Doctor's Thesis

Designing for Large Scale High Fidelity Collaborative Augmented Reality Experiences

Damien Constantine Rompapas

September 10, 2019

Nara Institute of Science and Technology
Graduate School of Information Science

This paper is a doctoral dissertation submitted to the Graduate School of Information Science and Technology, Nara Institute of Science and Technology as a requirement for awarding a doctoral degree in engineering.

Damien Constantine Rompapas

Reviewing Committee :

Professor Hirokazu Kato	(Supervisor)
Professor Kiyoshi Kiyokawa	(Co-Supervisor)
Assistant Professor Alexander Plopski	(Co-Supervisor)
Associate Professor Christian Sandor	(City University of Hong Kong)

Designing for Large Scale High Fidelity Collaborative Augmented Reality Experiences*

Damien Constantine Rompapas

Abstract

In recent years, there has been an increasing amount of Collaborative Augmented Reality (CAR) experiences, classifiable by the deployed scale and the fidelity of the experience. In this thesis, I first explore the LSHF CAR design space, drawing on technical implementations and design aspects from AR and video games. I then create and implement a software architecture that improves the accuracy of synchronized poses between multiple users. Finally, I apply my target experience and technical implementation to the explored design space. A core design component of HoloRoyale is the use of visual repellers as user redirection elements to guide players away from undesired areas. To evaluate the effectiveness of the employed visual repellers in a LSHF CAR context I conducted a user study, deploying HoloRoyale in a 12.500m^2 area. The results from the user study suggest that visual repellers are effective user redirection elements that do not significantly impact the user's overall immersion. Finally this thesis focuses on the visual consistency component of fidelity, expanding on EyeAR: refocusable content on Optical See-Through Head Mounted Displays (OST-HMDs) by evaluating the fidelity of refocusable content displayed on a single plane OST-HMD via. a modified Touring Test. The results from the evaluation show that refocusable content improves the fidelity of OST-HMD experiences. This work is the first to explore the domain of LSHF CAR and provides insight into designing experiences in other AR domains.

Keywords: Large Scale, Augmented Reality, High Fidelity

*Doctor's Thesis, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology,September 10, 2019.

大規模な高忠実度の共同拡張現実感体験のための設計*

ダミエン　コンスタンチン　ロンパパス

概要

In recent years, there has been an increasing amount of Collaborative Augmented Reality (CAR) experiences, classifiable by the deployed scale and the fidelity of the experience. In this thesis, I first explore the LSHF CAR design space, drawing on technical implementations and design aspects from AR and video games. I then create and implement a software architecture that improves the accuracy of synchronized poses between multiple users. Finally, I apply my target experience and technical implementation to the explored design space. A core design component of HoloRoyale is the use of visual repellers as user redirection elements to guide players away from undesired areas. To evaluate the effectiveness of the employed visual repellers in a LSHF CAR context I conducted a user study, deploying HoloRoyale in a 12.500m^2 area. The results from the user study suggest that visual repellers are effective user redirection elements that do not significantly impact the user's overall immersion. Finally this thesis focuses on the visual consistency component of fidelity, expanding on EyeAR: refocusable content on Optical See-Through Head Mounted Displays (OST-HMDs) by evaluating the fidelity of refocusable content displayed on a single plane OST-HMD via. a modified Touring Test. The results from the evaluation show that refocusable content improves the fidelity of OST-HMD experiences. This work is the first to explore the domain of LSHF CAR and provides insight into designing experiences in other AR domains.

キーワード: Large Scale, Augmented Reality, High Fidelity

*Doctor's Thesis, 奈良先端科学技術大学院大学 情報科, 学研究科 情報科学専攻 2019 年 9 月 10 日.

Table of Contents

List of Figures

Acknowledgements

The creation of this thesis has been somewhat an intense journey over a large portion of my early lifespan, along the way several people have provided their support. As such, there are a massive number of people who contributed both to my progress and morale during my thesis, and as such I would like to thank. First and foremost, I would like to thank my superviors Professor Hirokazu kato and Associate Professor Christian Sandor for the valuable advice given to me during my PhD degree. More importantly I wish to express a detailed thanks to Associate Professor Christian Sandor who, over the period of the last 6 years has taken the time to support me during all of my endeavours and push me to my limit (and then further past it). It is thanks to his input (and at times intense arguments and discussions) that the projects described in my thesis were transformed from ideals to a reality. Second, I would like to thank Assistant Professor Alexander Plopski for both his advice and for the late night discussions, supporting me during the most pressing hours. Third, I would like to extend my thanks to Professor Daniel Saakes for his valuable input during the design of HoloRoyale and his industrial insight. Fourth, I would like to thank both my friends and family back home (down under) for cheering for me. Fifth, I'd like to thank one of my now close friends in Japan, Sofia Onyshko, for providing a comfortable environment for me to vent my stresses during the most pressing hours of this thesis, and being there for me during the times of celebration. Sixth, I would like to thank the members of the Interactive Media Design laboratory for making the research environment both fun and engaging. Seventh, I would like to extend my thanks to the people at SCAPE technologies for hosting me during my short, but very productive internship. Finally, I would like to thank the rest of my friends in Osaka, Japan. No matter how insignificant you believe your interaction with me might have been, there was always a lesson to be learned or a warming moment shared.

Chapter 1. Introduction

Augmented Reality (AR) is the technique of embedding Computer Graphics (CG) into the user's view of their surrounding environment. This is done by either projecting CG images onto the environment, or through a display medium that shows the user the virtual and the real world simultaneously. The idea of the ultimate display that provides perfect visual and interactive feedback was originally devised in 1965 by Ivan Sutherland [93]. He described the experience which a display provides should have virtual objects that behave like their real life counterparts. As such a defined aspect of the immersion an AR experience provides is the fidelity of the content presented to the user. Kruijff et al. [53] identified several issues that affect the fidelity of an AR experience. From the identified issues, I describe the fidelity of an AR experience by the following metrics:

- **Virtual-Real interactions:** Does a virtual object behave like its real world counterpart? This includes physical collisions and occlusions. An example would be a virtual ball interacting with a real wall.
  High Fidelity: A real wall in front of a virtual ball will occlude it. When the ball is thrown at the wall, it will bounce off of the wall.
  Low Fidelity: The virtual ball will always be visible, contradicting the depth placement of the two objects. When thrown, the ball will pass through the real wall.
- **Accurate content registration:** Is the placement of virtual content consistent with the real world context? An example is a virtual statue being placed on a bust.
  High Fidelity: If correctly registered the statue will appear on the bust.
  Low Fidelity: A bad registration will lead to the statue visually floating in the air.
- **Spatio-temporal consistency:** When an interaction occurs, do all users see the action at the same time and place? For example, a user

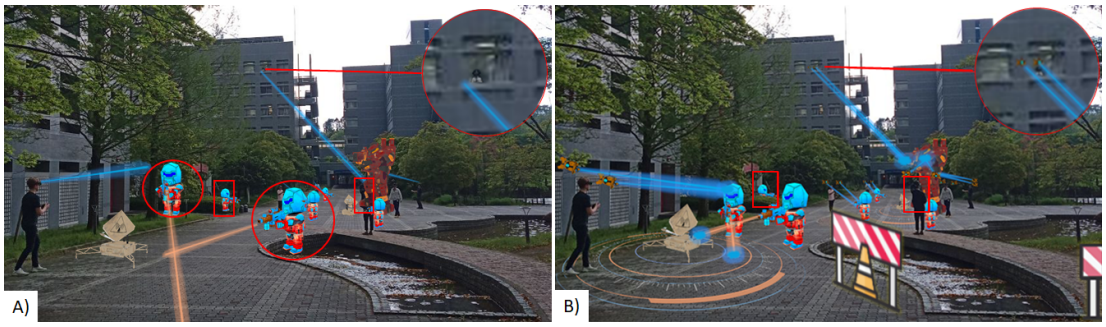

Figure 1.1: The aim of this work is to create and deploy a Collaborative Augmented Reality (CAR) experience on a university sized area, with high fidelity (HF) AR content. A) Current large scale (LS) CAR experiences only exhibit low fidelity content such as: inaccurate registrations (zoomed section, red circles), missing occlusions (red squares) and no interactions between the real and virtual environment. B) I achieve accurate content registration over large distances (zoomed section), correct occlusions (red squares) and interactions between the real and virtual environments. I hide spatial and temporal inconsistencies by representing users as remote avatars (drones). I also incorporate the following user redirection elements: attractors (highlighted satellites) to guide users towards key locations and repellers (roadwork signs) to keep users away from areas that are dangerous/prone to system failure.

throwing a virtual rock.

High Fidelity: When there is spatio-temporal consistency all users observe the virtual rock being thrown at the same time, with the origin of the thrown rock at the users hand.

Low Fidelity: Spatial inconsistencies cause the virtual rock to appear to be thrown from a visually incorrect origin, or hit an incorrect target. Temporal inconsistencies cause delays between the user's movement and the virtual rock being thrown.

- **Visual consistency:** Does a virtual object have the same visual properties as the surrounding environment and their real world counterparts.

This can be considered from three aspects, geometrical quality, optical consistency and lighting consistency. An example is a virtual copy of a real statue being shown side by side.

High Fidelity: The virtual object is indistinguishable from its real world counterpart as the geometry of the virtual object is dense and able to represent the smooth surface of the real world statue accurately, and the lighting condition of the virtual environment is accurate to the real world. Additionally it exhibits the same optical effects of the viewer (for example correct distortion/ Depth of Field (DoF))

Low Fidelity: The virtual object can be easily identified by either a mismatch in lighting, DoF/optical distortion or by deformations in the geometrical model used to render the virtual object.

Although visual consistency plays an important role in achieving perfect fidelity, it presents significant challenges such as accurate estimation of the illumination and the scene reflectance [110, 111, 112], transparency [113], realistic rendering of shadows [114], and replication of other visual effects exhibited when observing real objects, such as the depth of field [115]. While an experience that does not replicate photorealistic elements can still be high fidelity, if users are aware that non-photorealistic rendering is justified by story and artistic elements. Conversely, a photorealistic experience will not necessarily be high fidelity, for example if it has significant temporal inconsistencies.

The absence of such visual consistency qualities also has an impact on the depth perception of users, which is critical to interactions in LS AR scenarios. We rely on several depth cues in large scale environments to determine the visual placement of content in our environment, such cues include shadows, lighting, occlusions, and depth of field. In my previous work EyeAR [115], I created a system that accurately modelled the visual properties of the eye in order to re-create the Depth of Field for AR through CG rendering. This compensates for the optical subsection of visual consistency.

In this thesis I aim to address both the interaction related components of

fidelity (Chapter 3 through to 6) and the refocusable component of visual consistency (Chapters 7 and 8).

Experiences that target large areas [1, 20, 78], commonly have only rudimentary interactions with the physical world, suffer from content registration errors, or exhibit spatio-temporal inconsistencies and therefore do not cover many of the fidelity issues described by [53]. I classify these experiences as large scale and low fidelity (LSLF). On the other hand, various room sized experiences [2, 6, 7] satisfy all of the fidelity metrics. I classify these experiences as RS and high fidelity (RSHF). Although it's technically possible to track multiple users with high accuracy in a large scale environment using Simultaneous Localization and Mapping (SLAM) [24], there are several technical challenges, such as the accuracy of synchronized poses between users and network latency.

My goal is to create the first LSHF CAR experience by addressing these challenges. In particular, I aim to create a multiplayer AR game deployed in a suburban area larger than 10,000m^2, featuring high fidelity content (Figure 1.1). To achieve this, I have to not only address the technical challenges, but also consider additional design issues unique to LSHF CAR. I helped organize a week long workshop between NAIST and KAIST. Seven researchers from HCI, Augmented Reality, Game and Industrial Design backgrounds gathered together to discuss these design issues. We reviewed prior implementations [20, 117, 120, 119, 118] and identified the following core design issues:

- Users will be moving over a large area, and can potentially move into hazardous areas (such as a busy road or a staircase) or areas that the utilized system may not function within (dark areas). [28]
- Interactions will occur over large areas, within a single contiguous instance. Additionally, the actions of one user will affect the global state (An example is a sniper taking a virtual shot over a long distance, or a user activates a virtual button at one location, triggering a door in a separate location to open).
- Users will be distributed over the large area and will need an under-

standing of their environment, the situation within the experience, and the intentions of non co-located users (An example is a team of users working together at separate locations to achieve a goal).

- The input device used to interact with the virtual content can exhibit errors in tracking, making the interactions difficult, especially over large distances [78]

Then, through affinity diagramming [116], we grouped these challenges into the following four clusters:

- **User redirection**: How to move users around the play area, directing them towards key locations (attractors) and away from dangerous/unplayable areas (repellers).
- **Inconsistencies:** How to handle spatio-temporal inconsistencies during runtime, providing a consistent experience for all users.
- **Spatial awareness:** How to provide the users with information about the surrounding real/virtual environment, and the location of other users.
- **Communication:** How to provide a means of communicating between non co-located users.

Although AR research has extensively explored communication and navigation in LS CAR environments [28, 79, 96], user redirection and handling spatio-temporal inconsistencies have yet to be addressed. I can adapt game design elements to specifically address these design issues as they share the same design issues [63]. Ng et al. [68] utilized video games elements to navigate users within a room scale environment. Although, they did not consider the use of game elements outside the game context, they highlight the necessity of user redirection elements.

In this paper, I derive the requirements needed to achieve my target LSHF CAR experience. Based on these requirements I explore the LSHF CAR design space, drawing on technical implementations and design aspects from both AR and video games. I then present a software architecture and technical implementation that improves the accuracy of synchronized poses between multiple track-

ing systems. I apply my target experience and technical implementation to my established design space, creating Holoroyale, the first instance of a LSHF CAR experience. One of the most pressing concerns I identified during the workshop is keeping users away from potentially hazardous areas or areas that the system cannot be used in. Because of this, a core design component of Holoroyale is the use of visual repellers to guide players away from dangerous/unplayable areas. While demonstrating HoloRoyale in smaller scale demonstration venues, I found that users became frustrated with the placement of the repellers, but respected their boundaries. This raised the question if this was due to the scale of the venues and what other effects repellers could have on users immersed into a LSHF CAR. This prompted us to evaluate the effectiveness of the employed visual repellers in a LSHF CAR context. To do this, I conducted a user study, deploying Holoroyale in a 12.500m^2 area. The results confirm that visual repellers are effective user redirection elements that do not significantly impact the user's overall immersion. My results also show that peer and time pressure can lead to users ignoring repellers, which requires their effects to be reinforced by means of additional design elements. Furthermore, I found that repellers complicate communication between users as they make it more difficult to maintain a mental image of the environment layout.

From the discussion above, I summarize the challenges of creating a LSHF CAR experience as a series of requirements categorized by the scale of deployment, the fidelity, and the between user collaboration as follows:

1.1 Requirements

My aim is to create a high fidelity AR experience that is deployed on a university/suburban sized scale with multiple simultaneous users. As my target experience covers the fidelity challenges I derived from [53] and the unique LSHF CAR challenges ascertained during the workshop described in Section 1, I can consider it an experience that encompasses all challenges expected in a

LSHF CAR experience. Through affinity diagramming [116] I categorize these challenges as requirements based on the component of the experience that they affect. The result is the following list of general requirements for a LSHF CAR experience (Figure 3.1).

Scale

- The system must be deployable in areas up to and beyond a maximum size of 10,000m^2, to cover the target university sized area.
- Due to users moving around a larger area, the system must be able provide users with information about their surrounding environment.
- To assist users' movements over the larger area, the system must provide navigation cues to assist players when moving between key locations.
- Since it's expected that user's encounter dangerous situations, move into areas where the system may no longer work, or be unaware of the next destination, the system must provide the following user redirection elements:
  - Repellers to deter users from entering dangerous/unplayable areas.
  - Attractors to highlight key locations, motivating users to move towards them.

Fidelity

- To provide visually realistic CG the system must:
  - Render high density geometry
  - Accurately model the lighting and visual conditions of the real environment
- The system must produce a 3D model of the environment for virtual-real interactions and visual occlusions.
- The display and input must have a total motion-to-photon latency no larger than 20ms to prevent motion sickness while moving around the large area [19].
- The system must be able to render convincing geometrical models of virtual objects, whenever applicable.

- To ensure that the virtual content appears consistent within the environment its displacement in the user's view must be less than 1 arcmin [49]

Collaboration

- To enable a collaborative environment, the system must share the pose and logical state of several clients.
- To provide a consistent experience between clients, the system must handle:
  - Inconsistent between-client temporal states.
  - Erroneous between-client pose synchronization.
- To support collaboration between users,the system must provide a means of communication between users.

1.2 Overview and Contributions

This section provides an overview of my research hypothesis, the research approach and highlights my research contributions. Then I provide a quick overview of the format of my thesis.

1.2.1 Research Questions

During the time spent on the work described in this thesis I had the following research questions.

Q1 With the current state of technology, is it feasible to create a system capable of a LSHF CAR experience?

Q2 Do video game design concepts assist with addressing fidelity requirements in LS CAR contexts?

Q3 Can game design concepts also address the user redirection requirements necessary for LS CAR contexts?

Q4 Does refocusable content improve the fidelity and realism of AR content on OST-HMDs?

1.2.2 Approach

The research methodology in this thesis is a combination of literature analysis, experience design, software development and experimentation. First by describing the target experience I derive several expected use case scenarios. From these scenarios I define a series of requirements that need to be addressed.

As part of the interactive components of fidelity, I develop a design space, expanding on the literature review and drawing on technical implementations and design aspects from both AR and video games. From the design space I selected hardware most appropriate for creating a system capable of LSHF CAR, then present a software architecture and technical implementation that improves the accuracy of synchronized poses between multiple tracking systems.

I then precisely define the description of my target LSHF CAR experience, and apply both the target experience and teachnical implementation to my established design space. The result of this application is HoloRoyale, the first instance of a LSHF CAR experience. This experience was demonstrated at several conferences, allowing us to observe how users interacted with the design concepts from my design space. One of the key observations was the interactions between users and the virtual repellers placed into the environment. I investigated the interaction observed during the demonstrations further in a user study with a controlled environment. To evaluate the effectiveness of the employed visual repellers in a LSHF CAR context I conducted a user study, deploying a modified variation of HoloRoyale in a 12.500m^2 area. The participants played two sessions of HoloRoyale in groups of three members, between each session I altered the appearance of the virtual repellers. The results from the user study suggest that visual repellers are effective user redirection elements that do not significantly impact the user's overall immersion.

Finally, addressing a small subsection of the visual consistency components of fidelity, this thesis expands on my previous masters work that aimed to address the fidelity limitations of single plane OST-HMDs by creating refocusable content based on measuring the focal properties of the user's eye. I follow up from

this work by evaluating the effects of the refocusable content that it creates.

1.3 Research Contributions

The work described in this thesis makes the following contributions:

1. My work is the first to explore the challenges of LSHF CAR.
   (a) I establish a design space that offers a new approach and perspective to handle the requirements of LSHF CAR experiences by adapting concepts from video game design.
   (b) I improve the accuracy of synchronized poses between multiple SLAM systems compared to the out-of-the-box hololens by aligning several smaller SLAM maps, creating a global coordinate system. I track each user relative to the smaller SLAM maps, avoiding pose drift over large areas. My framework enables the creation of future LSHF CAR experiences on a global scale.
   (c) I create the first instance of a LSHF CAR experience by applying my technical implementation and my target LSHF CAR experience to my established design space.
2. The results from my evaluations show:
   (a) Virtual repellers can be effective user redirection elements in LSHF CAR contexts. This leads to new research questions on the benefit of user redirection elements and how to reinforce the effect they provide.
   (b) Rendering CG based on measurements of the user's eyes improves the perceived realism of the observed graphics. In particular, CG rendered with EyeAR were always perceived to be more realistic than CG rendered with the pinhole eye model.

The work in this thesis is a first step into the previously unexplored domain of LSHF CAR, opening up several new avenues for future work. Besides investigating the effects of adapted game design elements on users in LSHF CAR scenarios, there are several research questions that can now be investigated.

What other AR spaces can benefit from the adaptation of game design into AR? How does hiding spatio-temporal inconsistencies impact the performance of users in LSHF CAR scenarios? What are the psychological impacts of diegetic repellers when represented as dangerous obstacles?

1.4 Thesis Outline

The work in this thesis is outlined as follows. In Chapter 2 I discuss related work. In Chapter 3, I analyze related work based on the technical implementation and design aspects and explore feasibility of various hardware implementations and interface designs to establish a design space for LSHF CAR. In Chapter 4 I describe the selection of hardware used for implementing a system capable of LSHF CAR experiences, and discuss it's limitations. In Chapter 5, I describe my target experience in detail, and apply both the experience and the implementation described in Chapter 4 to the design space established in Chapter 3. The result is the first instance of a LSHF CAR experience. In Chapter 6 I evaluate the effectiveness of a subset of the design elements established in Chapter 3. In Chapter 7, I redirect my focus to the visual consistency components of fidelity, extending on the work of my Masters thesis, breifly re-describing a system that creates refocusable AR content on a single plane OST-HMD and evaluating the realism via a turing test. I then conclude and re-list the contributions of this thesis in Chapter 8.

Chapter 2. Related Work

The work in this thesis is related to Collaborative AR experiences, OST-HMD design, DoF effects, and Visual Turing Tests that are commonly applied to CG. In this Chapter, I categorize other known CAR experiences by the deployed scale and the fidelity of the experience, discussing the limitations of each experience. Then I only give a short overview of related work regarding collaborative experiences in this section as Chapter 3 describes and explores the related work in extra detail. Then, as I intend to evaluate the visual qualities of the renderings produced by the work conducted during my masters thesis, I discuss previous studies that evaluate the impact of DoF effects on user perception and experimental protocols for visual Turing Tests that validate the realism of CG.

2.1 Collaborative Augmented Reality Experiences

Since AR technology has improved, several collaborative AR applications have been introduced both in research and in the marketplace. As described in my introduction these experiences can be classified by the scale which users are distributed and the fidelity the experience achieves. One of the recent and most prominent applications is Pokemon Go [1] and Human PacMan [20]. These experiences utilize GPS and gyroscope for tracking the user, and composites CG onto a video feed [1] or an OST-HMD [20]. Because they don't use an environment model, there are seldom interactions between the real and virtual environments. Additionally the sensors used for tracking the user are largely inaccurate, leading to bad content registration and content jitter. Furthermore, although the experiences are multi-user, users do not interact within the same environment but separate instances that affect a global state.

