter and static errors such as spatial distortion, calibration errors, and stability errors such as slowly accumulated drift. These are persistent problems in tracking systems, which can limit the usability of mixed reality applications. Therefore, researchers are working hard to overcome these problems [2, 3, 13, 23].

While there is no single best solution for motion tracking on mobile smart devices in unprepared environments, a possible design goal for any tracking system, according to Welch and Foxlin [20] is that: "Tracking artifacts remain below the detection threshold of a user looking for them." However, to the best of our knowledge, the field of human perception of tracking errors is still largely unexplored. Swan and Gabbard survey user-based experimentation within augmented reality (AR) [18], while mentioning nothing of studies or reports on user evaluation of current tracking or estimation accuracy. Also in a review by Zhou et al. [23], several studies with research in tracking techniques as well as hybrid tracking systems are surveyed. However, nothing on evaluating human perception of tracking accuracy is mentioned within these surveys of the field.

We present a study of human perception of static orientation tracking errors in video see-through augmented reality (classical AR) and indirect augmented reality (indirect AR), with an experiment designed to uncover the lower boundaries for human registration of tracking errors, specifically static orientation errors, in both a classical AR and an indirect AR setup. As any of the problems inherent to tracking systems affect this study as well, we present an experimental setup that to the best of our abilities attempts to overcome these problems in a laboratory setting.

In this paper, the term indirect AR is used to indicate the presentation of a purely virtual scene in a physical setting that matches the displayed virtual representation from some viewpoint, as defined by Wither et al in [21]. They present a system for displaying pre-captured panoramas to the user instead of the real camera feed. With this "indirect augmentation", it is only a convincing augmentation when viewed at its corresponding real location. In some cases, indirect AR is also known as a situated simulation [15].

Human perception of orientation tracking errors in a tracking system for commodity hardware may be useful in guiding the design of tracking systems for classical and indirect AR purposes. Knowing these boundaries might even relax the demands on the tracking system for some applications.

This paper is organized into the following sections: In Section 2, an overview of related works is given. In Section 3, the hypotheses that formed the basis of the experiments are presented and motivated. This is followed up by a detailed description of the main experiment and its follow-up in Sections 4 and 5. After this, the results of both experiments are presented together in Section 6 along with a discussion of the significance of these results. Finally, a conclusion is presented in Section 7.

2 Related works

Surveys of MR indicate that, in general, tracking systems are a broad and well-researched field, and motion tracking is a hard problem with no fixed solution for all cases, as examined by Welch and Foxlin, who explain the problems of motion tracking in great detail in [20]. This is also adressed by Azuma [2, 3] and van Krevelen [13], both of whom state that this is a complex problem.

Multiple studies have evaluated and analyzed tracking performance from a technical perspective [4, 10–12]. E.g. Gilson et al. [10] performed a quantita-

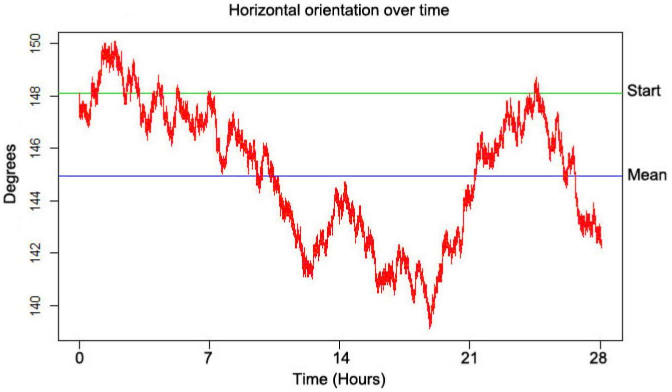


Fig. D.1: Illustration of yaw-drift from an iPad3's gyro, where the device is standing still over the course of 28 hours.

tive analysis of tracking systems and found that tracking performance deteriorates when the tracked object's speed increases. Others have attempted to estimate and fix these errors, such as Caarls et al. [6] who presents a framework for leveraging multiple sensors simultaneously to achieve a more precise and robust tracking system, providing 1 cm z-accuracy for distances up to 120 cm and small roll errors at distances under 70 cm.

In non-optical tracking systems for indirect AR or virtual reality (VR), inertial sensors are in many cases used for orientation estimation, where 3 degrees-of-freedom (DoF) tracking is sufficient, with examples being situated simulations for tablets/phones or VR environments and games for modern head-mounted displays (HMD) such as the Oculus Rift [1]. As depicted in Figure D.1, the gyroscope in consumer products is prone to drift over time. Thus, many researchers have focused on minimizing this problem [5, 22]. One example is Won et al. [22] who presents a tilt angle correction method for handheld devices that detects if the system is stationary, and uses the gravity vector to stabilize and correct the yaw component, whereas the roll and pitch angle change in relation to the acceleration values. For dynamic movements, the tilt angles are corrected accurately, but the yaw angle shows no significant improvement with the proposed method.

In any tracking system, latency can be a problem, unless all tracked objects remain stationary. The size of the problem depends on the nature and size of the latency, as explained by [1, 8, 17, 19]. A small and constant latency might not be a problem, whereas any non-constant latency or a large latency poses a problem in any setup. Mania et al. [17] presents the results of an experiment of users experiences and perception of latency with varying de-

---

[1] http://www.oculusvr.com/

grees of scene complexity, and finds that the just noticeable difference (JND) is 15 ms or less. This is in line with earlier investigations, such as Adelstein et al. [1], who measured 8-17 ms and Ellis et al. [8], who found an average of 11.6 ms to be the JND.

Swan and Gabbard [18] and Kruijff et al. [14] review literature on user-based experimentation and perceptual issues respectively. In [18], the authors describe human perception and cognition research within AR, and note relevant examples of depth perception research to be taken into consideration when designing the experiment presented in this paper. In one example of depth perception and distance perception research by Ellis and Menges [9], the authors found that objects in the near field tend to suffer from perceptual localization errors in x-ray or monoscopic setups. Even though depth estimation is not a part of this study, it should be noted that there are no real or virtual objects in the field between the viewfinder and the objects of interest, i.e. the objects that participants are asked to align. Kruijff et al. [14] show that there is a current lack of research within the evaluation of human ability to notice tracking errors. This is in spite of the fact that human performance on occlusion handling, x-ray rendering, visual quality, depth perception, and accommodation are all areas of interest within the research community.

Livingston and Ai [16] present a user study on registration errors, i.e. latency, noise, and orientation errors. Their experiment focused on evaluating user performance in an AR environment using an optical see-through HMD. By adding high or low error variables to the system parameters (the latency, noise, and orientation), they found noise to have a limited impact on user performance, despite being displeasing in a subjective sense. Latency was shown to have a significant impact on localization performance, with users being slower under high latency in comparison to low latency. Orientation errors did not present a significant difference in localization accuracy. This provides a great step in the direction of expanding the knowledge of user performance in relation to registration errors. In this study, we look into the lower boundaries of orientation errors visible to the users. Setting a lower boundary might influence the level of artificially added offsets in future performance studies on user performance and registration errors.

3 Hypotheses

The experiments presented in this paper are based on a desire to test two main hypotheses:

1. It is more difficult to perceive static orientation errors in indirect AR than in classical AR.
2. There is a difference between the perception of static orientation errors w.r.t. yaw, pitch, and roll.

(a) There is a difference between mean errors w.r.t. yaw, pitch, and roll.

(b) There is a difference between error variance w.r.t. yaw, pitch, and roll.

Hypothesis 1 can be reasonably justified by two facts: 1) A person viewing an indirect AR scenario will *not* have the direct, pixel-to-pixel correspondence between the virtual and real worlds produced by having a live camera feed on the screen. This means that the viewer will have to resort to using his/her spatial abilities and a mental mapping between the virtual and the real world. 2) In the most common form of indirect AR, there is no tracking of the viewer's position relative to the screen. This means that the view of the virtual world seen on the screen will only be absolutely correct from a single viewpoint. This is furthermore complicated by the fact that the user does not know where this virtual sweet spot is. Instead, the only option is to rely on head motion to find the sweet spot, if the viewer does not want to rely purely on spatial imagination to mentally blend the virtual and real scene.

With respect to hypothesis 2, we believe that to be supported by the fact that the human sense of balance is governed by an external reference force, i.e. gravity. This allows people to gauge their own orientation as well as the orientation of other objects in relation to the local direction of gravity. This makes us well-equipped, even in the absence of technical equipment, to deal with tasks involving corrections on the roll and pitch axes. E.g. most people can tell if a picture hangs reasonably straight on a wall, without the use of a spirit level. Similarly, the sense of balance reliably tells people if they are falling forwards or backwards. This is contrary to the yaw axis, where there is no absolute, external reference that can be sensed by humans to use as guide. E.g. people in a windowless room will likely not be able to tell their absolute heading, without the use of a compass. For these reasons, we predict that the perceptible errors on the yaw, pitch, and roll axes will not be the same, neither in terms of accuracy (mean error), nor precision (error variance).

4 Experiment 1: Classical and indirect AR

The first experiment is focusing on both classical AR and indirect AR. It investigates the lower boundaries for human perception of static orientation errors in both scenarios, and simultaneously investigates the effects of the error on the 3 local rotational axes. The purpose of the first experiment is to find the absolute minimum boundary for human error perception when calibrating an offset orientation in one axis. Thereby we also hope to find the lower threshold for user perception of registration errors.

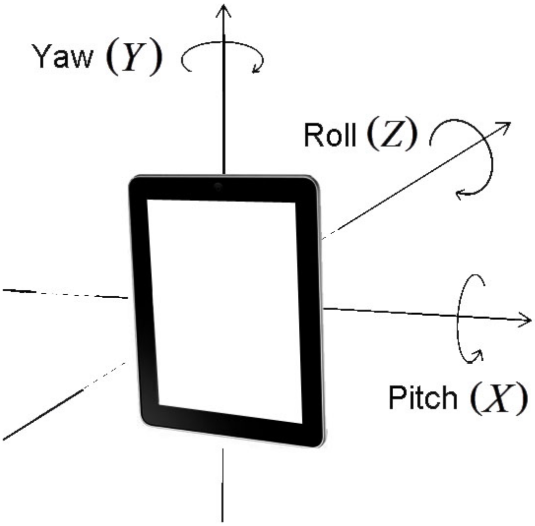


Fig. D.2: An illustration of the rotation axes relative to the tablet used in the experiment. Rotation around the local x-axis is named pitch, local y-axis rotation is named yaw, and local z-axis rotation is named roll.

4.1 Method

In evaluating user perception of static orientation errors, the attitude of a tablet device is used to describe the orientation, i.e. yaw, pitch, and roll angles, as depicted in Figure D.2. After initial calibration, the attitude displayed is offset from the calibrated attitude to simulate a static orientation error. In order to simplify the experiment, and to make the task easier for the participants, only a single axis is offset in a single trial, since early pilot testing revealed that simultaneous calibration of multiple axes greatly increased the overall difficulty of the task.

The error term is defined as the difference between the final user input and the actual offset in the system from the calibrated ground truth. This implies that an error of 0° is a perfect solution to the task. It seems reasonable to assume that the these errors will have a mean value of 0° across all participants and conditions. Any departure from this assumption in the experimental data will indicate that results are somehow biased. However, it is not useful to analyse the raw errors to find the desired lower boundary for perceptible errors. In order to find the desired bound, we instead use the absolute values of the errors, since errors of e.g. -1.35° and 1.35° both are equally wrong in terms of magnitude, and are equally far from the calibrated ground truth. Both raw errors and absolute errors are logged, such that it is both possible to detect any bias and the desired perceptual bound.

We suggest that the task of asking users to correct an artificial, static tracking error can be expected to uncover where the lower threshold of error perception is. If the user can detect an error, then he/she is expected to continue to correct it, until no error is perceived anymore. When the user stops correcting the error, it is therefore resonable to assume that the current error is below what can be perceived by that person. Furthermore, in a realistic usage scenario, it is expected that an error of similar magnitude will also not be perceived by the same person. This is especially true, if the user has not been instructed to be alert of any errors in the realistic usage scenario.

To reduce any initial estimation error in the optical tracking system, the system was manually calibrated by the experimenters prior to any participant interaction. This calibration step only took place once, after which the calibrated setup was left unmoved.

The procedure is that each participant is placed in a chair in front of a tablet device (see Figure D.3). Then the participant is introduced to the purpose of the experiment, and the controls of the application. We embedded a control system for this experiment within the application, allowing the user complete control of the rotation on a single predefined axis according to the current task. I.e. the user is only able to adjust the axis currently offset from the calibration. The control system allows for adjustment of angular orientation in increments down to $\pm 0.0001^\circ$. The controls are shown and explained in Figure D.4.

Prior to the actual experiment, the user goes through a training stage, in which random trials are presented in a configuration which does not occur in the actual experiment to mitigate any learning bias. Once the user is confident with the controls and the purpose of the experiment, the actual experiment commences. During the experiment, the participants were allowed to take breaks or ask questions, if needed.

All participants were told to solve the given tasks as well as possible, not paying any attention to the time spent. This was done in order to get the participants to emphasize quality over speed in their responses.

After completing all trials, the participants responded to a small questionnaire surveying their subjective evaluation of their performance along with explaining their strategies for solving the tasks for both the classical AR and indirect AR scenarios.

4.2 Materials

The experimental software is developed for tablet devices. The experiment is carried out using an iPad3[2] (2048x1536 resolution at 264 pixels/inch) using the back-facing camera (5 MP, 54.4° vertical field of view and 42.0° horizontal

---

[2] Featuring an Apple A5X chip (Dual-core 1 GHz Cortex-A9 processor with a PowerVR SGX543MP4 GPU)

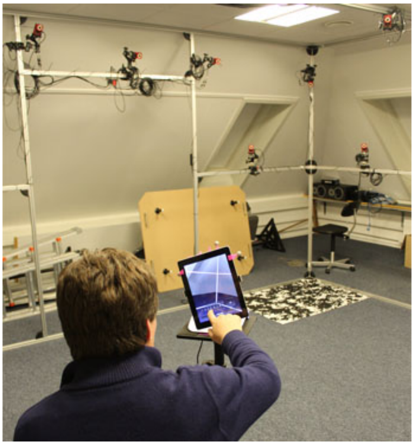


Fig. D.3: Example of the lab setup. The participant is currently working on an indirect AR task, hence all of the contents on the screen are virtual. The iPad stand is placed near the center of the room on a platform, facing one corner in which the trackable marker is placed, to ensure correct and stable tracking. The marker is printed on A0 format paper (841mm x 1189mm canvas paper). The user must then re-orient the offset rotation of the visualization towards the calibrated setting.

field of view) for tracking (indirect and classical AR) and background display (classical AR only). We generated the classical AR and indirect AR scenes using Unity3D[3] as the rendering engine, and Vuforia[4] for camera background (for the AR part) and as marker detection and tracking system. A trackable surface was created for the Vuforia tracking system, and printed as an A0 non-glossy poster (841mm x 1189mm canvas paper) placed at a 2 meter distance to the iPad stand (Figure D.3).

The laboratory is specifically chosen for this experiment in order to provide a manageable setting with many options for comparing orientations displayed on the device with the setup of the interior of the laboratory. The main feature of the laboratory is a $4.5\text{m} \times 4.5\text{m} \times 2.5\text{m}$ aluminium cage used for mounting lab equipment. This cage serves as the main feature of the virtual scene in both scenarios. For the classical AR scenario the image captured by the built-in camera is rendered as background with the aluminium cage augmented on top, and in the indirect AR scenario, a simplistic model of the room is rendered in addition to the cage. To make this model feasible to produce, it only included the major features of the room, i.e. floor, ceiling, walls, windows, door, and large pieces of furniture, but not any smaller objects lying around the room. We have not investigated the consequences of this choice of 3-D model, but it may be an interesting subject of study in the

---

[3] http://www.unity3d.com

[4] http://www.vuforia.com

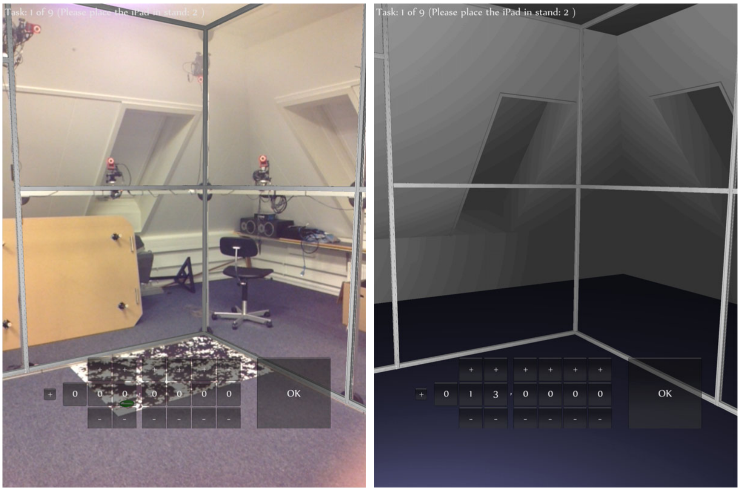


Fig. D.4: Illustration of the two scenarios, [Left] classic AR and [Right] indirect AR. Note that in the classical AR scenario, only the grey aluminium cage is virtual, whereas in the other scenario, all screen contents are virtual. The GUI controls used by particpants for re-calibrating the errors are also visible in both screenshots. For each digit of the user input, there is a plus and a minus button, allowing for precise control of every digit from ±10° down to ±0.0001°. Once the user is satisfied that the calibration error is gone, the large OK button is pressed.

future.

The cage was chosen as the main object, as the beams are linear objects connected at right angles, allowing for many opportunities in choosing which parts of the scene to use as calibration targets, such as horizontal/vertical beams, or the joints of beams at the corners of the cage. Modelling the cage is simple, since it is of known dimensions, cut to a precision of ±0.5mm. The room and its major features were modelled in Google SketchUp and Autodesk Maya to be imported into Unity3D. Figure D.4 presents screenshots of the two scenarios tested in the experiments, i.e. classical and indirect AR.

