Doctoral Dissertation

Context integration and geometrically correct rendering on Video See-through Augmented Reality displays

Geert Lugtenberg

Program of Information Science and Engineering
Graduate School of Sciene and Technology
Nara Institute of Science and Technology

Supervisor: Hirokazu Kato
Interactive Media Design Lab. (Division of Information Science)

Submitted on September 3, 2024

A Doctoral Dissertation
submitted to Graduate School of Science and Technology,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of Engineering

Geert Lugtenberg

Thesis Committee:

Supervisor	Hirokazu Kato (Professor, Division of Information Science)
	Kiyoshi Kiyokawa (Professor, Division of Information Science)
	Masayuki Kanbara (Affiliate Professor, Division of Information Science)
	Yuichiro Fujimoto (Affiliate Associate Professor, Division of Information Science)
	Taishi Sawabe (Assistant Professor, Division of Information Science)

Context integration and geometrically correct rendering on Video See-through Augmented Reality displays[*]

Geert Lugtenberg

Abstract

When we envision performing tasks supported by augmented imagery, display devices that are worn on the head often come to mind. However, with the advent of powerful consumer devices such as smartphones and tablets, it has become increasingly common to use these devices for displaying augmented instructions. They are flexible, accessible, easy to use, and do not cause the fatigue or visual discomfort associated with head-worn displays. This makes such *magic lens* (ML) displays an excellent candidate for training or simulation of quick, close-range operations.

The challenge with video see-through ML displays as a medium for augmented imagery is that they do not present the physical world in a natural way, which complicates interaction with the physical environment seen through the display. This issue also hinders the effective integration of the ML display with the real environment. Achieving a natural viewing experience involves making the display appear transparent.

In this dissertation, I present two technical contributions towards this goal: first, I propose a prototype for geometrically correct rendering on off-the-shelf devices. This prototype tracks the user's eye to align the view on the ML display with the surrounding environment, making it appear transparent and restoring motion parallax. Second, I enable stereoscopic and varifocal capabilities on the ML display to match its vergence and accommodation distances with those of the

[*] Doctoral Dissertation, Graduate School of Science and Technology,
Nara Institute of Science and Technology, September 3, 2024.

surrounding environment. In two user studies, I investigated how these prototypes perform regarding spatial awareness and context integration.

In the first study, I found that geometric correction immediately improves the accuracy of haptic interactions, particularly in scenarios where depth information from the surrounding visual context or tactile feedback was absent. In conventional ML displays (without geometric correction), a learning effect on depth accuracy was observed, indicating that the prototype display is particularly beneficial for tasks requiring immediate precision.

In the second study, I employed two viewing strategies to integrate the ML with its surroundings: rapid switching and viewing them as a cohesive whole. In a visual-acuity experiment, I found that minimizing the accommodation difference between the ML display and its surroundings is crucial for rapid gaze shifting, whereas minimizing the vergence distance is more important when viewing the ML and its surroundings as a single context. Conflicting vergence and accommodation distances did not significantly affect cognitive task load nor did they play a pivotal role in the accuracy of context integration.

The results of the two studies have several implications for using ML displays with AR to enhance close-range tasks. First, they demonstrate that off-the-shelf devices can facilitate more natural viewing experiences using the framework proposed here, without the need for specialized AR hardware. Second, developers should focus on integrating user-perspective rendering, maintaining environmental context, and including real-time visualization of hand-guided tools or hands for best performance. Third, they can cause a physical obstruction and therefore best when handheld or moveable.

Keywords:

augmented reality, video see-through, human-computer interaction, perspective, computer graphics, optics

Contents

List of Tables

List of Figures

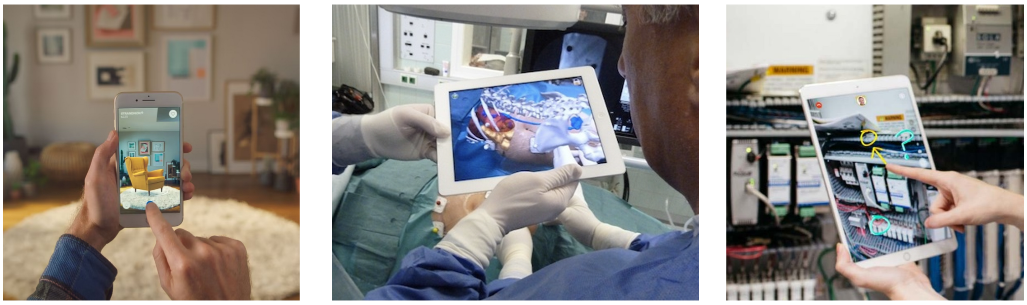
1

Introduction

The integration of Augmented Reality (AR) into various domains has revolutionized how users perceive and interact with digital information superimposed on their physical environment. By addressing the challenges in achieving true transparency and accurate contextual rendering in video see-through AR displays, this work aims to advance the capabilities and applications of AR technology on accessible and commonplace devices. This introductory chapter provides an overview of the background and motivation for the research presented in this dissertation, outlines the problem statement and the scope of the studies to come, and describes the contributions of this dissertation.

1.1 Background and Motivation

AR technology has rapidly evolved, reaching a popularity comparable to that of Virtual Reality (VR). A significant factor contributing to this surge is the ease with which augmented imagery can now be visualized on mobile devices such as smartphones and tablets. Since the mid-2000s, advancements in graphical processing units and the integration of onboard cameras in consumer devices have democratized AR, enabling virtually anyone to view and interact with augmented content in their real environment. These *video see-through* AR displays use a camera feed to overlay virtual objects onto real-world scenes and have found applications in such sectors as entertainment, healthcare, and industrial task support (see Fig. 1.1). The widespread availability and usability of these mobile devices make them an attractive platform for AR applications.



(a) (b) (c)

Figure 1.1: Examples of AR applications on video see-through handheld devices: (a) Home furnishing with IKEA Place, (b) AR-guided abdominal surgery [1], (c) AR annotations with Vuforia Chalk.

In the context of this dissertation, the term ‘Magic-lens displays’ (ML displays) will refer to video see-through AR devices that are either handheld or spatially fixed, distinguishing them from head-worn AR devices. ML displays act as a dynamic ‘lens,’ allowing users to view augmented digital content superimposed on the real world, thus aiming to facilitate a seamless integration of virtual and physical elements.

Despite their potential, ML displays face several significant challenges, particularly in achieving a true transparent effect, providing accurate context integration, and ensuring geometrically correct rendering. Therefore, they often fall short of providing adequate spatial understanding and depth cues, which are crucial for *effective interaction*. Moreover, it is not known how much each limiting factor affects effective context integration and performance. This dissertation addresses these gaps by leveraging user-perspective rendering and novel display technologies. As depth cameras and display technologies continue to advance, there is an opportunity to refine ML displays further to deliver more accurate, contextually integrated, and user-comfortable AR experiences. This advancement is essential for applications where precision, depth perception, and visual comfort are paramount, such as in task support systems and detailed visualizations in fields like healthcare and industrial training.

This dissertation aims to address these challenges by investigating how user-perspective rendering can enhance the spatial understanding and usability of ML displays, and how human ocular systems, specifically eye vergence and accommodation, impact the effectiveness of integrating AR content with the surrounding context.

1.2 Problem statement and Scope

This dissertation addresses the fundamental challenge of achieving true transparency in ML displays for AR. Traditional ML displays typically present a monocular, device-perspective view of the physical environment. A truly transparent magic lens, however, should align more closely with natural human vision, incorporating the following aspects:

1. Geometric correction: The display should correctly render the physical environment from the user's point of view.

2. Stereoscopic and varifocal capabilities: The display should support natural depth perception and focus adjustments as seen from the user's perspective.

These problems are shown in Fig. 1.2. In the *traditional ML display* setup (a), a video image captured by a camera on the back of the display is rendered resulting in a magnified view. Both eyes focus on the physical screen and are presented with the same image. When the user's perspective changes (second row), the image on the ML display stays the same. The distorted view and missing parallax depth cues from changing perspectives decrease spatial awareness.

In a true transparent ML display (Fig. 1.2b), there is a seamless transition from the physical environment to the screen. The eyes focus on the content (e.g. the red ball) rather than the physical screen, and each eye is presented with a stereo disparity image. Upon perspective change (second row), the image on the ML display changes, and depth cues from motion parallax become clear.

Solving these problems is crucial for enhancing task performance when using ML displays, particularly concerning *context integration* and *spatial awareness.*

1.2.1 Stereoscopic and Varifocal Vision

In most cases, the physical ML display does not align with the intended distance of its virtual content or the actual distance of the physical environment. This

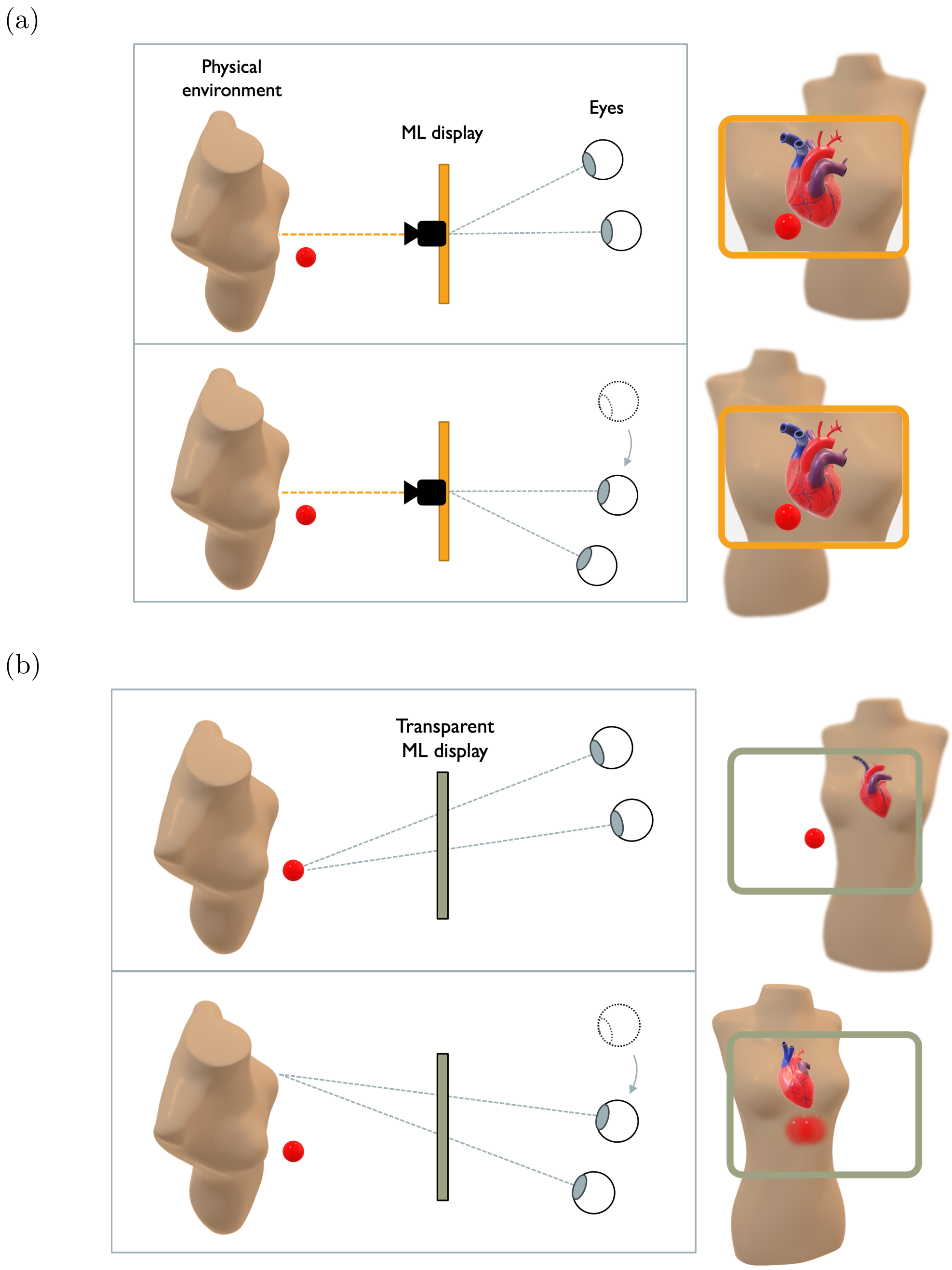


Figure 1.2: Problems in ML displays regarding perspective and the human ocular system. Left column: Schematic top-down view of the display technology. Right column: User's point of view. (a) Traditional ML display, (b) Transparent ML display.

misalignment causes a discrepancy that results in a blurry image due to mismatched eye accommodation and double vision due to misaligned eye vergence when viewing either the ML display or the physical environment (see Fig. 1.2). These issues complicate the integration of ML content with the surrounding real-world context.

A practical approach to mitigate double vision is to render content stereoscopically on the ML display, allowing the eyes to converge at the same distance as the surrounding physical context. However, this creates a new problem: the convergence distance is likely different from the focal distance required for clear vision. This well-documented phenomenon, known as the vergence-accommodation conflict, negatively impacts depth perception and visual comfort, particularly in head-worn devices.

To reduce blurry vision during context integration, the ML display would need to match the focal length of the physical environment (Fig. 1.2b), a task that is technically more challenging than stereoscopic rendering and is currently impractical with existing ML technologies. Therefore, this dissertation investigates the effects of blurriness, double vision, and the vergence-accommodation conflict on the user's ability to effectively merge ML display content with its surrounding environment, providing guidelines for future displays with varifocal capabilities.

1.2.2 Geometric Correction for User-Perspective Rendering

Achieving geometric accuracy requires an accurate reconstruction of the physical environment from the user's perspective, which often differs from the device's perspective. The main challenges include:

- Scene Reconstruction: Continuously updating the model of the physical environment as it changes.
- Texture Mapping: Correctly applying textures to the reconstructed model to mimic the real environment.

This dissertation examines the effects of user-perspective rendering on haptic depth accuracy in a *static scene* using an *interactive hand-guided tool.* We also propose a solution for reconstructing non-static (dynamic) scenes and real-time geometrically correct visualizations of the user's hands.

1.2.3 Research questions

While ML displays have the potential to transform user interactions with augmented environments, several unresolved issues hinder their effectiveness. The core problem addressed in this dissertation is the challenge of achieving true transparency.

Specifically, the following questions guide the research:

1. **User-Perspective Rendering:** How can we develop and implement a user-perspective rendering method on off-the-shelf video see-through devices (MLs) to create a truly transparent augmented view?

2. **Depth Accuracy in Close-Range Tasks:** How does user-perspective rendering impact depth accuracy in haptic interaction tasks viewed through the AR ML display?

3. **Integration with Surrounding Context:** What are the impacts of eye vergence and accommodation on the integration of the ML display with its surrounding context, and how do the artifacts from these physiological factors affect visual acuity?

1.3 Contributions

This dissertation makes two significant contributions to the field of augmented reality and magic-lens displays:

1. Improvement of spatial awareness using ML displays:
   - **Geometrically-correct ML prototype:** We developed a user-perspective rendering method tailored for off-the-shelf video see-through devices such as smartphones and tablets. This method operates without the need for specialized setups or additional hardware, leveraging existing mobile technology to deliver advanced AR experiences that are both accessible and practical.

- **User study and guidelines:** The prototype was employed in a user study to extend previous research by comparing performance with visual feedback of the interaction medium. The results of this study provide new insights into depth accuracy under varying conditions of perspective rendering, haptic feedback, and visual context awareness. Based on this we propose guidelines and recommendations for using ML displays in arm's length tasks.

2. Improvement of context integration using ML displays:
   - **Stereoscopic varifocal ML prototype:** We introduced a novel ML display capable of presenting content at variable vergence and accommodation distances, addressing the challenge of aligning virtual content with the user's natural vision.
   - **User study and guidelines:** A subsequent user study was conducted to evaluate context integration performance using this prototype. The study's findings provided guidelines for enhancing ML display technology, particularly in terms of integrating virtual content seamlessly with the surrounding physical environment.

These contributions advance the capabilities of ML displays by enhancing visual and haptic interaction and improving the integration of virtual content with real-world contexts, addressing critical challenges in current AR technology. We hope that our contributions will increase the popularity of AR in close-range tasks because low-cost, off-the-shelf devices could be effectively used for visualization, software execution, and sensor input.

1.4 Outline of this dissertation

This dissertation consists of five chapters. In this first chapter, we introduced the background of augmented imagery on ML displays and identified two problems in the state-of-the-art. We also defined the research questions, scope, and contributions. Chapter 2 will review AR displays and applications, identifying their limitations and gaps in the literature. Chapter 3 will focus on the first problem

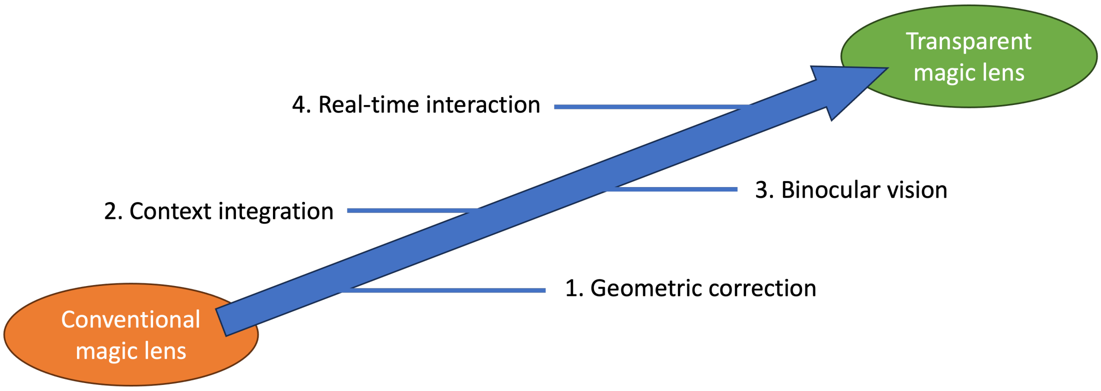


Figure 1.3: Planning towards a true transparent magic lens over the course of the doctoral dissertation.

of geometric corrections on the ML display, guided by RQ.1 and RQ.2. A prototype is proposed and evaluated in a user study. In Chapter 4, we focus on the second problem: unnatural eye vergence and accommodation and their artifacts, guided by RQ.3. We further improve our transparent ML prototype with stereoscopic and varifocal capabilities and evaluate it in two user studies. The results of the two topic studies (Chapter 3 and Chapter 4) are then combined in Chapter 5 where we will discuss their implications. In this chapter, we also propose the next steps for the transparent display prototype. Finally, we summarize our research findings and contributions. We describe its limitations, make recommendations, and reflect on the lessons learned.

2 Literature Review

In the upcoming chapters, we will deep-dive into spatial awareness (Chapter 3) and context integration (Chapter 4) using AR ML displays. These studies each present their topic-specific related work. In this chapter, we discuss the literature encompassing both studies to provide a basis of knowledge.

2.1 Overview of Augmented Reality displays

For us to see virtual imagery, we need a display medium. Bimber and Raskar [7] defined visual AR displays by their position relative to the human body and their visualization technology, subdividing them into head-attached, hand-held, and spatial displays. For this dissertation, we slightly redefine these to displays within the head-periphery, within the arm periphery (at arm's length), and spatial displays (shown in Fig. 2.1). Head-periphery displays are worn by the user and include retinal-, near-eye-, and head-mounted displays. In some specialized systems a projector is head-attached to visualize imagery not (only) on a planar screen, but directly on the physical object. Similarly, such projector systems can be hand-held or mounted in the environment (spatial). Next, arm-periphery displays include hand-held devices such as smartphones and tablets, and displays that are integrated into the environment at arm's length, such as monitors or projector surfaces. Last are spatial displays that are beyond the periphery of the arm. We define arm-periphery- and spatial see-through AR displays as *magic lenses* (green in Fig. 2.1). All of these display types have their advantages and limitations, which are discussed hereafter.

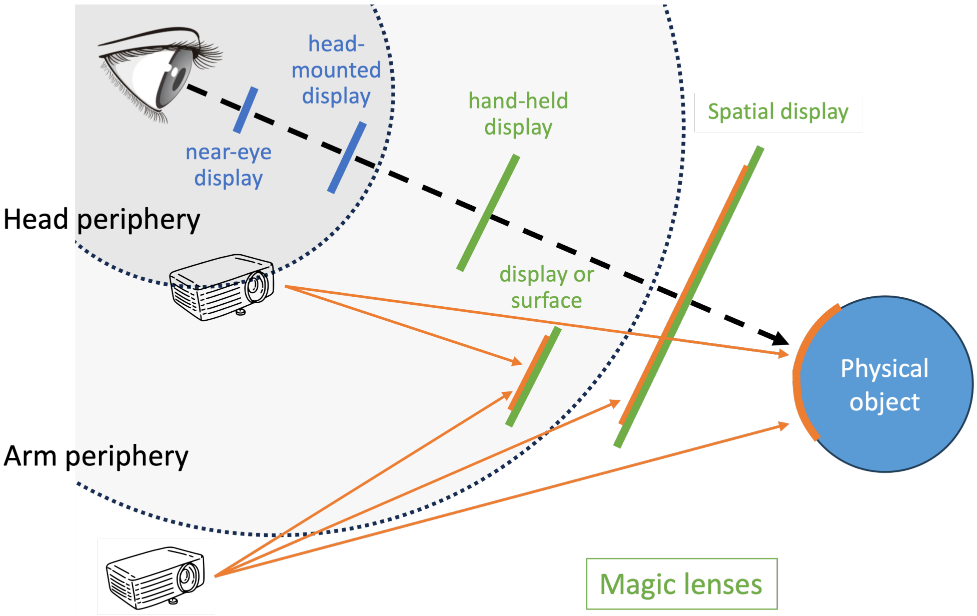


Figure 2.1: Types of AR display and their proximity to the human body. Displays in green are Magic lenses.