Alternatively there are experiences which are limited by the scale they are deployed. One example is SHEEP [86] that allows multiple users to interact within the same virtual space. It uses outside-in sensors to track user's and

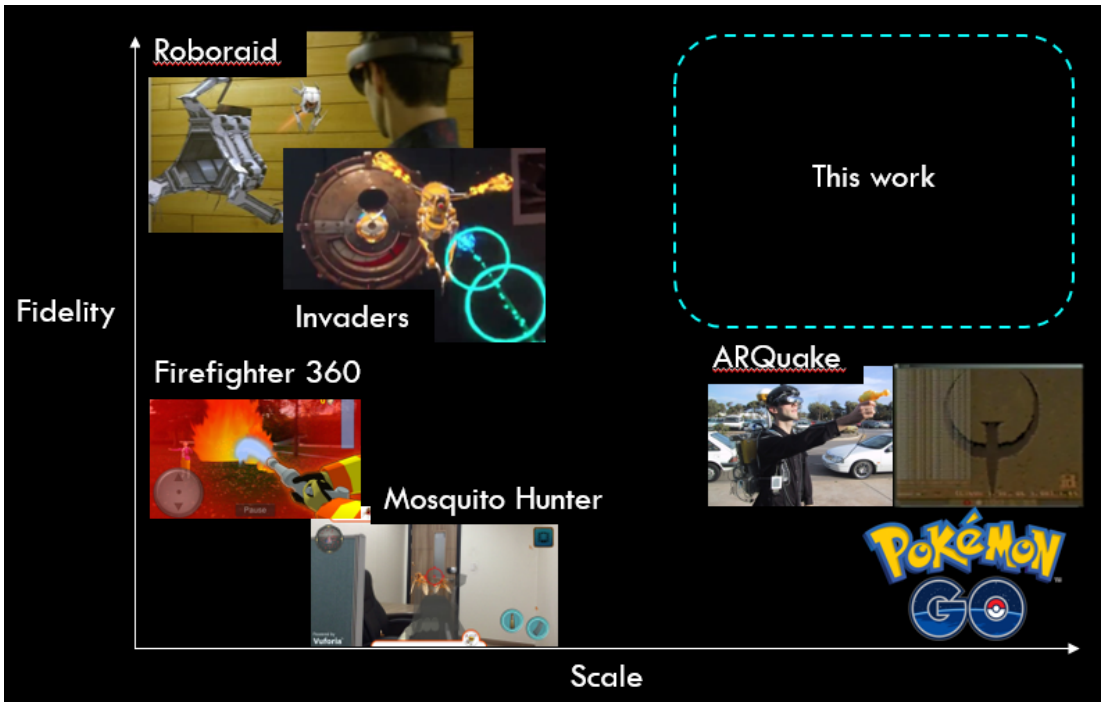

Figure 2.1: A summary of related work, classified by the scale in which the work is deployed, and the fidelity of the AR content experienced. The work in this thesis aims to expand into the previously un-explored domain of Large Scale, High Fidelity Collaborative Augmented Reality

objects in the real world allowing for high fidelity interactions. The tracking hardware is limited in scale however. There are other non multi user experiences that also provide high fidelity content such as Roboraid on the HoloLens [2] and Invaders on the Magic Leap [7]. Both utilize a form of SLAM [24] to obtain the pose of a user and obtain a 3D model of the environment for occlusions and virtual-real interactions. Although it's feasible to expand the room scale capabilities of the HoloLens or Magic Leap, it has yet to be done.

Finally, the largest dominance of applications on the consumer-market (in particular applications that are on mobile) are limited to the smaller scale areas for interaction, feature rudimentary interactions with the physical environment and suffer extreme content registration issues [122, 123]. A classification of the

related work for this segment and where my work aims to make a contribution can be seen in Figure 2.1. The work in this thesis aims to create first Large Scale multi user experience that features High Fidelity content.

2.2 Displays for Refocusable AR Content

A lot of research has focused on the development of displays capable of reproducing focus cues. Kramida [51] and Hua [42] provide a detailed overview of different approaches and technologies developed to enable refocusing onto virtual content presented at different depths and natural DoF effects.

A common approach to present depth cues is by displaying virtual content on multiple transparent planes [11, 44, 67]. Although MacKenzie et al. [59] have shown that five focal planes are enough to produce an acceptable range of real-time accommodative cues, multiple focal planes lead to a bulkier HMD design. My method differs from these in that I target off-the-shelf OST-HMDs that have a single focal plane that can't present refocusable CG.

Another approach to create refocusable OST-HMDs is to use a refocusable lens, either in combination with a static or movable display [23, 57]. This enables adjustment of the distance at which the virtual content appears. However, the lens refocuses the entire display at the same time, thus providing a realistic DoF that requires a very high update rate, which has not been achieved yet, and must be synchronized with the display to prevent unintended effects. In [23] the authors also note that it is important to evaluate the effectiveness of rendered DoF compared to DoF generated by optical elements. I applied a similar approach to my method [84]. By always matching the position of the display with the distance on which users were focused on I created realistic DoF effects at different focus distances.

McQuaide et al. [64] use a method that is very similar to ours. They combine a laser projector that shows the user an image that is always in focus with deformable micro-electromechanical system (MEMS) mirrors. By manipulating

the convergence of the MEMS elements they alter the focus of the laser beam. This requires users to refocus to see a sharp image. The idea of my work is very similar in that I assume that the user always observes a sharp image. However, my method differs from theirs in that I present the DoF not through optical means, but with CG.

Finally, the method I use to re-create the DoF properties of the users eye is also very similar to that of Kán and Kaufmann [47]. They obtain the camera's lens parameters and render CG using these camera parameters. The resulting CG's DoF matches the DoF properties of the video in each frame. My work crucially differs from [47], I created an OST system and also obtain the parameters of the user's eyes instead of a camera.

2.3 Impact of DoF Rendering

The impact of DoF rendering and refocusable displays has been studied in AR, VR, and gaming applications. Padmanaban et al. [74] used a display with refocusable lenses to investigate the impact of refocusable graphics on users to go even further and personalize the accommodation to each individual to correct vision impairments.

Eye-gaze tracking (EGT) has been used in combination with a variety of VR systems to estimate the focus position and generate gaze-contingent DoF effects. Hillaire et al. [37] found that users preferred graphics rendered gaze-contingent DoF effects over the CG in focus. On the contrary, just applying DoF effects [36] did not improve user performance. In particular, some users expressed fatigue and discomfort caused by the DoF effects. Mauderer et al. [61] found that gaze-contingent DoF effects increase realism and help estimate the depth of virtual objects. Vinnikov and Allison [100] found that although DoF effects improve depth perception in monocular systems, users felt discomfort when these were applied to stereoscopic systems.

Hua and Liu [43] found that DoF helps estimate the depth of virtual content

in AR, which is similar to a VR scenario. In a similar study, Xueting and Ogawa [106] found that users preferred graphics with too much blur applied to them over the correctly applied amount of blur. My study differs from these works in that I do not investigate the user's depth perception, but whether they can distinguish between the CG and real objects through an AR Turing test. Furthermore, in [106] the evaluation was performed on images rendered on a display, and not on an OST-HMD. Therefore, users did not experience inconsistencies between the focus distance and the presented CG.

2.4 Visual Turing Tests

In 1950, Alan Turing [98] introduced the Turing Test. In this test he proposed an "Imitation Game" which can be used to test the sophistication of AI software. Human participants are asked to interact with a conversation partner in order to determine whether the partner was human or an AI simulation. The test was passed if the chance of choosing the AI correctly became equivalent to a random guess.

Similar tests have been proposed in a variety of fields. For example, for Visual Computing [88], where participants were tasked to distinguish photographs from renderings of 3D reconstructions of buildings. In Computer Graphics, McGuigan [62] performed a Visual Turing Test with the following hypothesis: "The subject views and interacts with a real or computer generated scene. The test is passed if the subject cannot determine reality from the simulated reality better than a random guess".

Several previous studies proposed restricted versions of this test; Meyer et al. [65] conducted a study which required participants to view a physical setup as well as a setup displayed on a color television. Although this study showed that subjects were unable to distinguish between the physical and displayed setup, this was largely because the physical setup was viewed through a camera.

Later, Borg et al. [17] created a practical CG Turing Test which made use

of a box enclosure. This box enclosure featured a controlled light source, a removable end which can be replaced with a screen, viewport consisting of a small pin hole and simple, real object (which is duplicated in a virtual scene). The participants were asked to look through a small hole using one eye for 10 seconds and then tasked to identify if the scene observed was rendered or real. Results showed that participants were generally unable to accurately distinguish the virtual scene from the real scene better than a random guess. A variant of this test was recently shown at SIGGRAPH by Nvidia [72]. They presented two box enclosures with controlled lighting; one scene was a physical scene featuring an electronic drill and the other one featured the same scene as CG. Participants did not view the boxes directly, but were only looking at prerecorded photos.

In spirit, the AR Turing Test later described in this thesis follows the protocol of McGuigan's Visual Turing Test. An important difference is that I perform my test in an OST AR situation. To the best of my knowledge, this is the first time that an AR Turing test was performed in such a situation. I have also reported the design and results of this user study in [85].

Chapter 3. Establishing a Design Space

In this Chapter, I discuss how the requirements derived in Section 1.1 can be addressed from a technological and a user interface standpoints. I first explore two key areas for the technical platform: How to display content to the user, and how to track the user in the real world. By categorizing related work by the technical implementation used, the fidelity it achieves, and the scale of deployment, I identify the most suitable display and tracking technology for my target experience. Then I explore several aspects of the user interface and user experience. Hereby, I gather insights not only from previous AR implementations, but also from different genres in video games. Video games are widely imagined in AR [99] and have to handle many of the design challenges present in LSHF CAR. Figure 3.1 shows the design requirements for my target LSHF CAR experience and the domains that have previously explored them. A summary of my established design space resulting from the discussion in this section can be seen in Figure 3.2.

3.1 Technical Platform

Various aspects of my requirements, such as display latency and tracking accuracy, can be addressed by selecting the appropriate platform. In particular, I consider the following areas: How to display content to the user, and how to track the user within the environment.

3.1.1 Displays

One crucial component of any AR system is displaying CG content to the user's view of the real world. In general, there are three ways to show AR content to users:

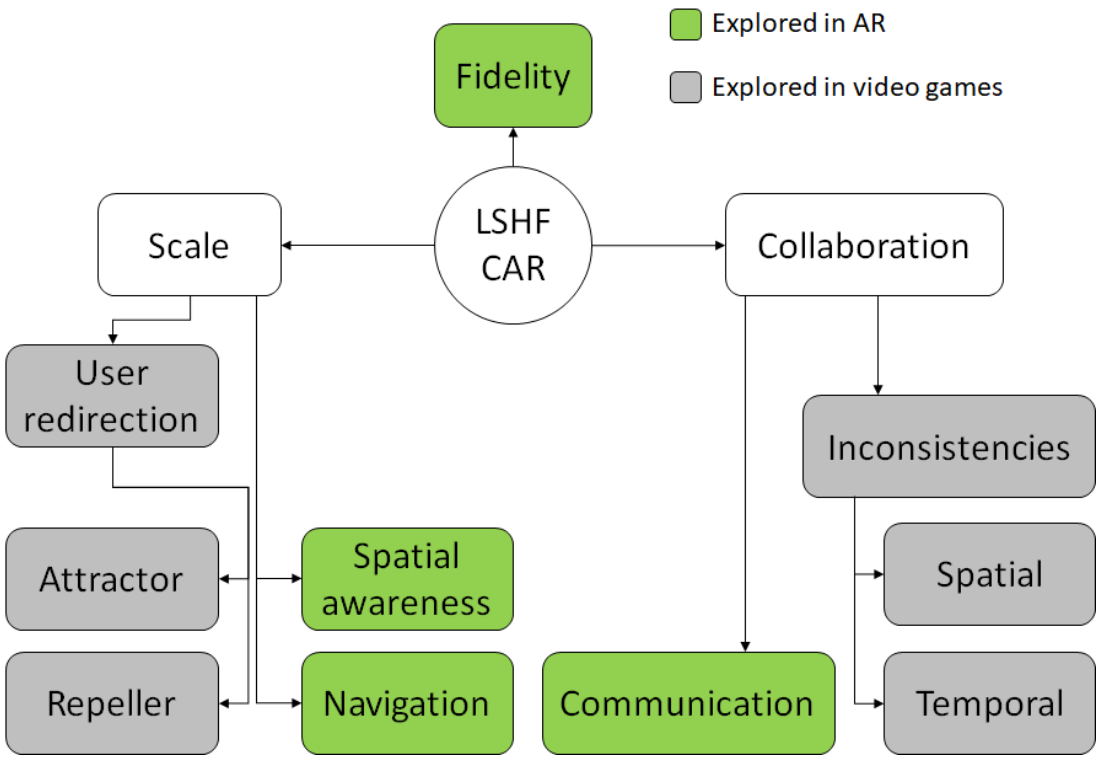

Figure 3.1: Taxonomy summarizing the dependency of system requirements for LSHF CAR applications and the domains that have explored methods to address these requirements.

Video See-Through (VST): This method composites CG content onto a video stream. This technology is commonly used in HMDs [20, 86, 45, 103] and hand-held devices [1].

Optical See-Through (OST): This method directly embeds CG into the user's view of the environment by reflecting a rendering from a screen off a transparent half-mirror, into the user's eye. Several experiences [2, 7, 28, 96, 78, 92, 3] utilize OST-HMDs.

Projection based: This method projects CG directly onto the environment. Although often used in AR [31], projector are typically statically placed, and are limited to presenting content onto the physical world (which limits the depth

perception of CG content). Therefore, projection based AR is not viable for my scenario.

My goal is to deploy my LSHF CAR experience in a suburban area. Therefore, I need to consider that users will be moving between indoor and outdoor areas. In this scenario, a display exhibiting a motion-to-photon latency larger than 20ms [19] can lead to dangerous situations such as users walking into an object or falling over due to motion sickness. There have been extensive comparisons of OST and VST-HMDs [80] that suggest VST displays are more restricted on the motion-to-photon latency. This is because users view the world through the video camera feed, with the CG composited, as opposed to OST-HMDs that directly render the CG content onto the user's view of the environment. Additionally, when a VST-HMD fails, users can no longer see their surroundings. This explains why most LS AR experiences utilize OST-HMDs [28, 96, 78]. OST-HMDs however rely on a half mirror to present content to the user, under bright lighting conditions the external light transmission causes the exhibited CG to appear more transparent, affecting the fidelity of the content shown. While it is possible to address this through brighter displays and occlusion-capable systems [121], currently no commercially available system provides this functionality.

Overall OST-HMDs are the best candidate for my LSHF CAR experience, as they satisfy the motion-to-photon latency requirement and are fail-safe. Additionally, hand-held VST can be used for non-immersive AR experiences [12].

3.1.2 Tracking

To place AR content and synchronize the poses of several users, I must obtain each user's pose in the environment. For this, there are three main approaches: *Sensor based tracking:* Uses the GPS, accelerometer, gyroscopic sensor, and compass on the device to obtain the position and orientation of the user, within the real world [1, 20, 28, 96, 78]. Although easily accessible, sensors are prone to

drift and inaccurate readings [105], which can cause severe content registration issues [78]. These sensors do not rely on any visual input for tracking and are therefore robust to differences in lighting conditions.

Outside-in tracking: Obtains the user's pose by utilizing external sensors placed within the environment. A common approach is to track fiducial markers placed on the user [86]. Although these systems can achieve high accuracy, the setup becomes excessively expensive when deploying over larger areas and requires careful calibration and preparation. Additionally, sunlight can negatively affect the tracking accuracy. However, since outside in tracking does not require natural features of the environment (and instead typically relies on retro-reflective markers that reflect IR light emitted from the mounted sensors) it performs very well under low light conditions.

Inside-out tracking: This method functions similar to outside-in tracking. However, the sensors (most commonly cameras) are placed on the user and track features within the environment. These features can be either fiducial markers placed throughout the scene [103] or natural features [24].

Although it is possible to use markers for LS environments [78], this requires careful between-marker calibration [87]. Furthermore, the user's pose can only be estimated when a marker is detected by the sensors.

The alternative utilizes natural features detected within the camera image to localize and track a user in the environment (Simultaneous Localization and Mapping, SLAM [24]). Recent improvements enable the use of SLAM on mobile systems [48] and track users even over large scales [25]. Nevertheless, pose drift occurs when tracking the user over large areas, even when using loop closure to minimize this error [25, 66]. These inside-out tracking methods are more robust in daylight scenarios but fail under low light conditions due to the lack of trackable features in the environment.

Hybrid: This method combines different tracking methods to leverage their advantages. RSHF experiences such as those shown on the Microsoft HoloLens [2] and the Magic Leap [7] utilize multiple carefully calibrated cameras for

visual SLAM, depth sensors for surface mapping, as well as gyroscopes and accelerometers to improve the tracking stability. Nevertheless, these systems also suffer the same pose drift issues over large scales. As this method still relies on some form of visual SLAM, it suffers from the same issues under low light conditions.

I assume that our target experience will only be played during daylight hours. With this assumption, although hybrid methods still suffer from pose drift issues, their improved accuracy and off-the-shelf availability makes them the prime candidate for my target LSHF CAR experience.

3.2 User Interface and Experience

Some of the requirements listed in Section 1.1 require careful design of the user interface and the CAR experience.

3.2.1 Representing the User in the AR environment

Many CAR experiences assume that users are in perfect sync both spatially and temporally [15]. However, the nature of distributed experiences means that spatial and temporal inconsistencies are present due to tracking errors and network latency. These inconsistencies can severely disrupt the fidelity of an experience. I can hide possible spatial and temporal inconsistencies by modifying how I represent the user in AR. I can represent users in the AR environment through:

Direct Representation: This representation is used by most AR applications. It utilizes the raw pose of tracked users and tools when placing CG into the scene. Although this is the ideal scenario, it is only viable if there are no spatial or temporal inconsistencies. An example of direct representation is the rendering of a gun over the controller in the user's hand [7].

Indirect Representation: This method represents the user/tool as an unattached

AR avatar. It, therefore, overcomes spatial inconsistencies by disassociating the user from the virtual environment. Furthermore, by interpolating and predicting the pose and state of the associated avatar [32], indirect representation hides temporal inconsistencies. For example, virtual wands could represent users in a magic game [69].

Since my target LSHF CAR experience features multiple distributed users, I expect temporal inconsistencies to occur. This favors indirectly representing users in the AR environment.

3.2.2 Interacting with Content in CAR

I need to consider how users interact with content, as this is a core component of any AR experience.

For any interaction to occur, users must first select a target for interaction. Fitts' Law [30] states that the time taken to select a target is determined by the distance from the user to the target and the size of the target. The shorter the distance and the larger the size, the easier it is to point at the target. Spatio-temporal inconsistencies vary the effective width and distance of a target increasing the difficulty of selection. Overall, users can interact with content in the following ways:

- Direct interaction: Direct interaction with virtual content appears to be most natural and is applied in a variety of AR experiences [15, 78, 7, 2]. However, as this interaction utilizes the user's raw input, it is highly susceptible to tracking errors, inaccurate pose synchronization, and network latency. Under such conditions, direct interaction can result in reduced efficiency and increased player frustration [78].
- Assisted interaction: Similar to direct interaction, assisted interaction uses the user's raw input for interaction. However, it improves the robustness to spatio-temporal inconsistencies by modifying the effective width and distance of a target without modifying its visual appearance. This

allows interaction to occur, even if it's not visually consistent but can cause frustration if the assistance constantly selects wrong target [35].

- Indirect interaction: This interaction technique is widely used in video games [9, 4, 8] and is only possible when users and their tools are represented by avatars. Hereby, the avatars always orient themselves towards the interaction target selected by users and perform the desired interaction. This interaction method can be further enhanced by applying assisted interaction techniques to the user's input for target selection. Indirect interaction is robust to pose synchronization errors and network latency, providing consistent interactions [22] while allowing incorrect selections if the user's aim is imprecise.
- Magnetized Interaction: This method is specific to projectiles. Hereby, the projectile acts like a 'heat seeking missile' continuously changing its flight path as it moves towards the intended target, independent of the user's input. Although it ensures consistent interactions, this method removes all challenge from the experience and can lead to dissatisfaction [35].

As I determined that I should indirectly represent users to overcome temporal inconsistencies, indirect interaction is most suited as it preserves spatial and temporal consistency. Although magnetized interaction could address this issue as well, it removes the challenge from the experience.

3.2.3 Communication Between Users

When users are distributed over large areas they need a means of communication with each other. From the grouping of related work, and examining the communication methods utilized in many video games, I identify four generally used types of communication:

Text based communication: Users communicate by sending a string of characters typed out on a virtual on-screen keyboard or a physical input device [101].

This method provides clear communication and requires minimal networking bandwidth. However, creating and reading a message is time consuming and causes an increased cognitive load [46]. Therefore, it should be avoided if possible.

Emoticon based communication: Instead of typing out messages, users can utilize a predetermined set of emoticon text messages or images based on the user's possible intentions. Emoticon-based communication is widely used in video games [50]. Emoticon messages have the benefit of being fast to send, are instantly understandable (requiring minimal cognitive load) and utilize minimal network bandwidth. However, due to the limited range of options, the intention that a user wants to portray can often be ambiguous.

Voice based communication: As an alternative to visual communication, many collaborative experiences use voice chat [15]. This offloads communication from visual to auditory, decreasing cognitive load [46]. It has also been shown to be preferred to text based communication in collaborative environments [46]. The drawback of voice based communication is the high networking bandwidth demand.

Video based communication: Instead of communicating only over voice, several games feature video based communication where users see either a first-person view or a view of the partner's face during communication [124]. Although such communication is often used in collaborative systems [15], it requires a much higher per-user networking bandwidth. Furthermore, it may not significantly improve the collaboration due to the limited size of the shared view and difficulties understanding what their partners mean when many users share their view at the same time. This makes it difficult to use in scenarios with more than 2-3 concurrent users.