A single iPad standing on a raised platform is used for all trials as shown in Figure D.3. All user translation and orientation of the device is disallowed. As the setup is static and never moves, the estimation performance should be static, and latency and jitter, as well as other dynamic errors, are eliminated.

The iPad3 provides an average angular resolution of about 70 pixels per degree at a viewing distance of 38 cm, which is a realistic viewing distance in the experiments of this paper[5]. This number can be calculated using Equation D.1. If the viewing distance is d and the height of the iPad is s, we get a total visual angle of $\alpha = 29.06°$. Using the iPad3's vertical resolution of 2048

---

[5] The angular resolution varies non-linearly across the field-of-view, because the screen is flat. However, this variation is quite small since the chosen field-of-view angle is also reasonably small.

pixels, we get 2048/29.06 pixels/° = 70.47 pixels/°.

$$\alpha = 2 \cdot \arctan(\frac{s}{2d})\frac{180^\circ}{\pi}\ [^\circ] \tag{D.1}$$

Given that the number of pixels per degree is higher than the human threshold according to [7], we assume that participants can, at least theoretically, detect a single pixel change from controlling the orientation in the application. The lower limit for in-app rotation needed for a single pixel change is more difficult to estimate, due to aliasing. Even the slightest rotation may contribute to one or more pixels changing value, if the involved pixels are currently located right on the verge of a change.

We therefore do not assume a minimum rotation threshold that changes a pixel value. However, for the virtual camera which has a FoV corresponding to the real camera (54.4° vertical field of view) and the iPad having an output resolution of 2048 pixels vertically, for a rotation of approximately $54.4^\circ/2048 = 0.026^\circ$, every virtual object will have shifted a minimum of 1 pixel on the screen. Allowing the user to control up to 4 decimals after the decimal point is considered reasonable to allow for the participants to accurately recalibrate the system.

4.3 Study design

Experiment 1 is a randomized, within-subjects experiment with 3 repetitions of 30 different trials for each participant, giving a total of 90 trials per participant. The factors of the experiment are scenario (2 levels; classical or indirect AR) and rotation axis (3 levels; yaw, pitch, and roll). Each combination of these factors are tested at 5 different, random initial offsets from the ground truth calibration. Inside each block of 30 trials, the order is completely randomized, meaning that classical and indirect AR trials are randomly distributed among each other. The offsets used are all in the range $\pm[10;30]^\circ$, implying that the smallest initial offset is 10°, and the largest possible initial offset is 30°. This is done to ensure that 1) there clearly is an error to correct and 2) the participant does not get confused about what part of the virtual cage to match to the real one (the cage is symmetrical).

4.4 Participants

A total of 30 (2 female, 28 male) unpaid participants took part in the first experiment. As such the data collected from experiment 1 comprises 2700 trials. All participants were students and staff recruited at the local university campus, which implies that all have a background in media technology. The mean age of the participants in experiment 1 was 24.07 years. The average total completion time was approx. 22 minutes in this experiment.

4.5 Issues

Both during and after the first experiment, it became clear from the observations and data gathered that the experiment's indirect AR scenario had an unfair advantage in comparison to a realistic indirect AR scenario. Even though the sequence of trials and conditions was completely randomized, the participants' memory of the final appearance of recent classical AR trials made them better at correcting the errors in any subsequent indirect AR trials. In the questionnaire, many participants mentioned that they learned the correct orientation after a while and as a consequence started to rotate the scene in indirect AR scenarios to match their memory of any previous classical AR trial solutions. This implies that the estimated perceptual threshold for indirect AR in experiment 1 is likely much too optimistic. However, the validity of the estimated thresholds for classical AR are unaffected by this problem. The classical AR scenario does not suffer from any such advantage, as the connection is happening directly in screen space, and the user simply has to match the virtual and the real representation on the screen. Furthermore, several participants mentioned that after the first few indirect AR scenarios, they stopped using the actual surroundings (i.e. the room) as a means of solving the task. Instead, they relied on their memory of the correct relation between the classical AR model and physical features on the iPad, or even the placement of GUI elements on the screen to complete the indirect AR trials.

Another weakness of the approach is that a static setup might result in participants using techniques for solving the task that would not be possible in a realistic usage scenario, e.g. with more degrees of freedom in handling the device. Such techniques include remembering settings from one trial to the next. For instance, given a yaw correction trial, the user might remember the correct floor placement when doing the next pitch correction task. Given that the user cannot employ these tricks in realistic settings, we believe that the experimental results will estimate the absolute lower thresholds for human perception of static orientation errors.

It is also a relevant concern that the accuracy of the experiment is limited by the calibration accuracy attainable by the experimenters and the tracking software used. As we cannot guarantee the entire setup to be perfectly calibrated, the calibration can only be performed to the level where the experimenters cannot reliably tell the difference in the best case (classical AR). Through early pilot testing, this limit was found to be somewhere around $\pm 0.01^\circ$. For this reason, any results of the experiment that go below this limit should be treated with caution.

One final difficulty in the experiment is the spatial, mental mapping in the indirect AR scenarios. In comparison to the classical AR scenarios, where the connection of the real and virtual scenes happens on screen, this does not

Fig. D.5: The lab setup for the second experiment. The iPad is placed on one of three platforms, all facing one corner of the room, in which the trackable marker is placed. In this example, the iPad is placed on platform #1. During the actual experiment it is moved between the platforms in random order.

happen in the indirect AR scenario. Here, the user must mentally connect the real and virtual scenes. The indirect AR scenarios are made even harder by the fact that there is no tracking of the viewing position of the user in relation to the screen. We see this as a necessity, as this is how most indirect AR experiences are presented to users in current, realistic usage scenarios. However, in future experiments, the inclusion of head tracking is definitely a worthwhile direction to take.

5 Experiment 2: Indirect AR

In this experiment, we attempted to eliminate the possibility of relying on memory and learning when solving indirect AR tasks. Furthermore, we sought to eliminate solution tricks involving fitting the position of physical iPad features or GUI elements to virtual features on the screen. This was achieved by using three different physical platforms for the iPad during the experiment on which only one set of yaw, pitch, roll trials would be performed, and by removing the classical AR scenarios from the experiment. The new setup is illustrated in Figure D.5.

5.1 Method

The experimental setup is very similar to the first experiment. The tasks of the participants are exactly the same as in experiment 1, and the tablet is still static during each trial, i.e. all user translation and rotation of the device itself is disallowed. However, the iPad will be moved from platform to platform by the experimenters as the experiment progresses. It is still assumed that other types of tracking errors are eliminated.

The user is placed in front of a fourth iPad platform during the training stage to eliminate the possibility of any memory of trials actually used in the test. In the training stage, random tasks are presented one at a time in a configuration which does not occur in the actual experiment, until the participant feels confident about the tasks and controls.

During the experiment, the user is placed in front of one of the three platforms at a time, chosen at random. Furthermore, the sequence of yaw, pitch, and roll trials at each platform is randomized.

After completing all of the trials, the participants responded to a small questionnaire surveying their subjective evaluation of their performance, along with explaining their strategies for solving the tasks.

5.2 Materials

The materials used for the second experiment were identical to those used for the first experiment, with the exception that three different iPad platforms were used for the trials of experiment 2, along with a fourth plaform for the training session.

5.3 Study design

Experiment 2 is a randomized, within-subjects design with one factor of interest, axis (3 levels; yaw, pitch, and roll) and one blocking factor, platform (3 levels; named 1, 2, 3). The platforms are used in random order, and three trials (one for each axis) is conducted on one platform in random order before proceeding to the next. This produces a total of 9 trials per participant. This design should ensure that the benefit of learning from trial to trial should be minimized.

5.4 Participants

A total of 16 (all male) participants took part in the second experiment, all recruited from the same population of students and staff as that of experiment 1. However, no participant took part in both experiments. Since each participant performed 9 trials, the data from experiment 2 comprises 144 observations. The average age was 23.25 years in experiment 2. Each participant spent an average of approx. 7 minutes on the trials of the second experiment.

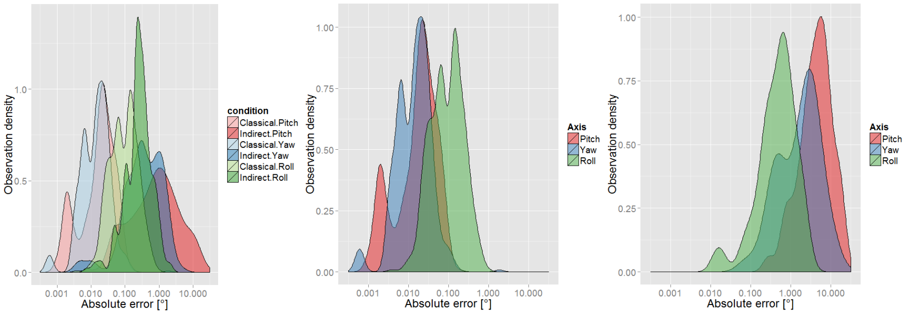


Fig. D.6: A density plot presentation of the angular error data from experiments 1 and 2. [Left] Density plot of the absolute angular errors in all conditions from *experiment 1 only*. The indirect conditions (darker regions) generally have larger absolute errors than the classical ones (brighter regions). [Middle] The classical AR condition absolute angular errors from *experiment 1 only*. The indirect AR results from experiment 1 have been removed from this plot, since these results are suspect. The roll errors (green) are larger than the yaw (blue) and pitch (red) errors. [Right] A density plot of the absolute angular errors from *experiment 2 only* (i.e. indirect AR). The roll errors (green) are smaller than yaw (blue) and pitch (red). Note that the horizontal axes are logarithmic in all the plots.

6 Results and discussion

All analyses have been carried out using the statistical software package R, using a significance level of $\alpha = 0.05$. The main analysis methods employed are type III ANOVA and the Friedman test. Post-hoc tests for pairwise comparisons following ANOVA has been performed using Tukey's honest significant difference (HSD) method. All reported confidence intervals have been computed using bootstrapping with bootstrap samples of size $n = 100000$.

None of the statistical analyses have used pooled data from experiments 1 and 2. Each experiment is separately analysed, and no cross-inferences about the results between the two are made. However, we do note that the results in the indirect AR condition in both experiments show the same general tendencies, which supports the validity of both experiments in spite of the flaws of experiment 1.

In both experiments, the collected absolute angular errors did not meet the standard assumptions of ANOVA analysis. Particularly, the normality of the residuals, and the homogeneity of the variance across the experimental conditions were found to be problematic. However, a logarithmic transformation of the angular errors completely solved this problem for the data from experiment 2, and greatly improved the situation for experiment 1. For this reason, the preferred scale for statistical analysis of the absolute angular errors is a logarithmic one. To further ensure that the conclusions were well-supported, a non-parametric Friedman test was run alongside the ANOVA to verify that no conflicting conclusions were found.

The data collected from experiment 1 and 2 is summarized in the density plots shown in Figure D.6. The difference between errors, absolute errors, and logarithmic absolute errors is illustrated in Figure D.7.

6.1 Hypothesis 1

The first hypothesis stated that people would be worse at detecting and rectifying static orientation errors in indirect AR than in classical AR, due to the missing information of the live camera feed. Running ANOVA on the data from experiment 1, this hypothesis is supported by the fact that scenario type is a significant main effect ($F_{1,29} = 677.1, p < 2.2 \cdot 10^{-16}$). This means that people are significantly worse at detecting and correcting errors in indirect AR than in a classical AR setup, in spite of the fact that the indirect AR errors from experiment 1 are likely too small to be realistic.

In experiment 1, the estimated mean absolute angular error in classical AR scenarios is 0.0610° with a 95% confidence interval of $[0.0559; 0.0670]^\circ$, whereas the same figures for indirect AR in experiment 1 are 1.06° and $[0.95; 1.19]^\circ$. In other words, the mean error is approx. 17.4 times larger with indirect AR than classical AR. As was previously explained, the conditions for indirect AR viewing in experiment 1 were in all likelihood unrealistically good, mainly due to the effect of learning from classical AR trials. This implies that the mean indirect AR figures above should be taken as an extreme best-case scenario for detecting errors. From experiment 2, the more realistic estimate of these figures in indirect AR scenarios are 3.11° and $[2.51; 3.78]^\circ$, implying that the realistic mean angular error is approx. 51 times larger in indirect AR than in classical AR.

Another way of checking this is to see the position of the least significant digit that participants chose to adjust in the two scenarios. This data is discrete by nature, so the two scenarios are compared using a Friedman test on the data from experiment 1 instead of ANOVA. The conclusion in this case is the same: People use more digits to adjust their estimate in classical AR than in indirect AR ($\chi^2_1 = 30, p = 4.3 \cdot 10^{-8}$). In the light of the analysis of the angular errors in the two types of scenarios, this means that not only are participants adjusting their responses more finely in classical AR, they are also reaching higher levels of accuracy by doing so.

6.2 Hypothesis 2

The second hypothesis stated that there would be a difference between the errors made on the three tested rotation axes, yaw, pitch, and roll. This difference was not only hypothesized to be a difference in mean error, but also in the error variance, as it seemed likely that, especially when adjusting

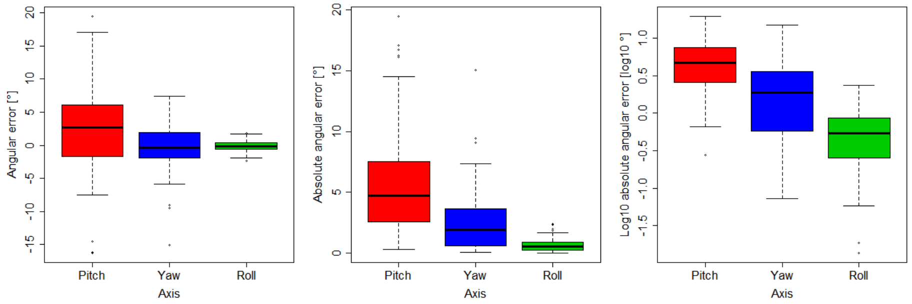


Fig. D.7: A boxplot presentation of the angular errors made in experiment 2. [Left] A boxplot of the untransformed angular errors, measured in degrees. This plot clearly shows that there is a difference in variance for the three different axes. Furthermore, the pitch axis is revealed to have an unexpected bias of about 2° that should not be there, if participants were equally likely to over- and underestimate the errors. [Middle] A boxplot of the errors transformed by taking their absolute values. This is a more useful representation, when the main concern is the magnitude of errors, rather than the direction of the errors. [Right] The absolute errors transformed by a $\log_{10}$ transformation that both has the effect of making the variance in all three axes equal as well as making the distributions on each axis closer to normal.

the roll axis, participants might get extra help from their sense of the direction of gravity.

Running an ANOVA analysis on the data from experiment 1 reveals that the axis is both significant as a main effect ($F_{2,58} = 38.56, p = 2.2 \cdot 10^{-11}$) and as an interaction with the type of scenario, i.e. indirect AR or classical AR, ($F_{2,58} = 81.1, p < 2.2 \cdot 10^{-16}$). The interpretation of this result must therefore be that the mean errors for the axes are different, and that these differences are significantly affected by the type of scenario. Following this result up by a Tukey HSD test on just the classical AR data from experiment 1, it is revealed that yaw and pitch are both significantly different from roll, but not from each other (all significant $p < 1.0 \cdot 10^{-4}$). In the classical AR scenario, the errors on the axes are significantly smaller for yaw and pitch than for roll.

With the data from experiment 2, the ANOVA results also show that there is a significant main effect of the rotation axis ($F_{2,30} = 65.81, p = 1.1 \cdot 10^{-11}$). The follow-up Tukey HSD test in this case shows that all three axes are significantly different from each other (all $p < 1.0 \cdot 10^{-5}$). In the indirect AR scenario, roll has smaller errors than yaw, which in turn has smaller errors than pitch. Thus, hypothesis 2 is strongly supported by the data from both experiments.

With respect to variance, the hypothesis that the variance would be different for different axes is also strongly supported by the data. This has been tested using Levene's test of equal variance on the absolute angular errors. The results all come out with $p \ll 0.001$. This implies that people are not equally consistent about their error estimates in all axes.

The estimated mean error responses for the three axes are presented in Table D.1, along with their associated 95% confidence intervals. The numbers in the table clearly indicate that the roll errors seem to be consistent across the two scenarios, whereas the errors on the other two axes are highly dependent on the scenario.

6.3 Other findings

There are several other interesting findings in the experiment that are not directly related to the hypotheses. First of all, when systematically asking all participants about their own perceptions of the experiment, many of them reported that they think that they somehow used a trick for some of the trials. When asked to elaborate, they told that they relied on help from aliased lines in the image (i.e. looking for the setting where a line no longer appears jagged because of the pixel grid), memory of the locations of specific linear features in the image relative to the frame of the tablet, by assuming that certain linear features on the screen had to be completely vertical or horizontal, or by expecting specific features to be centered on the screen. Most, if not all, of these tricks would not really be possible in a situation outside the lab, where the device is no longer stationary relative to the scene. For this reason, we believe that all estimated error tolerances are lower bounds, and that the error tolerance in a more realistic, dynamic situation might be somewhat larger. Furthermore, our results are unambiguous enough that these tricks have probably in most cases mainly helped in letting participants make more consistent responses, rather than more correct responses.

Another observation made in the data is that the mean (non-absolute) angular error on the pitch axis in the indirect AR condition in both experiments is approx. 2°. This is contrary to the expectation that the mean value should be around 0°, if it was equally likely to over- and underestimate the

Table D.1: A table of the estimated mean absolute angular error tolerance (with 3 significant digits), dependent on the two independent variables of axis and scenario. The estimated 95% confidence intervals for the mean values are given in brackets. All classical AR results are estimated from data gathered in experiment 1 only, and all indirect AR results are estimated using experiment 2 data only.