2.1.1 Head-mounted displays

HMDs are currently one of the most popular ways to display AR [8]. This is due to several key advantages:

- **Large field of view:** HMDs can have a relatively large field of view, increasing sense of immersion and interaction.

- **Natural interaction:** HMDs allow for natural head movements due to sensors on the device tracking its orientation, often incorporating additional sensors to allow hand interaction as well.

- **Movement freedom:** HMDs allow for hands-free interaction and relatively free movement without obstructing the user's immediate environment.

- **Stereoscopic vision:** HMDs incorporate stereoscopic vision by providing each eye with a separate screen or disparate graphics. This improves depth perception.

Commercial HMDs have been utilized in education and training [9], product design and manufacturing [10], and by healthcare and industry workers [11]. They differ in two key visualization technologies: *Video see-through* (e.g. Vision Pro, Gear VR, Pico 4, Vive Pro, Meta Quest) and *Optical see-through* (e.g. Magic Leap 1 & 2, most Xreal glasses, Hololens 1 & 2, Meta 2).

Video see-through HMDs

Video see-through (VST) HMDs block a user's view of the real world by positioning a video display centimeters in front of each eye, as part of the head-worn mount (see Fig. 2.2 right). Optical lenses are mounted in between so that a user can focus on the display. This type of display is often used in virtual reality (VR) since it can block the real world entirely to allow full control and immersion [12] into the virtual world. To enable see-through capabilities, cameras on the front of the HMD capture the real world to show on the eye displays. A major advantage of VST as an AR display is that the augmented imagery can completely occlude or replace the imagery representing the physical environment. This allows such techniques as X-ray [13] visualization (Fig. 2.3) and diminished reality [14] to add to the practicality of AR.

However, delays in visual feedback and conflicting multimodal feedback of movement cause *motion sickness* [15], adding to the difficulty in adopting this type of display to visualize AR. In recent years, VST HMDs became viable as AR displays as well, due to the latency of the video image representing the physical world decreasing into unnoticeable ranges. However, because of eye fatigue from conflicting focal distances (see Sect. 2.1.1 below), which is more prominent on VST than OST HMDs [16, 17], VST HMDs are still less popular for AR than OST HMDs.

Optical see-through HMDs

Optical see-through (OST) displays consist of a transparent medium to allow light from the physical environment and a display or projector to reach the eyes through

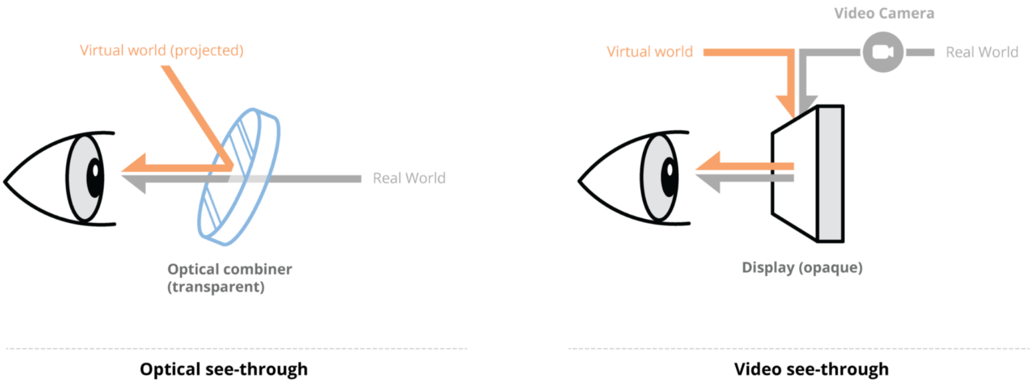
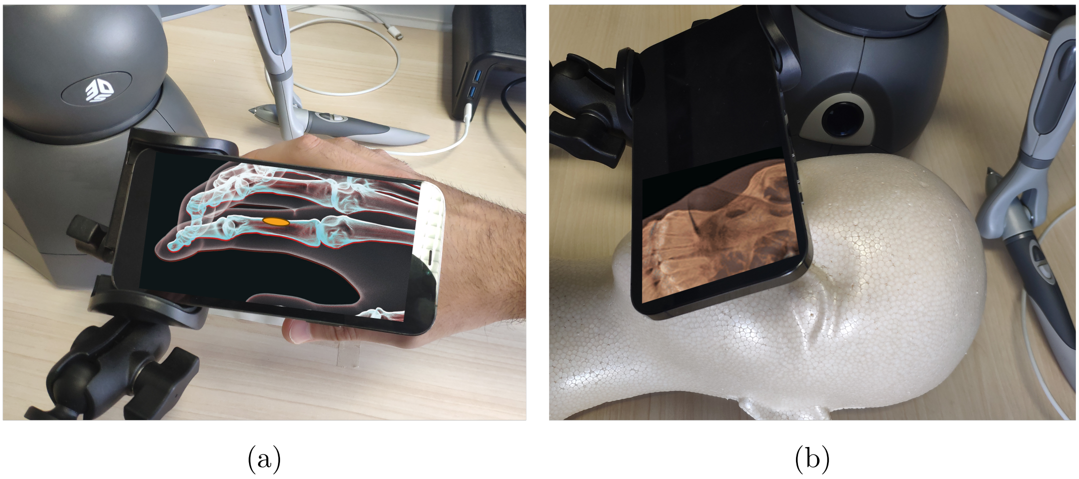


Figure 2.2: Display techniques of OST (left) and VST (right) displays. Image source: Niteesh Yadav [2].


(a)

(b)

Figure 2.3: X-ray visualization on a VST display of bones in (a) a human hand, and (b) a foam mannequin head.

an optical combiner, such as a half-mirror (see Fig. 2.2 left). This has the great advantage of a direct view of the real world; it does not have to be captured, processed, and then displayed on screens like VST displays. This means that, except for virtual content that is *added* directly onto this view of the real world, users have natural visual quality.

However, OST HMDs are impractical in scenarios where direct manipulation of video images is necessary, such as hiding real physical objects or manipulating their transparency (e.g. X-ray visualization in Fig. 2.3) [13, 18], limiting their practicality to only *adding* information to the real environment. Due to the nature of the optical combiner, virtual content on an OST display always has some transparency. In bright conditions, such as outdoors, the virtual imagery becomes even more transparent and difficult to see.

Vergence-Accommodation Conflict

Most HMDs pose discomfort to the wearer due to ergonomic factors such as weight and bulkiness, which can fatigue a user [19, 20], especially over long periods. Moreover, since HMDs present each eye with stereoscopic imagery at a fixed focal distance (see Fig. 2.4), eye vergence and accommodation distances often conflict, causing eye strain and fatigue [21, 22]. Numerous studies have focused on the fatigue and performance problems arising from the conflict between vergence and accommodation distances (VAC) [23]. This problem is particularly prevalent in mixed-reality contexts. Both VR and AR HMDs suffer from VAC [3, 24, 25, 26, 27, 28], which contributes to lower adoption rates in practical applications. Discomfort studies using HMDs [29] have shown that only focus-adjustable (varifocal) lens designs can accommodate simulated distances to improve comfort significantly. At current state-of-the-art, such HMDs are rare.

Other limitations

Many HMDs require peripherals to allow more sophisticated tangible input. To keep the user's hands free, instead, hand gestures or eye gaze are employed to interact with the system and allow input. However, the ease of use and precision of such input methods are much lower than tangible methods.

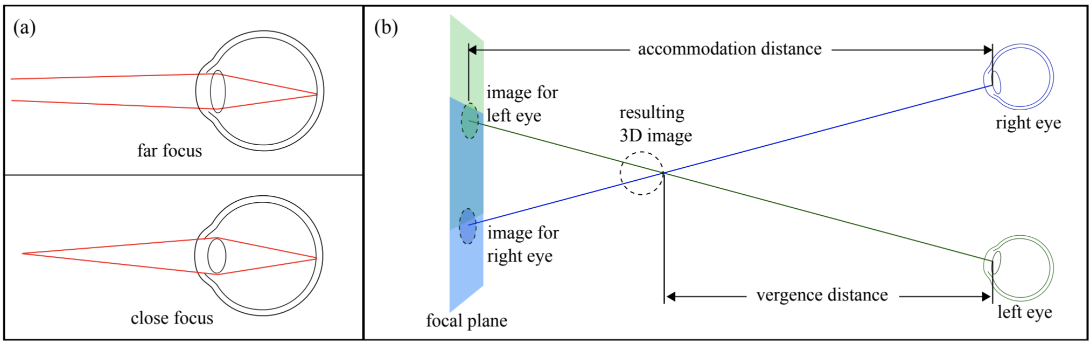
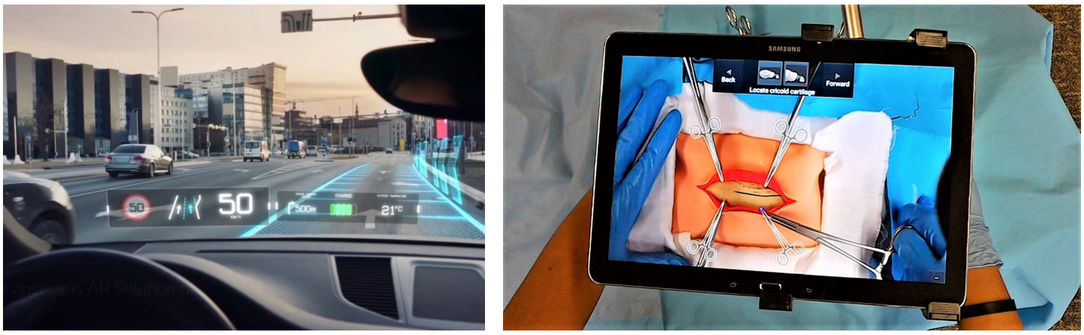


Figure 2.4: (a) Eye focus or accommodation in a far (light rays approximately parallel) and close (converging light rays) situation. (b) The focal plane is at a fixed distance, but the 3D image–created by stereoscopic disparity–is at a variable distance. This causes a conflict between the distances of accommodation (on the focal plane) and convergence (on the 3D image) termed VAC. Image source: Kramida [3].

HMDs can be impractical in some scenarios. They can be cumbersome and may get dislodged during vigorous activity, such as outdoor fieldwork with constant moving, climbing, or working in constrained spaces. They can also be a safety hazard if they *obstruct or limit vision* or get entangled with equipment, such as in medical procedures like surgery. Moreover, a surgeon would likely already wear other equipment on their head essential for the procedure, reducing the practicality of the HMD. Lastly, augmented imagery on the HMD is only visible to the wearer. Combined with the fact that the device occludes the wearer's eyes and part of the face, makes collaboration more difficult.

2.1.2 Handheld displays

With the advent of powerful graphical processors in smartphones and tablets, it has become possible and popular to use these handheld devices to visualize AR. This allowed developers to apply AR technology in many fields such as healthcare [30, 31], retail and e-commerce [32, 33], manufacturing [34] and maintenance [35], to name a few. They are significantly cheaper than HMDs and make AR accessible to common consumers. Handheld displays are usually small, light, and very portable. Due to their widespread availability, they are easy to use and



(a) (b)

Figure 2.5: (a) Head-up Display (HUD) projected on the windshield of a car. Image source: LG [4]. (b) Video see-through (VST) magic lens displaying augmented instructions for surgical telementoring. Image source: Andersen et al. [5].

adopt. For instance, they often have a tangible input interface, such as a touch screen or physical buttons, that make inputting information easy.

Limitations

Handheld displays require at least one hand free to hold and direct the device, limiting their practicality in interactive scenarios. However, if the device is light and small enough to be held in one hand, like a smartphone, the other hand could still interact and perform tasks with supported AR imagery. To our knowledge, only VST handheld off-the-shelf devices exist. They suffer from similar limitations as VST HMDs (Sect. 2.1.1). Surprisingly though, previous work on the existence or effect of VAC in handheld AR devices is lacking, motivating us to investigate this in Chapter 4.

2.1.3 Spatial displays

Other examples of magic lenses that are not handheld but in the spatial environment include head-up displays (HUD), used in the automotive [36, 37] (Fig. 2.5a) and air-flight industries, stationary displays [38] used in telementoring [5, 30] (Fig. 2.5b), and projector displays. Spatial projector displays, unlike the previously discussed display types, use the object surface or the physical environment to directly project augmented imagery onto. This has the benefit of a shared experience for collaborative interaction without any obstructing devices. They also do not suffer from the VAC, allowing for a comfortable experience no matter the usage duration.

Limitations

However, spatial displays are by definition not mobile. They have to be designed and set up with one augmentable task area in mind, which restricts their practicality and versatility. For this reason, they are often employed as interactive displays in museums [39] or entertainment. Furthermore, most projector displays utilized for AR are monocular, since both eyes view the same imagery projected on the surface. Binocular solutions exist, requiring glasses to filter the projected image based on the eye, usually through polarized light. In Chapter 4 we will utilize such a display in combination with active shutter glasses to obtain stereoscopy.

2.1.4 Magic-lens displays

The concept of a *magic lens* was first introduced by Bier et al. [40] which defined it as a see-through interface that could provide a customized view of the region behind a *lens area*. Magic lenses encompass both *hand-held* and *spatial display* types (Fig. 2.1), such as smartphones and tablets, or stationary displays mounted to the environment.

Previous studies on ML displays have predominately focused on handheld devices owing to their accessibility and technological advantages. Contemporary smartphones and tablets are equipped with a range of built-in sensors that are useful in AR applications. Often these VST ML displays are monoscopic, but there is research using stereoscopic MLs that use lenticular lenses to display disparity images [41, 42, 43, 44].

There is a difference between an ML that simply visualizes an on-device camera image as a background [40] (i.e. traditional AR ML displays) and an ML that visualizes geometrically correct views within the lens area, as seen from the user's

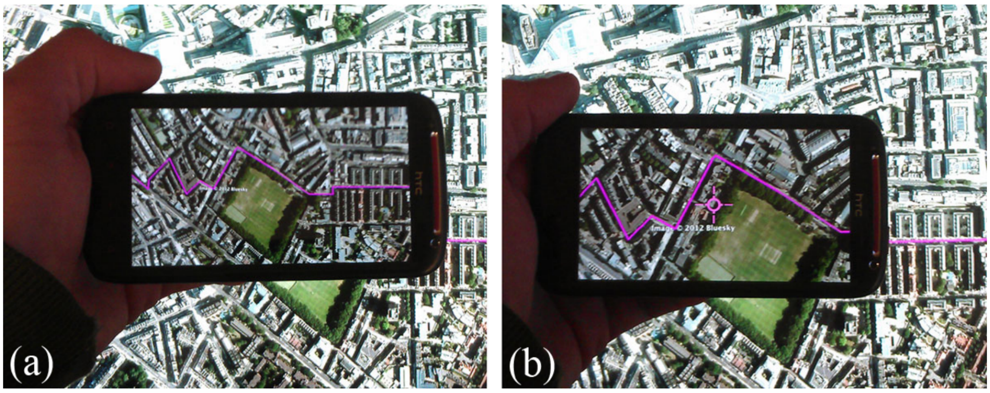


Figure 2.6: (a) Dual-view problem in video see-through MLs comes from the fact that the camera is positioned on the back of the device and thus has a different perspective than the observer. (b) When rendered from the perspective of the user, the dual-view is not present. Image source: Čopič Pucihar et al. [6].

perspective [6, 45, 46, 47]. See Fig. 2.6.

Limitations

ML displays that utilize handheld devices often offer a narrow field of view, dictated by the screen size. This limits the user's ability to perceive and interact with AR content within a broader context. User-perspective ML displays limit the field of view even further, though mitigations exist [47]. Čopič Pucihar et al. [48] investigated both conventional device-perspective rendering and user-perspective rendering and found that users consider the real environment and ML as separate views (Fig. 2.6) when using device-perspective rendering. This hindered context integration and affected performance.

2.2 Spatial understanding of content on AR displays

2.2.1 Depth perception

General theory [49] about human depth perception lists occlusion, binocular disparity, motion perspective, relative size, accommodation, and vergence as the most important depth cues for arm's length distances. A monocular ML display poses difficulties for users in accurately perceiving the depth of objects and the environment rendered on the display. This is due to several factors:

1. **Both eyes are presented with a singular view** (monoscopic rendering).
2. The camera is **not co-located with the eyes**.
3. **Eyes accommodate and verge on the display** while the display distance rarely matches the physical environment distance.

Whereas (1) and (2) are mitigated on HMDs by utilizing two cameras positioned at (almost) the same location as the respective eyes, and having a separate display per eye (stereoscopic rendering), (3) remains a shared issue for all types of see-through displays. OST displays may decrease the severity of all three issues by not having to re-render the real physical environment, but the issues persist for any virtual content [50, 51].

2.2.2 On ML displays

Čopič Pucihar et al. [6] and Baričević et al. [52] provided prototype user-perspective rendering ML displays and evaluated them in user studies investigating surrounding visual context [48, 53] and haptic interaction accuracy [6]. They found improved touch accuracy using user-perspective rendering, and in a virtual reality proof-of-concept also faster hand interactions. However, likely due to technological constraints at that time, they were not able to visualize the haptic interaction medium (i.e. hand) in AR, nor provide a *dynamic* user-perspective experience. Hueber et al. [47] developed the most recent user-perspective handheld ML at the time of writing, and in a task in which *virtual* object distances were *visually* *compared*, found that user perspective increased depth perception over traditional methods. Depth estimation by haptic interactions with physical objects or environment on AR ML displays remains uninvestigated.

2.3 Research gaps addressed by the studies

The related research presented in this chapter shows that there are gaps in the knowledge. We categorize the gaps as relating to spatial awareness and depth perception or relating to context integration, i.e. merging the magic lens with the physical environment. In the first study (Chapter 3), we evaluate depth accuracy with a magic-lens display and improve it by filling in the gaps in the knowledge. In the second study (Chapter 4) we investigate how human ocular systems affect integrating a magic lens with its surrounding context, evaluate four types of magic-lens displays, and pose recommendations of display type usage for different scenarios.

3

Effects of perspective, haptic feedback, and visual context on AR spatial understanding

Performing tasks at close range supported by augmented content or instructions visualized on a 2D display can be difficult due to missing visual information in the third dimension. This is due to the world on the screen being rendered from the perspective of a single camera, usually on the device itself. However, vision is supported by haptic feedback when performing tasks with our hands, and prior knowledge and visual context affect task performance as well. In this study, we re-enable depth cues from motion parallax by rendering the world on a display from the user's perspective and comparing it with the conventional device perspective during haptic interactions. We performed a user study consisting of 20 subjects and two experiments. In the first, touch point and depth estimation accuracy were measured under conditions of visual context and perspective rendering on a magic lens display. We found that user-perspective rendering slightly improves touch accuracy for targets on a physical surface but considerably so for interactions without tactile feedback. This effect is relatively larger without contextual information from the environment and diminishes as more haptic interactions occur. In a second experiment, we validated our results in a practical scenario of needle injection and confirmed that initial injections to virtual targets were more accurate using a user-perspective magic lens. The results suggest that in tasks with haptic interaction (that is, with hands or hand-guided tools), magic lenses with user-perspective rendering improve performance, especially when the physical environment frequently changes.

3.1 Introduction

Augmented Reality (AR) —enhancing reality by integrating virtual information— is known to benefit the execution of a myriad of tasks at arm's length. Visualizing relevant data, instructions, or visual cues onto objects within the user's immediate reach, aids efficiency in task fields such as assembly, repair, maintenance, and training. AR can simplify complex procedures and reduce the need to constantly move away from the task area to refer to user manuals, external devices, or other external consultations, improving productivity and accuracy.

A popular method of displaying virtual imagery is through the use of head-mounted displays (HMD). Typically worn on the user's head and covering their eyes, these types of displays have the advantage of keeping the user's hands free while providing virtual imagery in their natural field of view. However, at the current state of the art, these types of displays also have several drawbacks depending on the task context. Their input interface often has limited to no haptic interaction which can provide a steep learning curve to users accustomed to inputting information through touch screens, buttons, and controllers. Prolonged use of HMDs can lead to fatigue due to the weight of the hardware and ergonomics, or due to a phenomenon called vergence-accommodation conflict [23] where the focal distance of the displayed virtual imagery conflicts with the distance at which the eyes convergence, causing eye strain. Furthermore, it is necessary to have high-performance hardware inside the HMD due to the computational cost of tracking and low latency visualization of virtual content, and in the case of video see-through HMDs also visualization of the real environment. These requirements make HMDs generally expensive and lower their availability and appeal to the common consumer.