To my knowledge, there has been no study in a distributed AR context that directly compares video, voice, text, and emoticon based communication between users. Nevertheless, voice based communication is the prime candidate for my target LSHF CAR experience as it allows clear and fast communication.

However, if the available bandwidth is not sufficient enough to support voice based communication, its suggestable to add emoticon based communication as an alternative.

3.2.4 Providing Spatial Awareness

As users are exploring a large environment with AR content added to it, they require a method to obtain an understanding of their surroundings. This includes information related to the task, the environment, and the location of users. There are three ways to provide this information.

2D representations in the Heads Up Display: This method places spatial awareness cues into the 2D plane that lies in screen space (also known as the Heads Up Display or HUD). This representation can contain varying degrees of detail ranging from simple radars [18] to detailed maps of the environment [28]. The HUD can also contain cues for out-of-view points of interest [34]. However, adding too many elements to the HUD can also lead to visual clutter of the display [40].

3D representations in the user's environment: This method uses a 3D model representation of the environment, displayed within the user's viewport (instead of in screen space). Such a world in miniature (WIM) [91] shows the users' location within their environment [83] and any additional contextual information [14]. This technique is also commonplace in video games as both a symbolic and diegetic element embedded into the game environment [5]. The downside to this representation is that in order to provide detail it has to occupy a large portion of the screen space, potentially occluding the user's view of the environment.

Hybrid representations: Finally, there are hybrid implementations that combine both 2D and 3D representations of the environment. For example, when the user is looking at an AR scene it can be annotated by 2D labels shown on the HUD. Then, when the user views the WIM, the labels move to their corresponding

positions on the WIM [13]. Although hybrid representations retain the benefits of both 2D and 3D representations they require careful consideration on when the content should switch between its 2D to 3D representations.

Although all of the listed methods are viable, I utilize a hybrid representation as it is the most powerful of the three.

Navigation and User Redirection in AR:

When exploring large scale areas, navigation and user redirection cues become necessary. These elements help direct users towards intended areas and lead them away from areas that are hazardous or prone to system failure. These elements also function as navigation aids. There are two key types of user redirection elements:

Attractors: These elements highlight areas of interest, prompting users to move towards them.

Repellers: These elements highlight areas where users are not allowed to enter by either indicating danger, or inaccessibility.

Visual user redirection elements can appear as symbolic elements in the HUD [96, 79]. For example, an icon flashing on the screen is an attractor while a text prompt warning users if they enter an unwanted area is a repeller.

Instead of relying purely on symbolic in HUD elements, video games also employ diegetic user redirection elements to maintain the experience's immersion [27]. For example, a signal flare in the distance or a robot guide are both diegetic attractors. On the other hand, burning walls of fire or a closed door act as diegetic repellers by indicating that the blocked section is either dangerous or inaccessible. Ng et al. [68] utilized diegetic video games elements to navigate users within a RS game environment. However, they did not consider their use as user redirection elements outside the game context.

There are also several non visual cues usable for user direction. Audible voice feedback directly conveys necessary information to users, but can distract users from their current task [39]. Audible alarms are another alternative, but

are vague if there's no context for the alarm [104]. Finally I can consider vibro-tactile feedback that has been shown to be effective at navigating users with vision deficiency [60]. However, these cues are also vague without a given context.

Since our LSHF CAR experience targets a suburban environment, clear representation of navigation and user redirection elements is key. Although visual redirection elements have been shown to be most effective in similar environments within video games [125] it is unclear how effective they will be in LSHF CAR as the virtual object rendered on an AR display will not physically prevent users from entering the repellers bounds, their visibility may be obstructed by other elements in the environment, or users may plainly be distracted by other pedestrians and the immersive gameplay. At the same time, symbolic cues could be more obvious. This suggests that a combination of different cues should be used to overcome the limitations of each system.

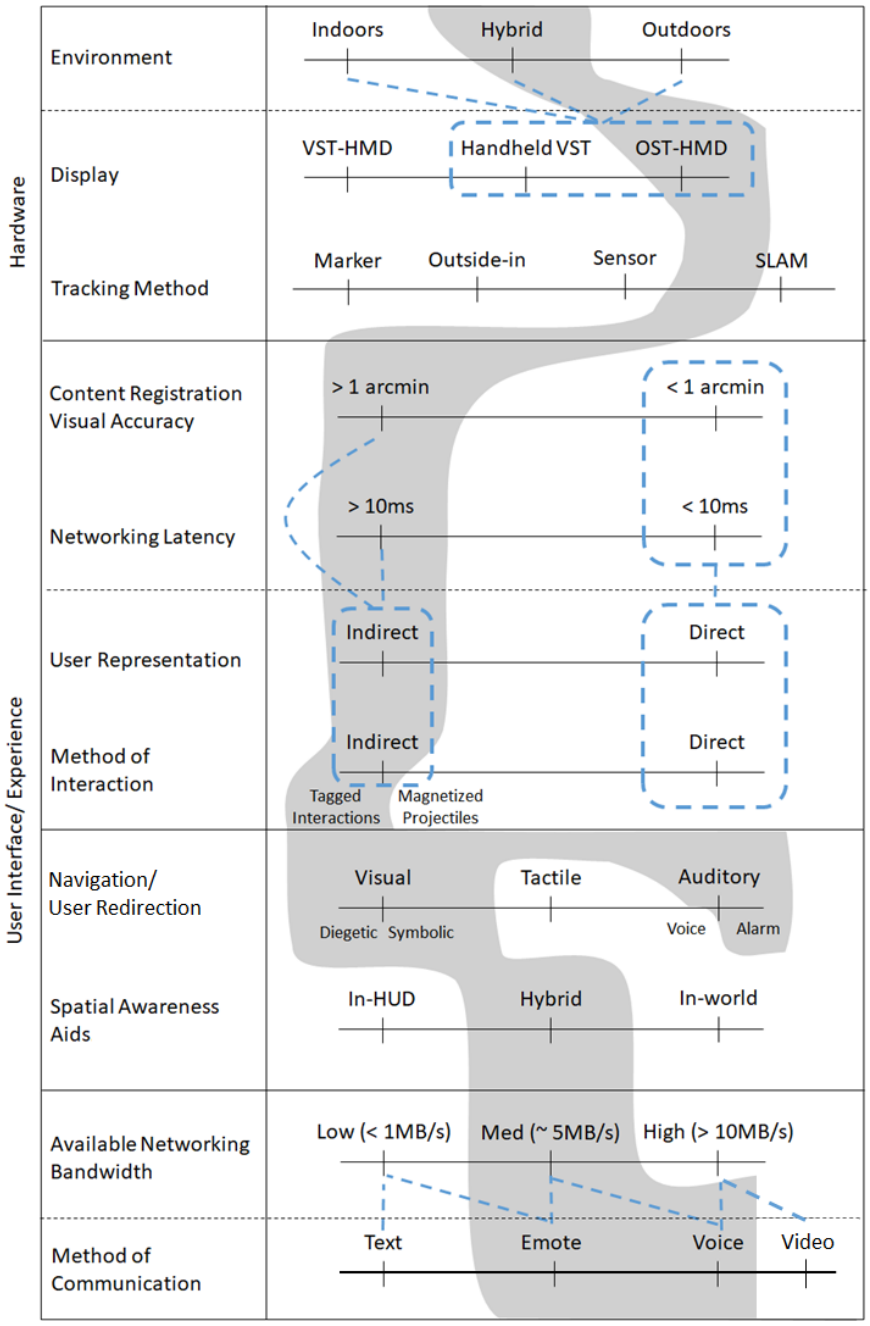

Figure 3.2: A morphological chart showing my established design space. The blue dotted lines indicate my general guidelines based on the discussion in Chapter 3. The grayed out area represents how my technical implementation and my target experience fit within my established design space.

Chapter 4. Creating a System Capable of LSHF CAR

Although combination of SLAM and sensor-based tracking offers the best approach for tracking users in large scale environments, system drift can lead to severe errors when sharing user poses (Figure 4.4c). In this section, I describe a client server architecture that improves the accuracy of synchronized poses between multiple users over large distances (Figure 4.4d).

4.1 Hardware Selection

From the technical analysis in Section 3, I find that currently the Microsoft HoloLens and Magic Leap One are the ideal hardware to deploy my experience on. Both devices feature compelling RSHF experiences [2, 7]. At the time of development, the Magic Leap was not commercially available, in consequence, I built my LSHF CAR system around the Microsoft HoloLens. The HoloLens is an OST-HMD with a motion-to-photon latency of less than 20ms [56] and has a microphone built into it, allowing voice communication. The Microsoft HoloLens contains a sensor assisted SLAM system for tracking the user and provides a 3D reconstruction of the surrounding environment that can be used for near-distance occlusions and virtual-real environment interactions. However, the tracking system inside the HoloLens can experience pose drift over large distances, limiting its deployable scale in CAR. Additionally the HoloLens has a limited FOV for augmented content, limiting the fidelity of the experience by deteriorating the visual consistency [53, 80]. Nevertheless, I opted for the HoloLens as my target platform as it addresses many of the requirements of my system and presents a fail-safe platform. The next section details how I extend the usable range of the HoloLens to satisfy the scale requirements for LSHF CAR.

4.2 Software Architecture

To satisfy my scale requirements for LSHF CAR, I must synchronize the pose of multiple users within a LS area. For this I have two options, using the HoloLens poses directly (which are susceptible to drift) or synchronizing via a cloud computed global coordinate system. There are already cloud solutions for localizing and sharing the pose of several clients in a single collaborative environment such as 6D.ai [126] and immersal [127], however, these only appear to work in single RS standalone instances, and do not collect poses over a contiguous global coordinate system. Furthermore, they are incompatible with the HoloLens. Instead of these cloud based solutions, I extend [70], taking several smaller mapped areas, but additionally computing transformations between the maps. This allows the poses of all clients and virtual actors to be synchronized into a single global coordinate system. I propose a client-server based architecture that performs the following steps to create a global coordinate system (Note that for the purposes of adaptability, I describe the design and implementation in abstract terms applicable to any Visual SLAM system, and mention the relevant HoloLens specific implementation terms in brackets):

Preparation: During preparation, I scan several areas up to 100m^2 using the Microsoft HoloLens. An origin of each mapped area is tagged (a HoloLens anchor is placed in the scene), and the 3D model, along with the tagged origin and binary data (HoloLens anchor data) that represents the VSLAM map, is uploaded to the alignment server. The alignment server then creates a global map, computing the transforms between each map origin (HoloLens anchor) by performing a series of bounding-box Iterative Closest Point (ICP) [89] alignments using the 3D models. This is done by attempting an alignment for each pairwise model along each side of a 6 sided cube, then accepting the alignment that contains the minimal amount of error, and saving the transformation for that alignment. The completed scene graph is stored in a database for later use. I later author AR content directly onto the aligned global 3D model. The

pose of the content is computed relative to the closest map origin.

Distribute poses: On system start-up I assume the HoloLenses start in an assumed starting location and query a web api with this location. The web api streams several candidate maps (HoloLens anchors) to the HoloLens. The maps (HoloLens anchors) are sequentially loaded into the HoloLens' internal tracker until it localizes a loaded map (Places the anchor into the scene). Once localized, I track the HoloLens relative to the localized map's origin (HoloLens anchor), sending the relative transform to a game server that then computes it's pose in respect to the global scene graph.

I then distribute the resulting updated global scene graph to all clients. To minimize networking bandwidth, the distribution is done in two parts. The static between-map (HoloLens anchor) transforms from the alignment are sent on demand. The computed poses of each HoloLens and computer-controlled virtual actors are synchronized at 15Hz. The poses are interpolated between frames as described in [32] (See Figure 4.3a). The 15Hz synchronization rate for poses can be extended up to 60Hz to provide higher precision, at the cost of an increased networking load.

A system decomposition that outlines the timing for sending subsections of the global scene graph can be seen in Figure 4.1. As the HoloLens moves through the area mapped out during preparation, additional maps (HoloLens anchors) are loaded and localized (placed into the scene). The HoloLens is always tracked relative to the closest localized map origin (HoloLens anchor). If a map cannot be localized, it is flagged on the server. Once a map is flagged by three separate clients, the origin (HoloLens anchor) is removed from the scene graph (with the 3D model retained). Then, a new map origin (HoloLens anchor) and 3D model of the surrounding area of a nearby client is captured and uploaded. *Place AR content:* AR content is placed according to the global scene graph. During runtime, as a rendering optimization I use Load on Demand to switch the model used for visual occlusions and interactions. I use a manually

prepared cube-based phantom model of the environment at distances larger than 5 meters, and the HoloLens spatial mapping model at distances shorter than 5 meters (the effective range of the depth sensor).
Offloading computation to the server: For computations I run a game engine on both the client and the server. During runtime I offload the majority of computation to the server. The client runs a minimized viewer, only interpolating the current local state based on incoming state updates from the server. The server processes the non user entities and transmits the states to the clients. To optimize collision detection, I utilize a combination of short-ranged collision detection on clients (as they contain the most recent model of the environment) and long-ranged collision detection on the server (as it contains a global map and can perform a higher rate of collision detection without impacting performance). The results of all interactions are reconciled on the server (See Figure 4.3c). A complete system decomposition with a focus on the offloaded core components can be seen in Figure 4.2.

4.3 Implementation

The following describes the specific hardware the system was implemented on and the software that the system was developed with.
Client:
The client runs on the Microsoft HoloLens, utilizing an XBox One S Controller for input and Mobile Wi-Fi networking. The software consists of the Unity game engine (2018.3.1f1) that comprises of C++ and C# code.
Alignment & Game Server:
Although it's possible to run the alignment and game servers on separate machines, I deploy both on a single Microsoft Surface Book 2 laptop computer with the following specs:

- Intel Quad-Core i7-8650U, @ 4.2GHz
- RAM: 16GB DDR4

- GPU: Nvidia GeForce GTX1060, 6GB

The alignment server utilizes a RESTful web api developed on Golang (9.2) and uses a PostgreSQL database to store static poses. The game server is built using the Unity game engine (2018.3.1f1) that comprises of C++ and C# code.

4.4 Visual Verification

I performed a visual verification to test the accuracy of AR content placement in screen space and pose synchronization using my system against using an out of the box HoloLens for pose synchronization. I placed two HoloLenses with infrared LEDs attached running my system in a previously mapped and aligned environment. Then augment the view from each HoloLens with a colored virtual crosshair placed according to the pose resulting from the synchronization system used (Red = my system, Blue = native HoloLens). I then oriented both HoloLenses so that they face each other roughly 5, 25, 50, and 75m apart. I compared the accuracy by estimating the distance between the infrared LED and the virtual crosshairs. At 5m both systems are at the maximum accuracy, as they are using the shared local map. As the HoloLenses were moved further apart, I synchronized using the native HoloLens tracker's global map and my system continued to load several local maps at 25m. I measured the error as the pixel displacement between the infrared LED and the virtual crosshairs in screen space (as the screen space visual consistency is all that is required). I then convert this pixel displacement to meters by comparing the known landmarks (two cement pillars) on either side of the dummy (that is placed 2.5m between each pillar). The results show that at 25m, the pose resulting from the HoloLens system begins to drift, causing a visual error of $\sim$1m at 50m and about 1.6m at 75m. Conversely, my system maintains an accuracy less than 0.5 meters at both 50 and 75m (Figure 4.4).

4.5 Limitations

The goal of my evaluation was to compare the quality of the alignment of virtual content in screen space over large distances. The results of my evaluation show that my system visually enhances the accuracy of synchronized poses (and therefore enhances the perceived accuracy of placed shared content) between multiple Microsoft HoloLenses in larger-than-room scale environments. However, it still does not achieve the visual accuracy required for LSHF CAR. This is due to two possibilites: innacuracies in the localization system between HoloLenses, and the accuracy of the ICP alignment that directly affects the accuracy of my system. Nevertheless, my improvements are enough to allow indirect representations of users in AR to hide this imprecision, providing the illusion of high fidelity at large scales. Another limitation is that each HoloLens must be initialized within a known starting location, but this can be easily addressed by using GPS to obtain a rough initial position, then loading candidate maps (HoloLens anchors) near the provided coordinates. Finally, I did not provide a complete analysis of the accuracy against a ground truth, because my focus was the quality of the alignment in screen space as users are unlikely to notice depth errors over large distances.

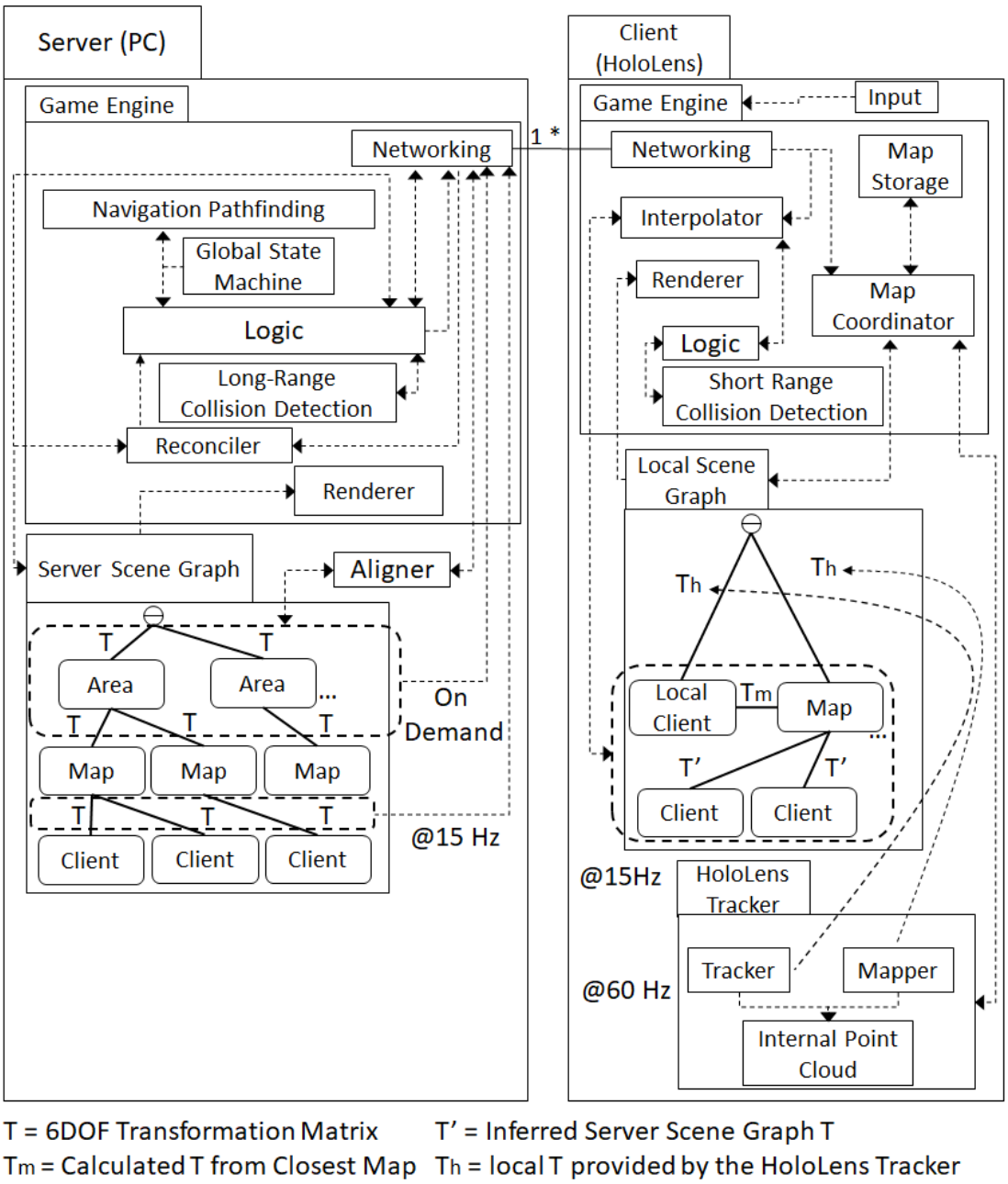

Figure 4.1: System decomposition with the scene graph sections expanded. The 6DOF transform matrices between maps are static, synchronized on demand. The client transforms relative to their closest tracked map are synchronized at 15hz.

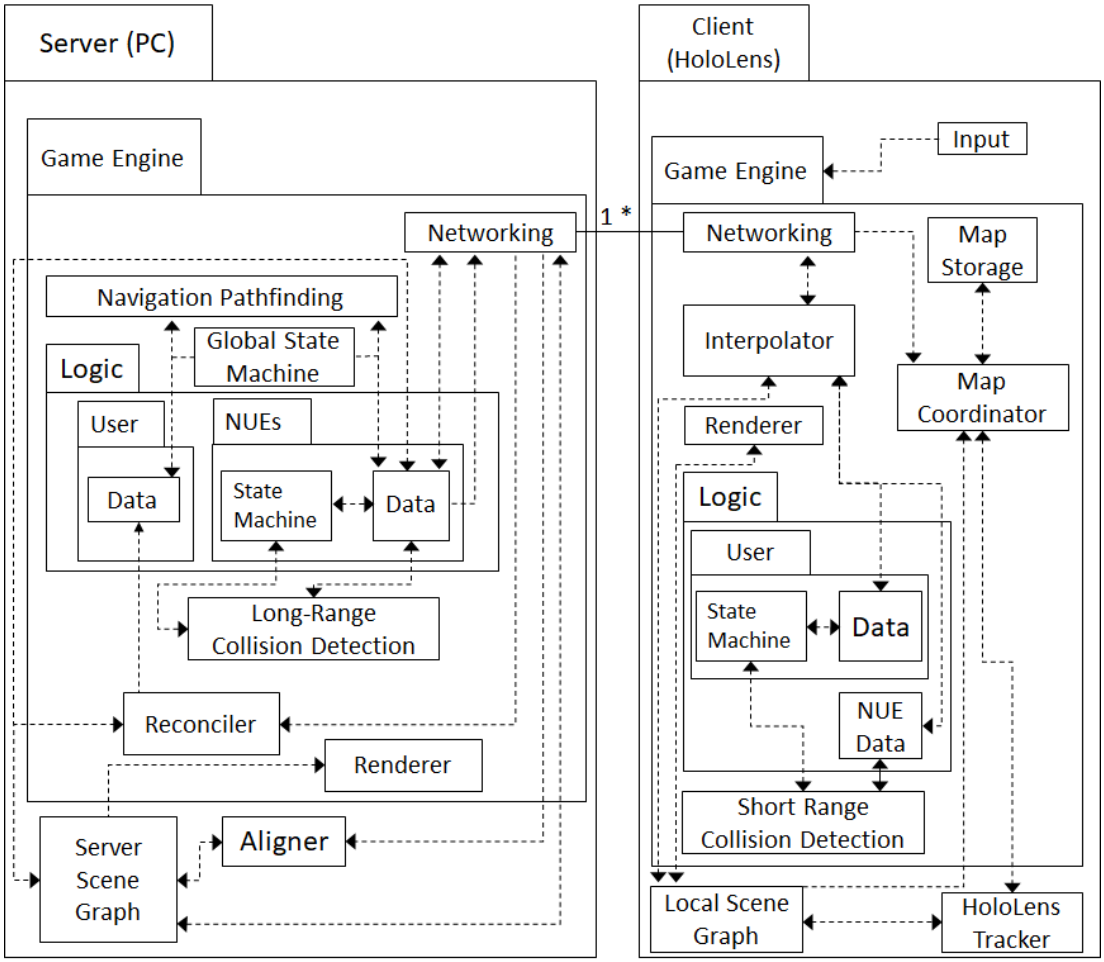

Figure 4.2: System decomposition with the game engine components expanded. This shows how the majority of computation and logic is offloaded onto the server. The server handles all of the experiences logic, including the state machines for all Non User Entities (NUEs). The client only runs a minimal viewer, processing the state of the local client, and short range collision detection.