Axis	Scenario	
	Classical AR (exp. 1)	Indirect AR (exp. 2)
Yaw	0.0235° [0.0185; 0.0329]°	2.74° [2.01; 3.63]°
Pitch	0.0254° [0.0234; 0.0275]°	5.88° [4.59; 7.32]°
Roll	0.134° [0.123; 0.146]°	0.680° [0.525; 0.857]°
Overall	0.0611° [0.0559; 0.0670]°	3.11° [2.51; 3.78]°

error. In other words, there is an unexpected bias of about 2° for indirect AR pitch tasks. We believe that this bias is caused by the combination of two conditions: 1) The tablet was pointing slightly downwards during both experiments, and 2) several people reported using a specific trick involving the expectation of seeing an equal amount of virtual floor and ceiling in the correct setting. However, since the tablet pointed slightly downwards, the correct setting shows more floor than ceiling. Therefore, these two facts will explain why there is a general tendency to overshoot on the pitch estimates in the experiments. Interestingly, many participants have also reported that subjectively, they found the pitch trials to be much more difficult than yaw or roll. Conversely, several participants stated that the roll tasks were the easiest, which is also supported by the data.

7 Conclusion

There are several conclusions to be made, based on the performed experiments. The two main lessons learned must be:

1. People are much more perceptive of static orientation errors in classical AR than in indirect AR scenarios.
2. The ability to detect static orientation errors is *highly* dependent on the rotation axis affected by the error.
3. Roll error perception seems consistent in indirect and classical AR, but yaw and pitch errors are perceived differently in indirect and classical AR.

These points can be elaborated further. Quantitatively, the size of static orientation errors that can be detected in indirect AR is somewhere between 1 and 2 orders of magnitude larger than those detectable in classical AR. This implies that designers of indirect AR systems need not worry as much about static orientation errors, e.g. errors similar to those produced by a slow drift caused by quantization and integration errors. The fact that the axes are differently perceived means that precise orientation tracking is not equally important for all axes. In classical AR, people are more critical of yaw and pitch errors than roll errors. Conversely, in indirect AR where there is no help from a camera feed, the roll axis is the one where it is easiest to detect errors, whereas yaw and pitch are less important. It seems reasonable to speculate that, in the absence of help from a camera feed, the roll errors are much easier to detect because the roll axis is also tied to the human sense of balance which works irrespective of the imagery on the screen.

Another important contribution of this paper is the first estimation of detection thresholds for static orientation errors on each of the three tested axes, and in both classical and indirect AR. These thresholds were presented in Table D.1. The difference in thresholds for the same axis depending on the scenario is quite large in some of the cases. For instance, in the case of the yaw, the threshold in the case of indirect AR is more than 100 times larger than that in classical AR. In the case of pitch, the difference is even more pronounced, the threshold being more than 200 times larger in indirect AR than classical AR. For the roll axis, the thresholds are much more consistent across the scenarios.

This study also leaves many open questions to be answered in the future, such as the effect of allowing the tablet and the participant to move around in a realistic manner and the effect of taking the study out of a controlled lab context and into a more realistic setting. Other interesting areas of future study might be incorporation of head tracking, a higher level of realism in the virtual model, and similar studies of static position errors.

Acknowledgment

The authors wish to thank Peter Skotte, Claus B. Madsen, and the anonymous participants.

References

[1] B. D. Adelstein, T. G. Lee, and S. R. Ellis, “Head Tracking Latency in Virtual Environments: Psychophysics and a Model,” in *Human Factors and Ergonomics Society Annual Meeting Proceedings (2003)*, 2003, pp. 2083–2087.

[2] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, “Recent advances in augmented reality,” *IEEE Computer Graphics and Applications*, vol. 21, no. 6, pp. 34–47, 2001.

[3] R. T. Azuma, “A survey of augmented reality,” *Presence*, vol. 4, no. August, pp. 355–385, 1997.

[4] M. Bajura and U. Neumann, “Dynamic registration correction in video-based augmented reality systems,” *IEEE Computer Graphics and Applications*, vol. 15, no. 5, pp. 52–60, 1995.

[5] J. Borenstein and L. Ojeda, “Heuristic Reduction of Gyro Drift in Gyro-based Vehicle Tracking,” in *Proc. SPIE Vol. 7305*, E. M. Carapezza, Ed., vol. 7305, no. April, May 2009, p. 11.

[6] J. Caarls, P. Jonker, and S. Persa, "Sensor Fusion for Augmented Reality," in *Ambient Intelligence*, E. Aarts, R. Collier, E. Loenen, and B. Ruyter, Eds. Springer Berlin Heidelberg, Jul. 2003, no. 1, pp. 160–176.

[7] R. Clark. (2013, Nov.) Notes on the resolution and other details of the human eye. [Online]. Available: http://clarkvision.com/imagedetail/eye-resolution.html

[8] S. R. Ellis, K. Mania, B. D. Adelstein, and M. I. Hill, "Generalizeability of Latency Detection in a Variety of Virtual Environments," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 48, no. 23, pp. 2632–2636, Sep. 2004.

[9] S. R. Ellis and B. M. Menges, "Localization of virtual objects in the near visual field," *Human Factors*, vol. 40, no. 3, pp. 415–431, 1998.

[10] S. J. Gilson, A. W. Fitzgibbon, and A. Glennerster, "Quantitative analysis of accuracy of an inertial/acoustic 6DOF tracking system in motion." *Journal of neuroscience methods*, vol. 154, no. 1-2, pp. 175–82, Jun. 2006.

[11] W. Hoff and T. Vincent, "Analysis of head pose accuracy in augmented reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 6, no. 4, pp. 319–334, 2000.

[12] R. L. Holloway, "Registration error analysis for augmented reality," *Presence: Teleoperators and Virtual Environments*, pp. 1–25, 1997.

[13] D. W. F. V. Krevelen and R. Poelman, "A Survey of Augmented Reality Technologies, Applications and Limitations," *The International Journal of Virtual Reality*, vol. 9, no. 2, pp. 1–20, 2010.

[14] E. Kruijff, J. E. Swan II, and S. Feiner, "Perceptual issues in augmented reality revisited," in *2010 IEEE International Symposium on Mixed and Augmented Reality*. IEEE, Oct. 2010, pp. 3–12.

[15] G. Liestøl, "Situated Simulations Between Virtual Reality andMobile Augmented Reality: Designing a Narrative Space," in *Handbook of Augmented Reality*, B. Furht, Ed. New York, NY: Springer New York, ch. 14, pp. 309–319.

[16] M. A. Livingston and Z. Ai, "The effect of registration error on tracking distant augmented objects," *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 77–86, Sep. 2008.

[17] K. Mania, B. D. Adelstein, S. R. Ellis, and M. I. Hill, "Perceptual sensitivity to head tracking latency in virtual environments with varying

degrees of scene complexity," in *Proceedings of the 1st Symposium on Applied perception in graphics and visualization - APGV '04*, no. study 1. New York, New York, USA: ACM Press, 2004, pp. 39–47.

[18] J. E. Swan II and J. L. Gabbard, "Survey of User - Based Experimentation in Augmented Reality," in *Proceedings of 1st International Conference on Virtual Reality*, 2005, pp. 1–9.

[19] R. J. Teather, A. Pavlovych, W. Stuerzlinger, and I. S. MacKenzie, "Effects of tracking technology, latency, and spatial jitter on object movement," *2009 IEEE Symposium on 3D User Interfaces*, pp. 43–50, 2009.

[20] G. Welch and E. Foxlin, "Motion tracking: no silver bullet, but a respectable arsenal," *IEEE Computer Graphics and Applications*, vol. 22, no. 6, pp. 24–38, Nov. 2002.

[21] J. Wither, Y.-t. Tsai, and R. Azuma, "Indirect augmented reality," *Computers and Graphics*, vol. 35, no. 4, pp. 810–822, 2011.

[22] S.-h. Won, N. Parnian, F. Golnaraghi, and W. Melek, "A quaternion-based tilt angle correction method for a hand-held device using an inertial measurement unit," *2008 34th Annual Conference of IEEE Industrial Electronics*, no. 1, pp. 2971–2975, Nov. 2008.

[23] F. Zhou, H. B.-L. Duh, and M. Billinghurst, "Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR," in *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE, Sep. 2008, pp. 193–202.

Paper E

Handheld Visual Representation of a Castle Chapel Ruin

Jacob B. Madsen
Claus B. Madsen

The paper has been published in the
ACM Journal on Computing and Cultural Heritage (JOCCH). Vol. 9(1), pp. 6:1–6:18, 2015.

The layout has been revised.

Abstract

We have experienced rapid development of augmented reality systems and platforms in the world of cultural heritage, namely in cultural settings and historical museums. However, we still face a range of challenges to design an AR system that meets the requirements for an augmented reality installation working autonomously in a cultural heritage setting for an extended duration.

This paper describes the development of two installations for visualization a 3D reconstruction of a castle chapel, running autonomously during opening hours on location of a castle museum. We present a convincing 3D visualization running at interactive frame-rates on modern tablets. In one installation, the tablet is connected to a large screen TV for an immersive experience, and in another the tablet is used handheld, facilitating translational freedom in the chapel. Both installations allow unsupervised usage during museum opening hours.

Based on in-field observations and on-device logging of application usage user behavior is analyzed and evaluated. Results indicate that users spent a limited amount of time using the application, and do not fully explore the visual area of the chapel. In order for the user to spend more time with the application, additional information must be presented to the user.

1 Introduction

Cultural Heritage (CH) represents an aspect of our collective history that over the years are subject to changes, reductions or destructions from different external or human factors. Due to these changes, the introduction of technologies and methodologies suitable to reproduce, restore and present our collective history to people are needed.

3D acquisition and modeling technologies represents a possible solution that allows us to generate the digital version of a real artifact to display for visitors of a museum. Furthermore, once an artifact is represented in digital form, management and conservation of the digital object is possible by field experts.

In recent years, CH institutions have begun to adopt computer-aided mediation of content, as well as computer-aided on-location experiences [12, 19, 26]. New computing systems allow for access to high quality re-constructed content in an ever changing and dynamic way, and it also allows content being mediated in a novel way for visitors [23].

In a series of collaborative efforts with Koldinghus Museum, we have investigated how interactive handheld devices can be introduced to a cultural heritage institution [14, 15]. The aim of the projects have been to introduce a cultural context aimed at guests of different ages, from primarily children [14] to primarily adults [15]. The focus of the collaboration is to attempt

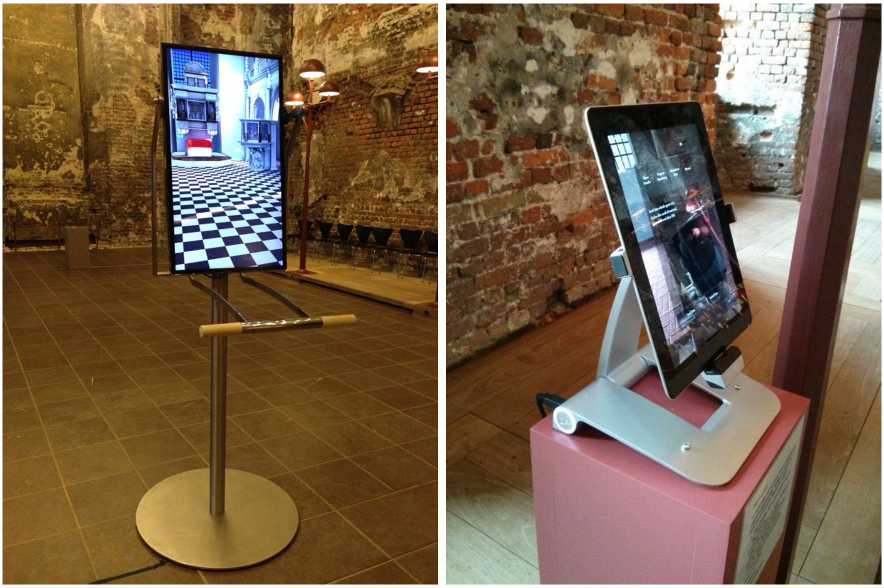


Fig. E.1: Presentation of the two installations for visualizing Koldinghus chapel anno 1604, in a static installation (left) and in a handheld version (right).

mediating content in a novel and interactive way. The work presented in this paper is a direct continuation of work presented in Madsen and Madsen [15]. We expand the original material with an additional installation and further findings from an almost one year long longitudinal study, and analyze data from both the original installation and the new installation presented in this work.

The main product is a physical interactive installation visualizing of the chapel at Koldinghus Museum in Denmark, anno 1604. The visualization attempts to stay close to the physical space (i.e. orientation of the visualization in the physical room, a realistic illumination in the chapel, See Figure E.1). On location, the installation works as a window into the past. It is necessary that this window be linked with the physical space so the visitor can fuse a perceptually coherent experience.

A development strategy is applied for the installations, utilizing off-the-shelves hardware, allowing for a convincing display of a 3D reconstruction visualization on current generation tablets. In one installation, depicted in Figure E.1 (left), a tablet is connected to a large screen TV, giving the visitors a larger field of view into the past. In another installation, Figure E.1 (right), multiple tablets are used in handheld mode, allowing for freedom of movement in the chapel area by selecting different viewpoints in the software and physically relocating one self. Both installations present the same

3D visualization of the chapel to the visitor, and the main difference is in the interaction method. Both installations allow unsupervised usage during museum opening hours. The current version of the handheld installation has logged user interactions via in-application logging successfully for almost a year.

The paper is structured as follows: Section 2 describes related work and sets the context for the project. Section 3 gives a brief overview of the history of the location. Section 4 describes the components of the system, with detailed descriptions of the static and handheld versions in Sections 5 and 6. An iterative design process lead to both hard- and software were changes during development and testing, during the last one and a half year of active usage. We evaluate the installations from the data gathered from both on-device logging and interviews with visitors and staff on location in the detailed explanations within these sections. Finally, we summarize the findings in Section 7

2 Related work

Museums provide a (semi-controlled) natural environment to investigate novel technology outside the laboratory. There is a vast body of work within the field of AR in museum contexts, ranging from museum guides [3, 17, 25], to building virtual and augmented installations and exhibitions for the museum [4, 16, 28, 29].

While most work presently have shifted to mobile devices [3, 10, 21, 25], there are recent examples of researchers developing static installations for augmenting information in a heritage context. Zöllner et al. [30] present systems for a museum setting allowing in one case a 1-degree-of-freedom (1-DOF) interaction by horizontal rotation of an installation stand (Movable-Screen) and in another case a handheld device (UMPC) with other affordances. The aim here was to present information from remote CH sites at museums. Pletinckx et al. presents in [19] TimeScope 1 for the Ename 974 project a church visualization for an archaeological park as a static installation with no interaction aspect, aiming at presenting an early medieval abbey on video shot of the foundation.

Within the field of CH, work is also being done using Virtual Reality, ranging from specialized work of 3D model reconstruction to user interaction and acceptance evaluations. Russo [20] describes two approaches to 3D modeling in CH. One is the representation of the moment "as is" through different approaches and technology, and the other is the previous hypothetical state through a scientific reconstruction process, and presents two examples of work in relation to this. Kersten [11] presents work on modeling a city based on a 3D scanning approach. Trapp [27] describes how the user is now able to explore CH artifacts in real-time, and presents the technical concept for implementing this.

Consensus in the field is that users appear interested in the novel experience of AR technology. Due to generally short-lived studies, it is not known whether this interest is an effect due to presented material or a "wow" effect of AR being novel to a majority of people. This study is, to the best of our knowledge, the first long longitudinal in-field evaluation of an AR installation in a museum context, and presents usage data for a longer period to indicate whether interest is just a "wow" effect or not. We conduct an unsupervised field trial of two implementations to visualize the chapel ruin. One static installation with 2-DOF for orientation inspired by [19, 30], and multiple 3-DOF handheld devices with additional limited positional movement. Both implementations have been in use extensively, the static installation for 45 days, and the handheld devices for more than a year.

In comparison to the related work, the presented installations in this paper differ in interaction methods to present an identical visualization of the chapel. One is designed as a multi-spectator viewfinder people can group around, and another as a series of handheld devices acting as personal viewfinders, allowing more freedom for individual use. By leveraging 2-DOF and 3-DOF orientations, monitored by hardware sensors in the systems, users are free to explore the chapel, and using the handheld device the systems facilitate translation changes in the application. This lets the user change the virtual position to match the geo-physical space. This implementation is a novel approach to presenting a visualization of a historical ruin to a visitor, while allowing the visitor interactive control of the experience.

3 Location overview

The Koldinghus castle dates back to the mid 1200s, and has historically been a place for protection from south, a residency for kings and the royal family, and has played a central role in Danish history.

The construction of the chapel and tower of the castle dates back to a start in 1597, when a fire destroyed that part of the building. This left the King with the opportunity to reconstruct and modernize this part of the castle. The king envisioned a bigger and more grandiose chapel, and saw the new chapel as a reflection of the king as God's representative in both ecclesiastical and secular affairs. The new chapel was finished in 1604 [18].

During the Napoleon wars in 1808 the castle burned to the ground due to a fire started in a chimney in a guards room. The fire destroyed the castle over a period of two days. It was left as a ruin until a basic restoration in the 1970s (Figure E.2). The current appearance of the castle chapel as Koldinghus is no more than a ruin of the former chapel, with restored floors and ceiling

Fig. E.2: Current Koldinghus exterior and chapel interior as it looks after the restoration in the 1970s. Images courtesy of Koldinghus Museum.

Fig. E.3: Overview of the physical castle chapel (left) and the 3D visualization as represented in the installation (right). Photo courtesy of Koldinghus Museum.

being the only addition to the ruin itself. The present chapel can be seen in Figure E.3 along with our rendered visualization.