An affordable and readily available alternative way to visualize augmented graphics is through off-the-shelf devices such as smartphones and tablets. In recent years these devices are commonly equipped with dedicated graphical processing units, high-quality cameras, and a myriad of other sensors that aid in tracking where the device is in relation to its surroundings. This allows such handheld devices to visualize virtual imagery together with the camera image of the physical world as if the device screen were transparent. Such a video see-through display is referred to as a Magic lens (ML) display [40]. ML displays also

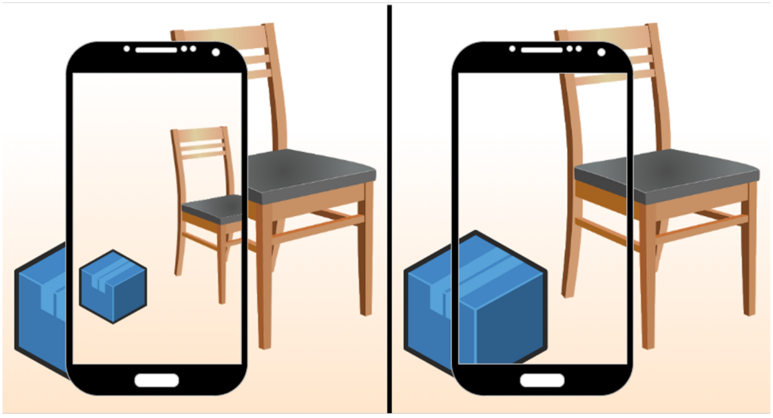


Figure 3.1: Visualization of augmented reality on a magic lens display using two methods: (Left) Device-perspective rendering. (Right) User-perspective rendering.

have practical downsides that can be significant depending on their task usage. They have to be held in hand which limits the user's freedom in hand tasks, making tasks that require two hands difficult and divides the attention of the user between positioning the ML and their task. To solve this, ML displays can be mounted in place to free up the user's hands, at the cost of mobility and some movement restrictions in the task workspace.

3.1.1 Problem description

Conventional fixed MLs visualize a view of the real environment as seen from a device camera, often positioned at the back of the display (see Fig. 3.1 left). This distorted view is in contrast to what the user would expect to see when the display is optically transparent (Fig. 3.1 right). We identify three problems of conventional mounted ML displays that affect spatial awareness and depth perception:

1. A magnified, distorted view of the real environment.

2. Missing depth information from motion/perspective parallax.

3. Missing depth information from binocular disparity.

In this chapter, we propose a solution to the first two problems by enabling *user-perspective rendering* of the ML display. Instead of the view frustum from the perspective of the device's camera, device-perspective (DP) rendering, we render it from the perspective of the user's (dominant) eye; user-perspective (UP) rendering, as shown in Fig. 3.1. This allows the user to regain a sense of depth by moving their head and looking around, through motion parallax. UP rendering also removes the magnification of the ML display area in relation to the surrounding physical environment, resulting in an apparently transparent lens. Mitigating problem 3 requires the user's two eyes to see two disparate UP views on the ML display. There are autostereoscopic displays that can present different graphics to the viewer's eyes based on their incident angle, which we will discuss in Chapter 5. However, combining the technology presented in this chapter will already relieve the effects of monocular vision drawbacks even without stereoscopy.

3.1.2 Motivation and Research Questions

The removal of depth cues from motion parallax as well as the distorted view is expected to lower the accuracy of hand-eye coordinated tasks that the ML display could otherwise support with augmented information. Previous studies [6, 48] investigated targeted touch accuracy using handheld ML displays with geometrically correct visualization of the real environment. However, these studies did not consider how touch accuracy might be influenced by proprioceptive depth cues from the hand (either holding the ML or performing the task) or visual depth cues from seeing the moving hand on the ML display. Furthermore, these studies only measured the accuracy of touching points on a planar, static physical surface, even though spatial awareness is known to be influenced by cognitive processes, particularly through visual context [54, 55]. In more common scenarios, users would likely interact with three-dimensional geometries and environments, and intangible AR content, where depth perception becomes crucial. In these dynamic and complex settings, users need accurate depth perception to judge distances correctly and interact effectively with virtual objects in the real world. However, the accuracy of depth perception while using ML displays has not yet been adequately measured.

To address these gaps in the previous research, we pose three questions:

- **RQ1** How does the proprioceptive and visual feedback of the users' hands affect targeted touch accuracy?
- **RQ2** How do perspective rendering and surrounding visual contextual information affect depth perception of AR targets in a changing environment?
- **RQ3** Does UP rendering offer a larger benefit when locating virtual target points in 3D space, i.e. without tactile feedback?

3.1.3 Contributions

We extend the methodology proposed by Čopič Pucihar et al. [6] by incorporating accurate hand-guided tool visualization and unconstrained UP viewing and testing them in a broader framework of *visual context* and *depth perception*. We demonstrated that incorporating visual feedback of the hand-guided tool significantly enhances touch accuracy, with an increase of 71% (1.15 cm) using DP and 47% (0.40 cm) using UP compared to the previous methodology. Without tactile feedback, we established baseline depth estimation accuracies of 2.1 cm for DP and 1.9 cm for UP on stationary ML displays. Notably, UP accuracy remained stable even without contextual information from the physical environment and was comparable to binocular performance observed in a similar study [56]. These findings indicate that UP ML displays, even without stereoscopy, are beneficial in scenarios where spatial awareness relies primarily on visual cues, such as guiding users to virtual targets or in the absence of contextual information from the real environment. We further validated these results in a simulated AR-guided needle injection task, where UP demonstrated a 38% improvement in accuracy (0.81 cm) for reaching virtual targets in initial attempts. These findings underscore the potential of monocular devices like smartphones and tablets as effective AR displays, offering spatial awareness similar to more sophisticated HMDs through perspective rendering. This advance makes AR guidance more accessible, especially for hands-free training or simulating physical tasks at arm's length.

In summary, this study contributes:

- A user-perspective rendering prototype tailored for off-the-shelf video see-through devices such as smartphones and tablets. The proposed prototype operates without the need for additional hardware.

- An empirical user study comprising two experiments, demonstrating the UP rendering prototype to improve the performance of locating target points compared to the conventional method.
- An experiment extending previous research demonstrating the effectiveness of mounted ML displays and proving that providing visual feedback of the interaction medium enhances targeted touch accuracy.
- An experiment showcasing a practical scenario of AR-guided needle injection that validates our results and further shows UP rendering to decrease injection attempts.
- Design recommendations for AR-guided tasks using ML displays to improve haptic interaction accuracy.

3.2 Related Work

Video see-through technology that allows for the experience of augmented content has been a rapidly evolving field since the advent of smartphones equipped with an onboard camera and graphical processing unit. This allowed developers to apply AR technology in many fields such as healthcare [30, 31], retail and e-commerce [32, 33], manufacturing [34] and maintenance [35], to name a few. Difficulty in understanding and matching the distorted visualization of the surroundings rendered on the video see-through device with the real physical surroundings has spawned investigations into its causes and possible solutions. Specifically, in this section, we focus on previous work concerning geometrically-correct (UP) rendering and investigations into depth perception using such video see-through displays.

3.2.1 Magic lenses and user perspective

The concept of a *magic lens* was first introduced by Bier et al. [40] which defined it as a see-through interface that could provide a customized view of the region underneath a *lens area*. By this definition, a video see-through display that is not head-worn and augments the physical world in some way is also a magic lens [57]. These magic lenses can be either directed by a user, such as hand-held smartphones and tablets, or stationary and mounted to the environment.

The effect of geometrically-correct rendering on hand-held devices has been previously studied. Baričević et al. [53] simulated UP rendering of a magic lens device in virtual reality with various sizes of the display to compare performance in a selection task. They found higher performance when combining UP rendering with a large display (tablet-sized) than the conventional DP. In this and their subsequent works [52, 58] the authors provided a proof-of-concept prototype made of off-the-shelf devices to provide UP rendering capabilities to hand-held devices. These works showed that future commercially available hand-held devices are likely to be capable of rendering a natural view of the physical world, especially given additional technology such as stereoscopic displays. Subsequently [6] and [48] evaluated perceptual issues on hand-held devices and confirmed the existence of the “dual-view problem"; misalignment and/or incorrect scale of content on the device screen compared to the surrounding physical content. They found that users of an AR-enabled ML *expect* UP rendering when presented with a touch interaction task. It outperforms the conventional DP in touch accuracy and task completion time. In both studies however, the user’s hands (or hand-guided tools) were not visible on the ML display and they were restricted in movement of their head in relation to the ML device, due to hardware restrictions. Conflicting depth information from not seeing your hand together with the proprioceptive sense of where the hand should be is expected to affect performance, especially in 3-dimensional situations. Additionally, these previous works mainly focused on task surfaces that are consistently flat, motivating us to investigate 3-dimensional surfaces in combination with *visible* hand-interaction.

Stationary or environment-mounted magic lenses have the benefit of keeping the user’s hands free while being able to augment a predetermined working environment or region of interest behind them. Examples include head-up displays (HUD) used in automotive [36, 37] industry and stationary displays on a stand [38]. In these cases, environment or object depth information cannot be gathered by moving the ML display (motion parallax) and this is expected to further hinder spatial interaction. [59] investigated using a stationary ML to trace an object held in hand versus a hand-held ML tracing a static object. Contrary to the authors' expectations, users performed better using the stationary ML, potentially due to powerful proprioception depth cues from two hands. Another possible reason is the creation of a cognitive map [54, 55] by fusing all our senses, that aids multi-modal interaction even when one of our senses is disrupted (e.g. finding a keyhole in the dark). A similar study [60] investigated the manipulation of a virtual object held in hand, visualized on a UP ML, and found that UP rendering significantly decreased task time compared to conventional camera perspective. However, to our knowledge, the effect of perspective with stationary magic lenses on an object not held in hand or on the physical environment has not been studied yet, especially concerning the existence and quality of a cognitive map.

A downside of stationary magic lenses is that the display can physically obstruct the working area which is expected to hinder haptic interactions. Related work has mitigated this problem by creating a *tool-mounted* interface, to support spine surgery [61] and visualize a planned trajectory on a surgical drill [62]. The latter found that such a display yields more accurate tool placement than the conventional guidance on a separate monitor. However, in an AR condition, they found no significant improvement due to the lack of surplus information. While UP rendering is expected to perform better on tool-mounted ML displays, such a ML would have to be designed for each tool, limiting the benefits of using off-the-shelf devices.

3.2.2 Depth perception on video see-through displays

General theory [49] about human depth perception lists occlusion, binocular disparity, motion perspective, relative size, accommodation, and vergence as the most important depth cues for arm's length distances. A monocular ML display poses difficulties for users in accurately perceiving the depth of objects and the environment rendered on the display. This is due to several factors: (1) *Presenting both eyes with a singular view* (monoscopic rendering) from the perspective of (2) *a camera that is not co-located with the eyes.* (3) *Eyes accommodate and verge on the display* while the display distance rarely matches the physical environment distance. Whereas (1) and (2) are easily mitigated on HMDs by utilizing two cameras positioned (almost) at the same location as the respective eyes and having a separate display per eye (stereoscopic rendering), (3) remains a shared issue for all types of see-through displays. Optical see-through displays may decrease the severity of all three issues by not having to re-render the real physical environment, but the issues persist for any virtual content [50, 51]. In a featureless black environment, Gao et al. [56] found that patients reached, on average, 2.8 cm short under monocular and 2.11 cm short under binocular conditions without haptic feedback or prior calibration. Similarly, Swan et al. [63] reported that users initially undershot by ∼4 cm using an optical see-through HMD but improved with proprioceptive and visual feedback. Issue (3) also negatively affects depth perception [23] as both eye mechanisms work together to provide powerful depth cues.

Some works [42, 64] have proposed a ML display utilizing stereoscopic rendering capabilities, restoring binocular disparity thereby solving issue (1). Their results on depth accuracy are however inconclusive or show that it is more dependent on monocular depth cues such as occlusion, motion parallax, and shadows. Similarly, Kerber et al. [65] found no significant effect of using an autostereoscopic display over a monoscopic display in a 2AFC depth task. For these reasons, we do not investigate stereoscopic displays further in this work, though it should be noted that binocular disparity depth cues play a larger role as objects get closer to the viewer [49] (see future work in section 3.9.1).

3.2.3 Magic lenses versus head-mounted displays

Most HMDs pose discomfort to the wearer due to ergonomic factors such as weight and bulkiness, which can fatigue a user [19, 20], especially over long periods. Moreover, since HMDs present each eye with stereoscopic imagery at a fixed focal distance, eye vergence and accommodation distances (issue 3) often conflict causing eye strain and fatigue [21, 22]. This further adds to VR sickness, which is more prominent in video see-through (VST) than in optical see-through (OST) HMDs [16, 17]. However, OST HMDs are impractical in scenarios where direct manipulation of video images is necessary, such as hiding real physical objects or manipulating their transparency (e.g. X-ray visualization) [13, 18], limiting their practicality to only *adding* information to the real environment. This suggests that AR magic lenses warrant further investigation, as they are less likely to induce the vergence-accommodation conflict [66] and are not known to induce VR sickness. Furthermore, smartphone-based (ML) AR has demonstrated comparable performance to HMD-based AR in percutaneous needle insertion tasks [67, 68], showing a significant improvement over conventional visualization methods that use CT-scan images on a separate display. This motivates us to employ the prototype proposed in this work in a needle injection task (see Sect. 3.6). Finally, while research has been conducted on AR/VR depth perception with HMDs [69, 70], studies exploring depth perception using ML displays remain limited and warrant further investigation.

3.3 User-perspective rendering prototype

There are several ways to accomplish rendering a geometrically-correct view of the environment through a ML display, as seen in Fig. 3.1 (right), that differ in their method for tracking the user's eye position relative to the display. Outside-in methods rely on hardware such as cameras that are (statically) positioned in the environment and aimed at the user's head. Inside-out methods affix hardware, such as cameras, to the user's head, or utilize glasses equipped with environment-facing cameras. Both methods have their upsides and downsides, based on the specific context of the task performed. Because of our motivation to use off-the-shelf devices, we decided not to track the user's eye position with external hardware.

We chose an iPad Pro (11-inch, 2nd generation) as our ML display and application platform. This choice was driven by the iPad's ability to simultaneously track its own pose and user facial features, by using Apple ARKit 5 in conjunction with Unity 3D (2011). ARKit is an SDK that leverages the device's sensors to provide AR experiences. We considered a phone-sized display, but the reduced augmentable workspace due to the smaller field-of-view was too restrictive, and UP rendering is shown to be more effective on larger-sized displays [53]. We utilize ARKit's Face Tracking functionality that uses the device's front-facing TrueDepth camera to detect face features, such as the user's eye positions relative to the camera. This has the additional benefit of naturally having a free, unoccluded, view of the user's eyes at any time that the user can see the device's display. We measured approximately 60° horizontal field of view of the TrueDepth camera for effective eye tracking. ARKit simultaneously tracks the device pose through Visual-inertial odometry (VIO) using the device's back-facing camera and inertia sensors, as depicted in Fig. 3.2 (top). The back-facing camera image is furthermore used in the software as the background image to create the AR experience. Uncorrected, this image represents the view as seen from the device's perspective (Fig. 3.1 left). For UP correction, we represent the user's dominant eye as a camera in Unity (denoted e in Fig. 3.2 bottom). Due to the monoscopic nature of our ML display, the same perspective imagery is viewed by both eyes and the sighting-dominant eye is suggested to be used for these monocular tasks [71]. In every frame, this eye camera position is set by the aforementioned Face Tracking system. We then calculate an off-axis perspective projection matrix of the eye camera as follows:

$$
\begin{aligned}
h\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} &= \begin{pmatrix} 2\dfrac{f}{W} & 0 & 2\dfrac{c_x}{W} \\ 0 & 2\dfrac{f}{H} & 2\dfrac{c_y}{H} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_e \\ Y_e \\ Z_e \end{pmatrix} \\
&= \begin{pmatrix} 2\dfrac{\vec{ea}\cdot i_n}{W} & 0 & -\dfrac{\vec{ea}\cdot i_r + \vec{eb}\cdot i_r}{W} \\ 0 & 2\dfrac{\vec{ea}\cdot i_n}{H} & -\dfrac{\vec{ea}\cdot i_u + \vec{ec}\cdot i_u}{H} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R & T \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}
\end{aligned} \tag{3.1}
$$

A schematic overview with the variables used in Equation 3.1 is depicted in Fig. 3.2 (bottom). The extrinsic matrix in the bottom-right of the equation consists of a rotation R and translation T from world space to back camera space (obtained from the VIO tracking framework) combined with a static transformation from back camera space to the image plane center (combine Fig. 3.2 top and bottom).

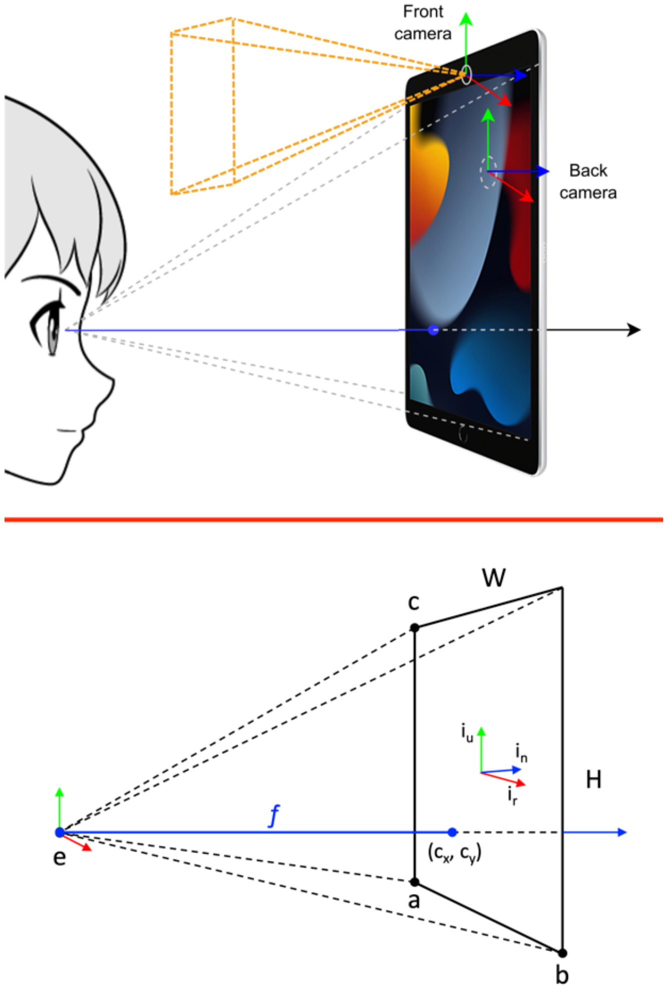


Figure 3.2: UP rendering prototype overview. (Top) The tablet display features a back-facing camera on the rear and a front-facing camera on the top edge. The user perspective view frustum (grey dotted lines) is calculated by tracking the user's eye position with the front camera (orange dotted lines). (Bottom) Schematic depiction of the image plane i and the eye camera e. The principal point c given e is often not centered on the image plane, resulting in an off-axis projection.

3.3.1 Foreground and background environments

The visible portion of the physical environment from the user's perspective often differs from what the device's camera captures. When the device camera functions as a ML display, DP rendering frequently presents a geometrically distorted view of the environment. The degree of distortion depends on the specific camera's intrinsic properties, the size of the display, and the distance ratio between eye-to-display and display-to-environment. When this ratio is less than 1 (when the viewed contents are further away from the display than the user's eye) the view is perceived as demagnified. This demagnification can be leveraged to achieve realistic graphics in the UP view by projecting the DP camera image onto the geometry of the environment in a process termed projective texture mapping. In instances where the user's viewing angle to the display is large or when the eye is very close to the display, the UP view reveals regions of the physical environment not visible in the DP. Graphics for these regions must be generated independently, as they are not based on the live DP camera image.

Projective texture mapping necessitates knowledge of the physical environment geometry by the 3D software, which is feasible for static environments. This geometry in the task area that remains unchanged is termed the background environment, and we capture and detail this background environment in a pre-processing step. Interactive elements, such as the user's hands, arms, tools, and any dynamic objects in the task area, are referred to as the foreground environment. To produce a realistic and accurate representation of the real environment, the foreground geometry needs to be highly detailed.

In our efforts to reconstruct the foreground interactive elements, we utilized depth cameras that are increasingly common in consumer handheld devices. By sampling every pixel on the depth image and representing them as small user-facing quads in the 3D software, we reconstructed and visualized an interactive physical environment in real time. However, this method exhibited some inaccuracy at the edges of the foreground geometry due to noise and the resolution and precision limitations of the depth camera. In a pilot study where participants used their index fingers to touch augmented target points, we observed that these edge cases required more accuracy than our real-time depth reconstruction method could achieve. Consequently, we opted to use a pointer tool with a known geometric shape and pose information available through tracking software and hardware. As shown in Fig. 3.7 and Fig. 3.10, this method allows for accurate and fast visualization of the haptic tool and touch-tip, while the user's hands are not rendered.