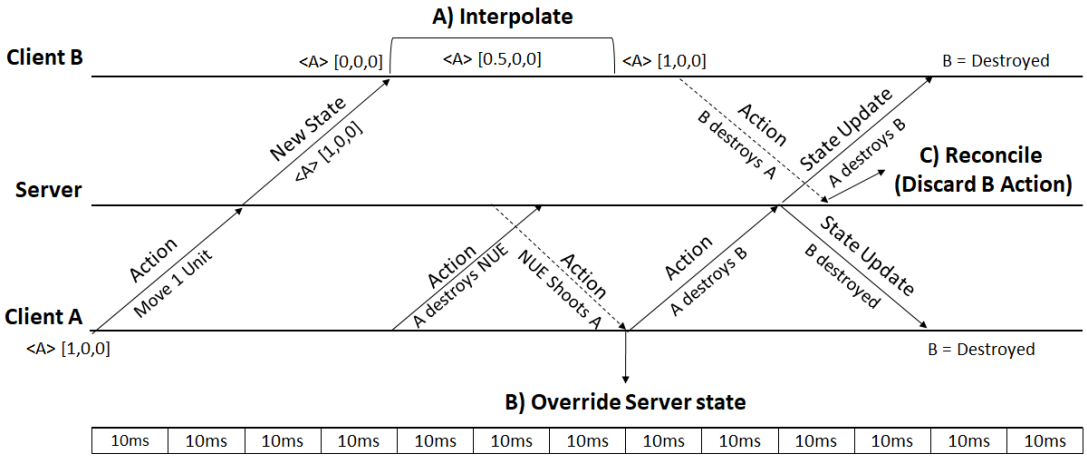

Figure 4.3: A network action sequence showing techniques used to share actions between clients [32]. In this scenario, two clients perform actions in 10 millisecond (ms) intervals, and each client experiences 20ms latency. A) Client side interpolation. When client B receives an update of client A moving. Rather than instantly updating the state on client B, I interpolate between the current and new state over time. This causes smoothing to occur, hiding jumps in poses. B) Client side overriding. A destroys a Non User Entity (NUE) the same time the NPE attacks client A on the server. Since A's time stamp is placed before the server action, the server reconciles, with client A overriding the state received by the server. C) Server side reconciliation. Client A destroys client B (sending the result to the server) and 10ms later, client B destroys client A. The server collects, and reconciles both actions according to the network time stamps (since A happened before B, the server discards B's action).

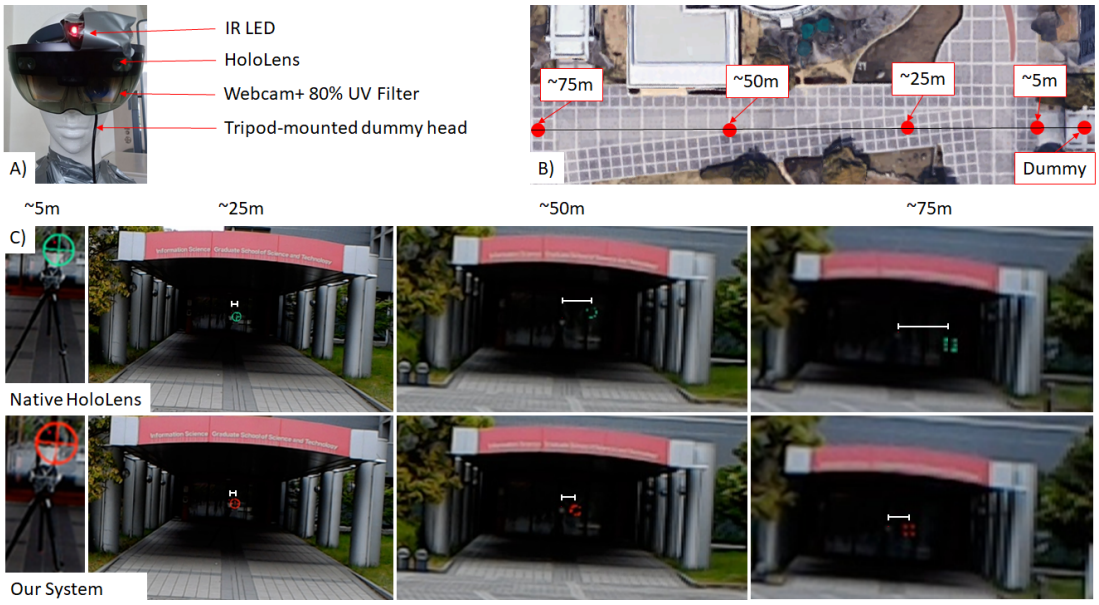

Figure 4.4: Improved accuracy of synchronized poses between multiple HoloLenses. A) I mounted two HoloLenses onto dummies, with a webcam embedded into the dummies' eye to take pictures through the HoloLens. B) I place the dummies in a previously mapped environment and oriented both HoloLenses so that they face each other at distances 5m, 25m, 50m, and 75m. C) With increasing distance, the error of the shared position of the out-of-the-box HoloLens system becomes very large, while my system maintains a higher accuracy.

Chapter 5. HoloRoyale: the First Instance of a LSHF CAR Experience

In this chapter, I fit my envisioned LSHF CAR experience and my current implementation into the design space outlined in Section 3. HoloRoyale is the first instance of a LSHF CAR experience where several users work together to defend key locations placed in urban areas against an invasion of virtual robots. Users had to form teams and defend several communication satellites distributed in the environment by destroying robots that attack the satellites in waves. After several waves a boss robot appears. The game ends when players destroy the boss robot or the robots destroy at least one base station. This experience leverages the high fidelity features of my system, including visual occlusions and real-virtual world interactions. This experience is also deployable in larger areas and is specifically designed for distributed interactions. One key limitation when applying the gamespace was that the original design of HoloRoyale had to be modified in order to fit my design space. As such, it is likely that when applying other experiences to this design space, their narrative will also need to be modified. This chapter details these modifications and the applications of the elements within the design space to create the experience. Then, in order to validate the experience, I demonstrated it at several conferences and describe the observations made during the demonstrations.

5.1 Fitting the Experience to the Design Space

By fitting HoloRoyale to the established design space (Figure 3.2), I apply the suggested configurations, addressing the challenges unresolved by the platform that I implement my experience on.

Interaction via remote AR avatars: Although the implementation presented in Section 4 improves the accuracy of pose synchronization over large distances, this error is still noticeable, and can be further impacted by the network latency.

To overcome this limitation, we modified how users interact with AR content. Instead of via a virtual hand held pistol-like controller per original design, I represent the users as remote avatars. Each user has two virtual drones that follow them (Figure 5.1a). Users interact with the virtual environment through these virtual drones by firing virtual lasers in the direction the user is facing. These avatars provides several key benefits:

- Hide any inaccurate pose synchronization while still keeping the illusion of perfect tracking between users.
- Hide temporal inconsistencies by utilizing client side interpolation (Figure 4.3a) & overriding (Figure 4.3b) [32].
- Provide targeting assistance for users and consistent interactions by tagging the target for interaction, and orienting the virtual avatars towards the target of interaction on all clients (Figure 5.1c).

Additionally server side reconciliation [32] allows us to resolve conflicting states between users (Figure 4.3c).

Spatial understanding: I provide a minimal interface (Figure 5.2a) to assist with spatial understanding. I place 2D symbolic attractors in the upper compass bar to highlight key gameplay objective locations. These serve two purposes, the first as a directional awareness aid, the second to provide additional information of the game context such as, the distance to the location and the direction relative to the user. I also provide several variations of the WIM [14] that can be zoomed by holding down one of the buttons on the gamepad.

Communication: To facilitate communicate between non co-located users, I provide a voice and emoticon communication system. Users can select an instant message by holding one of the buttons of the gamepad, and using the thumbstick to select one of the available messages, then releasing the button to send it. Users can also use voice chat by holding another button and speaking. The UI shows any instant messages sent, and which users are utilizing the voice chat system (Figure 5.2a).

User Redirection: To navigate users towards key locations, and away from dan-

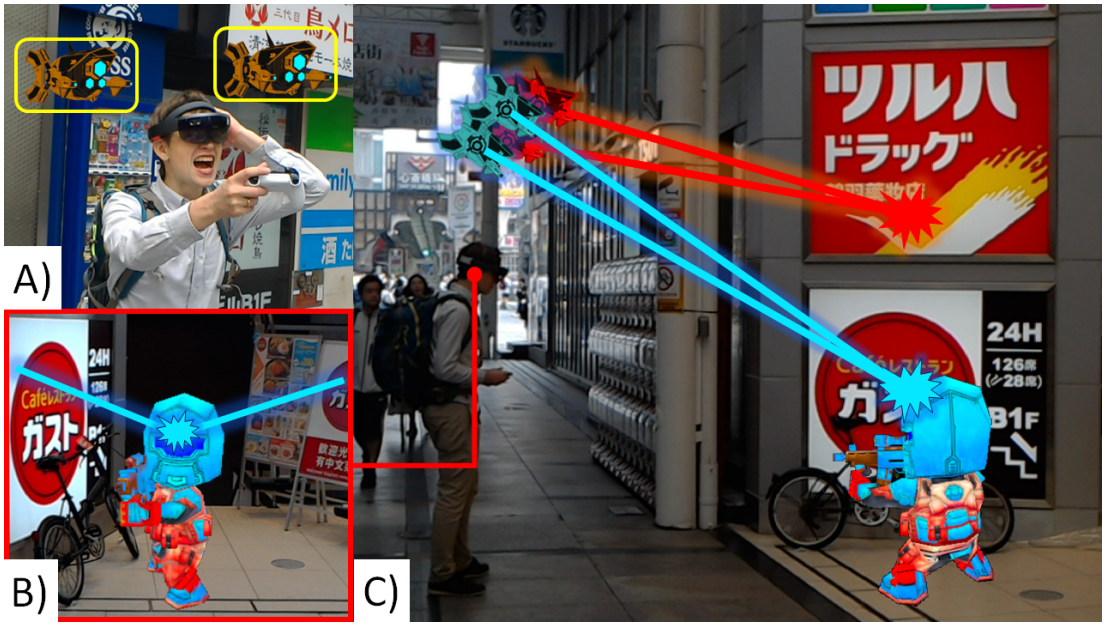

Figure 5.1: Unattached virtual avatars A) provide a means of hiding spatial and temporal errors in pose synchronization. B) If a user interacts with a target C) by tagging the target for interaction and orienting the virtual avatars towards the target, other users observe a correct interaction (blue) instead of a miss due to spatio-temporal inconsistencies (red).

gerous areas, I provide both 2D symbolic elements in the UI and 3D diegetic user redirection elements. The 2D symbolic elements in the HUD's compass bar flash to remind users of their objective, attracting their attention and guiding them towards their target. A 3D radio portal functions as a diegetic attractor, highlighting where users should be standing. The 3D symbolic navigational cues highlight a suggested pathway towards a target, functioning as both a navigation assistance tool and as a user redirection element (by guiding users towards key locations while avoiding areas tagged as dangerous, or likely to cause my system to fail). The 3D diegetic repellers are synonymous to roadwork barriers, blocking pathways to areas I don't want users to be in (Figure 6.1b).

5.2 Demonstrations

To obtain some early observations and feedback, I deployed my experience at ISMAR and UIST [81, 82], observing how users played the game. The demonstration area within these conferences was indoors, less than $250m^2$ in size, not isolated from pedestrian foot traffic, and featured occlusions from both other demonstrations and the surrounding environment. As the conferences were in the fields of AR and VR, it is safe to assume users who played HoloRoyale at these conferences were familiar with AR technologies, moreso with the Microsoft HoloLens. Each group of 3 users would play the game for 2 minutes, then return and if desired, provide feedback. The feedback given suggested the game was both fun and the interactions were natural. I noticed that users instinctively respected the user redirection elements with little instruction about them. Even so, some users reported being frustrated at the placement of the repellers, this is likely because of the small area of movement was being restricted further. There were limitations in the venues; The play areas were small ($< 250m^2$), significantly crowded, did not distribute users over the play area and therefore was not highlighting the collaboration of the large scale. The sessions were also restricted to 2 minutes. Because of these limitations the demonstrations did not target my envisioned scenario. The feedback from the participants and the limitations of the demonstration venues raised the question on the effectiveness of repellers and their effects on the user's enjoyment of the experience in LS environments. As such, I conducted a controlled user study, eliminating all possible limitations. I describe the study in the next chapter.

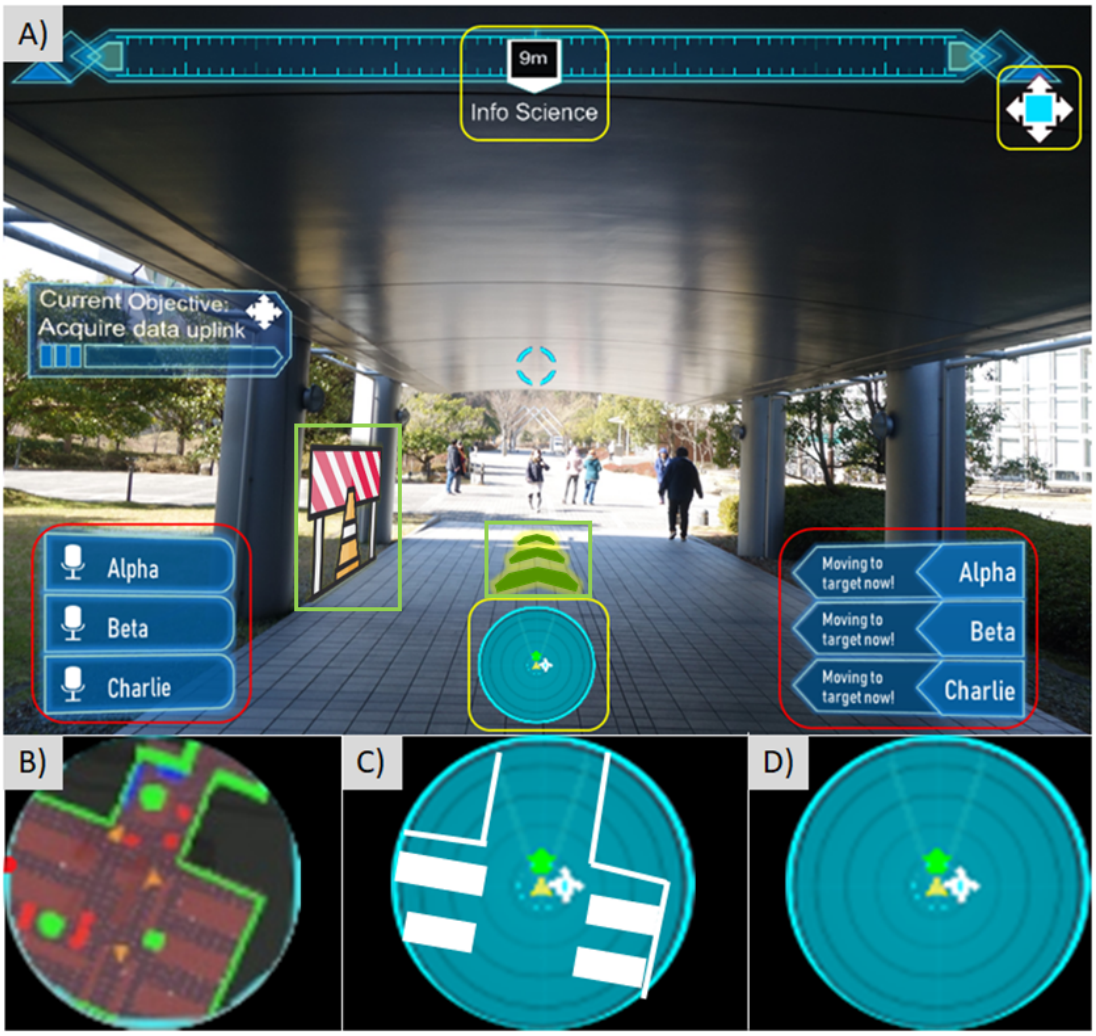

Figure 5.2: My user interface and experience provides the user A) Spatial understanding of their environment (yellow rounded squares), information related to the communication between users (red rounded squares), and elements for navigation and user redirection (green squares). The compass bar at the top of the user's view shows the relative rotational difference from the user's view angle. The arrows highlight the suggested pathways for users, while the roadwork signs indicate an impassable area. The map tool at the bottom provides limited spatial information. The map has three variants: B) World In Miniature [14], C) Simplified world in miniature, D) Radar showing only relative positional information.

Chapter 6. Evaluation: The Navigation Effect of Diegetic Repellers

I expected that a user's instinctive reaction to a virtual diegetic repeller will be analogous to a real wall, inciting them to find an alternate path to their goal. I also expected the virtual diegetic repellers to have no significant impact on the user's enjoyment because users would view the diegetic elements as part of the game experience[27]. During my demonstrations, users obeyed the boundaries created by the virtual diegetic repellers but reported frustration due to the restrictions they created. I hypothesized that this was due to the limited demonstration area. As I conceptualized repellers as a means of user redirection in LS environments, I conducted a user study focusing on the effect of virtual diegetic repellers on user navigation in a LSHF CAR context.

I deployed a variation of HoloRoyale in a 15,625m^2 area on my university campus (see Figure 6.1c) and recruited participants to play it in groups of 3 members at a time. For this user study, I removed navigation cues, and restricted spatial understanding tools to the compass bar (for showing attractors, Figure 5.2a) and the radar representation of the environment (Figure 5.2d). I also slightly modified the system that HoloRoyale is built on, increasing the client pose synchronization rate to 60Hz. I represented the virtual diegetic repellers as a construction roadwork sign (See Figure 1.1b) and the diegetic attractors as a highlighted radio box. I limit their visibility to 8 and 5 meters, respectively. The 2D symbolic in-HUD attractors were visible at all times. I had the following hypotheses:

H1 Participants will respect the barriers formed by diegetic repellers.

H2 The virtual diegetic repellers will not significantly impact the participant's enjoyment.

6.1 Participants

I recruited 24 participants (17 male, 7 female) between 22 and 34 years (mean 25.5, standard deviation 3.8), from students within my university via email, poster, flyer, and social networking. Participants selected their preferred time slots, creating 8 groups of 3 participants each from overlapping time preferences. Among them, 17 participants had not used a HoloLens before, 8 participants had not played a location based game before, and 15 participants rated their ability to use a map tool as above average.

6.2 Procedure

My study consisted of two phases, a preparation phase and the study itself. I show a timeline in Figure 6.2.

Upon their arrival, participants listened to a brief explanation of the experiment procedure, signed consent forms, and filled in a pre-study questionnaire. The participants then took part in an interactive tutorial of HoloRoyale that explained the gameplay, tasks, and user functions (5 mins). The tutorial had several paused sections, allowing participants to familiarize themselves with all game functions.

I assigned each participant one of three bases to defend (see Figure 6.1c). Once all participants arrived at their assigned base the game was started.
Participants played two sessions of HoloRoyale with the following flow (45 minutes each):

1 **Preparation Phase (Defend):** Participants defend bases by shooting virtual robots. The phase is completed once thirty robots are destroyed at each base. This phase ensured participants were at their respective starting locations before starting the next phase.

2 **Trial Phase (Upload at target location):** One of four statically placed target points appears in a random order within the play area. Participants converge to the location of the target, standing within 2

meters of it. Once all participants arrive at the target point, a progress timer starts to count down, with the phase ending after 10 seconds.

3 Participants return to their assigned bases and repeat phases 1 and 2 for all four target locations.

4 **Final Phase** After all locations were visited, the final boss appears at a static location. Participants converge to the boss' location and destroy the boss, ending the session.

After each session, everyone returned to fill in a post-session variant of the Usability Metric for User Experience [29] (See Figure 6.4) (4.5 mins).

Between the two sessions, participants took a 15 minute rest. After both sessions, participants were free to provide free-form feedback. The total time for each group was approximately 2 hours.

For safety reasons, during the user study each participant was shadowed by an assistant. The assistant did not interact with the participant, unless the participant reported something wrong with the system during play (for example, a system failure). This happened during 12 trials and the data for those trials was discarded. I compensated each participant for their time (∼10 USD per hour). This study was approved by the institutional review board of [Removed for Anonymity]

6.3 Variables

My experiment was a *within-subjects* user study with the following independent variables:

Repellers $\in$ { Displayed, Hidden }

This describes if diegetic repellers were present in the session. I counterbalanced the order this variable was chosen.