4 System descriptions

The main task of the system is to facilitate information to visitors, who can range from families with children to elderly couples. It is particularly interesting for the younger audience, who might be interested in a digital platform [24]. The system in itself should be self-explanatory and easy to use for the average visitor, who are assumed to have limited experience with any type of technical products. Design requirements for this user-base is based

Museum guests	Museum personnel	Technical
Light weight (Static)	Flexible placement	Low power consumption
Easy to handle	Low daily maintenance	Long running time
Visualization true to world	Easy to setup and use	High frame-rate
Self-explanatory	Low cost	Limited polygon count
		2-DOF or 3-DOF orientation
		Translation of user position (handheld)
		Robust

Table E.1: System requirements

on the knowledge and feedback of the museum personnel. Based on their a-priori knowledge of the population visiting the museum and these visitors expectations. As a design requirement, users should be able to interact with the system, without any instructions of how the system works, let alone how to use it. Furthermore, the museum personnel require a system with low cost and low daily maintenance as well as a transportable system, as the chapel area, which the system will augment, is regularly a forum for exhibitions and events that require the floor space of the chapel. In these events, the system might need to be set-aside for a small period of time, such as an evening or for a couple days, after which the installation is setup in the chapel again. During an initial discussion with the museum, a requirement list for the design was developed, factoring in requirements of the visitors, the staff and technical requirements for the installations, which are listed in Table E.1. This requirement list sets an initial design model for the visualization, as a first iteration of the interaction requested by the museum for their visitors.

The system itself should be able to process the visualization of the chapel ruin at interactive frame-rates, and be able to run during museum opening hours and with extensive usage. The processing power of the system should be sufficient to handle these requests, in addition to the visual processing of the visualization. Technical aspects such as polygon count of the model, and the shader performance for the visualization must be optimized to allow fulfillment of these requests.

The following sections, 5 and 6, will present the installations. Design of hardware and software is explained in detail, and field evaluations are presented. Both installations were designed to present the best user experience for the novice visitor, and allow staff of the museum an easy handling and daily maintenance of the installations. We elaborate on some of the problems encountered during the development and how to overcome these problems in the areas within hardware, software, reconstruction and realistic daylight simulation, as well as evaluate each design. This visualization does not contain any storytelling elements, animations or planned events. Reasoning for this is that the museum was interested in a first glimpse into how this system in itself would facilitate the novel idea of presenting a past historical point in time through an interactive installation. Following this, additional storytelling elements can be implemented. This is presented in more detail in the Future works, Section 8.

5 System description: Static Installation

The static installation (Figure E.1 (left)) allows a single visitor to operate the installation, while multiple other visitors are able to experience the visualization as over-the-shoulder spectators. When a visitor enter the chapel, the installation is present at the center of the room, inviting the guest to try interacting with the installation. A user interacts with the installation using the handle in front of the installation to manipulating the orientation of the screen and correspondingly the orientation of the visualization. The visualization is rendered in real-time from a static point in the virtual chapel, corresponding to the physical placement of the stand in the chapel room.

5.1 Hardware

In order to facilitate visualization to multiple visitors simultaneously, and to immerse the operator of the viewfinder in the virtual world, a large 40" flat screen TV[1] is used as display screen for installation.

The TV in itself is of a considerable size and weight, which suggests that the base of the installation must be designed to accommodate the weight and not feel flimsy to the visitor. At the same time it should not be too heavy to maneuver for the museum staff. A mobile tablet is running the visualization and is mounted on the back of the TV. This gives the installation a smaller design, while shaving a few kilos of the installation weight. The visualization is rendered using an iPad[2] tablet computer mounted on the back of the 40" TV. The output from the tablet to the TV is via HDMI interface. For a detailed overview of how the hardware is mounted on the installation, see Figure E.4. With this setup, the installation only requires a single source of power for both the TV and iPad, and the weight of the complete installation is roughly 70 kilos, for which a majority of the weight represented by the circular steel base-plate.

The installation provides the operator with 2-DOF for manipulating the viewing direction, and pan/tilt orientation. In this design, rotation is limited in both axes to allow safe operation of the installation and to not tangle or tear internal cables in the base of the installation. The limit for panning is 350° and for tilting the limit is $+/-45^\circ$.

---

[1] Model no: Samsung UE40ES6305UXXE

[2] Apple's 3rd generation iPad featuring an Apple A5X chip (Dual-core 1 GHz Cortex-A9 processor with a PowerVR SGX543MP4 GPU), and a 2048x1536 (264 ppi) resolution display

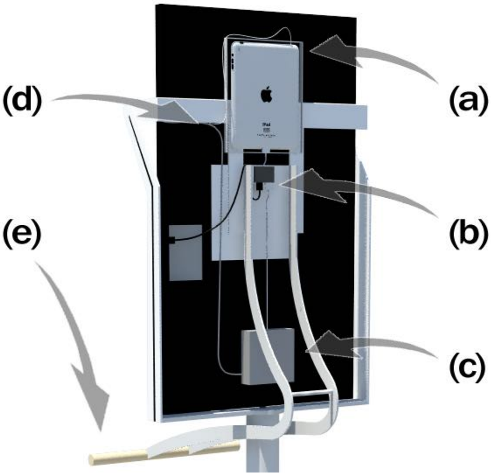


Fig. E.4: Visual render of the interactive TV installation from back, mounted with handle for manipulating the viewpoint and depiction of the tablet placed and connected to the TV. a) The iPad mounted, connected to an orientation tracker (explained later in the section) in the base of the installation using the microphone-in port, and outputs the visualization and draws power via the 30-pin connector. b) Signal splitter that outputs the HDMI signal to the TV and draws 5V power from (c), the power utilities box, which powers the TV with 220V and the tablet with 5V. d) The cables run from the base of the installation and out through the top, power for (c) and the orientation tracker to the tablet. e) Installation manipulator enabling user interaction with the installation.

As the chapel ruin provides a large open space for use, the chapel is often used for temporary events and exhibitions. Mobility of the installation is therefore a necessity in order to move it from the center of the room to a corner or the balcony, when such events take place. In order to facilitate re-positioning of the visualization, the iPad can be un-mounted from the back of the TV and the touch interface allows for relocation of the visualization viewpoint position within the chapel.

Orientation Tracking

We initially tracked user interaction via the gyroscope and compass contained in the tablet, in order to re-orient the virtual viewing direction and re-render the scene to match the new perspective. However, early experiments showed this to be sub-optimal, as the compass is far too inaccurate in these indoor locations of the chapel, with experiments showing changes in heading from -14° to $+56^\circ$ from magnetic north. The gyroscope posed the problems of having a large amount of drift over time over the horizontal axis (See Figure E.5).

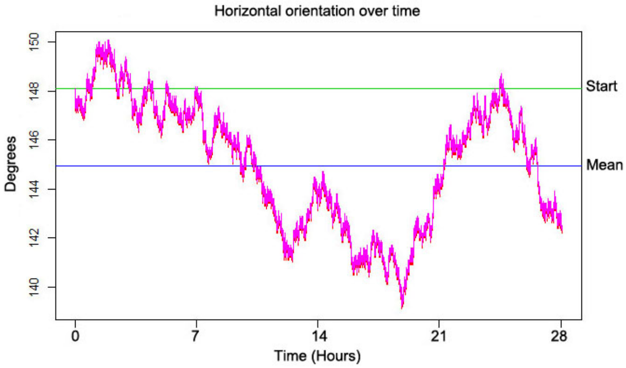


Fig. E.5: Example of horizontal gyroscope drift in a static placemetn over the course of 29 hours.

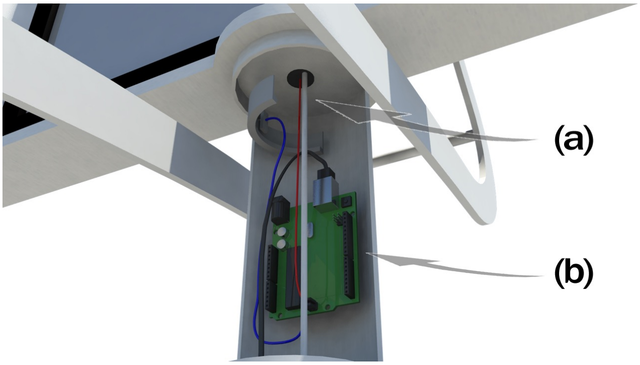


Fig. E.6: Visual explanation of the horizontal calibration using an Arduino in the base of the installation. a) The white and red wires are based through the base of the installation to the TV and tablet (see Figure E.4), and the blue wire illustrates how the orientation data measured is passed to the Arduino (b).

To overcome drift, an Arduino[3] [2] was mounted with a light-sensor in the base of the installation (See Figure E.6), designed to signal the iPad every time the TV rotated past a specific point. This signal is then used to re-calibrate the gyroscope and the orientation of the virtual chapel. As the charging input connector on the iPad were in use for charging and HDMI out to the TV, we opted to use an audio jack connector for sending data to the application, which in turn monitors audio input for the re-calibration signal from the Arduino.

---

[3] http://arduino.cc/

5.2 Software

The software is developed in the Unity game engine[4], allowing for efficient and rapid development for multiple platforms, as well as a component-based architecture for modular development. This modular system allows each component (i.e. input components, viewfinder components, etc.) to be toggled on/off for each platform. Furthermore, it eased adapting to requirement changes during running evaluations of the platform.

In order to process the visualization on the tablet at interactive framerates, we set strict limits on polygon count for the visualization, limit the amount of draw calls to be rendered for each frame, and pre-compute most of the lighting information to light maps split into direct and indirect lighting. Splitting the direct and indirect illumination to different light maps allows the engine to use the direct illumination map to compute specularity of any surface hit by direct illumination, in real-time, for only the places that are hit directly by the light. Small objects are combined to form many larger vertex objects, to reduce individual calls to draw objects in the rendering engine, resulting in fewer calls and better performance of the visualization. As a technical limitation, Unity supports a maximum texture size of 2048x2048 for mobile devices. This sets a natural restriction in the size of objects, without having to use multiple texture maps per object. In a trade-off between detail and real-time visualization of the model, the polygon count for the model was reduced to an acceptable level, which allowed for a high amount of details from all possible viewport positions. The overall polygon count for everything in the viewport never exceeds 150.000 polygons at any one time.

3D Visualization

To develop the 3D visualization of the chapel, information was gathered in three ways. 1) research in literature, 2) scanning and modeling from artifacts available at Koldinghus and 3) informed guesses to fill in the gaps.

Concerning literature research on the chapel, there were no paintings or detailed depictions of the chapel prior to the fire in 1808. The main part of the gathered information stems from books written on the subject of the churches of the Kolding area [18], informational drawings and posters on location, and information from museum experts.

In the ruins of the church today, it is still possible to see few sculpted columns that survived the fire. These artifacts and the chapel room itself was laser scanned to obtain room dimensions and placement of artifacts. The remaining artifacts from the chapel: sculpted columns, stone ornaments, etc., which survived the fire and are currently in archives at the castle, have been

---

[4] http://www.unity3d.com

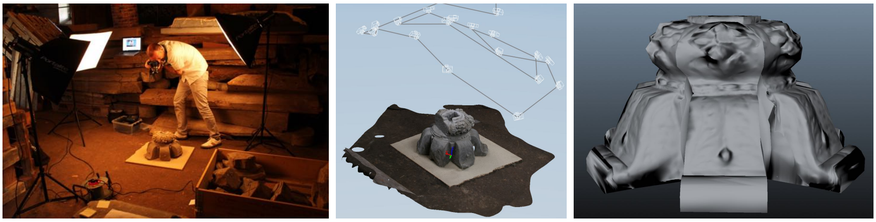


Fig. E.7: Process of gathering visual content from image capture using 123D Catch software, from capturing (left) to automatic 3D model generation (center). The generated high-res model is then manually remodeled into a low-res object plus a texture and normal map (right).

scanned to digital models using 123D Catch[5] from Autodesk. This involves a process of manually photographing objects from multiple angles, as illustrated in Figure E.7 (left). With multiple exposures, this allows generation of a 3D model with amount of details (Figure E.7, center) using an SLR off-the-shelf camera. This follows the principles of Kanade et al. [9]. The model is then cleaned up in post-processing, in a 3D modeling software, Autodesk Maya[6] (Figure E.7, right). The remaining details from the research have been manually modeled.

Unfortunately, there is a gap of knowledge, and lack of paintings or detailed drawings of the chapel from the period between 1604-1808, and lot of valuable information of the chapel has been lost. In order to compensate for this, we have filled in the gabs of this missing information with information from multiple sources, and the result in what is considered the most plausible appearance. Frederiksborg Castle Chapel was built shortly after Koldinghus Chapel, ordered by the same king, and designed by the same architect. In an attempt to further expand the knowledge of the most probable interior of Koldinghus Chapel at the time, Frederiksborg Chapel was used as inspiration for the generation of the virtual model for areas in which there was limited or no information available of the true decor at Koldinghus Chapel. Additionally, at Vor Frue Kirke in Aalborg, the old alter, which unfortunately has also been lost to a fire in 1902, was sculpted by the same sculptor as the original alter in Koldinghus. However, pictures of Aalborg alter exists, which has been heavily used as inspiration of the modeling of the chapel alter for Koldinghus.

To sum up the process of pipeline for generating the visualization presented, the following steps was taken: 1) literature research and inspection of buildings from the same architect at the time, 2) laser scanning and modeling of the chapels current appearance, 3) semi-automatically capture and remodel artifacts and ornaments from photographic capture on location, 4)

---

[5] http://www.123dapp.com/catch

[6] http://www.autodesk.com/products/maya

Fig. E.8: Photos of the static installation in use, as a staged press photo (left) and one taken from the launch of Koldinghus's King Christian IV exhibition (right), during which the static installation was also launched and used by the public for the first time with no guidance in usage whatsoever. Images courtesy of Koldinghus Museum.

manually model remaining ornaments and artifacts from researched literature, 5) combining all information, 6) expert review and testimonials, and 7) using the expert testimonials to construct remaining ornaments and clear up details in the visualization. The finalized visualization is modeled based on factual information of the interior of the chapel, and what experts in this field identify the most plausible appearance of the chapel at the time.

The visualization draws similarities to Simeone et al. [22], who presents a substitute reality, with contextually appropriate virtual objects, of a real setting. In this visualization, the substituted reality is the past, matched to the appropriate location of artifacts within the ruin. For the visitors at Koldinghus, it is presented as the best informed guess of the historical appearance of the chapel, supported by a physical note at the location, providing visitors the option to choose between a factually known visualization and a best guess options in the application is considered a future implementation.

5.3 Evaluating the static installation

The static installation was launched to the public on November 3rd, 2012, along with a new general exhibition on King Christian IV. During this event, the visitors and guests at the museum could try the installation by themselves, while we observed the installation being used by the public (Figure E.8). This section is based on first hand data from the initial launch and a further half a day of observation of usage. Additionally, second hand data from the staff at Koldinghus was taken into consideration for the evaluation, both regarding their experiences with the daily usage from visitors, but also from their experiences with moving and maintaining the installation.

In order to log usage information of the installation, we decided on an anonymous approach where the visitors are not interviewed or surveyed directly, an approach other authors mention briefly [21, 29]. Instead our approach was to log usage information from the device itself, and use this information to gain understanding of how the installation is used by the visitors, i.e. how long they use the installation and to what degree they explore the chapel [27]. We believe that adapting an anonymous logging approach will provide information about usage. In this way the user opinion or experience of the installation should be present in the gathered data.

During the launch event, the installation was in constant use by visitors who actively used it, and was also an attraction from even more visitors who only spectate. One interesting observation, which is in line with previous findings, was that initially the people in the room gathered around and looked at the installation, but no one would touch or interact with it. This went on for a few minutes until a small boy and girl went up and tried to steer the installation. This shows that the children are more willing to dive-in to the technology, as earlier research have found [6, 8]. Upon their success, the onlookers gave it an attempt and actively and enthusiastically used the installation. Several studies describes this as the "honey pot" effect, in which people become interested in an interactive system when they can see other people actively using it [1, 6]. The visitors expressed enjoyment of the installation and its facilitation of the visual past of the chapel. Similarly to results presented by Miyashita [17], visitors appeared to be pleasantly surprised by the visualization.

The installation was operational and available for the public over a period of 45 days, during which it has been monitored for performance issues as well as general issues expressed by the museum staff. During this period, the staff experienced problems with the installation, resulting in orientation errors in the visualization. An application reset and re-calibration of the system would fix the orientation. The problem continued to randomly appear for the duration of the time the installation was operational. After the initial 45 days of operation and usage gathering, operation of the installation was discontinued, due to a few factors: 1) The orientation-sensing Arduino component in the installation proved to be unreliable and could not satisfactorily calibrate the orientation, leaving the visualization to drift or in worst case re-calibrate at the wrong position. 2) The staff found the installation to be too heavy for usage in a chapel that changes exhibits very often, and expressed a general unhappiness with the heavy installation being moved every so often. 3) The physical stop on the installation responsible for assuring the horizontal rotation not exceeding 350°, broke during operation, leaving the installation vulnerable to being rotated excessively.

6 System description: Handheld Installation

The design idea behind the handheld installation is to allow a multiple visitors to each operate their own viewfinder, and use it as a window into the past, experiencing the visualization and the history of the chapel at their own pace, planning their own tour of the chapel [7]. The visualization on the handheld installation is identical to that on the static installation. When a visitor enter the chapel, the handheld installation is present near the wall next to the door into the room, inviting the guest pick up the device and explore the chapel. The user interacts with the device merely by orienting it physically in the desired direction of orientation, and the hardware with ensure the matching view in the visualization. Throughout the virtual floor, red markings (Figure E.9) allow the visitor to relocate the virtual viewpoint to a new place within the chapel. The visitor must then re-position himself accordingly for a coherent experience. After end usage the tablet must be returned to the station at the entrance to the chapel.

6.1 Hardware

Section 5.1 described how the gyroscope sensor for estimating orientation is not feasible as the only sensor, due to accumulated drift over time (Figure E.5). It also described how the imprecision of the compass meant it was unreliable for measuring magnetic north in this environment. As a substitute for the compass, and for calibration of the gyroscope, the docking station was used to calibrate the iPad to the orientation in the geo-physical world. The use-case dictates that the viewfinder can be in two different modes of operation: 1) The iPad is in the docking station, is static and charging, and 2) The iPad is held in the hands of a visitor and will move around for some period of time. As software detection of the charging state is simple, the static docking state is easy to detect in the application. The application behavior in the two states is thus:

Charging:
Reset model orientation to the calibrated orientation

Not charging:
Gyroscope readings to track the orientation

An added benefit to using the docking station is the ability to estimate the total number of uses by assuming that each usage is occurring when one user takes an iPad from the charging station (start) until it is returned to the charging station (end). It furthermore allows us to only log interaction data during this time period on the device.