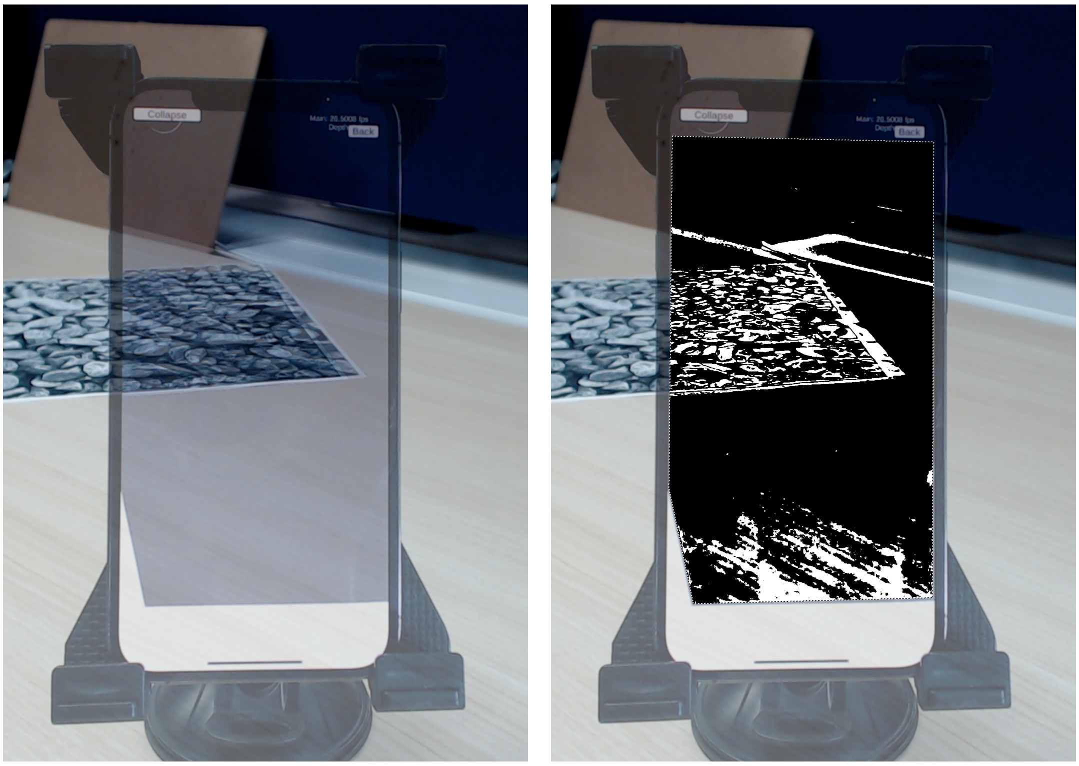
3.3.2 Prototype performance

At the core of the proposed user-perspective rendering algorithm is the detection of the user's eye position relative to the display. Small inaccuracies of the detected eye position have exaggerated effects on the rendered image on the ML display. Errors in eye detection in the camera's XY-plane result in an offset image on the ML display compared to its surroundings, whereas errors in the Z or depth axis result in an incorrectly scaled image, see Fig. 3.3.

A study by Nissen et al. [72] explored the accuracy of ARKit's technology in which they measured the distance from the user's eye to an iPhone X's TrueDepth sensor. Using identical technology and devices as proposed in our work, they found an eye-tracking positional error of approximately 4.8% at 30 cm away (M=1.45 cm, SD=0.11 cm). They found similar performance on other Apple devices with TrueDepth camera and ARKit technology. However, during testing, they maintained a fixed horizontal and vertical position of the eye relative to the sensor. Since positional errors in the XY-plane of the camera have a larger impact on user-perspective rendering, we evaluated our proposed prototype using a similar experimental design but additionally measured the accuracy of ARKit eye tracking with horizontal and vertical movements.

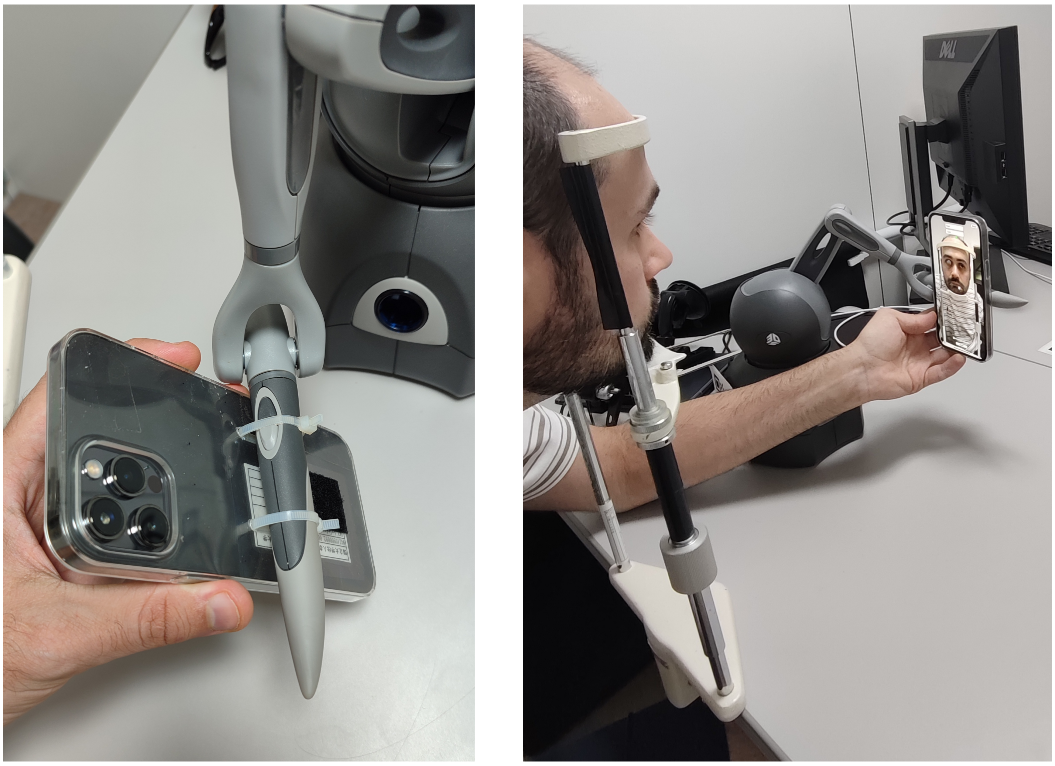
Test design and procedure

To evaluate eye tracking performance, we compare the camera's known movement distance to the detected eye movement distance in three dimensions. For our tests, we use an iPhone 12 Pro Max equipped with a TrueDepth front-facing camera and ARKit tracking technology to determine eye position. To accurately track the iPhone's position, we mounted it on the arm of a 3D Systems Touch [73] device, a haptic tool that provides precise 3D position data via the OpenHaptics 3.4 SDK running in Unity (2021). The setup is illustrated in Fig. 3.4.



(a) (b)

Figure 3.3: Result of the proposed user-perspective rendering prototype, displayed on an iPhone 12 Pro Max. (a) The ground truth image and UP rendering image are overlaid with 50% transparency. (b) The difference image between the ground truth and the UP rendering on the display. White regions denote pixel intensity difference above 20%.



(a) (b)

Figure 3.4: Setup for eye tracking performance testing. (a) An iPhone 12 Pro Max is affixed to the arm of a 3D Systems Touch haptic device. (b) The user's head is stabilized using a mount so that the real eye position does not change. The user's relative eye movement is then captured while the iPhone moves in 3 dimensions on the haptic arm.

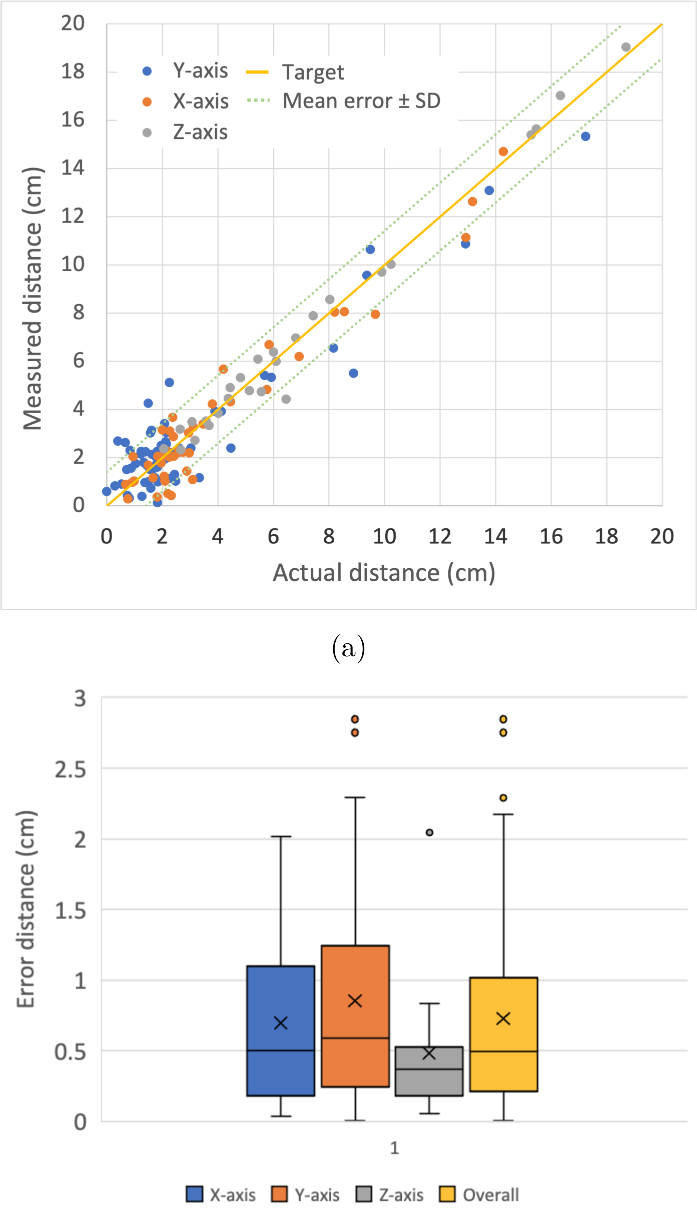
During the test, the iPhone mounted on the haptic arm was moved manually, ensuring movement was restricted to one of the three axes in the device's front-camera space (see Fig. 3.2, top). The device was moved distances ranging from 0.5 cm to 10 cm along the specified axis (X, Y, or Z), with 20 repetitions for each axis. For the X and Y-axis conditions, the device was kept 30 cm away from the eye. For each repetition, the difference between the start and end positions of the haptic arm tip (referred to as the *actual moved distance*) and the start and end positions of the user's right eye (referred to as the *measured moved distance*) were recorded. Data capture was triggered each time the button on the haptic device was pressed.

Results

The results of the eye-tracking accuracy test are shown in Fig. 3.5. The overall error is 0.72 cm ± 0.68 cm, representing a 2.42% error with the device positioned 30 cm away from the user's face. Eye position measurements in the XY-plane are less precise (M=0.78 cm) compared to measurements along the Z-axis (M=0.48 cm). Compared to the results reported by [72], we improved eye depth estimation by approximately 1 cm. However, our measurement method is less precise than theirs, as indicated by the difference in standard deviations (0.49 cm vs. 0.11 cm). The accuracy difference could further be attributed to advancements in Apple's eye-tracking technology and the variation in test devices (iPhone X vs. iPhone 12 Pro Max). The impact of the eye-tracking error on user-perspective rendering is illustrated in Fig. 3.3a, and is considered acceptable for tasks performed at arm's length.

3.4 User Study

A study was conducted to evaluate the proposed UP rendering and compare its performance to the conventional method for locating target points. The user study was divided into two repeated-measures experiments. In the first experiment, we assessed depth estimation and touch accuracy for targets on a physical tablet, which were only visible on a ML display. By mounting the ML display and visualizing a user-guided tool on it, we compared the results of the touch



(a)

(b)

Figure 3.5: Results of ARKit eye tracking accuracy test. (a) Moved (actual) distance vs. measured distance in the X-axis (orange), Y-axis (blue), and Z-axis (grey), with the mean error ± 1 SD shown as green dotted lines. (b) Error in the measured distance across axes and their mean (yellow).

task with those reported by Čopič Pucihar et al. [6] to address **RQ1**. We continuously altered the pose of the physical tablet while conditionally hiding it and the surrounding environment from view, allowing us to compare the depth estimation results under different perspective rendering conditions (UP and DP) and thereby address **RQ2** and partially **RQ3**. In the second experiment, users performed targeted needle injections into a physical mannequin head, relying solely on visual feedback of the targets inside the head. By comparing the number of injection attempts and their accuracy under perspective rendering conditions (UP and DP), we specifically addressed **RQ3**.

3.4.1 Hypotheses

When introducing a 2-dimensional display to support interaction tasks with augmented content, vision is (negatively) affected. While viewing the task area and objects through the 2D display, spatial depth perception is less accurate than it would be if there was no display. Given that our rendering method provides more visual depth cues than traditional 2D displays, we hypothesize that

> **(H1)** During haptic interactions with AR content on a magic lens display, UP rendering grants higher spatial accuracy than conventional DP rendering.

During close-range tasks, we frequently employ our hands or hand-guided tools to interact with objects or the environment. Given the prevalent literature concerning cognitive maps [54], we know such interaction is therefore a multi-modal process; guided by vision and adjusted by proprioception and tactile sensations. When our visual processes have fewer depth cues, as is the case using the conventional DP, it slows the creation, refinement, and accuracy of the resulting cognitive map. With continually changing environments or objects in the task area, this slow and less accurate creation of the cognitive map is expected to further decrease performance in hand interaction tasks. Building on the findings of **H1**, we hypothesize that

> **(H2)** The impact of UP rendering on a magic lens display is more pronounced while the physical environment is unknown and/or when tactile feedback of depth is absent.

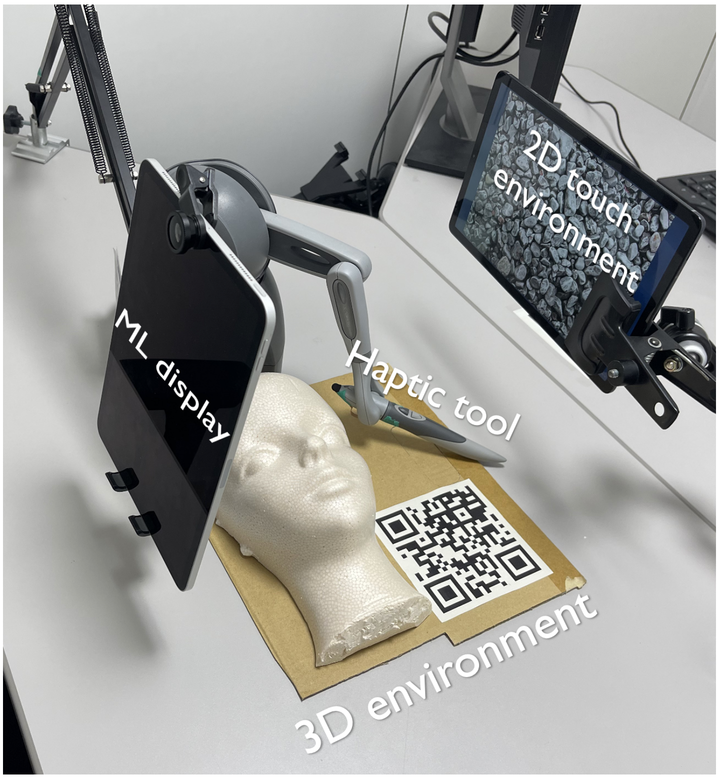


Figure 3.6: Experimental setup showing environments for both experiments. In *Exp. 1*, a touch tablet (labeled '2D touch environment') was used for haptic interactions. In *Exp. 2*, the touch tablet was swapped out for a foam human mannequin head (labeled '3D environment').

Specifically, we predict that higher accuracy in task performance will be observed under conditions of UP rendering, particularly when users navigate unfamiliar physical environments or when tactile cues for depth perception (i.e. touching a physical surface) are not provided.

We divided the user study into two experiments: in the first, we test the hypotheses using a 2D environment, and the second experiment poses a practical scenario in which the proposed system could be utilized.

3.4.2 Experimental environment

The ML display was mounted on a holding arm attached to a table, inside a brightly lit room. Positioned approximately 35 cm behind the ML display, we mounted a second display device (Samsung Galaxy Tab S5e) on an arm, serving as the 2D touch environment during the first experiment. This 2D environment provided a high-contrast background with a 'stones' texture for the task working area, as shown in Fig. 3.6.

To track the user's hand and provide a physical touch pointer, we employed a 3D Systems Touch [73] device positioned on the table behind the ML display. This haptic tool was connected to a PC (Windows 10 64-bit, i7 3.4 GHz processor, 16 GB RAM) running OpenHaptics 3.4 SDK in Unity (2021), located underneath the table. Communication between the devices was facilitated via Web sockets, with an average delay of 20 ms for transmitting haptic tool pose information (average 50 poses per second) and touch positions on the 2D touch environment. The delay was sufficiently low for the visuo-haptic system to be perceived as synchronous [74, 75, 76].

In addition to the 2D environment, we placed a foam mannequin head on the table, serving as a controlled 3D touch environment. In a pre-processing step, accurate 3D models of the haptic tool, mannequin head, and surrounding static environment (table and walls) were captured using the RGB-D camera of an iPhone 12 Pro Max and reconstructed using Scaniverse software (v3.0.3, Niantic Inc.). Refer to Fig. 3.6 for an overview of the experimental setup for the two experiments.

3.4.3 Participants

We recruited 20 participants between 23 and 38 in age (M=29.8, SD=4.59), 12 male and 8 female students of our graduate university. When asked to rank their experience with augmented reality devices between 1 (no experience) and 5 (daily experience), their average AR experience ranked 1.61 (SD: 1.03). 17 participants were right-handed, and 3 left-handed but could perform tasks and write with their right hand as well. We verified good visual acuity (with correction if necessary) in all participants. The total time per participant including demographic question-

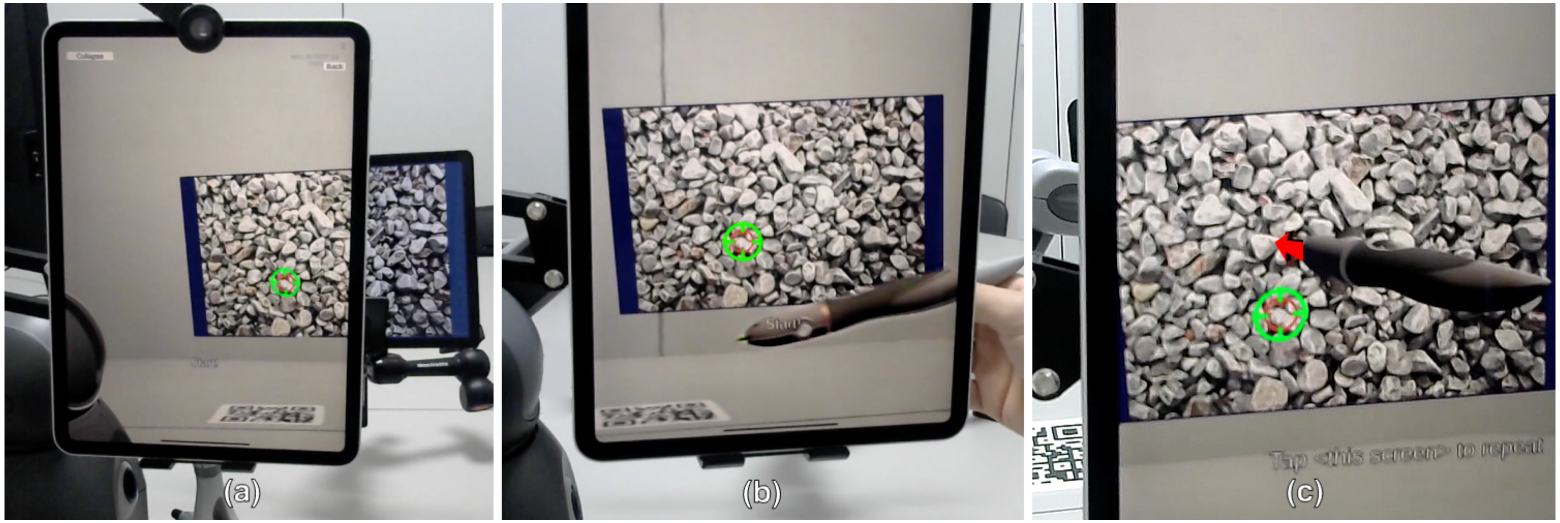


Figure 3.7: Setup and procedure for *Exp. 1.* (a) A virtual target (green) appears on the stones display. In this example, the system is configured to render in UP mode. (b) The user moves the tool tip as close as possible to the green target *without touching* the physical surface. (c) The tool tip physically touches the stones display, and a red arrow augmentation appears where the touch happened. The system waits for the user's button press to repeat the procedure. Note that the user's hand, arm, and the tool's arm are invisible while looking through the ML display.

naire, dominant eye test, instructions, practice, breaks, and post-questionnaire took approximately 1 hour and 15 minutes.