Target $\in$ { A, B, C, D }

Each session had four trials, one for each target. Repeller layouts were unique for each target creating the following situations: Barriers in open spaces, a

long hallway barricaded off, virtual navigation in narrow areas, repellers that are not visible until a participant is near the goal forcing a long redirect (See Figure 6.1). The order the target locations appeared in was randomized and counterbalanced between groups.

SessionNumber $\in \{ 1, 2 \}$

I include the session number to observe if there was a learning effect between sessions.

6.4 Results

H1 stated that participants will respect the boundaries set by the diegetic repellers. To investigate this I plot the pose and velocity data recorded during each trial (Figure 6.3). I estimate the total amount of poses at ∼1,382,400 (average time per target @ 4 mins * 4 targets * 30 poses per second * 2 repeller conditions * 3 participants * 8 groups). I plot the poses as a KDE heatmap with a kernel size of 3 meters for each target, except C that I use a 1 meter kernel size, due to the smaller viewport. They show that when repellers are present, participants walked through the barricated areas in 4/216 cases. **H2** stated that the existence of virtual diegetic repellers will not impact the participants' enjoyment. I investigated this by analyzing the results from the likert questionnaire participants answered after each session, as well as the amount of time participants took to complete the game. I use the criterion of $p < 0.05$ to determine statistical significance.

I show the results of my likert questionnaire in Figure 6.4. I compare the answers to my questionnaire with a Wilcoxon Signed-Rank test. The results showed that the presence of repellers had no significant impact on the frustration ($T = 39.5$, $p = 0.394$), ease of use ($T = 17.5$, $p = 0.94$), and how much participants enjoyed the game ($T = 39.0$, $p = 0.46$). On the other hand, participants reported that repellers significantly affected their ability to perform their intended actions ($T = 21.0$, $p = 0.0096$). The presence of repellers also

negatively affected the participants' mental image of their surroundings ($T = 19.5$, $p = 0.046$) and their ability to communicate with their partners ($T = 5.0$, $p = 0.008$).

To investigate if participants reached their targets faster as they became more familiar with the user interface and the game layout I compare the time it took them to finish each session. I show the time participants took to complete each session in Figure 6.6. As the Shapiro-Wilk test showed that the data was not normally distributed I used the Wilcoxon Signed-Rank test. The results show that participants completed the second session significantly faster ($T = 47$, $p < 0.001$). I checked how long participants spent looking at the zoomed map tool between sessions. The Wilcoxon Signed-Rank shows that during the second session participants spent significantly less time looking at the map ($T = 96$, $p = 0.005$). Finally, a Wilcoxon signed-Rank test showed that t_w, the time taken between sessions minus the time looked at the zoomed map tool, was significantly reduced ($T = 109$, $p = 0.019$).

I also investigated how the presence of repellers affected the time needed to reach each target location. As expected, participants took longer to reach the target when repellers were present (Figure 6.5). A Wilcoxon Signed-Rank test showed a significant difference in the amount of time taken for targets A, C, and D ($T = 0.0$, $p = 0.001172$) and no significance on target B ($T = 0.0$, $p = 0.093$).

6.5 Discussion

The results of the KDE plot visually *support* **H1**. When repellers were present in the scene, participants mostly respected the boundaries they set. This was the case even when users had to follow a complex pathway in wide areas (Target A), or a maze in a smaller area (Target C).

It is also worth noting that the repellers were not 100% successful. In the trials where repellers were not successful, one or two team members who had

already arrived at the target location continuously prompted the remaining participants to hurry. One participant even suggested to ignore the repellers and to walk through them, when his teammate could not immediately find the alternate path around the repeller.

This suggests that although virtual diegetic repellers present an intuitive barrier that is mostly respected, users may disregard them, e.g., due to peer and time pressure, frustration, or carelessness. When designing LSHF CAR experiences it is thus important to include reinforcing effects that prevent users from walking through diegetic repellers, e.g., by turning off the CG and prompting participants to return or by penalizing the crossing of diegetic repellers. Furthermore, when creating LSHF CAR experiences designers need to carefully consider the effects of collaborative mechanics as well as the placements of repellers, attractors, and areas of interest.

The statistical analysis of the likert questionnaire supports **H2**. Although participants did not report a significant impact on their enjoyment of the game or frustration, repellers significantly impacted the answers to questions Q3, Q4 and Q5. It is likely that participants felt that they could not complete the task as intended because the repellers blocked their way and they had to think of an alternative approach. This is supported by the KDE plot for target D, where the repeller in front of the target location forced participants to turn around to find an alternative path. Nevertheless, 81% of the participants stated that they could perform their actions as intended. We found no statistical difference between the answers of users who reported familiarity with the HoloLens vs those for whom this was their first experience, and therefore do not believe this was a factor in the overall results. This also suggests that the designed experience was very natural and could be used by both novices and those familiar with the platform.

The participants' difficulty to create a clear image of the environment when repellers were present could be due to a variety of factors. First, the repellers changed their location for each target. This could have confused the partici-

pants and made it more difficult to maintain a clear image of the environment. Second, the repellers were only visible when participants came close to them. This could have also contributed to the participants' anxiety when exploring a path. Furthermore, I did not provide navigation cues that could have helped participants efficiently navigate around the repellers. Nevertheless, 86% of the participants stated that they had a good mental image of the environment.

To my surprise, participants reported a significant negative impact on their ability to communicate with their peers. On follow up interviews several participants stated that it was more difficult to accurately portray and communicate the alternative pathways to reach a target when virtual repellers were present. This suggests that more detailed environment maps, e.g., WIM, that contain information about repellers and navigational cues could simplify communication in complex scenarios.

As expected, participants required significantly less time to complete the second session. This could be in part because participants become familiar with the layout of the environment and the UI. This is supported by 30% of participants with no prior experience reporting that they had initial difficulties understanding how to locate the target areas, but became adept at doing so very quickly. Another explanation of this finding could be the simple layout of my environment, which made it relatively easy for participants to find alternative routes to the target location. The simple layout could have also allowed participants to easily recognize the target location from the indication in the compass bar. These observations are supported by the reduced time participants spent looking at the map as well as the reduction of t_w during the second session.

When providing free-form feedback after both sessions were completed, overall participants stated they enjoyed HoloRoyale and liked having the ability to communicate with each other during the sessions. In addition, with HoloLens experience reported a feeling of a 'larger FOV' when playing HoloRoyale, compared to other applications they have tried previously. This could be because

during the preparation phase participants were actively engaged in the game. This focused their attention at the center of the screen thus effectively reducing the noticeability of partially rendered CG due to the limited field of view. At the same time, during the trial phase participants were asked to navigate through a LS environment whilst simultaneously being exposed to UI content being placed along the screen border. As the UI content was visible at all timed, this could have reinforced the illusion that the CG was not bound by the HoloLens' field of view. In the future, it is necessary to investigate what prompted this reply from my participants as it could provide means to create immersive experiences on OST-HMD with a limited field of view. Finally, although assistants were not allowed to interact with the participants participants, one of the assistants reported observing that a participant walked into a grassy area outside of the marked play area. This was later determined to be due to an error in the system's tracking during runtime. As a result that participants data was removed from the study.

6.6 Limitations

There were several limitations in this study. First, the study focuses primarily on a single representation of a virtual diegetic repeller, and did not evaluate the effects of all design elements adapted from the design space, such as the indirect representation. It is thus necessary to investigate the effectiveness of other design elements on navigation and interaction. For example, navigation cues and more detailed maps could help overcome the impact of repellers ob the user's mental environment model.

Second, due to the limited battery life of the HoloLens, I could not provide more than four target locations in a session. Furthermore, although I conducted my study in a large environment it was rather simple. As my study was conducted during class time the university campus was also largely without crowds. A more complex environments with many more distractions and

pedestrians could lead to different results.

Third, there are technical limitations of the HoloLens display, resulting in transparent rendered content, that could affect the fidelity of the overall experience and should be investigated in future studies.

Fourth, we only investigated one small component of the design elements described in our design space, and therefore should evaluate others, such as the effects of indirect vs. direct interactions in the presence/absence of spatio-temporal inconsistencies in the collaborative environment. Another possible future interaction is the effect of specific methods of communication within the LS interactions.

Fifth, I modified the original gameplay of HoloRoyale, removing any instances of virtual robots during the trial phase to prevent additional factors from affecting the navigational effect of repellers. Additional time pressure to rush to the target location and return to the bases due to the presence of attacking enemy robots, may have prompted participants to ignore repellers more often further underlining the need for reinforcement.

Finally, the environment does not completely match my target scenario. To ensure participants' safety there were no hazards in the area I deployed. It is thus unclear if in an AR context visual diegetic repellers could be sufficient to remind users of the danger thus keeping them out of harms way without the need of reinforcement.

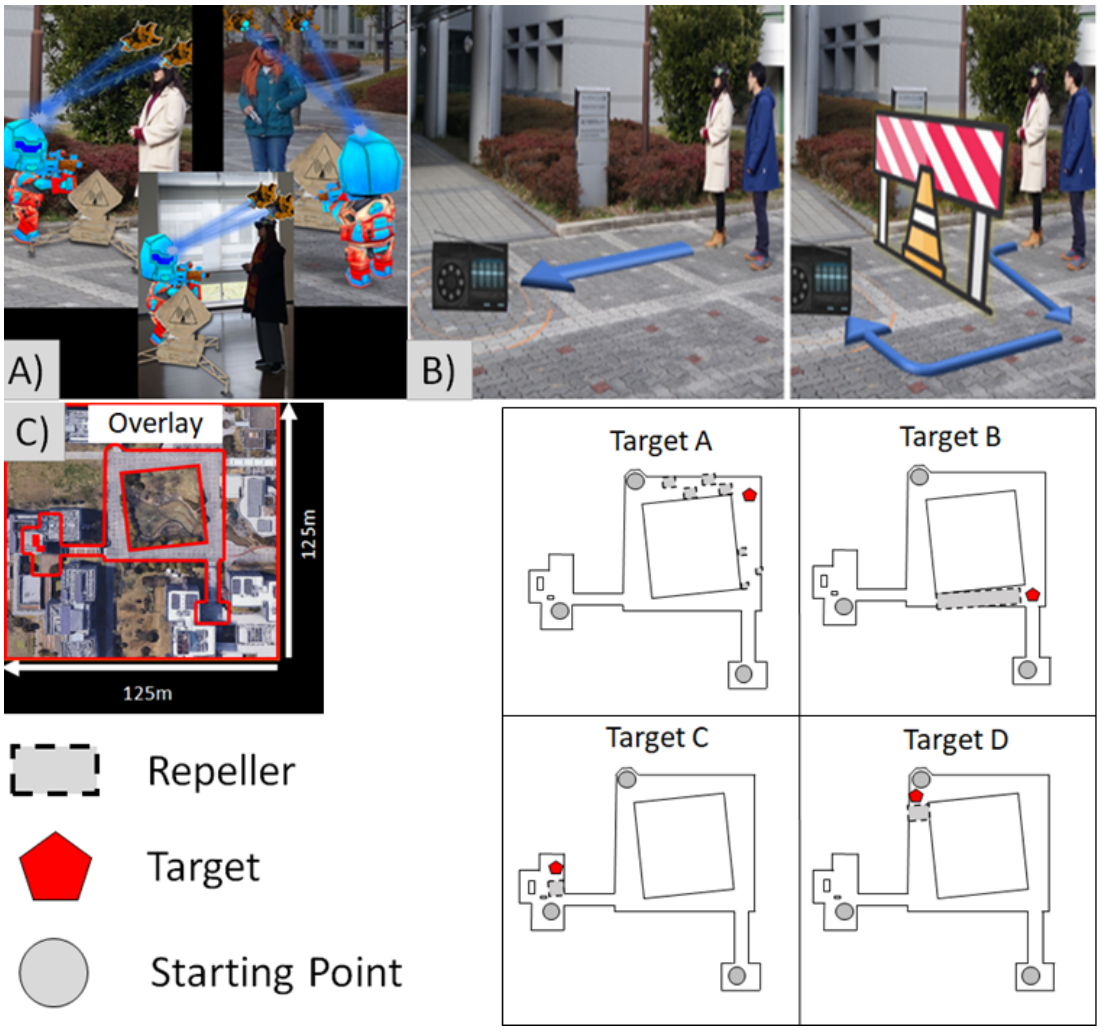

Figure 6.1: I deployed HoloRoyale on a university campus, spanning a 125x125m area, containing both indoor and outdoor areas. A) The first part of my user study has participants move to three statically placed bases, defending them against virtual robots. B) participants proceed to move to one of four target locations, the variable I introduce appears during this phase. This diegetic repeller is represented as a roadwork construction sign. C) The layout used in the study, I show a layout for each target location as different repeller layouts are setup to simulate different scenarios. [Target A] Virtual barriers in open spaces creating an obstical course. [Target B] A long hallway being barricaded off. [Target C] Virtual navigation in narrow areas. [Target D] Repellers not seen until the last approaching second to attempt a frustrated response.

2 Hours					
5min	45min	5min	15min	45min	5min
Tutorial	Trial #1	Q&A	Rest	Trial #2	Q&A

Figure 6.2: Experiment Timeline

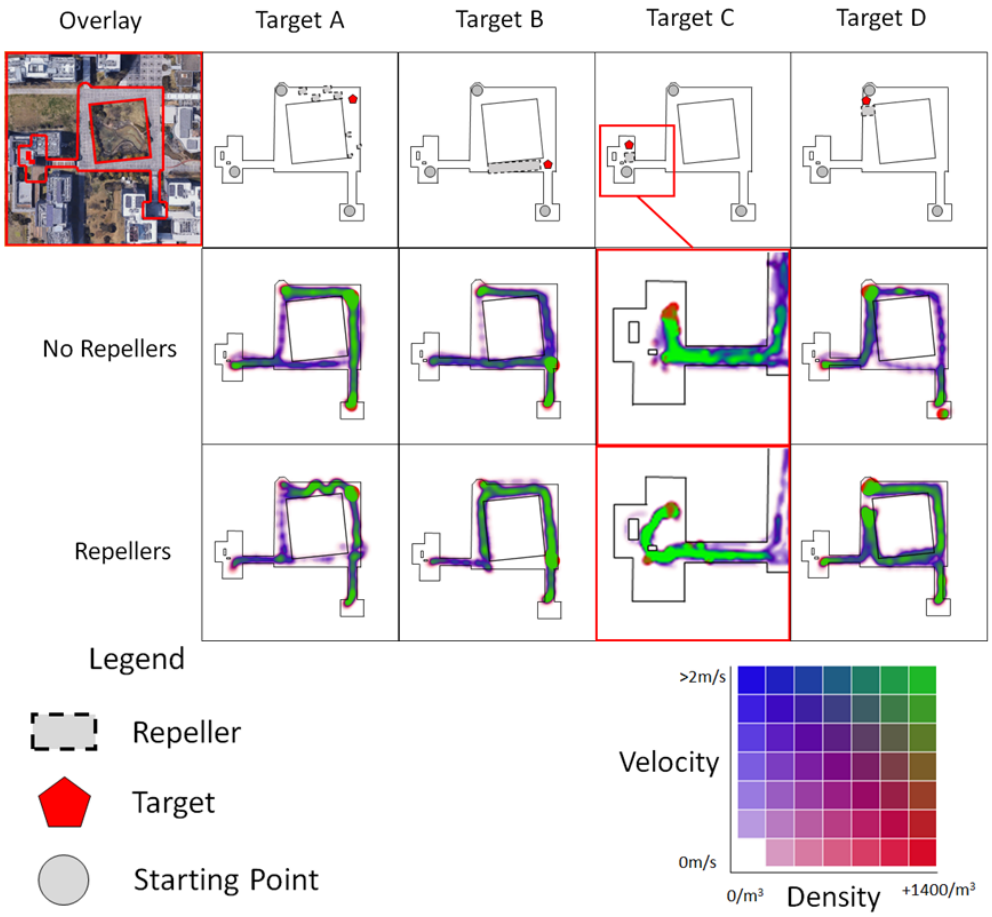

Figure 6.3: KDE heatmaps showing the density, and velocity of poses I collected during the trial phase of my study. Participants move from their starting position, to one of the target locations that appear in a random order. I use a kernel of 3 meters (with the exception of target C, for which I use a 1 meter kernel).

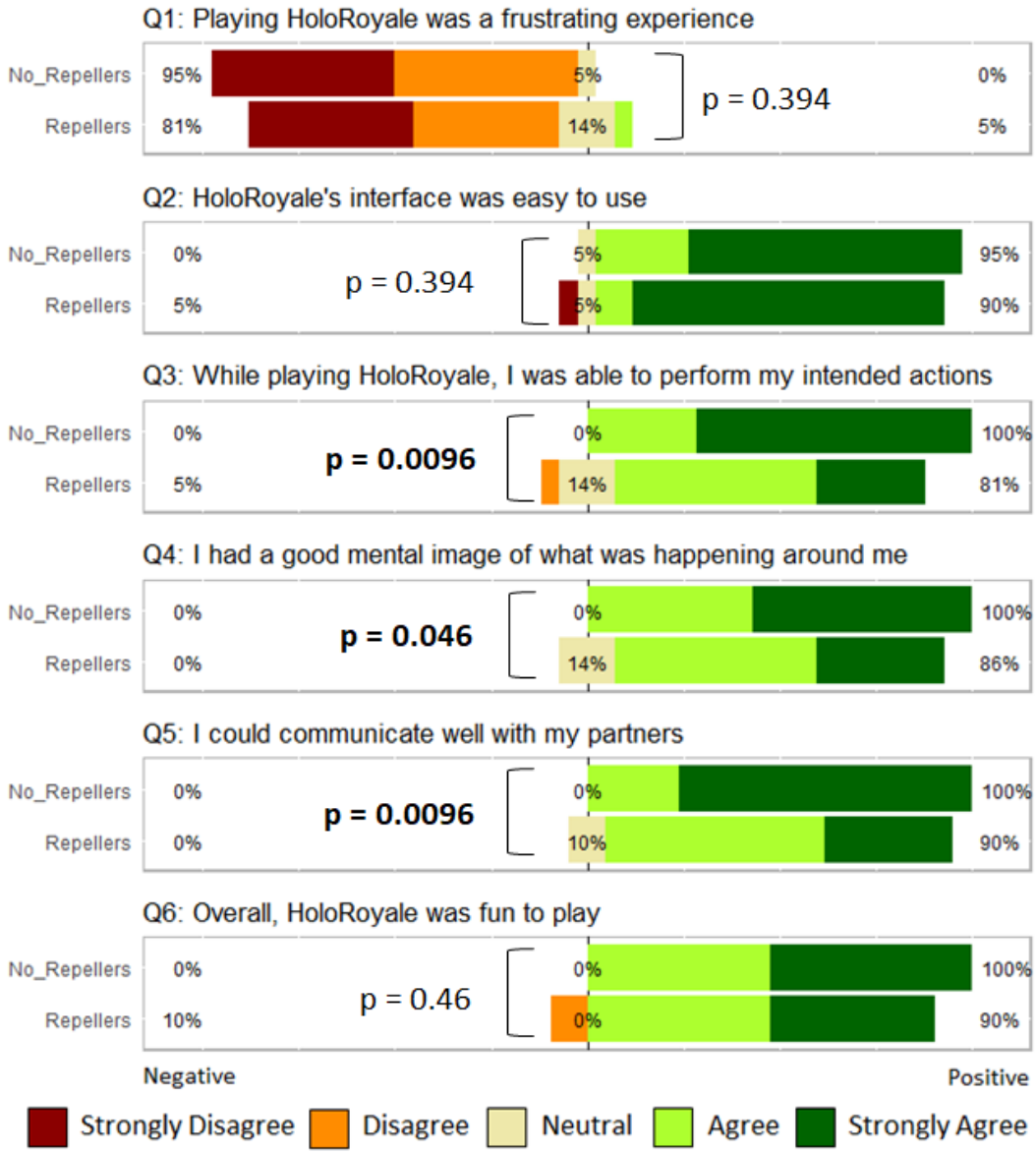

Figure 6.4: The results from my variation of the Usability Metric for User Experience [29].

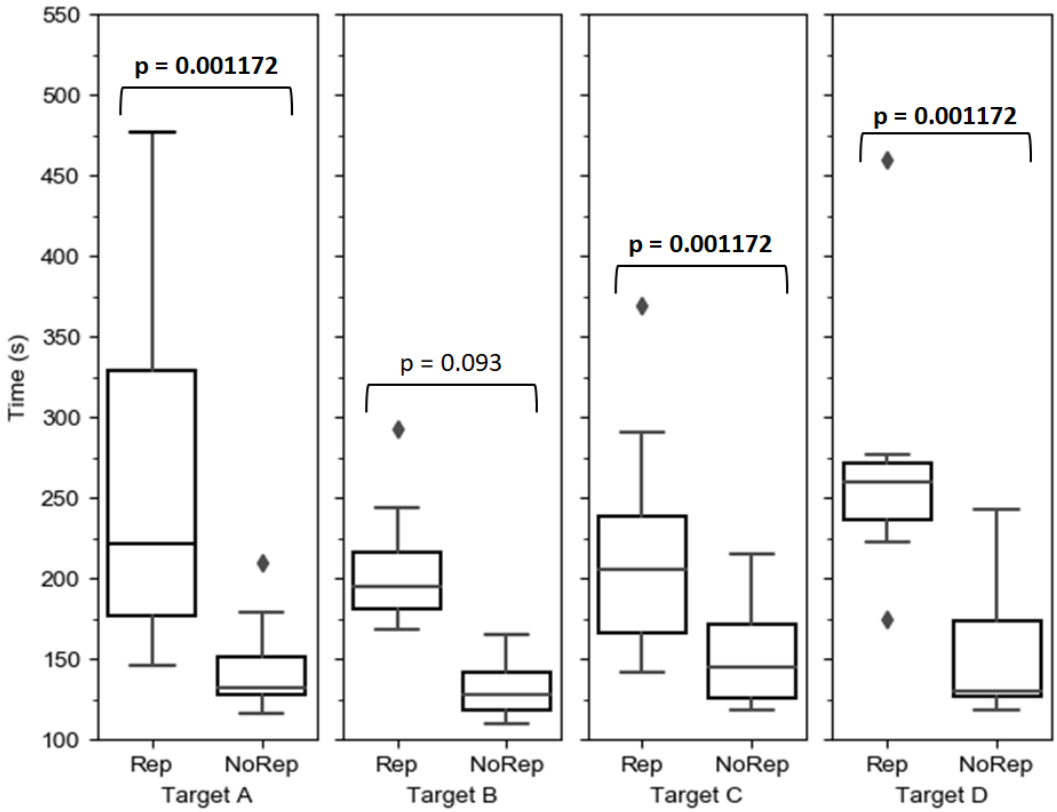

Figure 6.5: Box plots showing the time differences for each target with/without virtual repellers.