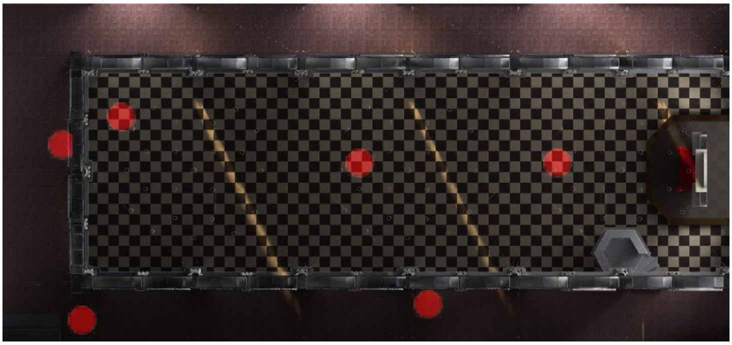


Fig. E.9: Depiction of the floor plan of the visualization, with red marking the areas in the visualization the user can translate to and experience the visualization from. The placement in the lower left corner is the entrance to the chapel, and the default starting position of the application. This is where the visitor picks up the tablet and returns it again.

In the final version of the installation, currently at display and in active use on location, the application is able to function autonomously for an entire day after being setup by the museum staff. This setup and calibration procedure has three steps: 1) Start the application if it is not already running (most of the time it will be running continuously for days). 2) In handheld mode, touch and drag on the screen to calibrate the horizontal orientation by orienting the visualization to the desired orientation to match the physical space, and 3) Disable the touch dragging via a hidden in-application menu. Following this, the iPad will work autonomously during the day, and should there be any problem due to drift, the museum staff can repeat the procedure again, or restart the application to reset everything, if needed.

6.2 Software

In the handheld version, the installation does not limit the user to a single location, as with the static version. Leveraging this fact, we wanted to allow the users freedom in navigating the chapel with the viewfinder, both in the physical and the virtual space. This of course requires position tracking, which is deemed technically and economically unrealistic for an off-the-shelves solution. We opted for a technically simple solution, namely to re-locate to predefined viewpoints [13], marked in the visualization as circles the floor (Figure E.9), allowing the users to move and experience the model from multiple locations. The user must re-position accordingly in the physical space, to gain a coherent experience of seeing through a windows to a historic place in time.

The in-application translation changes are implemented as method for transitioning between pre-defined locations in the virtual chapel via the user interface. The interaction between the physical and virtual space of course means that there is no true positional tracking, and the user must place himself in the physical one accordingly to the perceived position in the virtual space in the visualization. It forces the user to mentally make a connection between the real chapel and the virtual model. With this interaction method, the user is able to experience the chapel from multiple locations using a low-tech translation approach.

6.3 Evaluating the handheld installation

The handheld application has been in active use since February 10th 2013, and is still in active use today, with multiple iPads available at the museum, and with a new option from July 2014 to download the application for your own iPhone or iPad[7] (with Android coming eventually).

The logging framework logs the anonymous usage data as described for the static installation in Section 5.3. For the handheld version, the logging has been expanded to include the position changes within the software, as well as the time in action and time charging (considered non-usage).

This section reports evaluation of logged data from June 7th 2013 until April 30rd 2014. During this period, we have data from tablets being in active use for a total of 111 days of the 327 days from first to last log entry, meaning that roughly 2/3 of all days are without usage, or that this information is missing. The museum reported that this loss of data was due to tablets being broken by being dropped by accident, and one had the charging port broken from continued daily use. The tablets were discarded before any data could be retrieved, as it was only logged on-device due to wireless and network security protocol in place at the museum. As unfortunate as that is, the data retrieved show generally similar usage from the logged events. Data from the tablets have been logged at 0.5-second intervals, throughout the museum opening hours. The orientation log is using a resolution of 0.1°. This gives us an idea of the areas of interest to the visitors with sufficiently high accuracy. In order to filter out any invalid data from the log, we initially remove all data entries in which the gyroscope has not sensed movement between two consecutive logs, i.e. the tablet has not been placed properly in the charging station or left somewhere in the chapel area. We furthermore remove all entries outside opening hours of the museum, as we assume that all logs from this period if by staff of the museum, and not be a visitor.

The data shows that the two remaining tablets have been used during the logged days for a total of 87 hours and 7 minutes, corresponding to just over 47 minutes of active use per day (Usage per day is shown in Figure E.10). Contrary to intuition, it does not appear that users are mainly using the application in weekends or during holidays. As Figure E.10) illustrates, there

---

[7] Itunes link: https://itunes.apple.com/app/slotskirken/id665119672?ign-mpt=uo%3D5

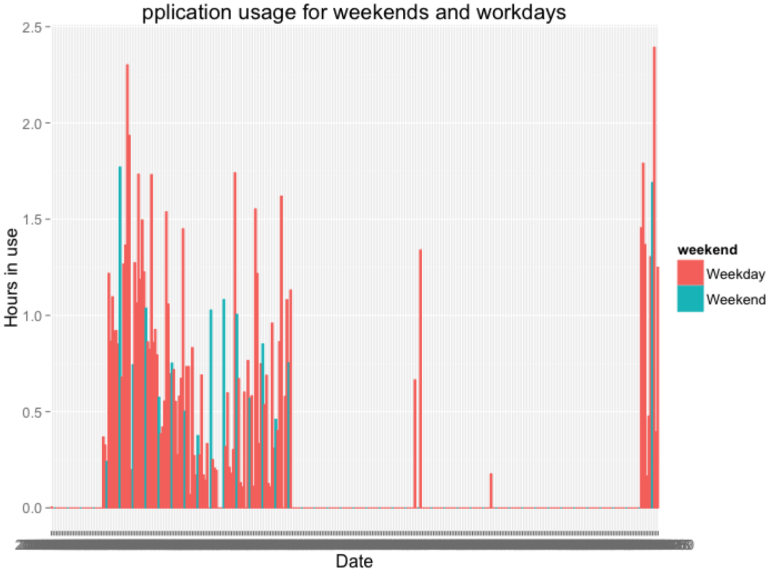


Fig. E.10: Histogram of tablet usage per day, illustrated as weekend and workdays (left), and school holidays and workdays (right). Usage of the application is fairly well spread over the course of the period, with marginally higher usage during the intro phase during the summer.

is no clear indication of this to be true. It was found through interviews with museum staff, that the main usage of the applications were from visiting school classes from the area, so most of the usage was from children. Given the estimate of 3470 uses, this corresponds to an average of roughly 1.5 minutes of use, each time the tablets are taken from the charger. We reason that the lack of continued interest in using the application is due to the limited amount of additional information available within the application, which could create additional usage value for the visitor.

From the central position in the visualization we were interested in exploring the logged orientation data, to investigate what the visitor is actually interested in, when looking around the chapel area. We use this position as it is somewhat center in the room, with equal distance to both ends of the room while having nothing obscuring the view.

Figure E.11 illustrates a heat map (b) of user view-direction from this position, displayed as an environment map in longitude-latitude format, compared to the rendered visualization (a), both also illustrated as environment maps. As illustrated in the heat map, users are mainly holding the tablet vertical or slightly tilted down. They do for example not visually study the ceiling or the galleries, and misses a large part of the visualization. We can at this point only speculate why this is. This might be caused by people being

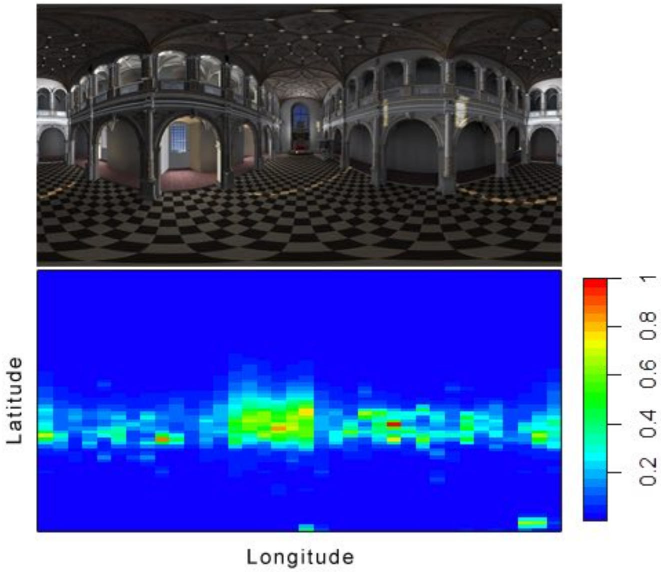


Fig. E.11: Environment map representation of the castle chapel area from the center position in the room, with (top) the rendered map, and (bottom) a heat map indicating user interest in certain areas.

uncomfortable using such a device in public, or that they simply do not see a reason to use it otherwise.

The data presented in the figure is only of value if we can assume the users physically position themselves to the virtual position in the visualization. Observations from the museum staff informs us that the visitors are in fact placing themselves in the physically correct area, corresponding to the virtual location, and does this to get a visual coherent experience.

7 Conclusion

In this article we have presented two installations for Koldinghus museum, and presented the design, development and evaluation process of these installations with a similar purpose, though very different in appearance and interaction. We have evaluated the usage of the installations and used this to present design criteria, which might aid and inspire researchers and developers of novel interaction technology for a CH context.

Using commodity hardware and software we have created a visualization of the castle chapel as it appeared more than 400 years ago, in 1604, a chapel which only have been restored to the bare walls. In both installations, visitors are presented with a window into the past, and by interacting with the installation, visitors can experience the past chapel, as it most likely appeared at the time, with realistic lighting estimated for a single point in time.

The main contributions of this article are the design and evaluations of the installations, with proposed methods for overcoming some of the production problems developers may face, in novel ways.

The main contributions are:

1. A static installation which presents visitors with a large immersive display for visually presenting a historical chapel
2. A handheld installation providing additional interactivity in the form of limited translation through the chapel
3. A first long longitudinal study reporting insights of in-field usage for almost a year

We have furthermore described the creation process of a low cost installation for reliable estimation of viewpoint and rendering of a 3D reconstruction of a historical chapel in a realistic lit environment. The implementation process has been described in detail from gathering of relevant information to the construction of the virtual model in 3D. The early evaluation provided lessons for providing a pleasant operational experience for the staff, as moving the stand around was a source of frustration. The field evaluation process of almost a year worth of logged data is presented. This data set and the lessons learned from the data can serve as prior knowledge and inspiration for researchers and developers of novel interaction installations for CH contexts.

One benefit of using a digital approach such as this is the option for the visitor to get an interactive glimpse into the past appearance of the chapel, as it appeared in the past. This would not be possible in any static 3D rendering of the same environment, and presents an alternative to the static material at the museum. Additionally, a digital installation in a CH context allows the museum to attract another kind of visitor [23], of the younger generation, who are not interested in a museum where artifacts are locked behind glass and nothing can be touched. It allows us to gain an insight into what such people are interested in, when visiting a museum. However, developing a digital installation comes with its downsides as well. As the technology has not fully matured yet, there are potential pitfalls between the technology and the usability of the installation, which has not been fully explored yet. As the detailed visualization in itself is not enough to keep visitors interested for long, additional information must be presented to the visitors. Suggestions for this is presented in the Future works, Section 8.

It appears that visitors are mostly interested in looking at the virtual scene horizontally despite efforts to create a full implementation of the chapel itself.

The findings point out a couple of obvious considerations and opportunities for further studies in the area of user exploration of virtual scenes, such as how to design an interaction model to direct the users focus to specific interesting artifacts or other points of interest.

It is still in active use at the museum today.

8 Future work

While the installations currently presents visitors at the museum with an visualization of the past appearance of the chapel, there there are still many avenues for future work. One such avenue is to add storytelling elements to the interaction. As mentioned, the interactive visualization in itself is not sufficient to keep visitors interested for long. Adding storytelling and information to the installation is the obvious next step. Similarly to the Ename 974 project [19], where the TimeScope 1 installation is accompanied by information about the site, its discovery, and the social and economic life, the installation presented here could be extended with information of both the visualization process as well as information related to the time of the visualization, the social and economic life at the castle and surrounding area. Given the interactive nature of the installation, such additional information could take appearance as animation and sound effects to retain the interest of the public and appear less static.

Furthermore, events of the time such as sermons and weddings of the time could be introduced in the installation, allowing visitors to get a glimpse into the life that the time. Such events could also bring returning visitors interested in specific holidays and events into the museum, and might increase the perceived usefulness of the installation [5].

Regarding gaining information related to the user of the installation, the presented heat map (Figure E.11) can be used to inspire further development of the application, and inform about which areas could be augmented with additional information [27]. Based on the initial user interest, the museum can use this information to inform or inspire visitors to learn historical facts from the chapel using other means for knowledge facilitation in the castle chapel, as well as for further development of the application based on actual visitor interest. One effect of such additional data gathering is in informing the museum of areas of interest of the user, in order to target development effort. Areas of large interest can be expanded with additional information to the user, or under-represented areas can be improved to attract more interest.

Acknowledgment

This project has been partly funded by Koldinghus Museum. The authors would like to thank Curator and head of exhibitions Axel Johnson, Ph.D., and Koldinghus Museum for their collaboration and efforts in undertaking and realizing this project, and Peter Skotte for his invaluable efforts and collaboration in creating the 3D models and textures used in the project.

References

[1] H. Brignull and Y. Rogers, "Enticing people to interact with large public displays in public spaces," in *In Proceedings of the IFIP International Conference on Human-Computer Interaction (INTERACT 2003)*, no. c, 2003, pp. 17–24.

[2] J. Carmigniani, B. Furht, M. Anisetti, P. Ceravolo, E. Damiani, and M. Ivkovic, "Augmented reality technologies, systems and applications," *Multimedia Tools and Applications*, vol. 51, no. 1, pp. 341–377, Dec. 2010.

[3] A. Damala, P. Cubaud, A. Bationo, P. Houlier, and I. Marchal, "Bridging the gap between the digital and the physical," in *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts - DIMEA '08*. New York, New York, USA: ACM Press, 2008, p. 120.

[4] C. de Bérigny Wall and X. Wang, "Interactive Antarctica," in *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts - DIMEA '08*. New York, New York, USA: ACM Press, 2008, pp. 319–325.

[5] A.-C. Haugstvedt and J. Krogstie, "Mobile augmented reality for cultural heritage: A technology acceptance study," in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Nov. 2012, pp. 247–255.

[6] U. Hinrichs and S. Carpendale, "Gestures in the wild," in *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. New York, New York, USA: ACM Press, 2011, p. 3023.

[7] B. Hodgson, M. Megliola, S. Toffolo, M. Zöllner, J. Dearlove, and J. Keil, "Encouraging Cultural Exploration: a Dynamic Tour Guide Based on Personalised Contextual Information and Ad-hoc Planning," in *Collaboration and the Knowledge Economy: Issues, Applications, Case Studies. Pt.2 : eChallenges e-2008 Conference*. IOS Press, 2008, pp. 1331–1338.

[8] E. Hornecker, ""I don't understand it either, but it is cool" - visitor interactions with a multi-touch table in a museum," in *2008 3rd IEEE International Workshop on Horizontal Interactive Human Computer Systems*, vol. 21. IEEE, Oct. 2008, pp. 113–120.

[9] T. Kanade, P. Rander, and P. J. Narayanan, "Virtualized reality: Constructing virtual worlds from real scenes," *IEEE Multimedia*, vol. 4, no. 1, pp. 34–47, 1997.

[10] J. Keil, M. Zöllner, M. Becker, F. Wientapper, T. Engelke, and H. Wuest, "The House of Olbrich — An Augmented Reality tour through architectural history," in *2011 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities*. Basel, Switzerland: IEEE, Oct. 2011, pp. 15–18.

[11] T. P. Kersten, "3D Scanning and Modelling of the Bismarck Monument by Terrestrial Laser Scanning for Integration into a 3D City Model of Hamburg," in *Third International Conference, EuroMed 2010, Lemessos, Cyprus, November 8-13, 2010. Proceedings*. Springer Berlin Heidelberg, 2010, pp. 179–192.

[12] G. Liestøl and A. Morrison, "Views, alignment and incongruity in indirect augmented reality," in *Proceedings of 2013 IEEE International Symposium on Mixed and Augmented Reality – Arts, Media and HUmanities (ISMAR-AMH)*, vol. 22, no. 3, Sep. 2013, pp. 174–186.

[13] G. Liestøl, T. Rasmussen, and T. Stenarson, "Mobile innovation: designing and evaluating situated simulations," *Digital Creativity*, vol. 22, no. 3, pp. 174–186, Sep. 2011.

[14] C. B. Madsen, J. B. Madsen, and A. Morrison, "Aspects of what makes or breaks a museum AR experience," in *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*, no. November. IEEE, Nov. 2012, pp. 91–92.

[15] J. B. Madsen and C. B. Madsen, "An interactive visualization of the past using a situated simulation approach," in *2013 Digital Heritage International Congress (DigitalHeritage)*. Ieee, Oct. 2013, pp. 307–314.

[16] D. Michael, N. Pelekanos, I. Chrysanthou, and P. Zaharias, "Comparative Study of Interactive Systems in a Museum," in *Proceedings of the Third international conference on Digital heritage*. Springer-Verlag Berlin, 2010, pp. 250–261.

[17] T. Miyashita, P. Meier, T. Tachikawa, S. Orlic, T. Eble, V. Scholz, A. Gapel, O. Gerl, S. Arnaudov, and S. Lieberknecht, "An Augmented Reality mu-

seum guide," *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 103–106, Sep. 2008.

[18] Nationalmuseet, *Kirkerne i Kolding: Slotskirker på Koldinghus*, ser. Danmarks Kirker. Frederiksholms Kanal 12, 1220 København K: Nationalmuseet, 2009, vol. 11, (Book; In Danish).