3.5 Experiment 1 – Target point interaction

In the first experiment, we evaluated the accuracy of touching target points on the surface of a 2D touch display, as well as the accuracy of estimating the depth of those target points.

3.5.1 Experimental conditions

Participants underwent *touch point* accuracy and *depth* accuracy assessments under four distinct conditions of two independent variables: *perspective* (user or device perspective) and *environment visibility* (visible or hidden). Under the environment visible condition, participants were provided full visibility of the physical environment surrounding the testing area. Conversely, in the environment hidden condition, a curtain was positioned behind the ML display, obscuring participants' view of the physical environment (see Fig. 3.8). This manipulation aimed to isolate the impact of environmental cues on haptic interaction accuracy as well as the creation of a cognitive map.

Each participant completed target point accuracy tasks in all four conditions, with the order of conditions counterbalanced to mitigate potential order effects. In this experiment, the physical environment was limited to the 2D touch display, with the removal of the 3D environment mannequin head. For each condition combination, the orientation of the 2D touch display relative to the ML display was varied while keeping a fixed distance of 35 cm between them. This variation aimed to simulate a dynamic environment and minimize the formation of a cognitive map.

3.5.2 Task and procedure

Participants were tasked with touching the center of a target point *on the surface* of the 2D environment with the haptic tool tip. This target point augmentation was only visible when viewed through the ML display, and its location was

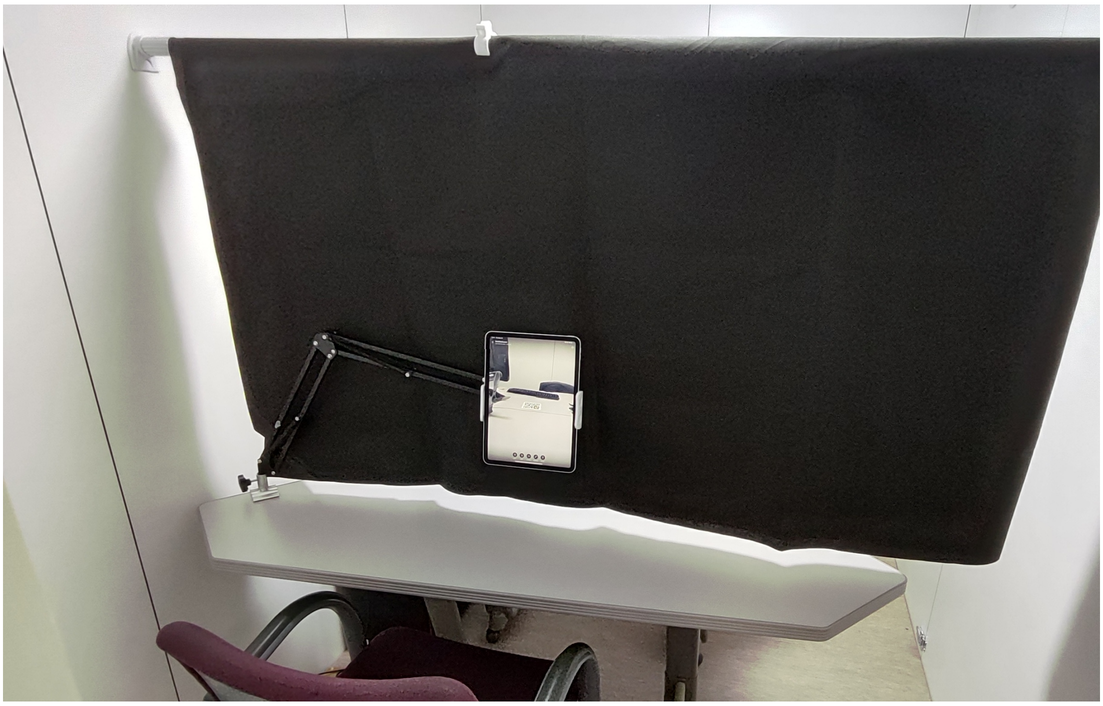


Figure 3.8: Experiment I set up in the hidden environment condition. A curtain occludes the user's view of the real environment while the ML visualizes it.

randomized every round. See Fig. 3.7 (a).

Prior to the main task, the examiner determined the participant's dominant eye through the Miles Test [77], and set up the apparatus perspective rendering according to the dominant eye (left or right). At the beginning of the task, participants were made familiar with the ML display and the haptic tool. They were instructed to grasp the haptic tool like a pen in their right hand and to move the tool behind the ML display repeatedly under the two *perspective* conditions. They also practiced moving their head to change their perspective view of the task area in the UP condition.

To initiate a touch interaction round, participants tapped anywhere on the ML display, triggering the appearance of a red target in the 2D environment. They were instructed to initially assess the depth of the target by moving the haptic tool pointer as close to the target as possible *without making contact* with the 2D environment display (Fig. 3.7, b). Failed attempts (they accidentally touched the 2D environment) caused the system to immediately advance to the next round. Once satisfied with the position of the haptic tool pointer, participants pressed the button on the tool pen so that the system registered their depth error. Lastly, participants aimed to touch the center of the red target point as accurately as possible by moving the haptic tool pointer (Fig. 3.7, c). Upon touching the 2D environment display, a red arrow appeared at the touch location, indicating the conclusion of input capture for that round. At the commencement of the experiment, as part of the practice phase, this process was repeated for 10 rounds under each combination of perspective and environmental conditions. Subsequently, participants performed 25 repetitions for each condition, with each condition executed twice, resulting in a total of 200 data pairs (depth error and 2D touch error) per participant.

3.5.3 Baseline accuracy

A control group of 9 participants performed the depth estimation and touch tasks *without* the presence of the ML display and its rendering conditions. The participants in the control group were distinct from those in the main experimental group. In place of the ML display, a cardboard cutout with identical dimensions was used to simulate an ideal transparent display. This can be seen in Fig. 3.9

under conditions of environmental visibility.

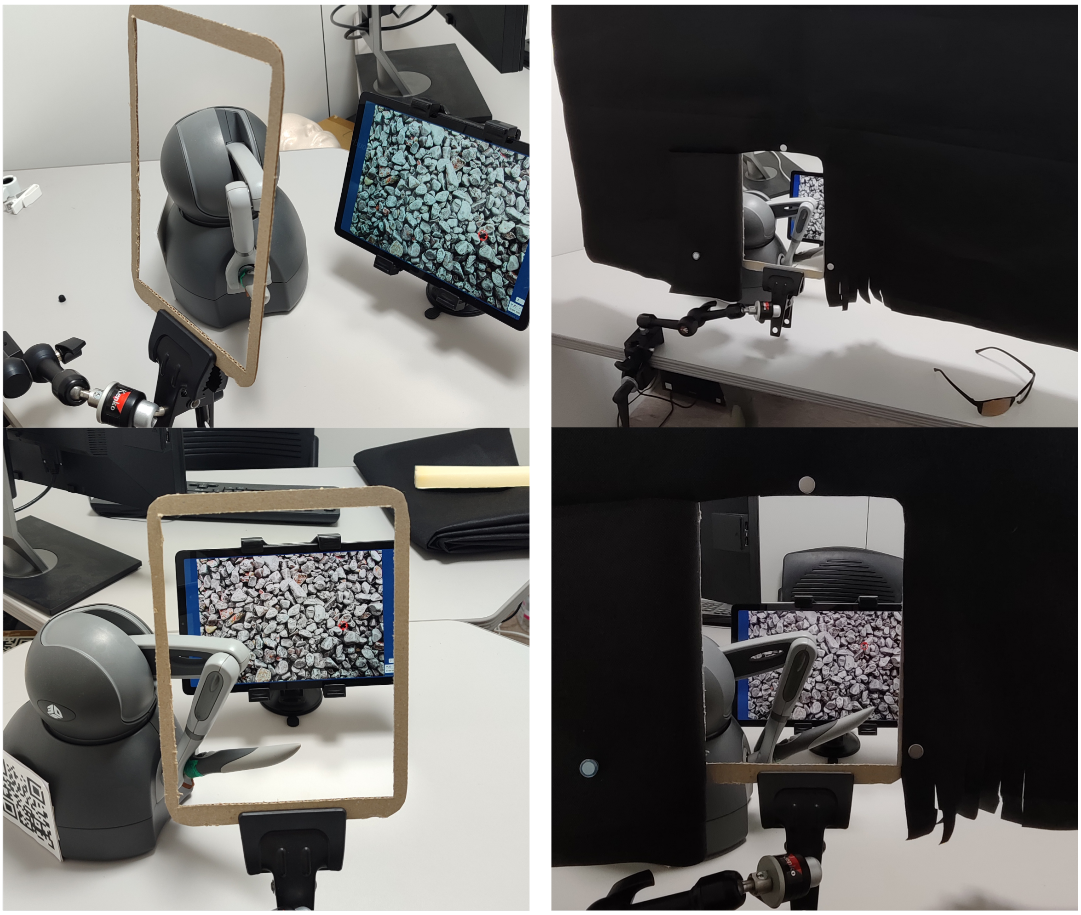
In all conditions of our main experiments, the participants had a monocular view of the physical environment behind the ML display (due to Sect. 3.2.2). To address stereoscopic vision, the control group performed the experiment under both *monocular* and *binocular* conditions. In the monocular condition, participants wore glasses without lenses, with a piece of plastic fitted into the frame to block the view of their non-dominant eye. In the binocular condition, they performed the tasks with vision from both eyes as they normally would.

3.6 Experiment 2 – Needle injection

In the second experiment, we assess the accuracy of target point localization utilizing the ML display within a simulated needle injection scenario. We anticipate that the developed prototype system will offer advantages in training and assisting healthcare practitioners in the execution of intravenous procedures, such as vaccinations and blood sampling, considering precision and comfort. In this context, an augmented target point guides users to the intended location for the injected needle tip. The virtual nature of the injection needle enhances safety, facilitates the creation of reusable training materials, and allows for seamless modification of training parameters such as needle dimensions and injection sites at minimal cost and effort.

3.6.1 Experimental conditions

The participants underwent an assessment of target point accuracy under two perspective rendering conditions: *Device* or *User* perspective. Analogous to the preceding task, the ML display portrayed imagery either from the viewpoint of the device camera (DP) or from the perspective of the user's dominant eye (UP). The experimental setup involved exclusively a physical foam mannequin head (labeled '3D environment' in Fig. 3.6), semi-transparently rendered through the ML display. This semi-transparency (X-ray visualization) improves depth perception of AR content due to structure occlusion cues [13] and is widely used with VST displays [18]. Error distance was recorded and quantified in centimeters as the spatial disparity between the virtual needle tip and the designated target



(a)

(b)

Figure 3.9: Experiment 1 setup for the control group, with (a) the surrounding environment visible, and (b) the surrounding environment obscured by a black curtain, except for a cutout region. The bottom row illustrates the participant's point of view.

point. Additionally, the number of needle injections (defined as instances of the virtual needle penetrating the surface of the mannequin head) was tallied and reset after each repetition.

3.6.2 Apparatus

The experimental setup remained consistent with the environment described in Sect. 3.4.2. However, in this task, we incorporated haptic feedback sensations simulating the act of needle insertion into flesh, using the haptic pen tool. As illustrated in Fig. 3.10, participants could feel the foam head surface through the tip of the virtual green needle, simulating the tactile properties of human skin. Applying pressure in the direction of the needle tip elicited a subtle haptic sensation akin to piercing the skin. Subsequently, the haptic tool solely allowed movement in the forward or backward direction of the needle, with a minor counter-force mimicking the sensation of moving a needle through flesh.

3.6.3 Task and procedure

Participants were tasked with touching a virtual target point as accurately as possible with the needle tip. To do this, they had to estimate the spatial position of the target point inside the mannequin head, grasp the haptic pen with their right hand, and inject the virtual needle with an appropriate entry direction. Participants were furthermore requested to make *as few injections as possible*, mimicking the real scenario of patient comfort in needle injection.

First, participants were familiarized with the new environment of a 3-dimensional foam human head. They were given the opportunity to investigate the physical head with the virtual needle by stroking it and probing it. Then, they performed 20 practice rounds per *perspective* condition, injecting and touching a randomized target point with the needle tip. During injection, when the participants felt that they were sufficiently close to the target point, they were asked to press the button on the haptic tool pen. Then their error distance was recorded and displayed on the ML display as performance feedback, and the round was repeated with a new randomized target position. They performed 30 repetitions per perspective condition, with a break in between conditions. The perspective condition was

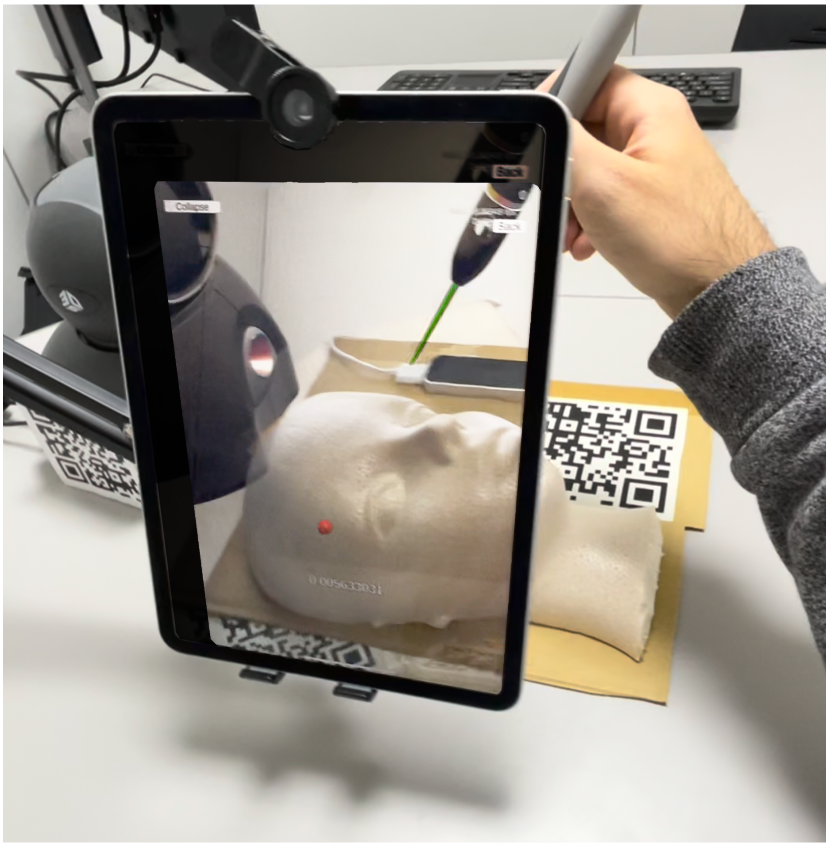


Figure 3.10: Setup for needle injection experiment (Exp. 2). A foam mannequin head is semi-transparently rendered on the ML display while a user is moving a tool equipped with a virtual needle (green). In the configuration shown here, UP rendering is enabled and a red spherical touch target is presented inside the mannequin head.

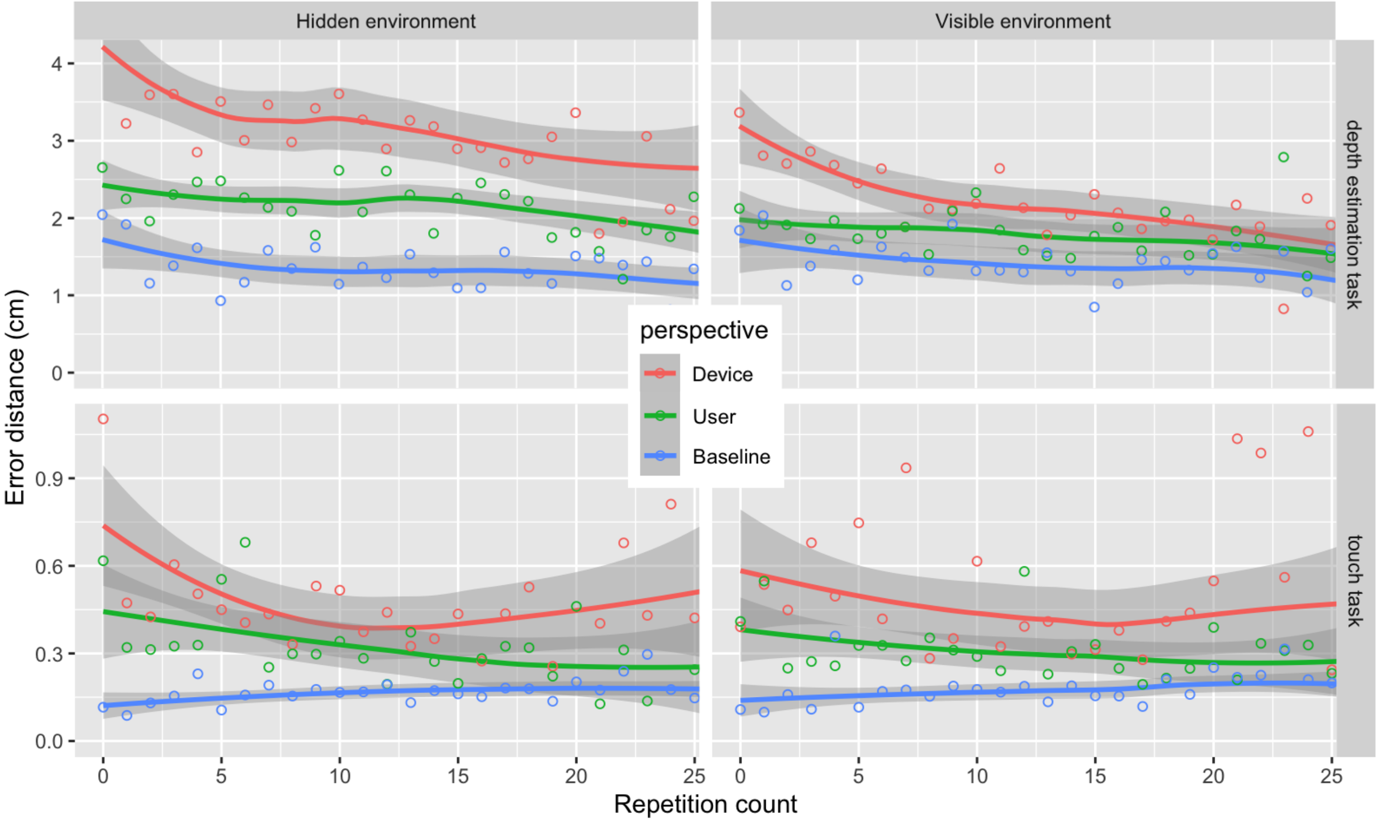


Figure 3.11: Results for Target point interaction experiment (Exp. 1). The top row shows the results of the depth estimation task, and the bottom row shows the touch task accuracy. Outlined dots represent the mean per repetition count. The lines are smoothed using 'loess' (fraction=0.9), with gray areas indicating confidence intervals. Note that the y-axis scales between the two tasks (rows) differ.

counterbalanced across participants.

3.7 Results

3.7.1 Experiment 1 – Target point interaction

The error distance (in centimeters) for the depth estimation task is shown in Fig. 3.11 top row, and the error distance of touch interaction on the 2D display is shown in the bottom row. The columns in the figure denote the environment visibility condition. Failed attempts of depth estimation (the participant accidentally touched the screen) were removed from touch interaction results. The depth estimation error for these failed attempts was set to the condition average (over 25 repetitions) during data analysis. We applied a two-factor (perspective×environment) repeated-measures ANOVA on the Aligned Rank Transform (ART) of the touch data and the depth estimation data. We used ART because the data did not meet the normality assumption.

The analysis of error distance in the touch task revealed a main effect for *perspective* ($p < 0.05$, $F_{1,18} = 7.497$) while *environment* ($p = 0.471$) and their interaction ($p = 0.946$) effects were not significant. Analysis of depth estimation errors showed main effects for both *perspective* ($p < 0.01$, $F_{1,18} = 15.713$) and *environment* ($p < 0.001$, $F_{1,18} = 39.599$). No significant interaction between the two factors was found ($p = 0.115$).

As shown in Fig. 3.11, error distance while UP rendering was significantly lower than using DP for three out of four conditions, partly supporting the first hypothesis **H1**. Only in a known, visible environment, perspective does not seem to have an effect on the accuracy of touching targets on a surface (bottom-right in the figure).

When comparing environment visibility conditions, we see a higher average error and a larger effect of perspective in a hidden (unknown) environment. Furthermore, the absolute error decrease using UP rendering for estimating depth (0.70 cm, 25.40%) is higher than the error decrease for the touch task (0.14 cm, 30.81%). These results support our second hypothesis **H2** stating that UP rendering has a larger impact in an unknown environment or while tactile cues of depth (touching a physical surface) are absent.