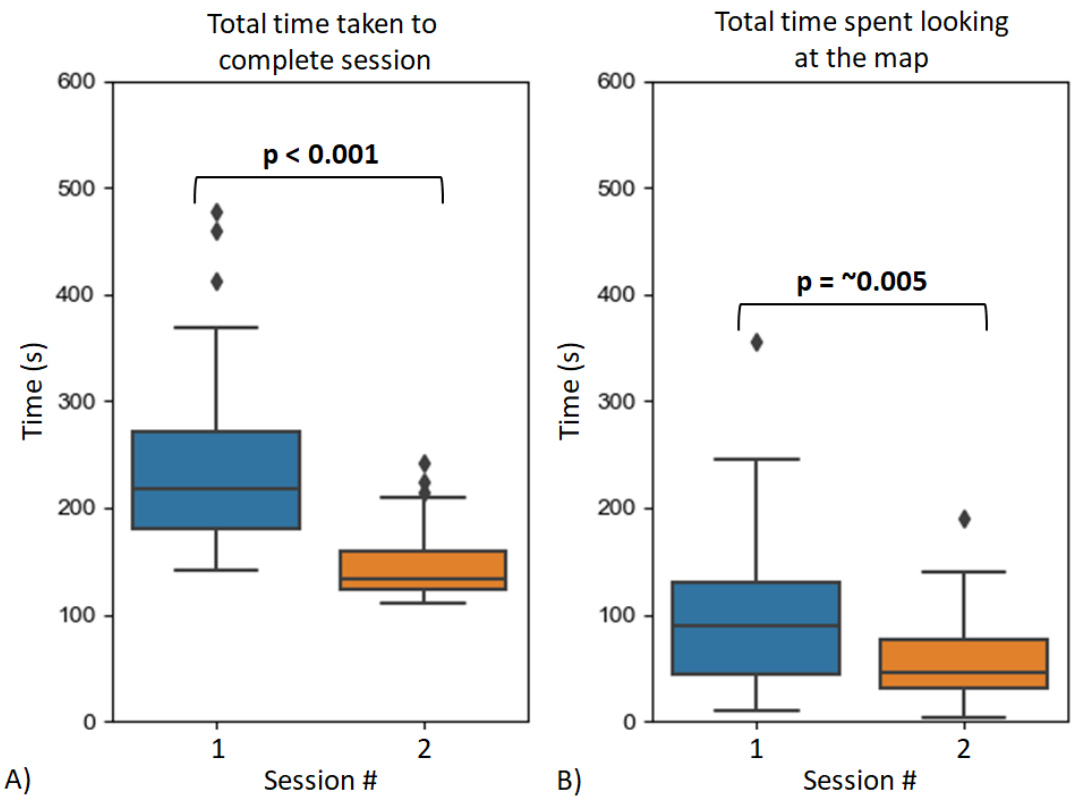

Figure 6.6: Boxplots showing A) the time taken between sessions and B) how long participants spent looking at the zoomed map of their environment between sessions.

Chapter 7. Expanding the Fidelity Capabilities of OST-HMDs

One core component of this thesis is the creation of high fidelity content, this includes producing AR content which closely resembles the human visual system. The system described in this thesis leverages a OST-HMD for compositing CG content onto the user's view of their environment, but due to the single focal plane of the OST-HMD, it is incapable of creating a rendering which can match the properties of the observers eye. This chapter briefly describes an extension of an OST-HMD to allow refocusable content (established during my masters thesis) then details an evaluation based on the first instance of an AR turing test focused on refocusable AR content.

7.1 System Design

I show the workflow of EyeAR in Fig. 7.1. EyeAR is a closed loop system with the following basic steps:

1. Measure the user's eye
2. Render CG with correct DoF
3. Correct screen-object disparity
4. User observes graphics
5. Goto 1

In the following subsections, I explain the design of each non-trivial step.

7.1.1 Measuring the Eye

Accurately measuring a user's focal distance is a very challenging task. Most available EGT devices, e.g. Tobii Glasses, accurately estimate the pupil size, but provide only inaccurate estimations of the focus depth from intersection

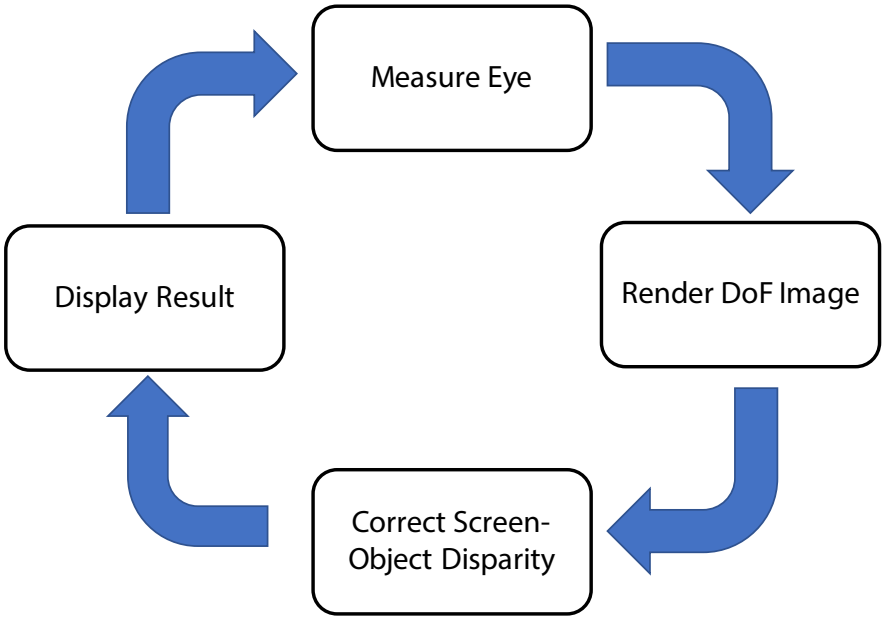

Figure 7.1: EyeAR is a closed loop system: Given eye measurements I generate a realistic DoF image. After correcting the offset between the screen and the virtual object I present the results on the screen, which triggers a response by the user's eye.

of the estimated gaze directions. Some methods use the variance in the reflection of infrared LEDs on the lens of the eye to estimate the focus depth [77]. Others distinguish between different focus planes [54, 97], or estimate the focus through intersection of the gaze with scene geometry [73]. However, they do not provide continuous measurements that are necessary for correct DoF representation. The most accurate solution to measure the focus distance is an autorefractometer. Although there are different designs, the majority measure the appearance of a ring emitted with near-infrared light on the retina. The autorefractometer adjusts the position of various focusing lenses until the image appears in focus to determine the focus distance of the eye. An autorefractometer can estimate the focus distance, with an average error of only 0.25D. It is often used in medical examinations, and was also used to verify the focus distance in refocusable HMDs [64, 57, 59]. Due to the required optical elements

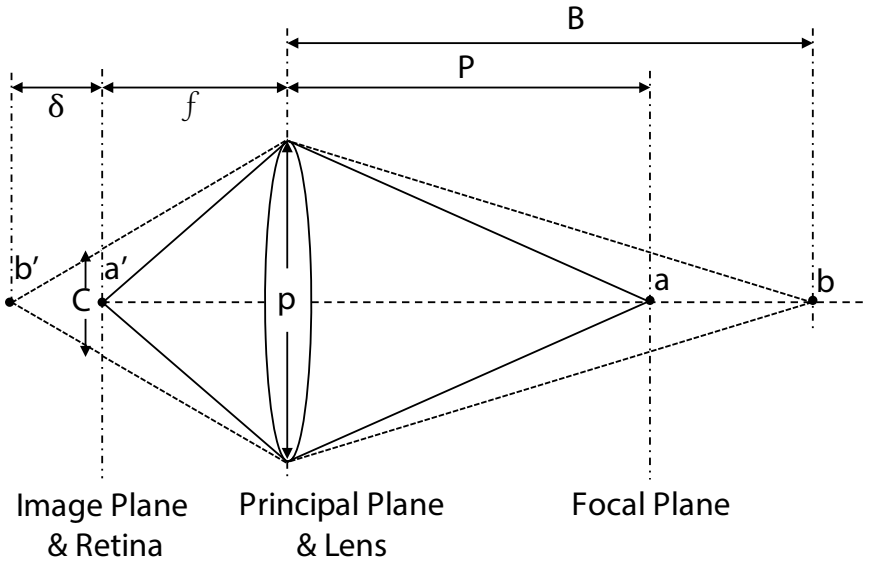

Figure 7.2: Camera model used in my renderer (adapted from [21, 33]). A point a lying on the focal plane, located P away from the eye, appears in focus as a' on the retina. Point b, located B away from the lens is focused onto b' and is blurred by the CoC, C.

even the smallest commercially available autorefractometers are too bulky to fit into an OST-HMD and have a low refresh rate of only about 5Hz. EyeAR could be applied with any of the methods mentioned above, as long as the system is capable of acquiring accurate measurements of the user's focus distance, e.g., through [77, 54]. Although it is possible to build a portable prototype of the system using eye tracking solutions, I opted to create a tabletop prototype that uses the autorefractometer as it offers the highest accuracy and reliability among the available options.

7.1.2 Rendering

The aim of my rendering component is to create CG with a DoF that matches the user's view of the real world. Technically speaking, I need to match the camera parameters used to create CG to the parameters of the user's eye.

The human visual system relies on several depth cues in order to distinguish which objects are closer to us than others. Teittinen [94] and Ware [102] discuss all depth cues in detail; one of them, Accommodation, is closely related to DoF and refers to the eye changing its shape to change its focal length, thus bringing objects at different distances into focus. The human visual system always focuses at a distinct depth. Therefore, objects that lie at different depths appear blurred (see Figure 7.2); They exhibit a circular defocus, commonly called the Circle of Confusion (CoC).

Figure 7.2 shows the camera model used in my renderer. The eye is represented as a camera with a single principal plane with focal length f and a circular aperture p. Point lights on the focal plane such as a are projected onto a point a' on the user's retina. Point b refers to the virtual image of a distant point light that would be focused δ behind the retina, resulting in a CoC with diameter C. C is proportional to both the pupil size p, as well as the ratio of distances of focal plane and point light. I can also express the second proportional term as the fraction of δ and $\delta + f$, with f being the focal length of the eye, resulting in:

$$C/p = \delta/(f + \delta) \tag{7.1.1}$$

Isolating C and substituting $\delta/(f + \delta)$ with ΔF, I obtain:

$$C = \Delta F p \tag{7.1.2}$$

Therefore, In order to create CG that matches the state of the user's eye, I must measure its focal length and pupil size to determine the CoC. I then feed these values into a distributed ray tracer [21].

7.1.3 Correcting Screen-Object Disparity

Most available OST-HMDs display CG on a fixed focal plane. In some models, the distance of this focal plane can be adjusted, for example in the Brother WD-100. However, the principle remains the same. When the user is not focusing on the same depth as the virtual display, the content on the HMD-screen becomes blurred. As my rendering algorithm assumes that the display position coincides with the focus distance, I attempt to correct this disparity.

In my demo at ISMAR2015 [84] I physically adjusted the position of the display to match the focus distance at any given moment. The moving rail in my prototype could adjust the screen position at a speed of 4cm/s, which I found to be too slow. This coincides with the observations made by Shiomi et al. [90], who found that users could accommodate on a target moving from a position of 50cm to 1m at a speed of 10cm/sec.

As an alternative, I tried SharpView [73]. SharpView estimates the amount of blur caused by the difference between a user's focus distance and the screen location. It applies a sharpening effect to the presented image, so that, when the image is blurred by the estimated amount, it coincides with the intended view. In my trials I found that although SharpView improves the visibility of the texture of the objects, it tends to create unrealistic sharpening artifacts that clearly distinguish CG from real objects. This is likely because the estimated amount of blur did not perfectly match the amount of blur perceived by participants. This coincides with the observation made in [73]. In my trials, participants could always find the virtual objects.

For the reasons explained above, I decided not to correct the disparity problem in this experiment. Instead, I carefully minimized depth variance in my experimental scene, in order to minimize the impact of the screen-object depth disparity (see Figure 7.6).

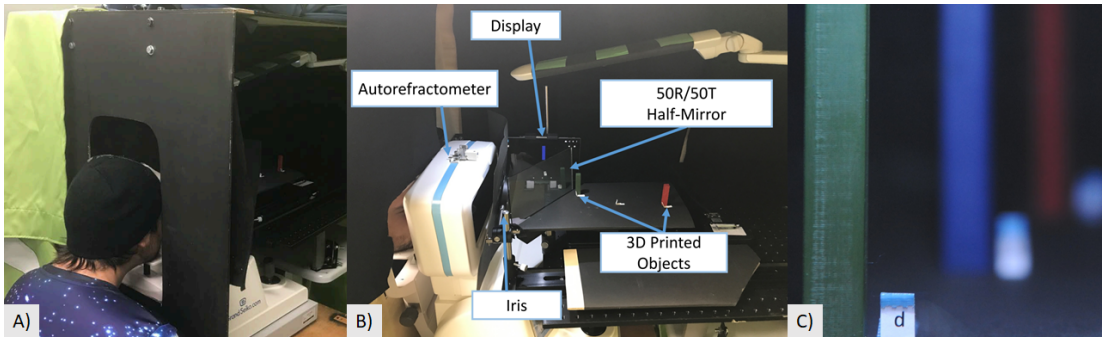

Figure 7.3: My experiment setup: (**a**) A user sitting in front of the autorefractometer and looking inside the box enclosure (the side panel was opened for illustrative purposes). (**b**) Internal view of the box enclosure. (**c**) User's view.

7.2 Experimental Platform

My experimental platform consisted of the following elements:

Measuring the Eye. I use the Grand Seiko WAM-5500 autorefractometer. It measures pupil size p and focal length in diopters D at 5Hz with an accuracy of ± 0.25D. It also allows us to read these measurements over a serial connection. I then calculate the focus distance as $P = 1/D$ and transmit p and f to the rendering component.

Rendering System. I implemented the rendering pipeline on a desktop computer with an Intel core i7 3790K processor, 8GB RAM and an NVIDIA GTX 980 graphics card.

My implementation of distributed ray tracing [21] is based on the NVIDIA's Optix [76] framework. For each point on the retinal image I combined ray tracing results from 32 samples uniformly distributed over the pupil. I additionally use bi-directional reflectance functions [71] for materials. To account for the slow measurement speed of the autorefractometer, I linearly extrapolate the values received over the past second to predict the next observation. I then interpolate from the current parameters towards the predicted values. I rendered all images at 30 fps and transferred them to the display in my experimental platform.

Physical Setup. My experimental setup is shown in Figure 7.3(a). The setup was surrounded by an enclosure to prevent variations of illumination of the scene caused by external light sources. Inside the box enclosure, an LED lamp was placed to provide a controllable area light source. I display the graphics on a Dell Venue 8 OLED tablet with a resolution of 2560×1600 and 359dpi and a 50R/50T optical mirror to create an optical-see-through display with a fixed display depth. The display was positioned 0.375m away from the participant. Inside the box enclosure I placed three 3D printed pillars painted in uniform colors. One was green, the second was blue, and the last one was red, and they were placed at depths of 0.25m, 0.375m, and 0.5m, respectively, on a tilted platform to provide a perspective depth cue (see Figure 7.3b). In each trial, one of the three real pillars was replaced with a computer-generated counterpart, which was rendered either in perfect focus or using the measurements taken from the autorefractometer, exhibiting DoF. Given the standard error of the autorefractometer of 0.25D, the expected focus measurements for each pillar were

- 23.5-26.6cm for the green pillar,
- 34.2-41.3cm for the blue pillar, and
- 44.4-57cm for the red pillar.

Liu et al. [57] have shown that measurements provided by the autorefractometer fluctuate near the ground-truth focus distance. I thus average the measurements over the past 2 seconds to account for this fluctuation.

Humans use multiple depth cues such as accommodation, vergence, overlap, and retinal image size to recover the depth of objects [94]. My setup was designed to eliminate shadows, aerial perspective, overlapping, and linear perspective cues. Additionally, to prevent movement and binocular parallax cues, participants had to wear an eye-patch over their non-dominant eyes and placed their heads onto a chin-rest. An iris behind the autorefractometer ensured that the participant's field-of-view was limited to the experiment setup. To prevent

texture depth cues the pillars had a uniform color. The position of the virtual pillars was manually aligned with the position of the real pillars. Whenever the pillars were replaced, their position shifted slightly. Therefore, participants would not have be able to distinguish between shifts misalignments caused by aberrations of their eye and pillar placement. To assist participants in refocusing at different depths the letters placed next to the pillars provided a texture cue, and the relative size of the pillars provided retinal image size cues. These cues allow participants to adjust the focal length of the eye to refocus at different distances. I could thus measure the participant's accommodation to determine the focus depth.

7.3 Experiment

The purpose of this experiment was to assess whether correctly rendered DoF computer-generated objects using EyeAR were more difficult to distinguish from the real objects than those generated without using EyeAR. For this I conducted a variant of the graphics Turing Test [17]. My hypotheses stated that:

H1 With the autorefractometer off, participants will guess the virtual pillar correctly more often than when the autorefractometer is on.

H2 With the autorefractometer on, participants will correctly guess the virtual pillar no better than random chance.

7.3.1 Participants

I recruited twelve participants (6 female, 6 male) between 19 and 45 years, mean 30.2, and standard deviation of 9.2, from both the students at the university and the general public. All participants claimed to have normal or corrected-to-normal vision with the use of contact lenses. I verified this with visual acuity tests under three conditions (see Section 7.3.2). The study was

conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Nara Institute of Science and Technology. Participants signed a consent form, and were monetarily compensated for their time.

7.3.2 Preliminary Tests

My system requires reliable measurements of the participants' dominant eye with the use of the autorefractometer. I verified that I could read data over a range of diopters. I also confirmed that they were able to focus on the objects within the range used in the experiment. All volunteers had to pass four preliminary tests before taking part in the experiment. The first test allowed us to verify the participant's dominant eye. The second and third verified the participants visual acuity for both far and near sight. The final test verified that the autorefractometer was capable of reading the eye diopters over the experimental depth range (0.25 - 0.5m). The total time needed for these tests was 10 minutes per participant. The tests are described as follows:

Eye-dominance test. Each participant stood 3 meters away from a marked object facing towards it. They were then asked to hold their hands 50cm away from their eyes with the thumb and index finger forming a connected arch, and looked at the marked object through the arch. This caused participants' to hold their hands biased towards their dominant eye. I consider the eye used to look at the object their dominant eye. The dominant eye was then used on the later preliminary tests, with the non-dominant eye covered by an eye patch.

Acuity test. Once the dominant eye was determined, each participant carried out two standard tests used in optometry to measure the visual acuity. In the first, they stood 2.8 meters away from a Snellen chart held at eye level. If the participant was not able to read a letter, they guessed it. The test finished when they failed to read more than half of a line, or was able to read the entire chart. I only accepted participants with a visual acuity of at least 20/30.

The second acuity test measured participant's short-distance visual acuity. Each participant sat down on a chair 40cm away from a Rosenbaum chart placed at eye level. They proceeded to read the numbers on each line. If they could not read the line, they guessed it. The test had the same ending conditions as previous one, with participants excluded if their determined visual acuity was not better than 20/30.

Operability test. The fourth and last preliminary test verified that the autorefractometer could accurately read the participant's dominant eye. The measurements can sometimes be affected by participants with corrected-to-normal vision with the use of either contact lenses or spectacles, or when the person is not able to focus on the objects placed within the range limit accepted by the autorefractometer, between 0.25m and 0.5m. Each participant was asked to look at a chart hanged from a horizontal beam. The dominant eye was measured while the card moved steadily along the beam, starting from 0.5m away and moving it towards the eye (up to 0.25m), and then back to the original position over an interval of 3 seconds.

7.3.3 Task and Procedure

Participants sat down in front of the autorefractometer and looked at the scene inside the box enclosure. The scene composed of three pillars with letters besides them. Participants were instructed to focus on the pillars and letters beside them for a total of 20 seconds per trial. At the end of each trial, a researcher occluded the scene and in a separate room the participant wrote on a sheet which pillar they considered to be virtual. While they answered, the researcher changed the pillars according to a sequence randomly generated in advance. The procedure was then repeated twelve times for each participant. From which each permutation of the experimental variables were repeated twice per participant. Each session took 40 minutes per participant in total.

7.3.4 Variables

My experiment was designed as a *within-subjects* experiment. I looked at whether participants correctly guessed the pillar that was computer generated as a binary outcome. I also collected general feedback from each participant with the goal of gaining insight on the results from the evaluation. The independent variables of my experiment were as follows:

VirtualPillar $\in$ { red, green, blue}

This refers to the pillar that was virtual.

Autorefractometer $\in$ { On, Off}

I evaluated two situations: one adjusting the blurriness of the virtual pillar according to the autorefractometer readings and the other without adjusting it, simulating a pinhole camera rendering.

TrialSequence $\in$ { 1 ... 12}

I included the sequence number in which the trials were carried out to observe whether there was a learning effect.

7.3.5 Results

Table 7.5 shows the results of the analysis of the recorded data. I use $p < 0.05$ as a criteria to determine statistically significant results.

The results of the regression *support* **H1**. The results show that the number of correct guesses was significant when the autorefractometer was off for all three pillars. I also found that there was a learning effect, making it easier for participants to guess the virtual pillar in later trials. The analysis also revealed that gender was also significant, which was not initially expected.