[19] D. Pletinckx, N. Silberman, and D. Callebaut, "Presenting a monument in restoration," in *Proceedings of the 2001 conference on Virtual reality, archeology, and cultural heritage - VAST '01*. New York, New York, USA: ACM Press, 2001, p. 197.

[20] M. Russo and G. Guidi, "Reality-based and reconstructive models: digital media for cultural heritage valorization," *SCIRES-IT*, vol. 1, no. 2, pp. 71–86, 2011.

[21] D. Schmalstieg and D. Wagner, "Experiences with Handheld Augmented Reality," in *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE, Nov. 2007, pp. 3–18.

[22] A. L. Simeone, E. Velloso, and H. Gellersen, "Substitutional Reality : Using the Physical Environment to Design Virtual Reality Experiences," in *CHI 2015 Proceedings*, 2015.

[23] S. Styliani, L. Fotis, K. Kostas, and P. Petros, "Virtual museums, a survey and some issues for consideration," *Journal of Cultural Heritage*, vol. 10, no. 4, pp. 520–528, 2009.

[24] W. Swartout, D. Traum, R. Artstein, D. Noren, P. Debevec, K. Bronnenkant, J. Williams, A. Leuski, S. Narayanan, D. Piepol, C. Lane, J. Morie, P. Aggarwal, M. Liewer, J. Y. Chiang, J. Gerten, S. Chu, and K. White, "Ada and grace: Toward realistic and engaging virtual museum guides," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6356 LNAI, pp. 286–300, 2010.

[25] A. B. Tillon, I. Marchal, and P. Houlier, "Mobile augmented reality in the museum: Can a lace-like technology take you closer to works of art?" in *2011 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities*, no. Figure 1. IEEE, Oct. 2011, pp. 41–47.

[26] J. Torpus and B. Tobler, "lifeClipper3 — An augmented walking experience field evaluation of an experience design approach for immersive outdoor augmented reality," in *2011 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities*. IEEE, Oct. 2011, pp. 73–82.

[27] M. Trapp, A. Semmo, R. Pokorski, C.-d. Herrmann, J. Döllner, M. Eichhorn, and M. Heinzelmann, "Communication of Digital Cultural Heritage in Public Spaces by the Example of Roman Cologne," in *Proceedings of the Third international conference on Digital heritage*. Springer-Verlag, 2010, pp. 262–276.

[28] W. van Eck and Y. Kolstee, "The augmented painting: Playful interaction with multi-spectral images," in *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*. IEEE, Nov. 2012, pp. 65–69.

[29] E. Woods, M. Billinghurst, J. Looser, G. Aldridge, D. Brown, B. Garrie, and C. Nelles, "Augmenting the science centre and museum experience," in *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Austalasia and Southe East Asia - GRAPHITE '04*, no. Figure 3. New York, New York, USA: ACM Press, 2004, pp. 230–236.

[30] M. Zöllner, J. Keil, H. Wüst, and D. Pletinckx, "An Augmented Reality Presentation System for Remote Cultural Heritage Sites," in *10th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, 2009, pp. 112–116.

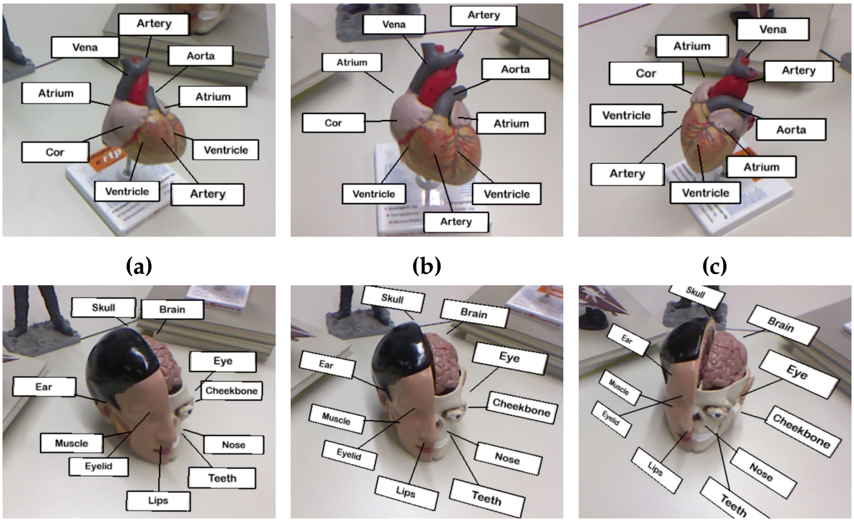
Paper F

Temporal Coherence Strategies for Augmented Reality Labeling

Jacob B. Madsen
Markus Tatzgern
Claus B. Madsen
Dieter Schmalstieg
Denis Kalkofen

The paper has been accepted to the
IEEE Transactions on Visualization and Computer Graphics (TVCG)

The layout has been revised.



(a) (b) (c)

(d) (e) (f)

Fig. F.1: (top) View management in 2D places annotations in image space and updates it in every frame. Conflicts between labels are resolved, but at the cost of an unstable layout. (bottom) View management in 3D places annotation in a plane defined in object space. This gives the option of making the layout stable by disabling updates, until the change of viewpoint becomes too significant.

Abstract

Temporal coherence of annotations is an important factor in augmented reality user interfaces and for information visualization. In this paper, we empirically evaluate four different techniques for annotation. Based on these findings, we follow up with subjective evaluations in a second experiment. Results show that presenting annotations in object space or image space leads to a significant difference in task performance. Furthermore, there is a significant interaction between rendering space and update frequency of annotations. Participants improve significantly in locating annotations, when annotations are presented in object space, and view management update rate is limited. In a follow-up experiment, participants appear to be more satisfied with limited update rate in comparison to a continuous update rate of the view management system.

1 Introduction

Annotations are commonly used in hand-drawn illustrations to add information, textual or pictorial descriptions to objects. Annotating elements of a map is a thoroughly researched area that has matured towards a common set of best practices in cartography [12]. By harnessing the computational power of modern computers, the placement of annotations can be fully automatized. View management algorithms automatically generate layouts of annotations for different application cases, such as maps [4] and 3D objects [5].

Numerous automatic view management techniques have been developed that mimic the annotation styles of traditional hand-drawn illustrations [1, 5] by enforcing a defined set of constraints. Hartmann et al. [10] provide guidelines after analyzing a wide range of medical and technical illustrations. They distinguish between two types of object annotations: Internal labels are directly overlaid onto the annotated object, while external labels are displaced from the object to avoid occlusions. Moreover, they identify three constraints for placing external labels: (R1) A label should be placed near the object it refers to. (R2) A label must not occlude another label or an annotated object. (R3) Leader lines connecting labels to annotated objects must not cross.

View management algorithms typically incorporate one or more of these constraints to create layouts of annotations. Aside from creating static layouts of annotations for print media, view management algorithms can also automatically accommodate a changing amount of information and even viewpoint changes of the annotated object. Hence, annotations can be used in Virtual Reality (VR) to interactively explore a virtual object. The layout adapts accordingly to enforce any constraints that are violated due to the changing viewpoints.

In Augmented Reality (AR), annotations provide additional information about real world objects. Due to the constantly changing viewpoint, the layout must be updated in every frame to resolve constraint violations, making it hard for users to keep track and to focus on single annotations. Therefore, *continuous* view management algorithms must also ensure that the layout respects temporal coherence.

A common approach to achieving coherence is to use hysteresis [5] to delay the positional updates of annotations. Alternatively, one can only update the layout when certain conditions are met, e.g., when the viewpoint of the object changes beyond a certain angular threshold [19, 20]. We refer to such approaches as *discrete* view management methods. Discrete methods trade potentially inferior layouts for improved temporal coherence.

In this paper, we perform a formal user evaluation to compare view management algorithms that continuously update the layout to algorithms that only update the layout at discrete points in time. Our focus lies on AR with permanent viewpoint changes. We limit ourselves to view management approaches that use external labels, since it was shown [6] that external labels are less ambiguous in case of tracking errors. We evaluate common force-based view management algorithms that work in *2D image space* [9], but also in *3D object space* [15, 19].

To the best of our knowledge, the behavior of labels over time has never been part of an evaluation of view management algorithms. Literature usually describes a set of constraints and methods to enforce these by continuously updating the layout. An open question is if such updates have a negative impact on the performance of a user during certain tasks, because they constantly change label positions.

Our intuition was that even though discrete view management algorithms cause violations of the layout constraints during viewpoint changes, they outperform the continuous versions in search-and-select tasks that are typical for AR applications using annotations. Based on our findings, we put forward design recommendations for view management systems.

2 Related work

Systematic annotation of objects have been discussed by cartographers since the 1970s. Imhof [12] generalizes a set of principles for annotating maps. In the 1980s, Ahn [1] presented an annotation algorithm for annotation of area features, line features and point features on cartographic maps.

With automatic generation of layouts and with digital representation of annotations on a computer, much research has shifted towards implementations of view management systems for digital media. Unfortunately, finding an optimal layout for annotations has been shown to be NP-hard even in 2D [13]. Practical view management systems focus on clustering or heuristic approaches [3, 5, 22], posing layouts as a constrained optimization problem for generating an optimal layout.

Hartmann et al. [9, 10] present guidelines for functional and aesthetic layout of external labels. They propose a set of metrics that can act as positive and negative constraints in a view management algorithm for automatic layout. These metrics are related to readability, unambiguity, aesthetics and frame coherence. Their 2D force-based method is probably the most widely used approach until today. Later work by this group introduced several automatic layout algorithms for external labels [2] and evaluated how coherency strategies can be used to annotate 3D animations of objects [8], such as moving the annotation itself or moving the reference line.

Azuma and Furmanski [3] presented view management techniques of 2D labels for AR based on various clustering strategies. They empirically evaluated user responses to the resulting motion. Further research using a stereoscopic HMD was done by Peterson et al. [14], enabling development of view management algorithms to leverage depth information of the scene to further separate annotations and creating new object annotation scenarios not possible in traditional 2D illustration. They describe a user study evaluating label placement techniques specifically developed for stereoscopic HMD usage, concentrating on how depth separation affects response time and errors.

Shibata et al. [18] describe the development of a modular view management system for mobile devices. This allows a developer to customize the view management system to target low powered mobile devices for use in mixed reality.

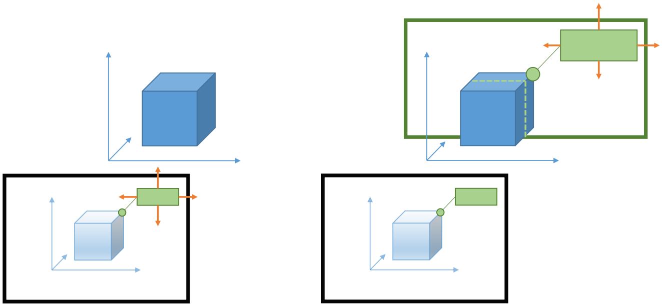
The first 3D view management system for AR was reported by Bell et al. [5]. Their system supports both internal and external labeling with greedy placement. Later work by Pick et al. [15] target an immersive multi-screen environment and resolves 3D occlusions using a technique similar to shadow volumes. Their approach uses a force based approach similar to the one of Hartmann et al., but in object space rather than image space. To ensure legibility of the system, a force is applied to the annotation, ensuring roughly constant distance in object space to the user. They evaluated their implementation using structured expert walkthroughs. Tatzgern et al. [19] developed a system for 3D view management of external labels in object space and addressed the problems of achieving temporal coherency, as the viewpoint changes.

Viewpoint changes trigger label layout updates to resolve violations of the layout rules. To allow users to follow these changes, the positional changes of labels are typically animated. Such animations are commonly used in information visualization to convey state changes to the user during interaction. Hence, animations have been used with the goal reduce the cognitive load when changing the visual states of hierarchical trees [16] or graph visualizations [21]. Heer and Robertson [11] studied animated transitions between different statistical data graphics and found that animated transitions improved graphical perception. The importance of animations is underlined in the design guidelines for fluid interactions in information visualization put forward by Elmqvist et al. [7]. The guidelines explicitly include smooth transitions to visualize transitions between different states so that potentially disorienting abrupt switches are avoided.

View management systems for AR should be able to display annotations coherently, as the viewpoint changes. Essentially, the update of the annotations must not confuse the user. This requirement adds complexity in comparison to static layout methods and has not received much attention in the literature. Ours is the first user study comparing the objective performance and subjective preferences of users exposed to view management in object space or image space. Additionally, we evaluate how discrete or continuous updates affect the users.

3 View Management Algorithms

We begin with a description of design options for view management and the associated advantages and disadvantages.



(a) (b)

Fig. F.2: Consider a scene with a blue 3D cube, which is projected to a light-blue image in the image plane (black frame): (a) View management in image space projects objects first and then places labels in 2D. (b) View management in object space places labels in 3D and then projects them to the image plane.

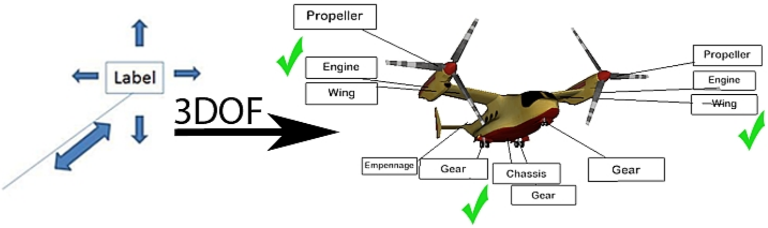
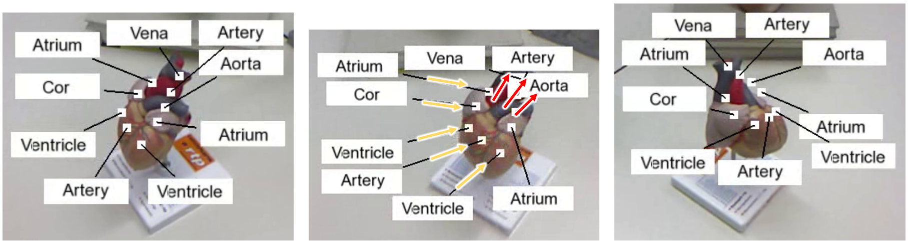


Fig. F.3: Placing annotations in 3D relative to the pole attached to the objects yields a useful constraint for temporally coherent viewing. All annotations have been adjusted along the pole first. The annotations marked with a green check-mark have also been slightly displaced in the viewing plane to avoid occlusions.

A two-dimensional label consists of a 2D billboard with text on it, a 3D anchor point on the annotated object, and a leader line connecting the anchor point and the billboard. Annotations can be placed with two degrees of freedom, namely the x and y coordinate in image space, since we do not allow label rotation.

A three-dimensional labels consist of similar elements: The billboard is given as a 3D polygon. The 3D anchor point is defined as before. The leader line becomes a 3D pole. Annotation can be placed with three degrees of freedom for position and one or two degrees of freedom for orientation, depending on whether one wants to allow a twist around the viewing vector or not.



(a) **(b)** **(c)**

Fig. F.4: Positional drift when using screen-space annotations. (a) Initially, labels are arranged around the object. (b) After camera movement, labels stick to their absolute image location. Note the long leader lines to the left. View management must move the labels back towards the object (yellow arrows) and also resolve the overlaps between the annotated object and the labels (red arrows) (c) Annotations can only stabilize their position relative to the object after the camera stops moving.

3.1 Image-Space vs. Object-Space Layout

Typically, view management techniques describe annotations in 2D relative to the object's projection into the image plane. More precisely, the anchor point is projected to 2D. After this projection, an image-space layout algorithm computes the 2D position of the label, as illustrated in Fig. F.2(a). The object (blue cube) is projected to the image plane (black frame), and the label (green) is placed in a 2D location near the object.

Conventional image-space algorithms, such as the one by Hartmann et al. [9] work with forces in 2D. A force resolves collisions (R2) between labels by pushing them away from each other using a direction vector that spans the centers of the 2D labels. The same applies to collisions between 2D labels and the projected 2D bounding box of the annotated 3D geometry. To avoid labels moving too far away from the annotated object (R1), another force pulls the labels back towards its annotated point. Crossing leader lines are resolved (R3) by switching the place of the labels that exhibit crossing leader lines. This is realized by applying a force that is orthogonal to the respective leader line.

Image-space algorithms were designed for producing static images. With camera motion, naive label updates, which re-use the absolute label position from the previous frame, lead to substantial positional drift. Labels stick to the screen, while the object moves away, as can be seen in Fig. F.4. Therefore, view management must actively apply rule R1 to move labels closer to the object (yellow arrows) and rule R2 to resolve the overlap between labels and object (red arrows). This positional drift is especially noticeable in AR, where viewpoint changes can include substantial translation.

A better approach for image-space label placement is to store label positions relative to the projection of the anchor point. This requires that the anchor points are re-projected into image space space in every frame, but eliminates the drift problem. Labels will not move further away from the object, but overlaps may still need to be resolved.

View management in object space treats labels as geometry placed relative to the anchor point in 3D. The label is part of the scene, and its projection to image space happens as part of the rendering process. Consequently, no drifting can occur.

Tatzgern et al. [19] propose to place labels in one or more 3D planes intersecting the annotated object, as shown in Fig. F.2(b). In this approach, updates to label positions are made in 3D and are guided by 3D rather than 2D constraints.

The pole's orientation is defined by the line connecting the centroid of the annotated object and the anchor point on the annotated object's surface. This ensures that leader lines cannot cross during initial placement, since they emerge radially from the annotated object's centroid.

The billboard must always touch the pole. This allows the following degrees of freedom for the annotation: The billboard can slide arbitrarily along the pole, as long as it does not penetrate the annotated object and does not move further away than a maximum distance from the anchor point. The billboard's center can be rotated arbitrarily about the chosen point on the leader line. Moreover, the billboard can be displaced arbitrarily in the chosen plane of orientation, as long as it still touches the leader line (Fig. F.3). This allows the billboard to be displaced by a maximum corresponding to half of its diameter.