We see a learning effect of DP in the depth estimating tasks (Fig. 3.11 top row). As the number of repetitions increases, the error distance gradually becomes lower. Fig. 3.12 shows the results for estimating depth without this learning effect by taking only the first 5 repetitions of each condition. A repeated measures ANOVA reveals again that perspective ($p < 0.01$, $F_{1,18} = 12.852$) and environment visibility ($p < 0.01$, $F_{1,18} = 13.99$) have a significant effect on depth estimation error. Using UP over DP rendering increased performance by 32.55% (0.93 cm) while the environment was known and 31.03% (1.08 cm) without *any prior knowledge*.

Finally, we confirmed that there was no significant difference in accuracy between left-handed and right-handed participants in either haptic task, likely because the left-handed participants were actually ambidextrous.

Baseline accuracy

The control group performed the experiment with a cardboard cutout in place of the ML display under monocular and binocular vision conditions. In this way, we measured a baseline accuracy for depth estimation and touch point interaction. Baseline accuracy for both vision modalities averaged for the two environment visibility conditions are shown in Fig. 3.13. On average, participants were better at estimating the depth of target points in the binocular condition (M=0.85 cm, SD=0.20 cm) than in the monocular condition (M=1.39 cm, SD=0.26 cm). We again see a learning effect in the depth task, in both vision modalities, with the error distance decreasing as the number of repetitions increases. For the touch task, participants also performed better in binocular viewing conditions (M=0.13 cm, SD=0.06 cm), slightly better than monocular viewing (M=0.25 cm, SD=0.18 cm).

The baseline accuracy using *monocular* vision is put into context with User perspective and Device perspective rendering accuracy in Fig. 3.11 and Fig. 3.12. Environmental visibility does not significantly affect the baseline results.

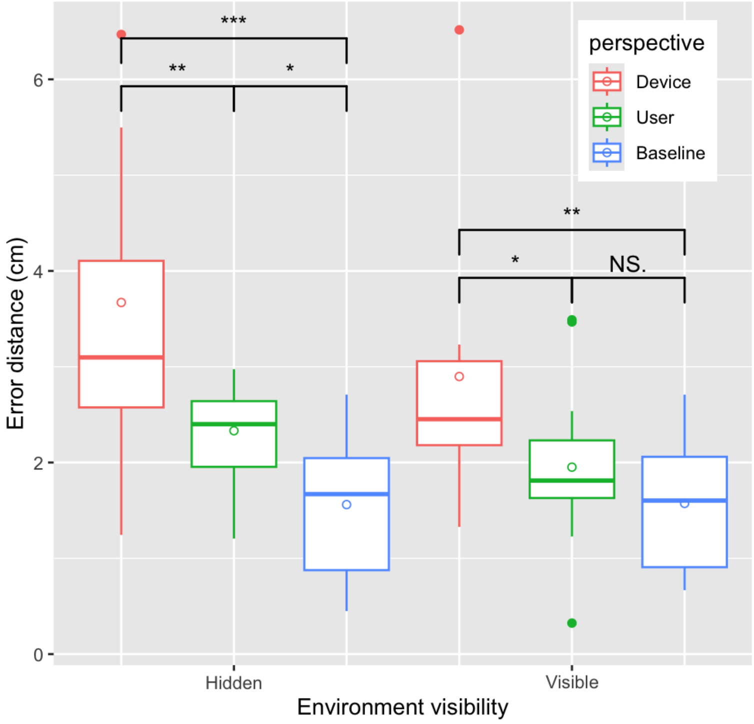


Figure 3.12: Depth estimation errors (Exp. 1) filtered by initial 5 repetitions per perspective condition, to adjust for the learning effect. The baseline results shown are with *monocular* viewing conditions.

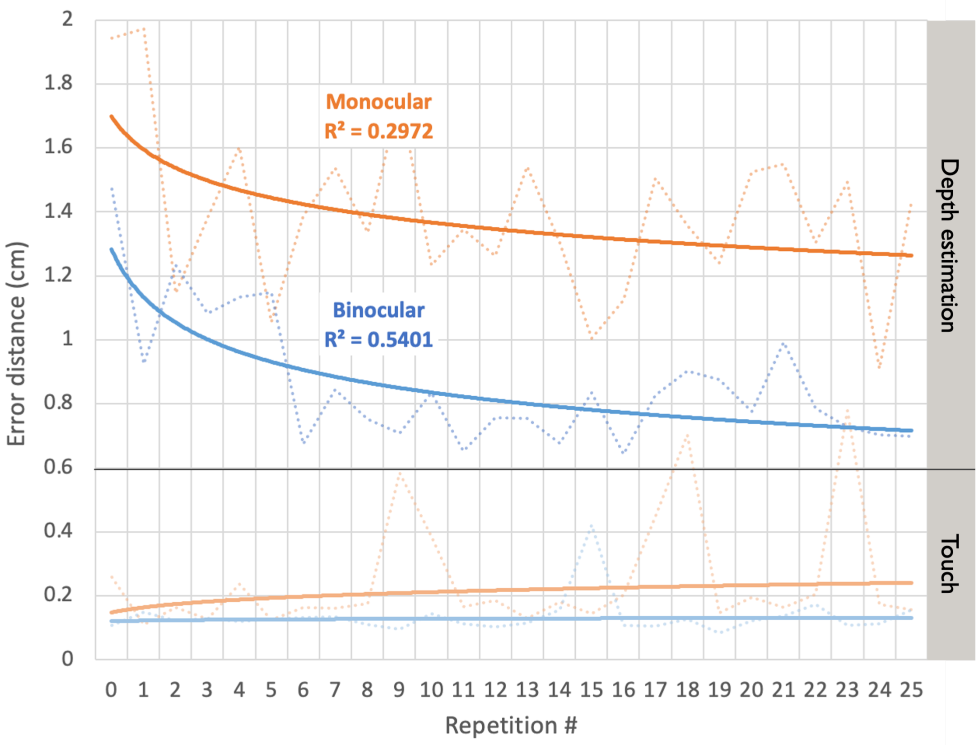


Figure 3.13: Baseline accuracy results of Experiment 1 separated by the depth estimation task (top) and touch task (bottom), under conditions of monocular vision (orange-dotted lines) and binocular vision (blue-dotted lines). Power curves ($f(x) = a * x^b$) are fitted to all conditions (solid lines).

3.7.2 Experiment 2 – Needle injection

The results of the needle injection experiment are shown in Fig. 3.14. The injection accuracy under conditions of UP or DP rendering is separated by the captured number of injections (pierced the physical surface) per repetition. We applied one-way repeated-measures ANOVA on the ART data including all injection counts and only the first injection. Including all injection counts revealed no significant effect of perspective on injection accuracy ($p = 0.055$, $F_{1,92} = 3.756$). However, since the number of injections per repetition varied among participants, the sample size was not uniform across conditions (see sample counts at the top of Fig. 3.14). When focusing only on the first injection per participant ($n = 20$), we found that perspective significantly ($p < 0.01$, $F_{1,19} = 15.322$) affects injection accuracy. As illustrated in the figure, the injection error distance decreases as the number of injections increases. Only during the initial injection, we observed a significant decrease in error (37.83%) between DP (M=2.14 cm, SD=1.18 cm) and UP (M=1.33 cm, SD=0.67 cm) rendering. This result supports the second hypothesis **H2** that states the effect of UP is more pronounced in the absence of other depth cues (haptic feedback, prior knowledge).

3.7.3 Preference questionnaires

After the completion of each experiment, participants were asked to fill in a questionnaire with their preference for perspective rendering. Specifically for Exp. 1, participants were asked for both visible and hidden environments in which perspective condition they preferred to perform the touch interaction and their optional reasoning. The results are similar to the quantitative results, with most (16) participants preferring UP over conventional DP in the environment hidden condition, but when the environment was visible their preference became divided across the perspective conditions (11 participants preferred UP versus 9 for DP).

After the needle injection experiment, participants were asked in which perspective condition they felt the most in control, and which perspective condition they preferred to perform needle injection, with optional reasoning or remarks. Again a majority (14) preferred UP over DP and similarly felt most in control in the UP condition (13). Some participants remarked that in the UP condition,

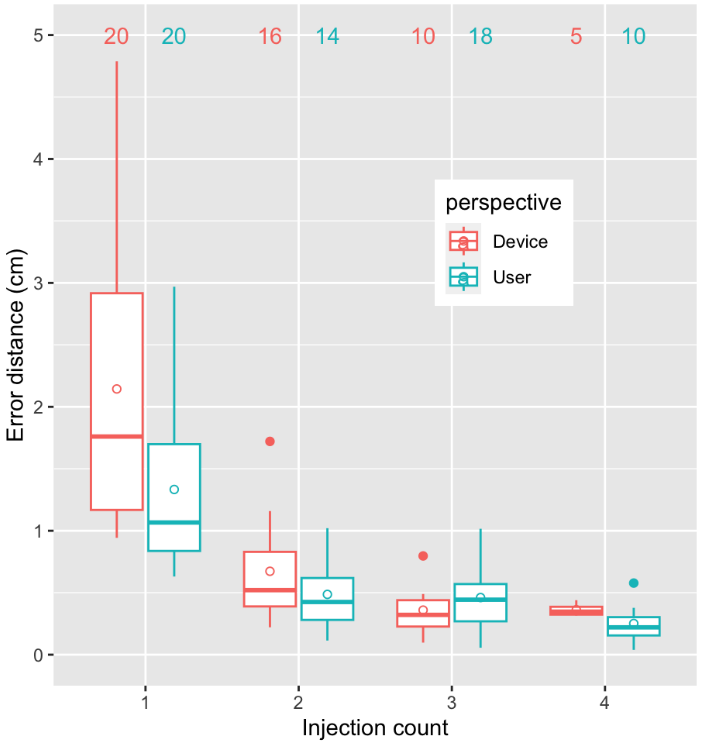


Figure 3.14: Exp. 2 results: *Error distance* (in centimeters) for needle tip-to-target, plotted against *Injection count* (the cumulative number of times the needle tip pierced the surface during a round) under two perspective conditions. Means are represented by outlined dots, outliers by filled dots, and the sample size for each condition is displayed at the top.

they expected that they could get closer to the ML display in order to see more details. This is a natural and intuitive way of working in close proximity. However, they were limited in minimum distance of eye accommodation for distinct vision.

3.8 Discussion

This study investigated how visual context (environment visible vs. hidden) and perspective rendering (user vs. device) impact depth estimation and touch accuracy when viewed through an AR magic lens display.

3.8.1 Baseline results

Baseline touch point accuracy (i.e. without visualizations on an ML display) under monocular and binocular viewing conditions are comparable with each other and similar to the accuracy found during normal (binocular) viewing in previous work [78]. This is because extra information provided by binocular disparity offers only a small advantage relative to the large impact of *monocular cues* like occlusion combined with *tactile feedback* of the target surface. However, the advantages of stereopsis are more pronounced without the tactile cues as can be seen in the depth estimation task results (Fig. 3.13). It is suggested that in AR-guided tasks that require accurate depth perception, designers of such systems should strive to present stereo disparity images to enable stereopsis.

3.8.2 Visualizing the haptic interaction medium

In extending the methodology by Čopič Pucihar et al. [6], we found that visualizing the hand-guided tool on the ML display significantly improved targeted touch accuracy by 71% (1.15 cm) using DP and by 47% (0.40 cm) using UP rendering. We confirm this in the control condition representing a monocular view of the environment and the (visible) interaction medium: Participants were approximately 50% more accurate in estimating egocentric depth than in related studies in which the interaction medium was *not visible* [56].

This result further highlights the importance of multimodal feedback in enhancing haptic interaction accuracy [79, 80, 81] and implies that rendering the hand or hand-guided tool on ML displays should be standard practice in applications requiring precise touch or manipulation. For instance, in AR-assisted surgery [82], visualizing the surgeon's hand along with the surgical tool can improve the accuracy of incisions or injections. Similarly, displaying the technician's hand can help align tools more precisely with the virtual overlay on machinery in AR-enhanced maintenance tasks.

3.8.3 The effect of tactile feedback

Comparing the performance between the depth estimation task and the touch interaction task (Fig. 3.11 top row vs. bottom row), we note a large absolute difference in accuracy for all conditions. The presence of tactile feedback significantly decreases the error distance in a haptic interaction task. This improvement due to visuotactile integration is consistent with the literature [49, 81, 83]. Contrary to the extended work [6], we found no learning effect in the DP condition, nor was there a significant effect of perspective on touch accuracy (see bottom right of Fig. 3.11), suggesting that visuo-haptic feedback of the hand-guided tool and visual context are sufficient to compensate for the distorted DP view of the real environment even during initial attemps. We hypothesize that users visually align the touch tip with the target on the ML display and continue moving forward until tactile feedback confirms target contact. The additional information provided by UP rendering does not seem to outweigh the visual precision advantage of DP (due to its zoomed-in perspective); the two cancel each other out. Therefore, conventional DP rendering may already offer adequate support for tasks involving touch feedback (while the touch medium is visible), such as interactions with physical surfaces. UP rendering provides only marginal improvements in the absence of visual contextual information.

3.8.4 Impact of perspective on depth accuracy

The performance of UP rendering is close to that of the baseline performance, and with environment contextual cues the difference between UP and the baseline accuracy is not shown to be significant (Fig. 3.12). This shows the performance of our UP rendering algorithm to be potent, approaching perfectly transparent monocular viewing conditions. DP rendering remains the worst provider of depth cues, significantly higher than UP and the baseline. These results follow the trend set by [6] but additionally provide insights into using AR on ML displays in close-range *depth* tasks. Aligning the rendered perspective with the user's natural viewpoint evidently offers a more intuitive and effective interaction with virtual content and spatial understanding. In line with prior knowledge [49], the motion parallax provided by UP improves depth perception *at close range* and falls off at distances over ∼3 meters [70]. Designers of AR-guidance applications on ML displays, such as smartphones and tablets, should consider implementing UP rendering when dealing with close-range or spatially complex content. While binocular viewing offers the best interaction performance, the difference compared to UP rendering is minimal (approximately 0.5 cm for depth targets), and UP rendering does not require specialized (auto)stereoscopic display technology. This significantly enhances the accessibility of off-the-shelf (VST) ML displays for supporting close-range tasks in AR.

3.8.5 Impact of visual context on depth accuracy

Placing an AR ML display within the context of a visible physical environment significantly enhanced depth accuracy (see Fig. 3.11 top row). This improvement is likely because the contextual information from the physical environment helps the creation of a cognitive map [54] that aids users in gauging distances and *spatial relationships* [49]. For AR applications that overlay virtual objects onto the real world, we have demonstrated that maintaining a visible environment context can improve spatial performance.

3.8.6 Interactions between perspective and visual context

In particular, for DP rendering, contextual information from the environment greatly increases depth estimation. Without environmental visual cues, UP provides much better depth estimation than conventional DP rendering (Fig. 3.11 top left); thus, users prefer UP under these conditions. This implies that when designing AR applications on ML displays for use in unfamiliar or visually sparse environments, UP rendering should be prioritized to mitigate the challenges caused by the lack of contextual cues. In practical applications, such as remote robotic manipulation [84] or virtual assembly tasks, UP rendering should be employed to ensure that users can accurately perceive and interact with physical objects and understand virtual depth information. However, there was a significant difference in accuracy in the hidden environment between UP and the baseline. It might be that the latency of video see-through UP rendering on the ML display or missing depth information from eye accommodation (see Chapter 4) is affecting performance that can be made up for with cognitive knowledge of the environment if available.

3.8.7 Learning effect on depth accuracy

We observed a learning effect, with depth accuracy improving over 25 interactions. This highlights the role of scene cognition [54] and practice in enhancing spatial perception in AR environments, suggesting that users initially require an adaptation time to interact effectively with virtual targets through haptic feedback.

The depth accuracy with DP rendering improved steeply during the first five interactions (top row of Fig. 3.11 and Fig. 3.12), then gradually increased. This pattern indicates that users initially struggle with DP because of deviations from natural viewing conditions; however, they can adapt with some practice when given alternative feedback of depth. This implies that training sessions can accelerate user adaptation to tasks using conventional DP rendering under the assumption that the *spatial layout remains unchanged*.

We demonstrate that UP rendering provides the highest depth accuracy even without scene understanding and learning because of the additional depth information provided by motion parallax. This depth accuracy is only marginally improved by subsequent haptic interactions, indicating that vision is the most important factor for spatial awareness. In AR-guided procedures, such as the needle injection task in our second experiment, the initial accuracy was also higher with UP rendering. For tasks requiring *immediate* spatial accuracy, such as AR-assisted assembly and surgical procedures, UP rendering should be prioritized to leverage natural spatial awareness.

3.8.8 Verifying results through Needle Injection experiment

The needle injection experiment confirmed our findings, showing higher initial accuracy with UP rendering. Over time, DP rendering performance caught up, reinforcing the advantage of UP for tasks requiring immediate precision, while indicating user adaptability to DP. For example, in medical training simulations requiring immediate accuracy, UP rendering should be used to facilitate early success. DP rendering can still be viable for tasks where users can afford a learning curve, such as ongoing training or non-critical interactions.

Using conventional DP rendering, Hecht et al. [67] achieved a consistent error of less than 0.10 cm for AR needle injection guidance on a smartphone. In contrast, our system averages a 2.14 cm error under similar conditions. This substantial performance disparity is likely due to the continuous availability of powerful *occlusion* depth cues in their system (by visualizing the full injection trajectory of the needle) which were intentionally not present in our implementation. We hypothesize that integrating these occlusion depth cues into our system would significantly enhance performance. This hypothesis is supported by our findings that, with multiple injections providing visual occlusion feedback (Fig. 3.14), we achieved an injection error of approximately 0.39 cm, comparable to the results reported by Hecht et al.. We anticipate that UP rendering in addition will decrease task time, as spatial awareness is better, while enabling hands-free operation.

3.8.9 Summary of Magic Lens key design principles

Our findings suggest several key design principles for AR ML systems:

- **User-perspective rendering** is preferable for tasks requiring immediate spatial accuracy and in the absence of visual context or tactile feedback.
- **Visible environment context** enhances depth perception and should be maintained whenever possible.

- **Visualizing the hand-guided tool** significantly improves haptic accuracy and should be integrated into ML displays.

3.9 Summary

We introduced a magic lens (ML) capable of rendering its display area from the user's perspective, leveraging off-the-shelf devices. Utilizing this prototype as a stationary medium for displaying augmented instructions, we conducted two experiments to assess user haptic interaction accuracy under conventional device-perspective (DP) rendering versus our proposed user-perspective (UP) rendering, coupled with cognitive awareness of the environment. The results demonstrated that UP rendering enhances both tactile interaction accuracy and depth estimation, particularly without visual context, such as visually concealed or dynamically changing environments. Additionally, we evaluated our ML display in a practical needle injection scenario, where users were tasked with accurately injecting a virtual target inside a physical dummy, using a hand-guided virtual needle tip. The findings from this second experiment corroborated the depth estimation results of the first experiment, indicating that UP rendering outperforms conventional DP in immediate spatial accuracy and while tactile feedback of the target is lacking.

These findings hold implications for leveraging off-the-shelf devices to support augmented reality applications in fields like assembly line work or medical procedures, where frequent changes in components and tools, along with accurate depth perception, are critical. We furthermore provided design guidelines for using ML displays in such AR applications.

3.9.1 Limitations

Throughout the experiments, we encountered several limitations associated with the use of our prototype stationary ML display, which were further validated by the quantitative feedback provided by participants in the questionnaires. Firstly, the physical presence of the ML display often imposed constraints on the user's movement range. On multiple occasions during the needle injection experiment, participants inadvertently collided with the ML display while attempting to locate an optimal injection angle. This limitation underscores the trade-off inherent in using magic lenses to facilitate tasks with augmented instructions: while handheld magic lenses offer greater freedom of movement, allowing users to allocate space for their off-hand to perform tasks, they come at the expense of the ability to use both hands simultaneously to execute tasks. Displaying virtual imagery using HMDs can be a valid alternative if only adding virtual information (OST HMDs) or if used for a short time (VST HMDs), see Sect. 3.2.3.

A second limitation arises from the natural inclination of users to move their heads closer to the ML display during high-precision tasks in order to examine details more closely. While the UP-rendered ML display maintains geometrically correct viewing, the proximity of the display itself can impede the user's ability to accommodate their eyes accordingly. Consequently, we observed several instances where participants inadvertently moved too close to the display, prompting us to remind them to maintain a suitable distance to allow their eyes to focus on the details presented on the ML display. It is worth noting that even if human eyes were capable of accommodating at shorter distances (<15 cm), the resolution of the display device would often need to be higher to visualize details beyond the current capabilities of off-the-shelf devices effectively. Utilizing an OST type ML could limit the focusing problem to only virtual content. Varifocal ML displays [66] could provide content at an acceptable focal distance but to our knowledge such a ML does not exist yet.

3.9.2 Future Work

While existing research suggests that stereoscopic displays enhance depth estimation and spatial awareness in arm's-length tasks, the combined effect of motion parallax from UP rendering and stereopsis remains unexplored in the context of stationary magic lens video see-through displays. Investigating whether the fusion of these depth cues further enhances touch interaction accuracy would be intriguing. Regardless, the UP rendering technique outlined in this study can be readily extended to accommodate dual-eye configurations.