H2 stated that with the use of the data collected from the autorefractometer, participants would correctly guess about 33% of the time, the same results as if they tried to guess by chance. This hypothesis was *rejected* for the green (58.3%) and red (41.7%) pillars. However, the results for the blue pillar (33.3%

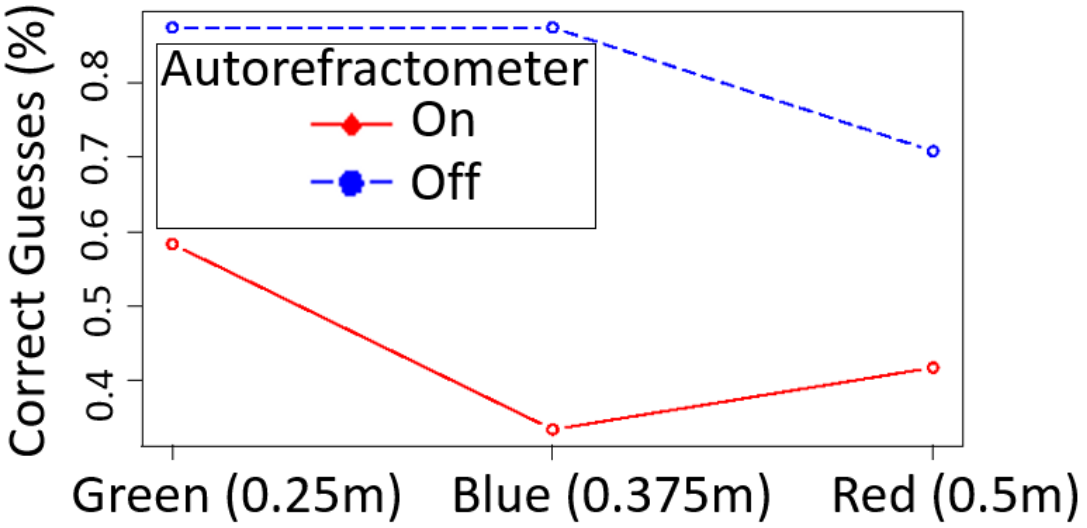

Figure 7.4: Overall percentage of correct guesses for each pillar when the autorefractometer was on (red line) and off (blue line).

	Estimate	Std.error	Wald	$Pr(>\|W\|)$	95% Conf. interval
Intercept	-0.561	0.565	0.99	0.3202	-1.67 to 0.55
TrialSequence	0.152	0.058	6.96	0.0083	0.04 to 0.27
Autorefractometer=on	-1.841	0.273	45.65	$< 10^{-10}$	-2.38 to -1.31
VirtualPillar	0.236	0.158	2.23	0.1357	-0.07 to 0.55
Age	0.010	0.013	0.58	0.4475	-0.02 to 0.04
Gender	0.872	0.304	8.23	0.0041	0.28 to 1.47

Figure 7.5: Coefficients and p-values of the experimental variables of a linear model fitting using GEE with correlation structure=exchangeable. The p-values show that Autorefractometer is the main contributor to the model but not the only one, also gender and TrialSequence are significant.

of correct guesses) were *compatible with* **H2**. In all cases, CG rendered in the "Autorefractometer On" condition were recognized less often than CG rendered in the "Autorefractometer Off" condition.

Figure 7.4 shows the percentages of success when guessing which pillar was the

virtual one for each experimental condition. At first sight, the plot shows that participants found it easier to guess the virtual pillar when the refractometer was off (81.9% of the times) compared to using the data from the refractometer (44.4%) to adjust the blurriness of the pillars in real time. Concerning the pillars, participants guessed it right more often when the green one was virtual compared to the other two. Looking at **H2**, the total number of correct guesses for "Autorefractometer=on" was 32, n=72. I computed the confidence intervals using different methods to test whether the probability to guess it correctly could be 0.33. Wilson score interval method returned a sample mean of 0.44 with a 95% confidence interval=0.335 to 0.559. This result rejects that 0.33 is the expected mean value of the population by a very narrow margin, therefore rejecting **H2**. Clopper-Pearson binomial interval returns a very similar value (95% conf. int.=0.327 to 0.566) but it does not reject the hypothesis, although it is still very unlikely that the true mean is 0.33.

I now take a look at the pillars separately. With Autorefractometer=on, the green pillar was guessed correctly 14 times (n=24, sample mean=0.583, 95% conf. int.=0.366 to 0.779), the blue pillar 8 times (n=24, sample mean=0.333, conf. int.= 0.156 to 0.553), and the red one 10 times (n=24, sample mean=0.417, conf. int.= 0.221 to 0.634). These results reject that green and red pillar were guessed only by chance, but provide evidence that the virtual blue pillar could not be distinguished from its real counterpart.

This was a within-subjects experimental design with a binary dependent variable. Then I decided to estimate a model using generalized estimating equations (GEE) [55, 10]. The GEE analysis provides statistical estimation similar to repeated-measures ANOVA, but can achieve higher power with a lower number of repeated measurements. Besides the two main independent variables, I also included the trial sequence (*TrialSequence*), gender, and age to test whether they could also be explanatory variables. A standard criteria to test the model fitting is the Akaike information criteria, but it has been argued that it is not applicable when using GEE, as this method does not provide models based on like-

lihood. [75] proposes that the quasi-likelihood information criterion (QIC) suits better for this case. Other criteria have been proposed such as the correlation information criterion (CIC) [38], although it is better take into account more than one criterion [26]. Table 7.5 shows that the main contributor of the linear model with no interaction terms was the experimental variable *Autorefractometer*, although it is not the only one. The trial sequence and gender are also contributors, though they have a lesser effect. On the other hand, VirtualPillar and age do not contribute. This model returned criteria values QIC=164.91 and a CIC=4.01. Removing the non-contributors and adding interaction terms for the predictors of the previous model returns that none of them are significant (Autorefractometer•Gender p=0.439; Autorefractometer•TrialSequence p=0.804; TrialSequence•Gender p=0.902). This second model returns higher scores on both QIC (169.14) and CIC (6.34), which means that the second model fitting is not as good as the previous one. One last model was tried with only Autorefractometer, TrialSequence, and gender, returns the lowest criterion scores, with QIC=164.54 and CIC=3.34, having again Autorefractometer ($p< 10^{-10}$) the strongest predictor, but both TrialSequence (p=0.022) and Gender (p=0.013) also need to be taken into account.

7.3.6 Discussion

The results of my experiment support **H1**. I found that participants were less likely to detect the CG content when rendered based on readings from the autorefractometer. I were surprised by the bad performance of the "Autorefractometer Off" condition for the central pillar. One would expect that if the pillar is rendered in a photorealistic manner, detection results would be identical to placing a picture at the position of the display. Upon further investigation I concluded that the pillar rendered under the pinhole camera assumption did not appear realistic enough as it looked too sharp. This also indicates that a pinhole eye model is not suitable for rendering realistic objects for AR. In the

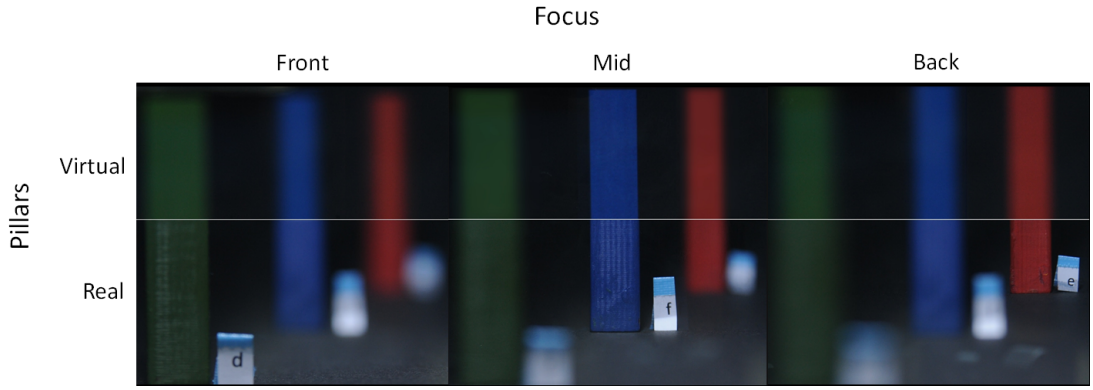

Figure 7.6: Limitation of my prototype: as the distance to the virtual image plane coincides with the blue pillar, it is the only pillar that can be rendered with only minor artifacts. Note how the red and green pillars exhibit noticeably greater artifacts. The top row shows all virtual pillars, while the bottom row only shows real pillars.

future, I want to investigate the improvement of realism taking only pupil size into account.

My results only partially supported hypothesis **H2**. This can be attributed to various factors, such as the disparity between the user's focus and display depth, appearance of the pillars, low update rate of the autorefractometer, and update delay of the resulting CG displayed. These issues could also explain the observed learning effect. Participants were more likely to detect the difference between the real objects and their CG counterparts as they became used to experiment. The results of my experiment suggest that future OST-HMDs, especially HMDs that have a single focal plane, should incorporate the concept of EyeAR to increase the realism of the rendered CG. This may range from using a non-pinhole eye model to generating CG based on measurements of the eye focal length and pupil radius, as reported in this paper.

The most likely reason that **H2** was not supported for the front and back pillars is the depth disparity between the position of the virtual pillars and the display. As discussed in Section 7.1.3 I did not correct the different depths of the participant's focus and display image, which may have caused unnecessary

blur when looking at the front or back pillars (see Figure 7.6). Although the same condition applied to the central pillar, these effects are more prominent for the front and back pillars, which led to the higher detection rate. In the "Autorefractometer On" condition, participants could not detect the blue pillar that coincided with the focal plane of the display and were less likely to detect the red pillar in the back than the green pillar in the front. I speculate that EyeAR can create realistic effects within a certain region around the actual focal depth of the display. In the future it will be necessary to investigate how large this region is, if it is ratio or diopter based, and at what point the benefits of rendering content with EyeAR are no longer detectable.

The higher detection rate could also be explained by the virtual pillars slight difference in appearance. The texture, environment illumination, refractive optics of the eye, and the sensitivity to different collars, to name a few. All affected the appearance of the virtual pillars. Although the design of my setup aimed to control the lighting conditions and I optimized the texture and appearance of the pillars to match that of their real counterparts, participants might have noticed slight differences in the color, hue, or texture of the pillars. This is also supported by the learning effects revealed by my analysis. Over the course of the experiment, participants might have become better at detecting the slight differences and cues that distinguished the pillars. I considered switching the pillars to counteract the learning effect and different sensitivity. However, this could have introduced other unintended effects. For example, artifacts introduced when optimizing the appearance of the various pillars could have influenced the results of the evaluation. Overall, my analysis did not reveal that the color of the pillar had a significant effect on the overall results ($p=0.1357$). While the color might have contributed to easier detection of the front pillar my results indicate that rendering the content based on the readings from the autorefractometer, or with the pinhole camera model of the eye had the strongest effect ($p < 10^{-10}$). In future experiments I plan to investigate what effects the pillar color had when content was rendered in the "Autorefrac-

tometer On" condition and if rendering content with a non-pinhole eye-model can achieve similar results without adjusting the blur effects of CG content.

The slow update rate of the autorefractometer is another factor that could have impacted the results of my study. For one, when participants refocused between the pillars it could take up to 0.2s for the system to register this change. Combined with the interpolation that was added to reduce the detectability of changes in the rendered content it created sufficient delay for participants to become aware of it over the course of the experiment. This could also explain the observed learning effects. At the same time, **H2** was supported for the central pillar. This suggests that the delay may not have been noticeable to the participants all the time, but potentially only when participants were observing large changes, e.g., refocusing from the front to back pillar.

Nonetheless, my results show that accounting for focus distance measurements improves the realism of the CG. I expect that faster autorefractometers and improvements to focus depth estimation from eye-gaze tracking cameras will enable focus updates at a speed that is not noticeable by participants. My results highlight the need for such technology. At the same time, the current measurement speed is enough when participants change focus between objects that are far away and very close by, e.g., in hand-reach. In that case, the maximum refocusing speed of the eye is less than 1.878 ± 0.625 diopter/sec [58], which is slower than the update rate of my system.

My analysis also revealed that the gender had a significant impact on the detection rate. My goal was to evaluate how likely users were to detect content rendered with EyeAR compared to content rendered with a pinhole camera eye model. Therefore, I did not counterbalance the participants in terms of age and gender.This might be a reason for the significance I detected and I will investigate how age and gender affects the results in future work.

Chapter 8. Conclusion

In this work, I describe a series of requirements for my LSHF CAR experience. From the requirements I established a design space drawing on both technical implementations and design aspects from both AR and video games. By applying the hardware aspect of my design space I created a software architecture and technical implementation that improves the accuracy of synchronized poses between multiple tracking systems. Then I apply my target experience to my established design space, creating HoloRoyale, the first instance of a LSHF CAR experience. I conducted a user study to explore how virtual diegetic repellers affect user navigation in a LSHF CAR context. The results from the user study suggest that virtual diegetic repellers are effective user redirection elements that do not significantly impact the user's overall immersion. I also quickly presented the design of a display concept based on the user's eye measurements as an alternative to light field displays. Then presented a tabletop prototype that emulates an OST-HMD setup and can accurately match the DoF of virtual objects to real objects. I then evaluated my prototype with a user study to verify my claims. My results strongly support **H1**, which stated that my closed loop system creates significantly more realistic renderings than a system that does not measure the user's eyes. On the other hand, my second hypothesis **H2** was rejected for the pillars in the background and foreground. This is likely due to the reconstruction error in CG caused by the screen-object disparity. Other aspects that could have contributed to this are the slow update rate of the autorefractometer, color of the pillars, or unintended artifacts.

8.1 Future Work

The work in this paper opens up several new avenues for future work. I will describe these in two sections, the first related to the EyeAR study, the second for the LSHF CAR experience created.

8.1.1 Exploring EyeAR Further

There is a need for refocusable CG for OST AR in order to create realistic scenes where objects are placed at different depths. Referring back to the previously published taxonomy of AR displays [108], EyeAR is an instance of a Personalized AR display, as opposed to the LFD approach, which is an instance of Surround AR. I are convinced that my approach is an interesting alternative to create realistic AR content. The main advantage is that EyeAR does not require complex optical systems and thus addresses many problems in existing refocusable concepts. On the other hand, EyeAR requires very accurate estimation of the user's focus depth and, in the best situation, a display that can adjust its location so that the image plane of the OST-HMD always matches the user's focus distance.

I have identified three main areas for future work: First, and most important, I need to improve the prototype to address the screen-object disparity. Second, I plan to integrate EyeAR into an OST-HMD by either miniaturizing, or replicating, the function of the autorefractometer, for example through use of eye tracking cameras as described in [77, 54, 73]. Third, I will refine my methodology for conducting AR Turing Tests and carry out several more.

My rendering algorithm assumes that the position of the display coincides with the position of the virtual object. This effectively limits its applicable range, as it can't correct large disparities between the screen and the focus distance. In the future I want to evaluate the effective range where these disparities become noticeable. Additionally, I aim to investigate how EyeAR could be combined with free-focus OST-HMDs. I imagine that EyeAR could be applied in retinal displays, or with refocusable lenses designed to always present content shown on the OST-HMD in focus. Determining the applicable range of EyeAR could also lead to a combination of EyeAR with multi-focal and varifocal displays in order to reduce the optical complexity of the system.

Alternatively, EyeAR could be combined with SharpView for single focal-plane OST-HMDs. However, I still need to investigate how to build user-specific

point-spread functions (PSFs), including 3D PSFs, and model the dynamic accommodation process. In order to achieve the required optical power, I are now considering programmable optical elements. This approach has been successfully demonstrated in the dual problem of increasing the DoF of projectors through coded apertures [109] and fast focal sweep based on a shape-changing lens [107].

In terms of portability, I aim to study how to reduce the size of my system to the point that it can be integrated into an HMD. I envision an EyeAR hardware module that can turn any legacy OST-HMD into a powerful display, perceptually equivalent to a LFD. Additionally, the update rate of my system (5Hz) can be sufficient to refocus between distances several meters away from the eye, but can lead to noticeable latency when quickly refocusing between objects placed near the eye as in my tabletop setup, where the farthest object was 0.5m away.

It is also part of my future work to improve the experimental design and standardize a methodology to carry out AR Turing Tests. In my experiment, I used three pillars with plain textures. I aim to study more complex scenes that include objects of different shapes, materials with different surface texture parameters, and models of light scattering. With increasingly complex scenes, experimental measurements could collect ordinal instead of binary answers from participants in order to provide more faceted results. I believe that CG rendered under the pinhole camera model appeared too sharp during my user study and were as a result easily detected. In the future I plan to compare the realism of CG rendered with a non-pinhole camera model versus EyeAR.

8.1.2 Exploring LSHF CAR experiences further

As this is work establishes the first design space and guidelines for LSHF CAR experiences there are several new avenues of future work that can now be explored.

The first avenue of future work is the application of my established design space into other LSHF CAR experiences, such as outdoor infrastructure planning. To create compelling LSHF CAR experiences over large distances it is also important to investigate the effectiveness of other design elements identified in my design space and their interactions. This includes the effect of indirect interactions on targeting assistance in the presence of temporal-spatial inconsistencies should be investigated. This is especially prominent in LSHF CAR scenarios due to the possibility of interacting with content placed at longer distances (which is highlighted as a problem by [78]). The amount of error users can adapt to when interacting with virtual content before experiencing difficulties is currently unknown.

I also plan to investigate the effectiveness of user redirection elements in urban scenarios with a large number of distractors and pedestrians, as well as smaller scale indoor scenes. my observations also raise questions about the effects the type and density of user redirection elements can have in different scenarios.

Third, to address the participants' comments about the perceived field of view of the HoloLens I need to investigate the effects of UI elements and user immersion on the perceived field of view of an OST-HMD.

Fourth, I evaluated the fidelity of refocusable AR content on an OST-HMD, although the study in this thesis showed that measuring the user's eye and creating CG content based on those measurements improves the realism of rendered CG content, this improvement is only significant at distances which coincide with the depth of the OST-HMD display. Future work involves using methods to address this screen-depth and content disparity.

Finally, this thesis established a crossover between LSHF CAR and video game design spaces. It's possible crossovers between video game design and other AR spaces exist. In the future, I plan to further investigate this crossover, applying the design concepts derived in this paper to other AR domains.

Publications

[1] Rompapas, D., Sandor, C., Plopski, A., Saakes, J., Shin. H., Taketomi, T., and Kato, H. Towards Large Scale High Fidelity Collaborative Augmented Reality Accepted Preprint In *Computers and Graphics* https://doi.org/10.1016/j.cag.2019.08.007.

[2] Rompapas, D. C., Rovira, A., Plopski, A., Sandor, C., Taketomi, T., Yamamoto, G., Kato, H., and Ikeda, S. Eyear: Refocusable augmented reality content through eye measurements. *Multimodal Technologies and Interaction 1*, 4 (2017).

[3] Rompapas, D., Sandor, C., Plopski, A., Saakes, D., Yun, D. H., Taketomi, T., and Kato, H. Holoroyale: A large scale high fidelity augmented reality game. In *Demo at ACM Symposium on User Interface Software and Technology* (Berlin, Germany, October 2018).

[4] Rompapas, D., Sandor, C., Plopski, A., Saakes, D., Yun, D. H., Taketomi, T., and Kato, H. Holoroyale: A large scale high fidelity augmented reality game. In *Demo at IEEE International Symposium on Mixed and Augmented Reality* (2018), IEEE.

[5] Rompapas, Damien C.; Rovira, Aitor; Ikeda, Sei; Plopski, Alexander; Taketomi, Takafumi; Sandor, Christian; Kato, Hirokazu. EyeAR: Refocusable Augmented Reality Content through Eye Measurements. In *Demo at the IEEE International Symposium on Mixed and Augmented Reality Best Demo Award* (September 2016).

[6] Rovira, Aitor; Rompapas, Damien C.; Sandor, Christian; Taketomi, Takafumi; Kato, Hirokazu; Ikeda, Sei. Eyear: Empiric evaluation of a refocusable augmented reality system. In *Poster in Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)* (2016), pp. 11–12.

References

[1] Niantic – pokemon go, 2016. Last accessed: June 26, 2017.

[2] Roboraid – microsoft, 2016. Last accessed: June 26, 2017.

[3] Star wars: Jedi challenges, 2018. Last accessed: April 22, 2019.

[4] 2k studios – xcom: Enemy unknown, Apr 2019.

[5] Electronic arts – deadspace 3, Apr 2019.

[6] Google – project tango (platform), Apr 2019.

[7] Invaders – magic leap, 2019. Last accessed: April 22, 2019.

[8] Microsoft studios – age of empires, Apr 2019.

[9] Nintendo – pikmin, Apr 2019.

[10] Agresti, Alan. *Categorical Data Analysis*. Springer Berlin Heidelberg, 2011.

[11] Akeley, Kurt; Watt, Simon J.; Girshick, Ahna R.; Banks, Martin S. A Stereo Display Prototype with Multiple Focal Distances. *ACM Transactions on Graphics 23*, 3 (2004), 804–813.

[12] Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and MacIntyre, B. Recent advances in augmented reality. *IEEE computer graphics and applications 21*, 6 (2001), 34–47.

[13] Bell, B., Feiner, S., and Höllerer, T. View management for virtual and augmented reality. In *Proceedings of the 14th annual ACM symposium on User interface software and technology* (2001), ACM, pp. 101–110.

[14] Bell, B., Höllerer, T., and Feiner, S. An annotated situation-awareness aid for augmented reality. In *Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2002), UIST '02, ACM, pp. 213–216.

[15] Billinghurst, M., and Kato, H. Collaborative augmented reality. *Communications of the ACM 45*, 7 (2002), 64–70.

[16] Roboute Guilliman. Codex Astartes. In *the grim darkness of the future* (021.M31).

[17] Borg, Martin; Johansen, Stine S.; Thomsen, Dennis L.; and Kraus, Martin. Practical Implementation of a Graphics Turing Test. In *In Proceedings of the International Symposium on Advances in Visual Computing* (2012), pp. 305–313.

[18] Bork, F., Schnelzer, C., Eck, U., and Navab, N. Towards efficient visual guidance in limited field-of-view head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics 24*, 11 (Nov 2018), 2983–2992.