Note that potential violations of the overlap constraint must still be determined in image space. In our approach, we do this by intersecting 2D bounding boxes of objects and labels after projection to the screen.

3.2 Continuous vs. Discrete Updates

Even after resolving the drift issue, perspective projection of labels can still lead to occlusion and crossing leader lines. Resolving these constraint violations requires updates to the layout after camera motion. In a desktop VR setup, this does not cause any problems, since the layout becomes stable, once the user lets go of the camera control. However, in AR, the viewpoint is constantly changing, because the camera is directly attached to the user's hand or head. Locally optimal placement may change from frame to frame even through small unintentional movements of the camera, leading to fluctuations, which are very unnerving for the observer (Fig. F.1(top)). Hence, labels never settle, which makes the aspect of temporal coherence a major issue for layout algorithms in AR.

Object-space algorithms can control fluctuations by switching from a continuous to a discrete update strategy. The layout is only calculated for an initial viewpoint. When the user changes the viewpoint, the labels retain their position in object space and remain temporally coherent. If they also retain their orientation (Fig. F.1(bottom)), label-to-label overlap cannot occur, but text readability may suffer from perspective foreshortening. If label orientation is re-adjusted to align with the image plane in every frame, label updates have to deal with more occlusions instead.

4 Experimental conditions

The considerations above suggest a design space for view management algorithms that has two main independent factors: the *space* in which the annotations are described and simulated algorithmically (2D, 3D), and the *update approach* (continuous, discrete). This implies four view management techniques, which we have evaluated in our experiments. We only considered drift-free methods in both 2D and 3D, since drift effects are clearly undesirable and would dominate the experience. For discrete methods, we only compute the layout once and free it after this initialization. For our test scenarios, which have a preferred viewing direction, this is sufficient and avoids handling update rate as a continuous variable.

4.1 Continuous Object-Space Labeling

The continuous 3D algorithm (**C3**) is derived from the force-based method of Tatzgern et al. [19]. Labels are placed in a 3D plane through the object center. Unlike the original method, we update the plane orientation to match the image plane in every frame. Consequently, this algorithm behaves similar to a 2D algorithm [9], and labels are always oriented towards the observer. However, updates of labels positions are made in object space, respecting 3D constraints.

4.2 Discrete Object-Space Labeling

The discrete 3D algorithm (**D3**) works like C3, but calculates the layout only once. After the initialization, the 3D label positions remain fixed. Eventual occlusions between annotations and annotations and the annotated object can be resolved by the user due to the parallax effect of the used 3D planes that place the labels.

4.3 Continuous Image-Space Labeling

The continuous 2D algorithm (**C2**) uses a force-based implementation similar to the one described by Hartmann et al. [9], with label positions stored relative to the anchor point projection (2D drift compensation).

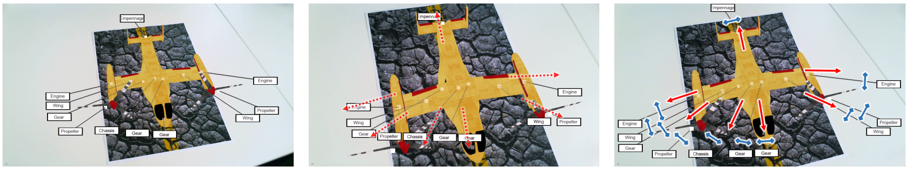
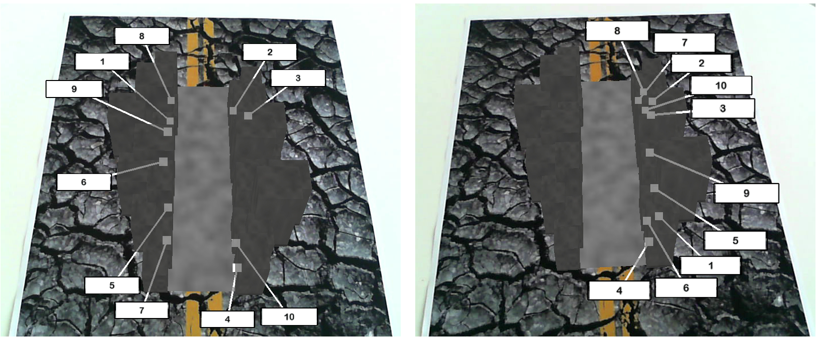


Fig. F.5: Discrete screen-space labeling. Unlike in the 3D view management, layout updates cannot be completely avoided. (a) The layout for the initial viewpoint. (b) If layout updates are stopped, labels overlap the annotated object when the viewpoint moves closer. The dotted arrows indicate the deactivated constraint for resolving overlaps. (c) The view management system continuously updates the layout to resolve overlaps (arrows). Another constraint ensures that labels stay close to their initial position on the bounding geometry of the object. Small movements are allowed to resolve eventual overlaps between labels (blue lines).



(a) **(b)**

Fig. F.6: (a) The distribution of anchor points is balanced over the object. (b) Anchor points are clustered on one side of the model.

4.4 Discrete Image-Space Labeling

The discrete 2D algorithm (**D2**) is a modified version of C2. If no further updates are made after computing an initial layout, the aforementioned 2D drift compensation can at least compensate for translational camera movements. However, rotations and scale changes quickly lead to substantial overlap between labels and objects, effectively rendering this method useless. We therefore decided to use a variant of C2, which retains the R1 and R2 forces. We only disable the R3 force resolving crossing leader lines, since, in practice, R3 is responsible for most of the disturbing non-continuous motions. Fig. F.5 illustrates these issues using the example of a motion bringing the viewpoint closer to the object.

5 Evaluation: Image vs. Object Space

In this evaluation, we investigate the task performance of the four implementations of the view management system.

5.1 Scenario

The experiment simulates a learning scenario, in which a user is confronted with an unfamiliar object. A user will typically get an overview of the parts from a more distant viewpoint, before stepping closer and investigating the details of the annotated object. Labels give the users an explanation of the annotated parts. They identify a part by following a leader line to the anchor point.

To avoid that participants rely on familiarity with the object of the study, we use an abstract shape that has no resemblance with any real world object. It consists of annotated blocks of approximately equal size and uniform color (Fig. F.7a). The objects lack any salient clues, which participants could use to memorize the location of the anchor points. Our intention was to make the perceived visual clutter in this experiment is consistent among the participants and avoid influence from prior knowledge of the object or any prior expertise [17] for the objects in this test scenario, given the users unfamiliarity with the scene, task and objects presented.

We use the same 3D positions for anchor points across all experiments to achieve consistent visual clutter. To minimize any bias of individual task and factor combinations, each participant is presented with a randomized selection of anchor points and associated labels, based on the task conditions.

The perceived visual clutter may differ between participants. However, we use an identical setup for each participant, so that the objectively measurable visual clutter in the scene is consistent.

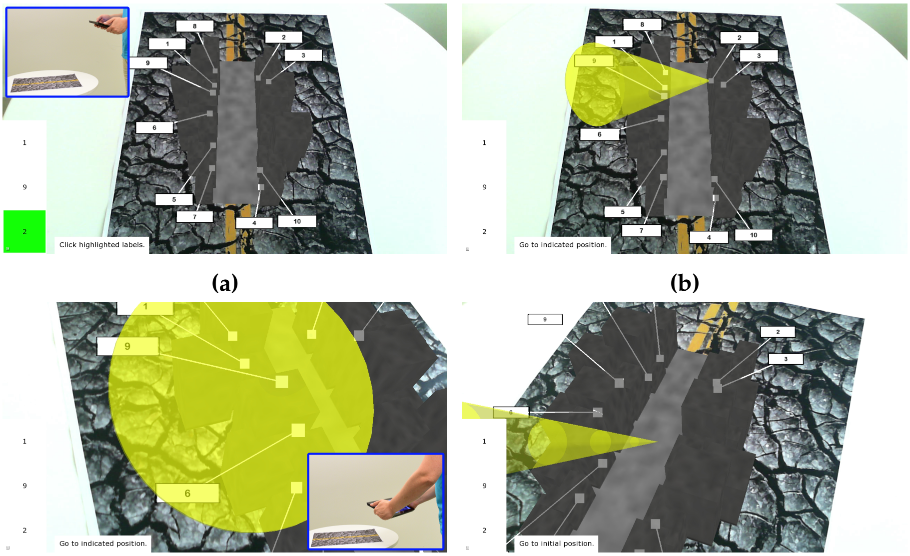
5.2 Apparatus

The experimental code is written in C++ using OpenSceneGraph[1] for task creation and scene display. Marker tracking is handled via a natural feature tracker. The trackable marker is printed on an A3-format paper (297mm × 420mm) and is placed as the sole item on a freely standing table, with sufficient room to move around.

The experimental application is deployed on a Microsoft Surface Pro 2 tablet running Windows 8, with an Intel Core i5 CPU, 4GB RAM and a 10.6" screen (1920 × 1080, 208 ppi, 16:9 aspect ratio). We use the tablet's rear-facing camera (1.2MPixels, 720p) for tracking. Input to the application is handled via the touch screen.

---

[1] http://www.openscenegraph.org/



(a) **(b)**

(c) **(d)**

Fig. F.7: Experiment participants had to perform the following steps. (a) From an overview position, the participant selects a sequence of labels indicated by the system by highlighting a number on the left. (b) After finishing the sequence, the participant must click on the labels. (c) For each label, the user is asked to align the viewpoint with a cone. (d) After exploring each labels up close, another cone guides the user back to the overview position.

5.3 Study design

We define three independent variables for this study: the update method (continuous, discrete), the spatial description of labels (2D, 3D) and the distribution of anchor points of the object (balanced, unbalanced). The variables regarding the update method and spatial description directly refer to the previously described view management algorithms: 2D image-space with continuous update (C2) and discrete update (D2); 3D object-space with continuous (C3) and discrete update (D3).

We included the distribution of anchor points as variable, because we wanted to investigate its effect on the behavior of labels during on the viewpoint changes. We speculated that multiple anchor points grouped very closely together on the reference object cause more violations of the layout constraints and stronger label movements in continuously updating view management systems. In contrast to such unbalanced layouts (Fig. F.6b), a more balanced distribution (Fig. F.6a) of anchor points causes less changes.

The experiment is a mixed methods design, using a randomized, repeated-measures design, with two factors being within-subject, and one factor being between-subject. The within-subject factors are update method (continuous, discrete) and the spatial description (2D, 3D), while distribution of anchor points (B=balanced, U=unbalanced) is a between-subjects factor. The within-subject factors correspond to the four view management systems. Each participant performed three repetitions, leading to a total of 12 tasks per participant. For each participant, the combination of factors and their order was randomized using Latin squares.

As dependent variables, we measured the duration of each task, and the duration of the full trial. Furthermore, we measured error rate metrics and layout statistics: the amount of wrongly identified labels, label order changes, leader line crossings, object space and screen space movement of the relevant labels.

5.4 Task

A task consists of the following steps (Fig. F.7):

S1 The participant must identify three labels of interest in a certain order in the overview by clicking on them, then

S2 physically move the viewpoint closer to each anchor point of the corresponding label of step 1.

S3 Repeat (S1) and (S2) three times for each factor-level combination

The purpose of the tasks is to simulate a learning environment, in which a user gets an overview of an object from a viewpoint from which the whole object and its annotations are visible. This is simulated with step (S1), in which the participant had to select a randomly generated sequence of three labels. The sequence was shown on the mobile device (Fig. F.7a). After identifying and clicking on all the relevant labels, the participants had to physically change the viewpoint of the device and move the viewpoint closer to each identified anchor point (S2). Participants performed the second step for each label in the same order as they were presented in the first step. Before moving closer to a label, they had to click on it again to select it. Clicking on a label would force participants to look for the label by moving the device, which would again trigger layout changes.

After clicking on the label, the system showed a transparent yellow cone, with which the participants had to align the mobile device in order to continue the study (Fig. F.7b). This additionally enforced movement of the mobile device. The bottom of the cone indicated the position the mobile device should move to, while the tip of the cone pointed to the anchor point of the identified label. The participant had to align the mobile device with the bottom and look at the tip of the cone (Fig. F.7c). The cone disappeared when the alignment was correct, which indicated that the participant could continue with the task. We introduced a positional and angular tolerance to the alignment, to avoid that participants spend too much time aligning the view. During the trials, we did not experience issues with participants having alignment problems.

After aligning the device with the cone, the task continued with the next label, until the task forced participants to go back to the overview to trigger the next iteration (Fig. F.7d), starting again with step (S1). Overall, each participant repeated the task twelve times, three times for each investigated view management system.

5.5 Hypotheses

We expected that user performance of task completion differs depending on the view management systems. We had two main hypotheses:

- **H1**: User task completion time differs between the view management systems.
- **H2**: Anchor point distribution has an effect between view management systems.

Regarding H1, we expected that the properties of the view management system influences the task performance during viewpoint changes. When a user changes the viewpoint, a continuously updating system constantly resolves layout constraints. Therefore label positions and their relationship to each other and to the annotated object change. We reasoned that such changes have a negative impact on repeatedly locating labels, as required by the task of this evaluation. In discrete setups, the relative label layout does not change, which makes it easier for users to keep track of the locations of labels during viewpoint changes and improves task performance.

In the discrete case (D2 and D3), labels will never change relative order. Like D3, D2 is prone to layout violations from crossing leader lines. Therefore, we expected the discrete update methods D2 and D3 to outperform both continuous update methods C2 and C3. Hence, the properties of the view management algorithm was considered to have an effect on task completion time.

Regarding H2, we expected that the distribution of labels relative to the object influences the view management systems in different ways. For this purpose, we defined balanced and unbalanced distributions of annotated labels. The unbalanced layout grouped anchor points and their corresponding labels closely together. We speculated that during viewpoint changes, this setup would cause more layout violations and more label updates than a balanced setup, where anchor points and labels are well distributed. We hypothesized that balanced and unbalanced layouts would lead to a performance difference, because the relative locations of labels changed to a different degree.

5.6 Procedure

Prior to starting the experiment, we asked the participant to fill out an informed consent form along with a demographic questionnaire, including questions of age, familiarity with technology, mobile devices and AR technology. We introduced the participant to the experiment with a thorough explanation of the purpose of the study and the used system.

Before starting the experiment, the participant performed a set of training tasks with the view management system of the current condition. During this task, the participants were free to ask any questions regarding usage and control of the system. The training task was a simplified version of the real task with only four labels, two of which were part of the selection and identification procedure. The configuration of labels of the training task was different than the configuration of the actual task to avoid unintended learning effects. Following the training task, the participant completed the task without interruption. After finishing the task, participants were allowed to take a break, before moving on to the next view management condition, which again started with a training task.

Participants were told to pay attention to solving the task to the best of their abilities, and not mind the amount of time spent on each task. We logged both completion time and error rate.

After completing all trials, the participant responded to a small open-ended questionnaire and gave additional verbal feedback in an interview with emphasis on the interface and their strategies for completing the tasks.

5.7 Participants

A total of 24 participants (6 female) were part of the experiment, aged 24-36 (M=29,3). All were recruited from on and off the campus area. All participants self-reported normal or corrected-normal vision. Participant familiarity with AR was self-reported to be average on a 5 point Likert rating, ranging from novice to expert user, and familiarity with handheld mobile devices above average using the same 5-point Likert rating scale. Data collected from the experiment comprised of data from 24 participants, each with 12 tasks, resulting in a total of 288 tasks. The average completion time of the experiment per user was 24 minutes (SD = 4 min).

5.8 Results

The analysis has been carried out using the statistical software R, using a significance level of $\alpha = 0.05$. The main analysis method were type III ANOVA and Friedman test, testing the overall difference between the three independent variables. Pairwise comparisons in post-hoc tests were performed using Tukey's honest significant difference (HSD) method for any interaction, while

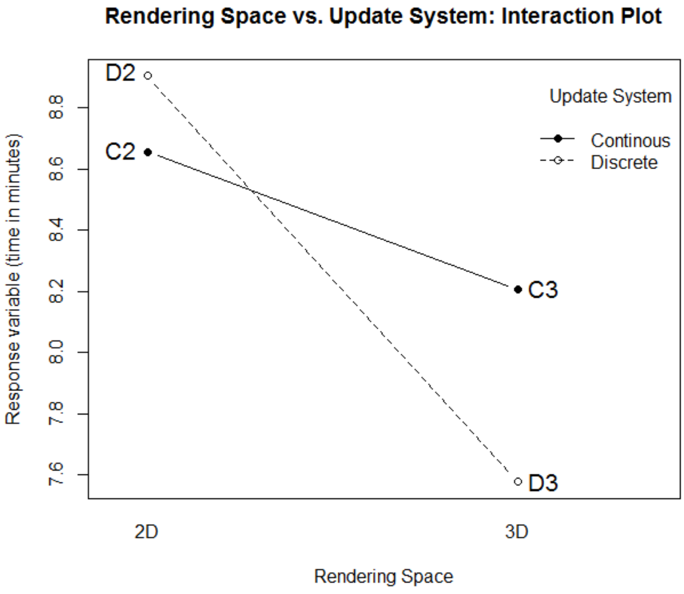


Fig. F.8: Interaction plot of factors rendering space and update method. Although significance can be reported, the data suggests an interaction is present between the two factors.

controlling the error rate. Near-significant results reported in the results are defined as being in the range of $0.05 < p < 0.10$.

This section describes different statistical tests throughout. The reason for using multiple different tests lies in the relationship amongst the data. Initially, we log transform the data in order to meet assumptions for normality. However, when splitting up the data or investigating interesting factors or interaction, this proved to be impossible. In these cases, as well as in cases for a single factor, a relevant statistical method has been chosen, according to the data level of measurement, as well as the parametric or non-parametric nature of the data.

In the experiment, the collected task performance time did not meet the standard assumptions of ANOVA analysis, as the normality of the residuals and the homogeneity of the variance for the factors were found to be problematic. A logarithmic transformation of the task time resolved the problem. Therefore, we used a logarithmic scale for statistical analysis of the task performance time. As verification of the conclusions, a non-parametric Friedman test was run in parallel with ANOVA to ensure that no conflicting conclusions were found and that the conclusions are well supported.