As outlined in Sect. 3.3.1, the real-time reconstruction of interactive elements, such as the user's hands, demonstrated limited precision for sub-centimeter tasks in our current study. Future studies could improve this with noise reduction algorithms, filters for edge smoothing, or neural networks for dense reconstruction of sparse point clouds. Additionally, with ongoing advancements in depth-camera technology found in consumer devices like tablets and smartphones, future iterations of our prototype could capitalize on these improvements. By integrating high-quality depth cameras, our system could achieve more accurate and geometrically-correct representations of the physical environment in real time, eliminating the need for additional tool-tracking hardware.

4 Effects of Vergence and Accommodation on AR Context Integration

Augmented reality (AR) magic-lens (ML) displays, such as handheld devices, offer a convenient and accessible way to enrich our environment using virtual imagery. Several display technologies, including conventional monocular, less common stereoscopic, and varifocal displays, are currently being used. Vergence and accommodation effects on depth perception, as well as vergence–accommodation conflict, have been studied, where users interact only with the content on the display. However, little research exists on how vergence and accommodation influence user performance and cognitive-task load when users interact with the content on a display and its surroundings in a short timeframe. Examples of this are validating augmented instructions before making an incision and performing general hand-eye coordinated tasks such as grasping augmented objects. To improve interactions with future AR displays in such scenarios, we must improve our understanding of this influence. To this end, we conducted two fundamental visual-acuity user studies with 28 and 27 participants, while investigating eye vergence and accommodation distances on four ML displays. Our findings show that minimizing the accommodation difference between the display and its surroundings is crucial when the gaze between the display and its surroundings shifts rapidly. Minimizing the difference in vergence is more important when viewing the display and its surroundings as a single context without shifting the gaze. Interestingly, the vergence–accommodation conflict did not significantly affect the cognitive-task load nor play a pivotal role in the accuracy of interactions with

AR ML content and its physical surroundings.

4.1 Introduction

Augmented reality (AR) has long been envisioned as a support system for both every day and specialized tasks [57, 85, 86]. The widespread availability of smartphones equipped with dedicated graphical processing units, high-quality cameras, and displays has enabled average users to augment the world using virtual imagery. By utilizing a smartphone camera image, the user can view the physical world together with the digital content as if the handheld device is a transparent screen or lens. Such video see-through displays are also referred to as a magic-lens (ML) displays [40].

The popularity of augmented content in optical see-through displays has recently increased. In contrast to ML displays, optical see-through displays use transparent or semi-transparent materials to superimpose virtual imagery. This method offers a natural way of viewing the real environment as light passes through the material and maintains the user's perspective. This is one of the major advantages of this type of display, because only augmented content requires generation and visualization. Thus, no computing power or time has to be spent on the visualization of the natural environment, and could be better spent improving the virtual imagery quality and update speed. However, similar to ML displays, virtual imagery is commonly presented in a fixed image plane. Therefore, if the surroundings and image-plane distances are not aligned, the image falls out of focus. Another drawback of optical see-through displays is that they are expensive and difficult to access. Typically worn on the head, these displays, known as head-mounted displays (HMD), offer the advantage of being hands-free. However, given the limited availability of haptic interactions, the provided input interface can be difficult to operate. Prolonged use of HMDs can lead to fatigue [19, 20] and might, in certain situations, prove cumbersome or impossible owing to the presence of other headgear or environmental constraints.

Displaying augmented content on a handheld device is affordable and provides users with a familiar touchscreen interface. However, this approach has several disadvantages related to vergence and accommodation when interacting with a display or its surrounding environment. *Vergence*, also known as binocular convergence, is a property of stereoscopic systems, in which both eyes rotate to allow light to converge at retinal centers, where vision is the sharpest and most detailed. While *accommodation* is a property of the eye that involves stretching the lens to adjust the focus and maintain clear vision at varying distances, both vergence and accommodation provide strong cues that are essential for accurate depth perception. A challenge arises when using the conventional monoscopic ML setup (Fig. 4.1.1), where the eyes converge and are accommodated at the same distance (B) on the display, irrespective of the actual distance to the real surface being visualized. Consequently, when shifting gaze from display (B) to its surroundings (A or C), both eyes must readjust their vergence and accommodation distances. This process is required to bring A or C into focus and remove double vision (top-right of Fig. 4.1.1). The same readjustment is required when shifting the gaze back to display (B). This process requires time and effort [87], potentially hindering smooth interactions with the augmented content displayed on the ML and its surroundings.

When shifting the gaze between the display and its surroundings, the optimal approach to alleviate eye effort is to have the vergence and accommodation set equally on the surroundings (Fig. 4.1.4). In such a system, gaze shift does not require eye adjustment, and A, B, and C are always perfectly focused, that is, without double vision. However, creating such a display is technically challenging. An alternative method to alleviate eye effort is to accommodate the display for the surrounding distance (Fig. 4.1.3), such as in varifocal displays, in which the AR content plane can accommodate any distance. When shifting gaze from the display to its surroundings, only vergence requires change (because the display is monoscopic) to resolve the double vision of A and C (Fig. 4.1.3 top-right). While this approach appears promising, it introduces a phenomenon termed vergence–accommodation conflict (VAC) [23]; a mismatch between the focusing distance (accommodation) and vergence distance of the eyes, which is common in mixed-reality displays (i.e., augmented and virtual reality). This conflict causes eye strain and is expected to hinder the ability to interact with AR content and its surroundings. Additionally, display systems, such as varifocal displays, require large, complex optics that pose another technical challenge. A

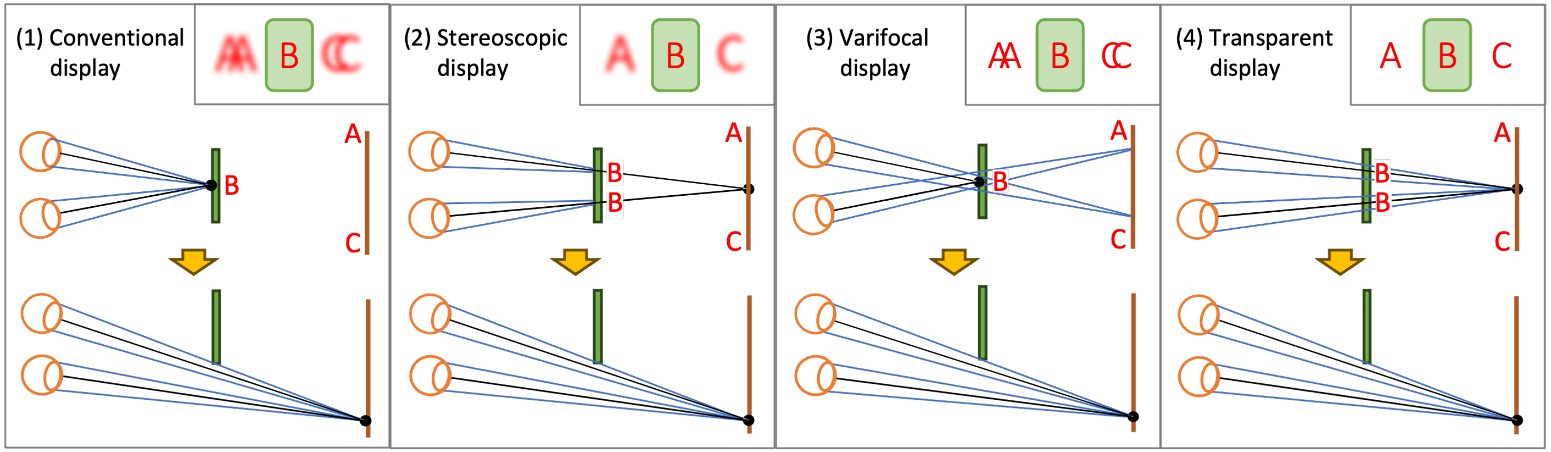


Figure 4.1: Depending on the type of magic-lens display (green), eye accommodation (blue lines) and/or vergence (black lines) distances must change as the observer shifts from a near display (green) to the far surroundings (brown), or vice versa. The top-right of each scenario shows the first-person view of letters "A," "B," and "C" when the gaze is fixed on letter "B." We considered four display types: (1) Conventional displays. "A" and "C" are out of focus and have double vision. Both the vergence and accommodation distances must change as the gaze shifts (yellow arrow) toward the surroundings. (2) Stereoscopic displays. "A" and "C" are out of focus; only the accommodation must change. (3) Varifocal displays. "A" and "C" are observed with double vision; only the vergence must change. (4) Transparent displays. All three letters in focus and without double vision; no accommodation or vergence changes are necessary.

final option to alleviate eye effort when shifting the gaze between the ML and its surroundings is to set the vergence distance of the display at or close to the surrounding vergence distance such that only accommodation has to change to achieve perfect focus for A and C (Fig. 4.1.2). Such a system only requires a stereoscopic display (such as a lenticular screen, multiview [44] or light-field [88] display); however, it also suffers from a vergence–accommodation conflict.

Despite the importance of being able to simultaneously see and understand augmented content on a ML display alongside its surroundings, the *individual* influence of eye accommodation, vergence and the existence of VAC on human performance has not yet been studied. In theory, users can use two different strategies when interacting with the display content and its physical surroundings, depending on the location of the physical content. The first is rapidly shifting gaze between the display and its surroundings (when the required physical content is far from the edge of the display), whereas the second strategy involves viewing the display and its surroundings in a single context (when the physical content is close to the edge of the display). In such cases, user performance and task loads can be influenced by the combination of changing vergence and accommodation (Fig. 4.1.1), changing accommodation (Fig. 4.1.2), changing vergence (Fig. 4.1.3), and when the vergence-accommodation conflicts (Fig. 4.1.2 in .3). If users do not accommodate or verge at the correct distance, the result may be a blurry image or double vision. However, it is unclear which of these has a greater effect on the user performance and task loads. Knowing which technology is better for assisting users interacting with AR content on the display and its surroundings within a short timeframe (while switching gaze or viewing the context as a whole), will increase the success rate of task completion and improve user experience.

To address this knowledge gap, we developed a system capable of recreating all four types of ML display (Fig. 4.1). We used this system to conduct two visual-acuity user studies ($n = 28$ and $n = 27$) under four vergence and accommodation conditions. During the first user study, we measured user performance and task loads in recognizing eye-test symbols, while rapidly shifting the gaze between the display and its surroundings. In the second study, users perceived the display and its surroundings as a whole and had to recognize eye-test symbols within a short viewing time. Our first finding indicates that when users rapidly shift their gaze between the display and its surroundings, the change in the accommodation distance (Fig. 4.1.1 and 2) significantly influences performance, whereas eye vergence does not play a major role. Our second finding indicates that when the display and its surroundings are perceived as a single context (as the gaze does not shift, the eyes always verge and accommodate the display), performance improves the most when the vergence distance of the ML display is close to that of its surroundings, such that there is no double vision (Fig. 4.1.2). The results indicate that stereoscopic displays provide an affordable solution for quick interaction between the display and its surroundings compared with other displays.

4.2 Background and Related Work

To investigate the impact of vergence and accommodation on the interaction between a display and its surroundings, it is essential to understand how these eye phenomena affect everyday vision.

4.2.1 Accommodation and eye strain

The process of bending light to focus it on the retina involves the contraction or relaxation of the eye lens, adjusting its convexity. When this accommodation process does not function optimally, objects near the eye (hyperopia) or at a certain distance (myopia) appear blurry. Myopia is a common phenomenon that occurs in approximately 23% of people [89], and varies with age, ethnicity and lifestyle. In addition, middle- and older-aged adults often have difficulty seeing things in proximity (i.e., presbyopia). Although these conditions can be corrected with convex or concave lenses, either mounted on spectacles or in the form of contact lenses, neglecting treatment may lead to eye fatigue, headaches, and overall impairment of daily activities. Moreover, the extended use of near-view displays such as computer monitors, tablets, and smartphones causes accommodation-related symptoms and subsequent eye discomfort [90, 91, 92].

In mixed-reality systems, virtual content is often displayed at a fixed focal distance, leading to a discrepancy in the accommodation distance between virtual and real environments, causing either to be out of focus. Existing research on AR systems that employ HMD focuses on mitigating this issue with varifocal technology [93], where the focal point is either mechanically changed or multiple focal planes exist. Near-eye light-field displays [94, 95] address the accommodation issue by rendering scenes from various viewpoints, resulting in different depths per viewing angle. Koulieris et al. surveyed state-of-the-art AR and virtual reality (VR) near-eye displays including those with unfixed focal distances[28]. They classified accommodation-supporting near-eye displays as varifocal/multifocal, multiplane, focal surface, and holographic, using a combination of lens optics and screen techniques. Techniques such as the Maxwellian view [96] project a virtual image onto a specific part of the eye retina, most commonly the fovea. This requires careful calibration, considering the positional relationship between the eyes and display and has been previously used in HMDs [97]. Interestingly, AR ML displays, which are commonly implemented in handheld devices or MLs, encounter similar accommodation challenges because the virtual content focal plane is fixed at arm's length. However, they received considerably less attention than their HMD counterparts, which partly motivated the proposed work.

4.2.2 Vergence and diplopia

Vergence is stimulated by the stereo disparity images created when the two eyes "collaborate" to converge images in a unified binocular vision. It primarily involves rotation of the eyes to ensure that the fixation area of the image falls precisely at the center of the retina in both eyes. However, vergence is also driven by blur and accommodation (see Sect. 4.2.4). Disparity in the images of objects that do not converge can be (voluntarily) perceived as double vision or diplopia. This effect is illustrated in Fig. 4.2 within the surroundings, when the gaze is fixed on the AR display. Although synthetic double vision has been used as a depth cue in AR [98, 99], HMDs commonly display disparate images to each eye to facilitate a sense of depth [44, 100, 101]. In mixed-reality studies utilizing projector displays, stereoscopic images are obtained with polarized glasses that filter frames intended for each eye [99, 102, 103]. Additionally, related works using (auto) autostereoscopic AR ML [41, 42, 43, 44] can produce similar depth perceptions.

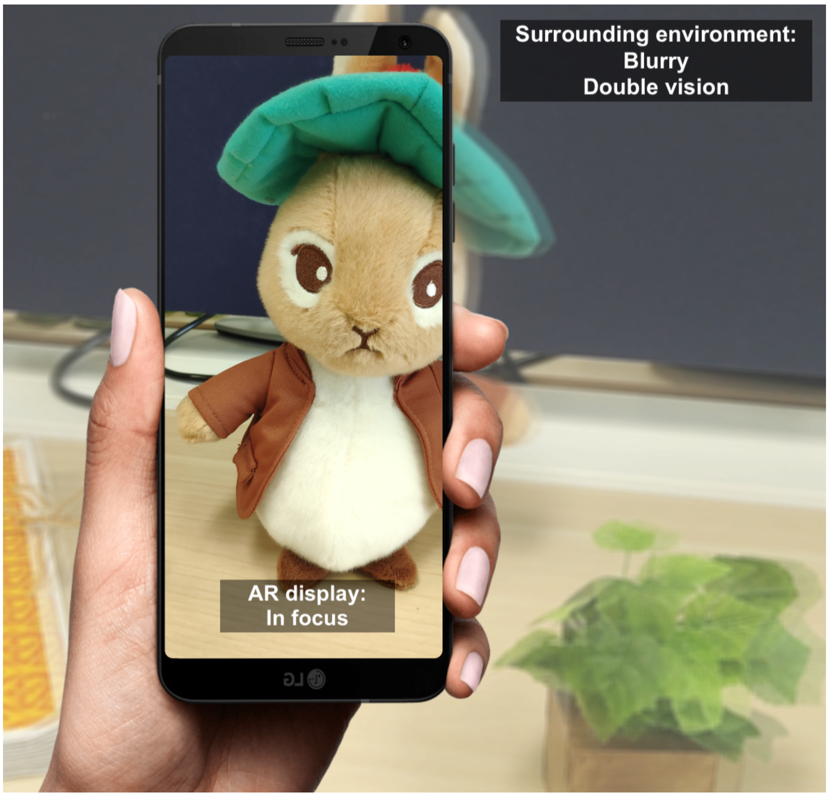


Figure 4.2: Display content as rendered from the perspective of the user. Focusing on the handheld display causes the surroundings to become blurred and doubled (diplopia) due to eye vergence and accommodation respectively.

4.2.3 Vergence-accommodation conflict

Numerous studies have focused on the fatigue and performance problems arising from the conflict between vergence and accommodation distances [23]. This problem is particularly prevalent in mixed-reality contexts, which are experiencing a surge in popularity. Both VR and AR using HMDs [3, 24, 25, 26, 27, 28] suffer from vergence-accommodation conflicts (VAC), which contribute to lower adoption rates in practical applications. The aforementioned study [28] addressed VAC by matching the binocular disparity of virtual imagery with optical focal cues at various depths. Discomfort studies using HMDs [29] have shown that only focus-adjustable lens designs can accommodate simulated distances to significantly improve comfort. In an ML system similar to that proposed in this study, researchers provided a 3D viewing experience through parallax images with directional rays coming from a Super multi-view (SMV) lenticular lens [104]. By equalizing the eye focus and convergence distance of the virtual image and measuring the accommodative response, they concluded that SMV can reduce the effects of VAC. However, the extent to which VAC would affect performance or task load in a pure stereoscopic ML (Fig. 4.1.2) or varifocal ML (Fig. 4.1.3), remains to be explored. Notably, researchers found minimal fatigue and discomfort with viewing distances at TV level (4.5 m) [105]. However, to the best of our knowledge, distances equivalent to arm's length have yet to be evaluated.

4.2.4 Interaction between Vergence and Accommodation

Eye vergence and accommodation form an interconnected visual system, meaning that changes in one influence or even drive changes in the other; this is the so-called accommodation-convergence reflex. The accommodation response can be driven by blur and stereo disparity between the two eyes. Similarly, accommodative vergence is a blur-driven response that converges or diverges from an eye. Studies have shown that, as accommodative responses deteriorate with age, the interaction in which vergence drives accommodation increases [106]. Therefore, it is expected that age has an impact on the ability to merge a near-view screen with its far-view surroundings, as well as any eye condition that influences the accommodation-convergence reflex. Blurry vision caused by insufficient accommodation affects tasks that require precision such as reading and writing, detailed work, and face (expression) recognition. However, double vision caused by incorrect vergence affects most tasks that rely on accurate binocular vision and depth perception, such as hand-eye coordinated tasks, reaching for objects, and driving. It was observed [107] that small-separation diplopia also negatively affects reading ability. Given that both phenomena affect reading, we decided to use the recognition of text characters to test accuracy in our forthcoming user studies.

Theories in the field of optics and ophthalmology describe eye vergence and accommodation reaction times to accommodative stimuli such as blur, apparent size, and distance. The authors of [108] reported an average reaction time of

0.62 and 0.56 s for “far-to-near” and “near-to-far” accommodation change, respectively. The convergence response times were considerably faster, averaging approximately 0.20 s. This observation is interesting, as it suggests that the accommodative response time is a bottleneck when the fixation point shifts between different distances. It follows logically that reducing the accommodation distance would make reaction times faster up to the point where the vergence time becomes an issue. To the best of our knowledge, these reaction times have not yet been studied in the context of AR ML displays. Therefore, in our first user study (Study A), we verify whether these response times are similar in the context of a rapidly shifting gaze between the ML display (near plane) and the surroundings (far plane). In the second user study (Study B), a fixed gaze was maintained to investigate the resulting effects from incongruous vergence and accommodation – blur and diplopia, respectively.

4.2.5 Magic-lens systems

Previous studies on ML systems have predominately focused on handheld devices owing to their accessibility and technological advantages. Contemporary smartphones and tablets are equipped with a range of built-in sensors that are useful in AR applications. There is a difference between an ML that simply visualizes an on-device camera image as a background [40] and an ML that visualizes geometrically correct views within the lens area, as seen from the user’s perspective [6, 45, 46]. Čopič Pucihar et al. [48] investigated both types, as shown in Fig. 2.6, and found that users consider the real environment and ML as separate views when using device perspective rendering. This perception also holds when a user-perspective ML is not sufficiently performant. In such cases, users will rapidly shift their gaze between the ML area and the surrounding real environment to interact with both rather than looking at the scene as a whole. This observation serves as the motivation for the first user study (Study A), in which the strategy of rapid gaze shifting was employed. Other optical see-through ML displays include heads-up displays (HUD) [109] used in aviation that allow pilots to see the runway even in bad weather conditions, and automotive industries [36, 37], to provide information such as speed or navigation on their windshield whilst driving. Particularly with the increasing prevalence of electric cars equipped with HUDs, there is a growing frequency of gaze switches between virtual information displayed at the HUD distance and real-world information. Consequently, understanding the impact of each visual system on performance becomes increasingly important.