[19] Carmack, J. Latency mitigation strategies. *Twenty Milliseconds* (2013).

[20] Cheok, A. D., Goh, K. H., Liu, W., Farbiz, F., Fong, S. W., Teo, S. L., Li, Y., and Yang, X. Human Pacman: A Mobile, Wide-area Entertainment System Based on Physical, Social, and Ubiquitous Computing. *Personal and Ubiquitous Computing 8*, 2 (2004), 71–81.

[21] Cook, Robert L.; Porter, Thomas; Carpenter, Loren. Distributed ray tracing. In *Proceedings of ACM SIGGRAPH* (1984), pp. 137–145.

[22] Depping, A. E., Mandryk, R. L., Li, C., Gutwin, C., and Vicencio-Moreira, R. How disclosing skill assistance affects play experience in a multiplayer first-person shooter game. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), ACM, pp. 3462–3472.

[23] Dunn, David; Tippets, Cary; Torell, Kent; Kellnhofer, Petr; Aksit, Kaan; Didyk, Piotr; Myszkowski, Karol; Luebke, David; Fuchs, Henry. Wide field of view varifocal near-eye display using see-through deformable membrane mirrors. *IEEE Transactions on Visualization and Computer Graphics 23*, 4 (2017), 1322–1331.

[24] Durrant-Whyte, H., and Bailey, T. Simultaneous localization and mapping: part i. *IEEE robotics & automation magazine 13*, 2 (2006), 99–110.

[25] Engel, J., Schöps, T., and Cremers, D. Lsd-slam: Large-scale direct monocular slam. In *European conference on computer vision* (2014),

Springer, pp. 834–849.

[26] Evans, Scott; Li, Lingling. A Comparison of Goodness of Fit Tests for the Logistic GEE Model. *Statistics in Medicine 24*, 8 (2005), 1245–1261.

[27] Fagerholt, E., and Lorentzon, M. Beyond the HUD-User Interfaces for Increased Player Immersion in FPS Games. Master's thesis, Chalmers University of Technology, 2009.

[28] Feiner, S., MacIntyre, B., Höllerer, T., and Webster, A. A touring machine: Prototyping 3d mobile augmented reality systems for exploring the urban environment. *Personal Technologies 1*, 4 (1997), 208–217.

[29] Finstad, K. The usability metric for user experience. *Interacting with Computers 22*, 5 (2010), 323–327.

[30] Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology 47*, 6 (1954), 381.

[31] Fujimoto, Y., Smith, R. T., Taketomi, T., Yamamoto, G., Miyazaki, J., Kato, H., and Thomas, B. Geometrically-correct projection-based texture mapping onto a deformable object. 540–549.

[32] Gambetta, G. Fast-paced multiplayer, 2017. Last accessed: June 27, 2017.

[33] Green, Daniel G.; Powers, Maureen K.; Banks, Martin S. Depth of Focus, Eye Size and Visual Acuity. *Vision Research 20*, 10 (1980), 827–835.

[34] Gustafson, S., Baudisch, P., Gutwin, C., and Irani, P. Wedge: clutter-free visualization of off-screen locations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2008), ACM, pp. 787–796.

[35] Gutwin, C., Vicencio-Moreira, R., and Mandryk, R. L. Does helping hurt?: Aiming assistance and skill development in a first-person shooter game. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (2016), ACM, pp. 338–349.

[36] Hillaire, S., Lécuyer, A., Cozot, R., and Casiez, G. Depth-of-Field Blur Effects for First-Person Navigation in Virtual Environments. *IEEE Computer Graphics and Applications 28*, 6 (2008).

[37] Hillaire, Sébastien; Lécuyer, Anatole; Cozot, Rémi; Casiez, Géry. Using an Eye-Tracking System to Improve Camera Motions and Depth-of-Field Blur Effects in Virtual Environments. In *Proceedings of the IEEE Virtual Reality* (2008), pp. 47–50.

[38] Hin, Lin-Yee; Wang, You-Gan. Working-Correlation-Structure Identification in Generalized Estimating Equations. *Statistics in Medicine 28*, 4 (2009), 642–658.

[39] Holland, S., Morse, D. R., and Gedenryd, H. Audiogps: Spatial audio navigation with a minimal attention interface. *Personal and Ubiquitous computing 6*, 4 (2002), 253–259.

[40] Höllerer, T., Feiner, S., Hallaway, D., Bell, B., Lanzagorta, M., Brown, D., Julier, S., Baillot, Y., and Rosenblum, L. User interface management techniques for collaborative mobile augmented reality. *Computers & Graphics 25*, 5 (2001), 799–810.

[41] Hatsune Miku A virtual synthesized voice *Audio Proceedings* , (2007).

[42] Hua, Hong. Enabling Focus Cues in Head-Mounted Displays. *Proceedings of the IEEE 105*, 5 (2017), 805–824.

[43] Hua, Hong; Liu, Sheng. Depth-Fused Multi-Focal Plane Displays Enable Accurate Depth Perception. In *Photonics Asia* (2010), International Society for Optics and Photonics, pp. 78490P–78490P.

[44] Huang, Fu-Chung; Luebke, David; Wetzstein, Gordon. The light field stereoscope. In *ACM SIGGRAPH Emerging Technologies* (2015), pp. 24:1–24:1.

[45] Itamiya, T., Tohara, H., and Nasuda, Y. Augmented reality floods and smoke smartphone app disaster scope utilizing real-time occlusion. In *IEEE International Conference of Virtual Reality (IEEE VR) 2019* (Jan 2019).

[46] Jensen, C., Farnham, S. D., Drucker, S. M., and Kollock, P. The effect of communication modality on cooperation in online environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (2000), ACM, pp. 470–477.

[47] Kán, Peter and Kaufmann, Hannes. Physically-based depth of field in augmented reality. In *Eurographics (Short Papers)* (2012), pp. 89–92.

[48] Klein, G., and Murray, D. Parallel tracking and mapping for small ar workspaces. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality* (2007), IEEE Computer Society, pp. 1–10.

[49] Klemm, M., Seebacher, F., and Hoppe, H. High accuracy pixel-wise spatial calibration of optical see-through glasses. *Computers & Graphics 64* (2017), 51 – 61. Cyberworlds 2016.

[50] Klimmt, C., and Hartmann, T. Mediated interpersonal communication in multiplayer video games. *Mediated interpersonal communication 309* (2008).

[51] Kramida, Gregory. Resolving the Vergence-Accommodation Conflict in Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics 22*, 7 (2016), 1912–1931.

[52] Tonshin, Dirk. Enriching Reality with Faces. *Journal of humour*, (2015).

[53] Kruijff, E., Swan II, J. E., and Feiner, S. Perceptual Issues in Augmented Reality Revisited. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality* (2010), pp. 3–12.

[54] Lee, Ji Woo; Cho, Chul Woo; Shin, Kwang Yong; Lee, Eui Chul; Park, Kang Ryoung. 3d Gaze Tracking Method Using Purkinje Images on Eye Optical Model and Pupil. *Optics and Lasers in Engineering 50*, 5 (2012), 736–751.

[55] Liang, Kung-Yee; Zeger, Scott L. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika 73*, 1 (1986), 13–22.

[56] Lincon, P. C. *Low Latency Displays for Augmented Reality.* PhD thesis,

2017.

[57] Liu, Sheng; Cheng, Dewen; Hua, Hong. An Optical See-Through Head Mounted Display with Addressable Focal Planes. In *Proceedings of the IEEE/ACM International Symposium on Mixed and Augmented Reality* (2008), pp. 33–42.

[58] Lockhart, Thurmon E.; Shi, Wen. Effects of Age on Dynamic Accommodation. *Ergonomics 53*, 7 (2010), 892–903.

[59] MacKenzie, Kevin J.; Hoffman, David M.; Watt, Simon J. Accommodation to Multiple-Focal-Plane Displays: Implications for Improving Stereoscopic Displays and for Accommodation Control. *Journal of Vision 10*, 8 (2010), 22–22.

[60] Mahmud, N., Saha, R., Zafar, R., Bhuian, M., and Sarwar, S. Vibration and voice operated navigation system for visually impaired person. In *2014 international conference on informatics, electronics & vision (ICIEV)* (2014), IEEE, pp. 1–5.

[61] Mauderer, Michael; Conte, Simone; Nacenta, Miguel A.; Vishwanath, Dhanraj. Depth perception with gaze-contingent depth of field. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2014), pp. 217–226.

[62] McGuigan, Michael. Graphics Turing Test. *arXiv preprint cs/0603132* (2006).

[63] McKenzie, G. Gamification and location-based services. In *Workshop on Cognitive Engineering for Mobile GIS* (2011).

[64] McQuaide, Sarah C.; Seibel, Eric J.; Kelly, John P.; Schowengerdt, Brian T.; Furness, Thomas A. A Retinal Scanning Display System that Produces Multiple Focal Planes with a Deformable Membrane Mirror. *Displays 24*, 2 (2003), 65–72.

[65] Meyer, Gary W.; Rushmeier, Holly E.; Cohen, Michael F.; Greenberg, Donald P.; Torrance, Kenneth E. An Experimental Evaluation of Computer Graphics Imagery. *ACM Transactions on Graphics*

5, 1 (1986), 30–50.

[66] Mur-Artal, R., Montiel, J. M. M., and Tardos, J. D. Orb-slam: a versatile and accurate monocular slam system. *IEEE transactions on robotics 31*, 5 (2015), 1147–1163.

[67] Narain, Rahul; Albert, Rachel A.; Bulbul, Abdullah; Ward, Gregory J.; Banks, Martin S.; O'Brien, James F. Optimal presentation of imagery with focus cues on multi-plane displays. *ACM Transaction on Graphics 34*, 4 (2015), 59:1–59:12.

[68] Ng, G., Shin, J. G., Plopski, A., Sandor, C., and Saakes, D. Situated Game Level Editing in Augmented Reality. In *In Proceedings of the ACM International Conference on Tangible, Embedded and Embodied Interaction* (Stockholm, Sweden, March 2018), pp. 409–418.

[69] Ng, K., and Ng, K. Wizard's duel ar (prototype): Cross-platform multiplayer augmented reality game using google cloud, Jul 2018.

[70] Nguyen, T., Sandor, C., and Park, J. *PTAMM-Plus: Refactoring and extending PTAMM*. PhD thesis, Citeseer, 2010.

[71] Nicodemus, Fred E. Directional Reflectance and Emissivity of an Opaque Surface. *Applied Optics 4*, 7 (1965), 767–775.

[72] Nvidia. How We Created Demos that Blur the Line Between Real and Rendered. Available online: `http://blogs.nvidia.com/blog/2015/08/11/photorealistic/` (accessed on 14 05 2017).

[73] Oshima, Kohei; Moser, Kenneth R.; Rompapas, Damien C.; Swan II, Edward J.; Ikeda, Sei; Yamamoto, Goshiro; Taketomi, Takafumi; Sandor Christian; Kato, Hirokazu. Improved clarity of defocussed content on optical see-through head-mounted displays. In *Proceedings of the IEEE Symposium on 3D User Interfaces* (2016).

[74] Padmanaban, Nitish; Konrad, Robert; Stramer, Tal; Cooper, Emily A.; Wetzstein, Gordon. Optimizing Virtual Reality for all Users through Gaze-Contingent and Adaptive Focus Displays. *Proceedings of the*

National Academy of Sciences of the United States of America 114, 9 (2017), 2183–2188.

[75] Pan, Wei. Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics 57*, 1 (2001), 120–125.

[76] Parker, Steven G.; Bigler, James; Dietrich, Andreas; Friedrich, Heiko; Hoberock, Jared; Luebke, David; McAllister, David; McGuire, Morgan; Morley, Keith; Robison, Austin; others. Optix: A General Purpose Ray Tracing Engine. *ACM Transactions on Graphics 29*, 4 (2010), 66.

[77] Patricia Rosales; Michiel Dubbelman; Susana Marcos; Rob van der Heijde. Crystalline Lens Radii of Curvature from Purkinje and Scheimpflug Imaging. *Journal of Vision 6*, 10 (2006), 1057–1067.

[78] Piekarski, W., and Thomas, B. ARQuake: The Outdoor Augmented Reality Gaming System. *Communications of the ACM 45*, 1 (2002), 36–38.

[79] Reitmayr, G., and Schmalstieg, D. *Collaborative augmented reality for outdoor navigation and information browsing*. na, 2004.

[80] Rolland, J. P., Fuchs, H., et al. Optical versus video see-through head-mounted displays. *Fundamentals of Wearable Computers and Augmented Reality* (2001), 113–156.

[81] Rompapas, D., Sandor, C., Plopski, A., Saakes, D., Yun, D. H., Taketomi, T., and Kato, H. Holoroyale: A large scale high fidelity augmented reality game. In *Demo at ACM Symposium on User Interface Software and Technology* (Berlin, Germany, October 2018).

[82] Rompapas, D., Sandor, C., Plopski, A., Saakes, D., Yun, D. H., Taketomi, T., and Kato, H. Holoroyale: A large scale high fidelity augmented reality game. In *Demo at IEEE International Symposium on Mixed and Augmented Reality* (2018), IEEE.

[83] Rompapas, D. C., Sorokin, N., Taketomi, T., Yamamoto, G., Sandor, C., Kato, H., et al. Dynamic augmented reality x-ray on google glass. In *SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applica-*

tions (2014), ACM, p. 20.

[84] Rompapas, Damien C.; Oshima, Kohei; Moser, Kenneth R.; Swan II, Edward J.; Ikeda, Sei; Yamamoto, Goshiro; Taketomi, Takafumi; Sandor, Christian; Kato, Hirokazu. EyeAR: Physically-Based Depth of Field through Eye Measurements. In *Demo at the IEEE International Symposium on Mixed and Augmented Reality* (October 2015).

[85] Rovira, Aitor; Rompapas, Damien C.; Sandor, Christian; Taketomi, Takafumi; Kato, Hirokazu; Ikeda, Sei. Eyear: Empiric evaluation of a refocusable augmented reality system. In *Poster in Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)* (2016), pp. 11–12.

[86] Sandor, C., MacWilliams, A., Wagner, M., Bauer, M., and Klinker, G. Sheep: The shared environment entertainment pasture. In *IEEE and ACM International Symposium on Mixed and Augmented Reality ISMAR* (2002), vol. 2002, pp. 1–5.

[87] Schmalstieg, D., and Wagner, D. Experiences with handheld augmented reality. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality* (2007), IEEE, pp. 3–18.

[88] Shan, Qi; Adams, Rene; Curless, Brian; Furukawa, Yudai; Seitz, Steven M. The Visual Turing Test for Scene Reconstruction. In *IEEE International Conference on 3D Vision-3DV* (2013), pp. 25–32.

[89] Sharp, G. C., Lee, S. W., and Wehe, D. K. Icp registration using invariant features. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*, 1 (2002), 90–102.

[90] Shiomi, Tomoki; Uemoto, Keita; Kojima, Takehito; Sano, Shunta; Ishio, Hiromu; Takada, Hiroki; Omori, Masako; Watanabe, Tomoyuki; Miyao, Masaru. Simultaneous Measurement of Lens Accommodation and Convergence in natural and Artificial 3D Vision. *Journal of the Society for Information Display 21*, 3 (2013), 120–128.

[91] Stoakley, R., Conway, M. J., and Pausch, R. Virtual reality on a

wim: interactive worlds in miniature. In *CHI* (1995), vol. 95, pp. 265–272.

[92] Sutherland, I. E. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I* (1968), ACM, pp. 757–764.

[93] Sutherland, Ivan E. The ultimate display. In *Proceedings of the IFIP Congress* (1965), pp. 506–508.

[94] Teittinen, Marko. Depth Cues in the Human Visual System. *The Encyclopedia of Virtual Environments 1* (1993).

[95] Solidus, Snake. Towards box-related stealth tactics. Stealth, 1974.

[96] Thomas, B., Demczuk, V., Piekarski, W., Hepworth, D., and Gunther, B. A wearable computer system with augmented reality to support terrestrial navigation. In *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No. 98EX215)* (1998), IEEE, pp. 168–171.

[97] Toyama, Takumi; Orlosky, Jason; Sonntag, Daniel; Kiyokawa, Kiyoshi. A Natural Interface for Multi-focal Plane Head Mounted Displays using 3D Gaze. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (2014), pp. 25–32.

[98] Turing, Alan M. Computing Machinery and Intelligence. *Mind 59*, 236 (1950), 433–460.

[99] Vaughan-Nichols, S. J. Augmented reality: No longer a novelty? *Computer 42*, 12 (2009), 19–22.

[100] Vinnikov, Margarita; Allison, Robert S. Gaze-contingent depth of field in realistic scenes: The user experience. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications* (2014), pp. 119–126.

[101] Wang, J., Feng, Y., Zeng, C., and Li, S. An augmented reality based system for remote collaborative maintenance instruction of complex products. In *2014 IEEE International Conference on Automation Science and Engineering (CASE)* (2014), IEEE, pp. 309–314.

[102] Ware, Colin. *Information Visualization: Perception for Design*, 3rd ed.

Elsevier, 2012.

[103] Weir, P., Sandor, C., Swoboda, M., Nguyen, T., Eck, U., Reitmayr, G., and Dey, A. BurnAR: Feel the Heat. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality* (2012), pp. 331–332.

[104] Wilson, J., Walker, B. N., Lindsay, J., Cambias, C., and Dellaert, F. Swan: System for wearable audio navigation. In *2007 11th IEEE international symposium on wearable computers* (2007), IEEE, pp. 91–98.

[105] Wing, M. G., Eklund, A., and Kellogg, L. D. Consumer-grade global positioning system (gps) accuracy and reliability. *Journal of forestry 103*, 4 (2005), 169–173.

[106] Xueting, Lin and Ogawa, Takefumi. Blur with Depth: A Depth Cue Method Based on Blur Effect in Augmented Reality. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality* (2013), pp. 1–6.

[107] Iwai, Daisuke; Mihara, Shoichiro; Sato, Kosuke. Extended Depth-of-Field Projector by Fast Focal Sweep Projection. *IEEE Transactions on Visualization and Computer Graphics 21*, 4 (2015), 462–470.

[108] Sandor, Christian; Fuchs, Martin; Cassinelli, Alvaro; Li, Hao; Newcombe, Richard A.; Yamamoto, Goshiro; Feiner, Steven K. Breaking the Barriers to True Augmented Reality. *arXiv preprint arXiv:1512.05471* (2015).

[109] Grosse, Max; Wetzstein, Gordon; Grundhöfer, Anselm; Bimber, Oliver. Coded aperture projection. *ACM Transactions on Graphics 29*, 3 (2010), 1–12.

[110] Gruber, L., Richter-Trummer, T., and Schmalstieg, D. Real-Time Photometric Registration from Arbitrary Geometry. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality* (2012), IEEE, pp. 119–128.

[111] Mashita, T., Yasuhara, H., Plopski, A., Kiyokawa, K., and

Takemura, H. Parallel Lighting and Reflectance Estimation based on Inverse Rendering. In *In Proceedings of the International Conference on Artificial Reality and Telexistence* (Tokyo, Japan, December 2013), pp. 102–107.

[112] Plopski, A., Mashita, T., Kiyokawa, K., and Takemura, H. Reflectance and Light Source Estimation for Indoor AR Applications. In *In Proceedings of the IEEE Virtual Reality* (Mineapolis, USA, March 2014), pp. 103–104.

[113] Kán, P., and Kaufmann, H. High-Quality Reflections, Refractions, and Caustics in Augmented reality and Their Contribution to Visual Coherence. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality* (2012), IEEE, pp. 99–108.

[114] Sugano, N., Kato, H., and Tachibana, K. The Effects of Shadow Representation of Virtual Objects in Augmented Reality. In *Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality* (2003), IEEE, pp. 76–83.

[115] Rompapas, D. C., Rovira, A., Plopski, A., Sandor, C., Takafumi-Taketomi, Goshiro, Y., Kato, H., and Ikeda, S. EyeAR: Refocusable Augmented Reality Content through Eye Measurements. *Multimodal Technologies and Interaction 22*, 4 (2017), 22:1–22:18.

[116] Hanington, B., and Martin, B. Universal methods of design: 100 ways to research complex problems. *Develop Innovative Ideas, and Design Effective Solutions: Rockport Publishers* (2012).

[117] Reitmayr, G., and Schmalstieg, D. Mobile collaborative augmented reality. In *Proceedings IEEE and ACM International Symposium on Augmented Reality* (2001), IEEE, pp. 114–123.

[118] Regenbrecht, H. T., and Wagner, M. T. Interaction in a collaborative augmented reality environment. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems* (2002), ACM, pp. 504–505.

[119] Reitmayr, G., and Schmalstieg, D. Scalable techniques for collab-

orative outdoor augmented reality. In *3rd IEEE and ACM international symposium on mixed and augmented reality (ISMAR' 04), Arlington* (2004).

[120] Miyake, M., Fukuda, T., Yabuki, N., Motamedi, A., and Michikawa, T. Outdoor augmented reality using optical see-through hmd system for visualizing building information. In *16th International Conference on Computing in Civil and Building Engineering* (2016), pp. 1644–1651.

[121] Kiyokawa, K., Billinghurst, M., Campbell, B., and Woods, E. An Occlusion-Capable Optical See-Through Head Mount Display for Supporting Co-Located Collaboration. In *Proceedings of the IEEE/ACM International Symposium on Mixed and Augmented Reality* (2003), IEEE Computer Society, p. 133.

[122] Presselite Firefighter 360. https://www.fastcompany.com/1451123/firefighter-360-augmented-reality-game-its-hot, Last accessed April 2019

[123] Kuznietsov Vadym Mosquito Hunter

[124] Barakonyi, I., Fahmy, T., and Schmalstieg, D. Remote collaboration using augmented reality videoconferencing. In *Proceedings of Graphics interface 2004* (2004), Canadian Human-Computer Communications Society, pp. 89–96.

[125] Adams, E. *Fundamentals of game design.* Pearson Education, 2014.

[126] 6d.ai – cloud based ar, 2018. Last accessed: July 9, 2019.

[127] Immersal – cloud based ar, 2018. Last accessed: July 9, 2019.