Results of type III ANOVA report the spatial description (2D, 3D) as a significant main effect ($F_{1,22} = 15.79, p < 0.001, \eta_G^2 = 0.084$). This means that participants showed significantly slower completion time in the image-

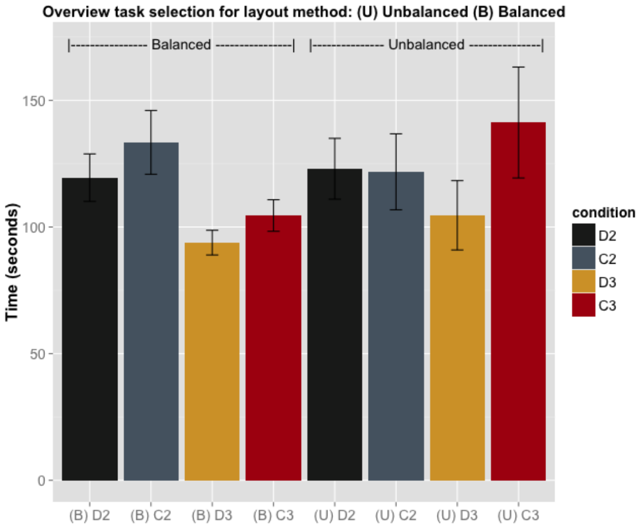


Fig. F.9: Isolated data of step S1 of the task for each condition. A bar represents the average time for the "selection" part (S1) of the full task (as is shown in Fig. F.7 (a)), reported with standard error.

space rendering condition. This is in line with the first hypothesis, H1. The second hypothesis, H2, is not supported in the $log(time)$ data of the full set of tasks, as the layout of anchor points does not show time performance difference between the layouts. No other main effects or interactions were found in the analysis of $log(time)$. However, the two-way interaction between spatial description and update method was near-significant $(F_{1,22} = 3.90, p = 0.061, \eta_G^2 = 0.018)$.

The near-significant interaction between rendering space and update method is illustrated in Fig. F.8. While the interaction was not significant, there was an interesting visual cue illustrated by the crossing (i. e., interaction), which lead us to investigate possible interactions using a follow-up Tukey-HSD test. The test on the within-subjects factors from the experiment showed that the 3D discrete system (D3) significantly differs from both 2D systems, C2 and D2, (both $p < 0.001$). D3 is faster than both C2 and D2. Furthermore, C3 is near-significantly different from D2 $(p = 0.085)$. Based on these results, D3 achieved best task performance in this experiment.

In order to explain the performance difference, we isolated portions of the task for further investigation. In step S1 of the task, participants had to find and select three labels, and repeated this step three times. Therefore, participants could potentially build spatial memory of the label locations, which would be evident in a better performance. To investigate a potential effect on spatial memory, we isolated the data of step S1 (Fig. F.9). To investigate the data, we hypothesized that all conditions performed equally, and used a Chi-Squared Goodness-of-Fit to test for consistency in the data, treated as a categorical variable. We found a significant difference from the expected values ($\chi^2(7) = 359.43, p < 0.01$). To determine which factors follow the expected performance, we used multiple pairwise comparisons with Bonferroni correction, which confirmed that only "B+D2", "U+D2" and "U+C2" follow the expected performance, while the rest either has a longer (B+C2, U+C3) or shorter average duration (B+D3, B+C3, U+D3) as Fig. F.9 indicates.

Visually inspecting the results, the difference between balanced and unbalanced C3 (red bars in Fig. F.9) is larger than for the other systems. We investigated the data of label order changes for C3 non-parametrically for difference in mean using a Mann-Whitney U test, because the isolated data does not meet the criteria for normality. The test indicated that the number of re-orderings of labels between the two conditions is a significant factor: The mean ranks of balanced and unbalanced groups were 25.5 and 47.5, respectively ($U = 251, n = 72, Z = -4.4713, p < 0.01$). This result indicates that anchor point distribution might have an effect on view management system in accordance with the H2 hypothesis.

6 Evaluation: Continuous vs. Discrete

We performed a follow-up study to collect more qualitative feedback on selected view management systems. Although participants of the previous study already filled out a questionnaire to collect qualitative feedback, the questionnaires did not yield any reliable results regarding the preferences of systems. Based on the feedback gathered from participants, we believe that participants could not distinguish between the four systems after completing the experiment.

To avoid users mixing up the different systems, we focused on only two view management systems. The first study identified D3 as the one achieving best task performance. Therefore, we removed the 2D conditions and compared only C3 and D3 in this study. Furthermore, the data from the first experiment indicated that unbalanced layouts cause stronger layout changes than balanced layouts. Therefore, we removed the independent variable regarding the distribution of anchor points of the object by focusing only on the more challenging unbalanced scenario. The apparatus, task and procedure were identical to the first experiment.

6.1 Study design

The study had a randomized, within-subjects design with one independent variable: update method (continuous, discrete). In this study, we only used 3D view management methods. Therefore, the method directly refers to the continuous 3D method (C3) and the discrete 3D method (D3). Participants performed the same task as in the first study, as illustrated in Fig. F.7. For the two conditions, this produces a total of six tasks per participant.

Data is procured from a 5-Likert scale questionnaire on the participant's satisfaction using the methods, immediately following the tasks. The scale for satisfaction ranges from dissatisfied to satisfied with the behaviour of the system that had been used. Participant preference data is gathered following the full test of both factors, and the participant is prompted to answer which methods was preferred for the given tasks, forcing the user to make a choice.

6.2 Participants

A total of 10 participants (all male) were part of in the second experiment. All participants where recruited from the same pool of participants as those who took part in the first experiment. No participant took part in both experiments. All self-reported normal or corrected-normal vision.

6.3 Results

The analysis was performed with the statistical software R, using a significance level of $\alpha = 0.05$. The analysis method for user satisfaction was a Wilcoxon signed-rank test. Analyzing participant preference scores was performed using the Exact Binomial test method.

We found a significant effect when analyzing the difference in the responses of the 5-Likert scale question of participant preference. The mean ranks of D3 and C3 were 14 and 7, respectively: $W = 3.5, Z = -2.21, p = 0.02734, r = -0.49$. For ties in the data, where two or more values are the same, we average the ranks for the tied values to compute the values. This is a strong indication that users were overall more satisfied with the discrete update system, in comparison with the continuous, despite a lack of significant differences in task performance.

For participant preference, users were asked to make a binary choice of preference, choosing between either the continuous or discrete update system. As one user did not have a preference, the outcome was eight for discrete, one undecided, and one for continuous. This indicates preference towards the discrete system, even though the small sample size does not allow to observe a statistically significance (8/10 votes for discrete updating is not significantly higher than chance at $0.5, p = 0.055$, 1-sided. We are confident that the binary choice of preference shows borderline significance due to the small number of participants in this second experiment. A follow-up study with a larger number of participants will provide valuable insights in user preference for continuously or discretely updating systems.

7 Discussion

The first study clearly showed that the view management system, which treats labels as 3D objects and creates a static layout (D3), significantly outperforms the 2D continuous view management system (C2). This is in line with our expectation, because the continuously updating layout seems to make it difficult for users to keep track of the labels. However, D3 also outperforms its 2D counterpart D2, which avoids strong label motions by preserving the order of labels. This difference can be attributed to the more stable placement in object space.

By inspecting the performance data (Fig. F.8), we can see that D2 and C2 exhibit very similar performance. This indicates that, even though D2 enforces a certain label order, the small motions of the simulation running in the background and the lack of a static 3D representation may have a negative impact on the ability of users to locate and interact with labels.

The difference between the 2D systems and D3 could also be explained by the way labels are implemented. Despite using force-based approaches for both systems, the spatial representation of labels clearly influences the implementation of the systems. To isolate this factor, we also included a continuously updating 3D layout (C3) in our study. Indeed, the near-significant difference between D2 and C3 hints at implementation specific differences (Fig. F.8). This supports our argument that a spatial representation of labels as 3D objects may be preferable. Even though the discrete algorithm D2 does not rearrange labels, it seems to perform worse than an implementation that continuously rearranges labels, but works in 3D space (C3). This aspect requires further investigation by unifying the behavior of 2D and 3D layout algorithms based on quantitative analysis of differences in motion patterns. However, due to the difference in the spatial representation, it may be extremely challenging to make the systems behave exactly the same.

Isolating the data of step S1 of the task, (Fig. F.9) shows a difference between D3 and C3, which is not present in the initial analysis of the first experiment. We performed a follow-up study, in which we compared D3 and C3 in order to collect more qualitative feedback to investigate this difference. For this study, we focused on an unbalanced label distribution, because this is the worst-case scenario for label placement. Small viewpoint changes can already introduce a reordering of labels. This study revealed that participants preferred a discrete 3D layout (D3) for the given task. Therefore, we can recommend 3D layouts, which are placed in static locations relative to the object, as the most suitable approach to present labels in a view management system.

Overall, we accept hypothesis H1. A combination of update method (continuous, discrete) and spatial description (3D, 2D) has an influence on task performance. The static layout of the discrete 3D view management system significantly outperforms the 2D versions. A visual inspection of the performance data also shows better performance, when compared to the continuous 3D version. Due to the small sample size of our second study, we refrain from reporting on the performance difference between C3 and D3. To gather reliable performance data, a follow-up study should investigate the performance aspect with a larger sample size.

Regarding the second hypothesis H2, we did not find any significant difference in $log(time)$ task performance between balanced and unbalanced layout systems. Nevertheless, we noticed a larger number of label changes in the data when the label layout was unbalanced. To compare this layout data across all systems, we limit the data set to logged data in which the users are in overview mode, i. e., the part of the task where all labels are visible and the participant must identify three labels correctly.

The number of label order changes in the unbalanced condition is 1151 changes, while the number for the balanced condition is 457. This leads to approximately four changes per overview in the unbalanced condition and between two and three changes per overview in the balanced condition (24 participants × 4 systems × 3 tasks = 288 overviews). Hence, the amount of layout change is higher, when algorithms create layouts for unbalanced label distributions, than for balanced label distributions. This objective measurement supports our assumption that layouts are prone to changes, when anchor points are not distributed well in the annotated scene. Furthermore, visual inspection of Fig. F.9 shows a strong performance difference of the continuously updating system C3 between the unbalanced and balanced condition. Nevertheless, the results of the first study indicate that avoiding constant label motion and placing labels relative to the annotated object in 3D, as done by system D3, has a larger impact on performance than avoiding label order changes.

8 Conclusion

Based on the results of the studies, we can give the following recommendations for designing a view management system. Given that users perform better with labels that are placed statically in the 3D space relative to the annotated object, view management systems should avoid updating labels after their initial placement from the current viewpoint. Furthermore, labels should be treated as 3D objects and part of the scene. Integrating 3D labels into the scene not only allows the AR system to naturally apply the camera pose to the labels, but also simplifies the design of 3D interaction methods. For instance, a method for manipulating a 3D object in AR can be directly applied to manipulating a label. To avoid frequent changes of the layout, which could influence the ability to interact with labels, the anchor points of labels and the labels themselves should be well distributed in the annotated space.

An additional advantage of a static 3D layout is that it can be calculated by optimizing the overall layout for a single frame. After the initial computation, no additional computational resources are required, because the layout is not continuously updated. This factor can be beneficial for the battery life of mobile devices.

For future work, it is interesting to isolate which factor influences the performance of the participants. Object-space labeling may have performed better due to stable label placement or due to the additional spatial cues from registering the labels as 3D objects in the scene. Another avenue of future work is the update method of the object-space labeling system. Currently, the system updates all labels, when the viewpoint moves beyond a certain threshold. The layout of the new viewpoint may be completely different from the one of the previous viewpoint. Such drastic changes may have a negative effect on relocating labels of the object. To achieve better coherence between these layouts, a better solution would take into account the layout of the previous viewpoint when calculating the layout of the new viewpoint.

Acknowledgment

This work was partially funded by the European Union (FP7-ICT-601139 "CultAR", FP7-ICT-611526 "Magellan"), by the Austrian Science Fund (FWF) under contract P-2402 and by the Christian Doppler Society ("CDL Handheld Augmented Reality").

References

[1] J. Ahn and H. Freeman, “A program for automatic name placement,” pp. 101–109, 1984.

[2] K. Ali, K. Hartmann, and T. Strothotte, “Label layout for interactive 3D illustrations,” *Journal of the WSCG*, vol. 13, 2005.

[3] R. Azuma and C. Furmanski, “Evaluating label placement for augmented reality view management,” in *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.* IEEE Comput. Soc, 2006, pp. 66–75.

[4] K. Been, E. Daiches, and C. Yap, "Dynamic Map Labeling," *IEEE transactions on visualization and computer graphics*, vol. 12, no. 5, pp. 773–780, 2006.

[5] B. Bell, S. Feiner, and T. Höllerer, "View management for virtual and augmented reality," in *Proceedings of the 14th annual ACM symposium on User interface software and technology - UIST '01*, vol. 2001. New York, New York, USA: ACM Press, 2001, pp. 101–110.

[6] E. M. Coelho, B. MacIntyre, and S. J. Julier, "OSGAR: A scene graph with uncertain transformations," *ISMAR 2004: Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 6–15, 2004.

[7] N. Elmqvist, A. Vande Moere, H.-C. Jetter, D. Cernea, H. Reiterer, and T. J. Jankun-Kelly, "Fluid interaction for information visualization," *Information Visualization*, vol. 10, no. 4, pp. 327–340, Oct. 2011. [Online]. Available: http://dx.doi.org/10.1177/1473871611413180

[8] T. Götzelmann, K. Hartmann, and T. Strothotte, "Annotation of Animated 3D Objects," in *In SimVis, SCS*, 2007, pp. 209–222.

[9] K. Hartmann, K. Ali, and T. Strothotte, "Floating labels: Applying dynamic potential fields for label layout," *Smart Graphics*, pp. 101–113, 2004.

[10] K. Hartmann, T. Götzelmann, K. Ali, and T. Strothotte, "Metrics for Functional and Aesthetic Label Layouts," in *Smart Graphics: 5th International Symposium, SG 2005*, 2005, pp. 115–126.

[11] J. Heer and G. Robertson, "Animated transitions in statistical data graphics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1240–1247, Nov. 2007. [Online]. Available: http://dx.doi.org/10.1109/TVCG.2007.70539

[12] E. Imhof, "Positioning Names on Maps," *The American Cartographer*, vol. 2, no. 2, pp. 128–144, 1975.

[13] J. Marks and S. Shieber, "The Computational Complexity of Cartographic Label Placement," Center for Research in Computing Technology, Harvard University, Tech. Rep., 1991.

[14] S. D. Peterson, M. Axholt, M. Cooper, and S. R. Ellis, "Evaluation of Alternative Label Placement Techniques in Dynamic Virtual Environments," in *10th International Symposium, SG 2009*, ser. Lecture Notes in Computer Science, A. Butz, B. Fisher, M. Christie, A. Krüger, P. Olivier, and R. Therón, Eds., vol. 5531. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 43–55.

[15] S. Pick, B. Hentschel, I. Tedjo-Palczynski, M. Wolter, and T. Kuhlen, "Automated Positioning of Annotations in Immersive Virtual Environments," in *EGVE - JVRC'10 Proceedings of the 16th Eurographics conference on Virtual Environments & Second Joint Virtual Reality*, 2010, pp. 1–8.

[16] G. G. Robertson, J. D. Mackinlay, and S. K. Card, "Cone trees: Animated 3d visualizations of hierarchical information," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '91. New York, NY, USA: ACM, 1991, pp. 189–194. [Online]. Available: http://doi.acm.org/10.1145/108844.108883

[17] R. Rosenholtz, Y. Li, J. Mansfield, and Z. Jin, "Feature Congestion : A Measure of Display Clutter," in *CHI 2005 Proceedings*, 2005.

[18] F. Shibata, H. Nakamoto, R. Sasaki, A. Kimura, and H. Tamura, "A View Management Method for Mobile Mixed Reality Systems," in *EGVE'08 Proceedings of the 14th Eurographics conference on Virtual Environments*, 2008, pp. 17–24.

[19] M. Tatzgern, D. Kalkofen, R. Grasset, and D. Schmalstieg, "Hedgehog labeling: View management techniques for external labels in 3D space," in *2014 IEEE Virtual Reality (VR).* IEEE, Mar. 2014, pp. 27–32.

[20] M. Tatzgern, D. Kalkofen, and D. Schmalstieg, "Dynamic compact visualizations for augmented reality," *2013 IEEE Virtual Reality (VR)*, pp. 3–6, Mar. 2013.

[21] K.-P. Yee, D. Fisher, R. Dhamija, and M. Hearst, "Animated exploration of dynamic graphs with radial layout," in *Proceedings of the IEEE Symposium on Information Visualization 2001 (INFOVIS'01)*, ser. INFOVIS '01. Washington, DC, USA: IEEE Computer Society, 2001, pp. 43–. [Online]. Available: http://dl.acm.org/citation.cfm?id=580582.857705

[22] F. Zhang and S. Hanqiu, "Dynamic Labeling Management in Virtual and Augmented Environments," in *Ninth International Conference on Computer Aided Design and Computer Graphics (CAD-CG'05).* IEEE, 2005, pp. 397–402.

SUMMARY

In Augmented Reality applications, the real environment is annotated or enhanced with computer-generated graphics.

This is a topic that has been researched in the recent decades, but for many people this is a brand new and never heard of topic.

The main focus of this thesis is investigations in human factors related to Augmented Reality. This is investigated partly as how Augmented Reality applications are used in unsupervised settings, and partly in specific evaluations related to user performance in supervised settings.

The thesis starts by introducing Augmented Reality to the reader, followed by a presentation of the technical areas related to the field, and different human factor areas. As a contribution to the research area, this thesis presents five separate, but sequential, papers within the area of Augmented Reality.

ISSN (online): 2246-1248
ISBN (online): 978-87-7112-307-4

AALBORG **UNIVERSITY PRESS**