4.2.6 Context and depth switching

Owing to the popularity of mixed reality, and its increasing usage for task support in the industry, previous studies have investigated the effect that switching depth layers (i.e. between the real physical environment and virtual imagery) has on human performance in visual tasks. Eiberger et al. measured task completion time and error rate on a combination of optical see-through HMD and a body-proximate display at 30 cm [110]. They found that during a visual search task, when content is on different depth layers (i.e. it is necessary to refocus and verge), performance significantly decreased compared to a visual search on a single depth layer. Using a monocular near-eye display, Gabbard et al. similarly showed decreased performance when focal distance needed to switch and that repetitions increased visual fatigue [111]. When replicated and extended these studies by using a binocular AR Haploscope, Arefin et al. additionally found that only increasing focal switching distance degraded performance in the binocular condition [112]. In these studies, researchers maintain consistent accommodation and vergence distances, likely due to the well-documented adverse effects of VAC. Consequently, the individual impacts of these visual systems (Fig. 4.1.2 and Fig. 4.1.3) on performance remain unclear, which is the main goal of this research.

4.3 System Design

We considered a typical scenario in which a handheld device was used as an AR system. The average near-working distance between the user's eyes and the smartphone screen is between 32 cm and 36 cm [113]. Accordingly, we fixed the user's head at 35 cm away from the ML *near display*. This ML *near display* was mounted and was not adjustable by the user. For the surroundings, we selected a

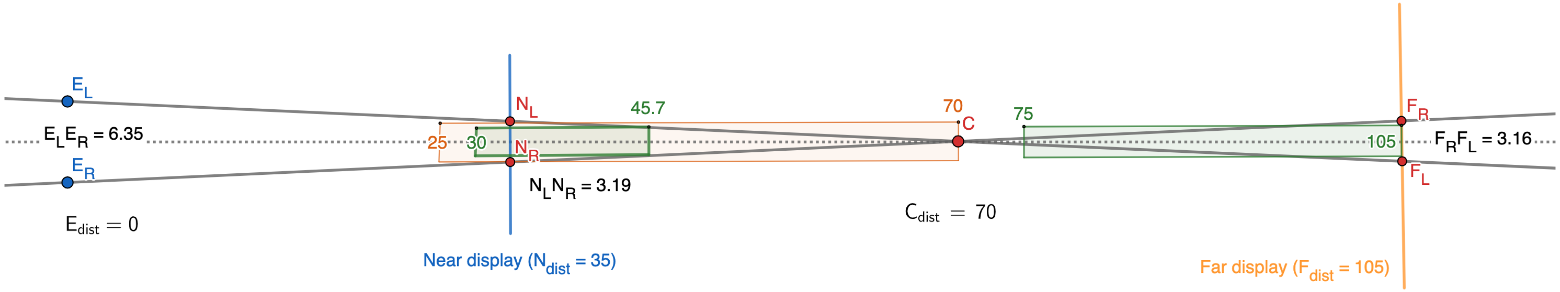


Figure 4.3: Binocular fusion limitations of stereoscopic images that contributed to how we built our proposed setup. Values in centimeters. Average human interocular distance is 6.35 cm between left eye E_L and right eye E_R, a *near display* is 35 cm away and a *far display* 105 cm away from the viewer. Green sections are the comfortable viewing ranges outside of which the viewer can experience difficulty merging the stereoscopic images. They are defined by the local minima and maxima of convergence point C where $N_L N_R < \frac{C_{dist}}{30}$ and $F_R F_L < \frac{C_{dist}}{30}$. The orange section is the stereoscopic distance range that could be obtained by participants in the user study using the proposed system.

plane 1 m away from the user, which is referred to as the *far display.* A top-down schematic is shown in Fig. 4.3.

4.3.1 Accommodation distance

Through defocused blurring, the eye can be accommodated on a surface that emits or reflects light. In the proposed setup, shown in Fig. 4.3, we were able to move the *near display* closer and further along the depth axis of the viewer (dotted line). Accommodation was controlled by altering this distance.

4.3.2 Vergence distance

In order to control the distance on which the user's eyes converge, we employed stereoscopic rendering. In this type of rendering, each eye is presented with a different disparity image. Binocular disparity links these two images for different eyes as one, forming a strong depth cue. One approach to achieve stereoscopic rendering is by using an autostereoscopic display that uses a parallax barrier, e.g. lenticular lens array, to visualize stereo images. However, when we want to adjust the accommodation distance, an autostereoscopic display poses problems because its viewing angle, in combination with the distance, is fixed. Furthermore, these displays are susceptible to crosstalk between eye images. An easier solution is to use a stereoscopic projector and adjust its focal length according to the *near display* (Fig. 4.4 and Fig. 4.5), in combination with active shutter glasses. By wearing active shutter glasses the user perceives a stereo image of the ML surface, allowing us to control the vergence variable.

4.3.3 Stereoscopic rendering

There are limitations to the depth effect created by binocular disparity images in stereoscopic rendering. As a rule of thumb, the interaxial distance should not exceed 1/30th of the convergence distance [114], assuming an interocular distance of 6.35 cm. In the proposed setup, this enables a range of 30 cm to 45.7 cm of comfortable stereo viewing for the *near-ML display* and a range of 75 cm to 105 cm for the *far display* (Fig. 4.3). The desired vergence distance of 105 cm away from

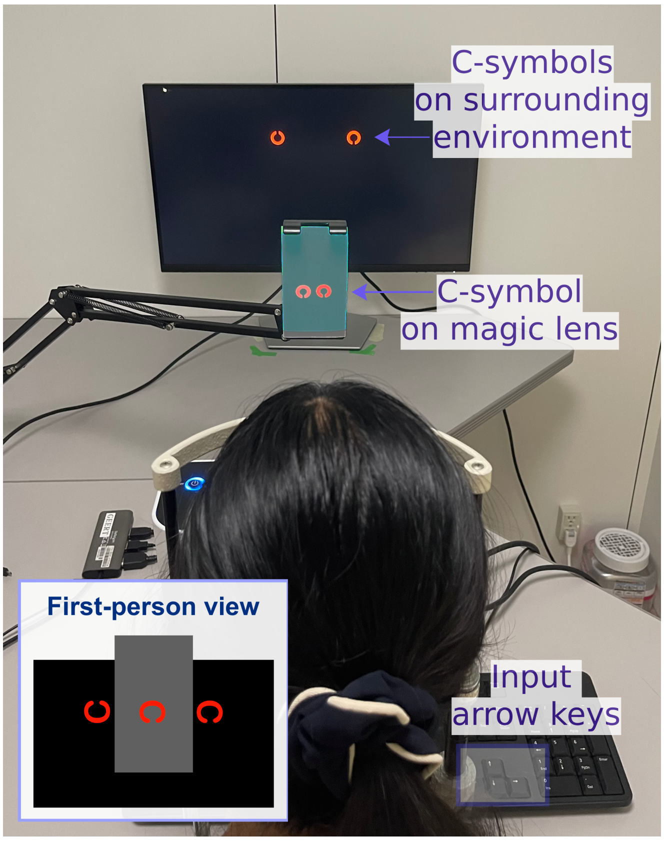


Figure 4.4: Experimental setup with the perspective from behind the participant, and their first-person view. The system is running a task for Study B, where three symbols are displayed: two outer symbols are displayed in the background on the surrounding display, whereas the middle symbol is stereo-projected onto the ML display. By wearing polarizing 3D glasses, the participant perceived only one symbol in the middle, with its depth determined by the amount of disparity. Arrow keys on a keyboard in front of the participant allowed for input of the direction of the openings of the C-symbols during the experiments. The diameter of the C-symbol on the ML display was 1.9 cm and the C-symbols on the surrounding display were scaled so that all three sizes appeared equal to the viewer.

the viewer projected onto the *near display* was not possible. Similarly, verging 35 cm away from the viewer while projecting onto the *far display* was not possible.

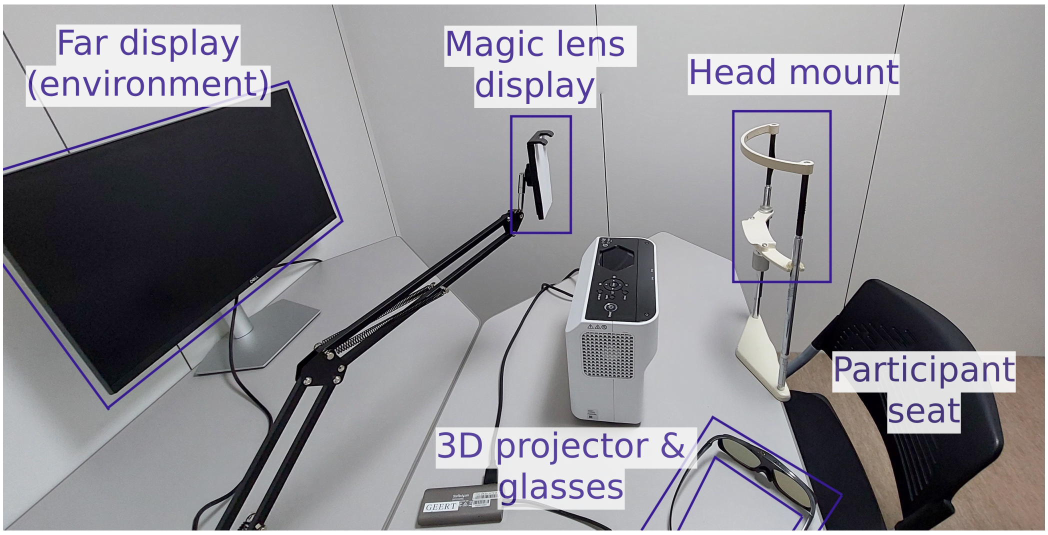


Figure 4.5: Experimental environment from the side. The distance from the *far display* (surrounding environment) to the ML was 70 cm, and the distance from the ML (*near display*) to the head mount was 35 cm. The short-throw 3D projector on the table projects a stereoscopic image onto a white sheet (6.5 cm × 12 cm) mounted on a movable clamp, representing the near display.

We performed a pilot study with nine participants using the described system, in order to determine acceptable values for eye-acuity symbol size and horizontal distance, as well as color and disparity capabilities. Using a trial and error approach in which the examiner continually adjusted the binocular disparity, it was found that the range in which the user could perceive a symbol clearly in a 3D space ranged from approximately 25 cm to 70 cm from the viewer in the *near*-accommodation *display* setup (orange range in Fig. 4.3). In the *far*-accommodation setup, the clear viewing range was approximately 75 cm to 105 cm from the viewer. These values varied slightly per participant, depending on eye health, age, ability to focus, and interocular distance and increased with familiarity and practice. Therefore, calibration of each user's stereoscopic ability and practice before using the system is recommended and is part of the forthcoming user studies.

4.3.4 Eye-acuity symbols

We measured eye acuity by the participant correctly discerning three symbols in a row. A common practice in eye examinations is to use Snellen chart symbols [115], which consist of capital letters progressively becoming smaller. In the aforementioned pilot study, where Sloan letters (a subset of the Snellen chart) were used, participants needed a considerable amount of time to input their answers on a keyboard and to take their head off the mount to have a clear view of the keyboard letters. When testing the verbal confirmation of the answer letters, accidental input of an incorrect answer was more likely due to confusion between participant's intended answer (pronunciation and shared attention) and the examiner's interpretation of the answer. Therefore, we used Landholt C-symbols [116] to test eye acuity. These consist of C-symbols that can have openings in one of the four directions: up, down, left, right.

We wanted a high contrast between the symbols and background on the surrounding display to make it easier to discern the openings of the symbols. Therefore, we dimmed the light and used a black background with red colour symbols. Red was found to be a good trade-off between contrast and the resulting ghosting effect from stereoscopic projection; when using white symbols (highest contrast), some users were able to perceive two images on the stereoscopic display instead of one 3-dimensional image.

When tasked with discerning symbols, their size is an important factor. Together with the viewing time, these two variables would make our user study complex and would require long experimentation times to gather sufficient repetitions per condition. Therefore, we used C-symbols of a fixed size (1.9 cm in diameter) that were determined in the pilot study to be just discernible (visual angle of approximately 5 degrees) on the far-surrounding display while focusing on the ML (see Fig. 4.4). Increasing this minimum size is expected to positively affect symbol recognition. This mimics a real scenario in which the user looks at the display and its surroundings in a single context.

4.4 Experimental Design

4.4.1 Hypotheses

Our objective was to measure visual-acuity under different conditions of vergence and accommodation distances, influenced by different types of ML displays located at arm's length and far surroundings. To interact with the (augmented) content on the ML display and its surroundings, we investigated two strategies: (1) rapid gaze shifting between the ML display and its far surroundings and (2) fixed gaze on the ML display while looking at the display and its surroundings in a single context. According to the literature [108], the accommodation response is significantly slower than the vergence response when gaze fixation moves from near-to-far or from far-to-near. Therefore, we hypothesize as follows:

> **(H1)** When rapidly shifting gaze, reducing the eye accommodation distance of a ML display in relation to its surroundings, results in faster interaction with AR content, more so than reducing the vergence distance.

Using the second strategy, eyes do not re-accommodate nor reconverge on different focal distances. However, artifacts resulting from misaligned accommodation distance (blur) and vergence distance (double vision) persist. Therefore, we hypothesize as follows:

> **(H2)** When gaze is fixed on a ML display, reducing its eye vergence and accommodation distance in relation to its surroundings will result in a more accurate merging of the AR content.

Given the prevalent findings in the literature (Sect. 4.2.3) regarding fatigue and eye strain in VR scenarios caused by vergence–accommodation conflicts, we also expect the following:

> **(H3)** A mismatch between the eye accommodation distance and vergence distance of a ML display requires the viewer to concentrate more, resulting in a higher perceived task load and less accurate interaction between the AR content and its surroundings.

4.4.2 Experimental conditions

In the upcoming user studies, we used a 2 × 2 within-subjects design with four combinations of vergence and accommodation variables, each of which could be *near* or *far*. In the accommodation-near condition, a stereoscopic image was projected onto the ML display (example seen in Fig. 4.4), and the vergence distance was either on the display (*near*) or 70 cm away from the participant (*far*). In the accommodation *far* condition, the stereoscopic image was projected onto the surrounding display, and the vergence distance was set to either the surroundings distance of 105 cm (*far*) or 75 cm away from the participant (*near*). These two conditions were counterbalanced among the participants.

4.5 User Study A: Shifting gaze

4.5.1 Apparatus

For the surrounding display we vertically mounted a 1 m high by 2 m wide white plastic panel on a table. Projection onto this panel was performed using an Optoma WU515ST projector. In front of this panel was positioned another 0.15 m × 0.15 m white plastic panel that functioned as the ML display. We used a Barco F50 WQXGA projector to display the stereoscopic side-by-side images. The 3D projector's image was projected onto the surrounding display in the accommodation-*far* condition. To observe the stereoscopic image, we used active shutter glasses with digital light processing (DLP) technology. See Fig. 4.6 for an overview.

4.5.2 Task

Participants were directed to use arrow keys on a keyboard placed in front of them to indicate the direction of the openings of three C-symbols, proceeding from left to right. For example, in the round shown in Fig. 4.4, the participant presses the up key, down key, and down key again. During this first experiment, the left and right C-symbols were placed horizontally 1 m (50.8 degrees) apart behind the ML display. This was determined to be sufficiently wide for the participants

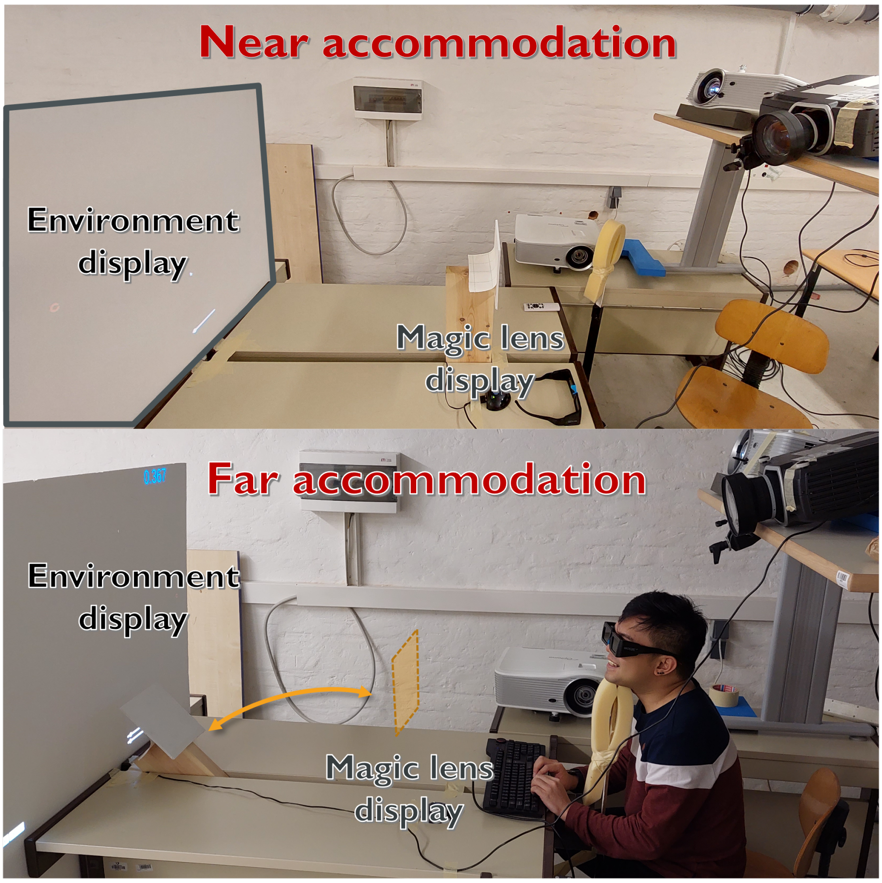


Figure 4.6: The experiment set-up for User study A. The top figure shows the *near* display configuration: One projector displays imagery on a far screen (environment display) and a second, side-by-side stereo projector projects imagery on a near screen (Magic lens display). The bottom figure shows the *far* eye accommodation configuration where the physical near display has been removed from view. Both projectors provide imagery on the far (environment) screen, but stereo projection creates the illusion of a near display (shown in orange).

to be forced to shift their gaze between symbols.

Following a three-second countdown shown in the top-center of the surrounding display, the three C-symbols would appear simultaneously in random orientations. Participants were asked to focus on the position of the left symbol during the countdown and, as soon as the symbols appeared, to shift their gaze from left to middle and then to right while pressing their choice of arrow key (three times). For each symbol, we captured response correctness and response time.

4.5.3 Participants and procedure

We recruited 28 university students (14 males and 14 females) aged between 18 and 32 (M=21, SD=3.25). Five participants had prior experience with mixed-reality systems, whereas 22 had no or limited experience. We verified good visual acuity (with correction if necessary) in all participants.

This study consisted of two phases: preparation and data-gathering. In the preparation phase, the participants were asked to sit on a chair in front of the experimental setup, and the intention of the experiment was explained. They were then asked to wear shutter glasses, and the light in the room was dimmed. The examiner checked their ability to see the stereo depth by displaying a C-symbol on the ML display and asking the participant to estimate their distance from them. This was performed with various degrees of stereo disparity and was repeated on the surrounding display.

In the data-gathering phase, the examiner adjusted the setup according to the counterbalanced conditions of accommodation and vergence (see Fig. 4.6), followed by a practice round. This round consisted of 10 repetitions of the task described in Sect. 4.5.2, during which participants were allowed to ask questions. Following sufficient practice, when ready, participants were asked to focus and complete the task 40 times as quickly as possible. When finished, the participants were asked to take off their shutter glasses, and rest their eyes, during which the examiner re-adjusted the experiment setup according to the counterbalanced condition of accommodation and vergence. Then the second phase (practice and task completion) was repeated.

The entire procedure lasted approximately 30 minutes including three breaks of 3 minutes. The study received prior approval from the institutional review board.

4.5.4 Results

In Fig. 4.7, the average response time per condition of vergence and accommodation is visualized for only rounds in which all three symbols are correctly answered. We applied a two-way–repeated-measures analysis of variance (ANOVA), which revealed no statistically significant interaction between eye vergence and accommodation ($F_{1,27} = 0.122,\ p = 0.730$). A simple main effects analysis showed that accommodation had a significant effect on response time ($F_{1,27} = 8.711,\ p < .01$), whereas eye vergence did not ($F_{1,27} = 1.768,\ p = 0.195$). The average round response time in the conditions in which the eyes had to re-accommodate when shifting gaze to and from the ML display (accommodation distance = NEAR) was 1.510 s (SD = 0.258), and when they did not re-accommodate (accommodation distance = FAR), the average response time was 1.431 s (SD = 0.203).

4.6 User Study B: Fixed gaze

4.6.1 Apparatus

We used a Dell desktop monitor in front of which was a device holder mount with a white surface that served as a ML display surface. We used a Ricoh PJWX4153N projector with stereoscopic side-by-side projection in combination with BenQ YDD3PG active shutter glasses that worked with the DLP-link technology of the projector. The two displays were placed in a row on two tables and a head mount was attached to the front table. The mount maintained the participants' head stability in a calibrated position throughout the experiment. The Dell desktop monitor functioned as the surrounding display and was replaced by the stereo projector in the *far*-accommodation condition. The setup for the *near*-accommodation condition is shown in Fig. 4.5.

4.6.2 Task

Participants were instructed to input the opening directions of three C-symbols using arrow keys on a keyboard in front of them. In this study, the symbols on the surrounding display were placed horizontally close to the ML (from the

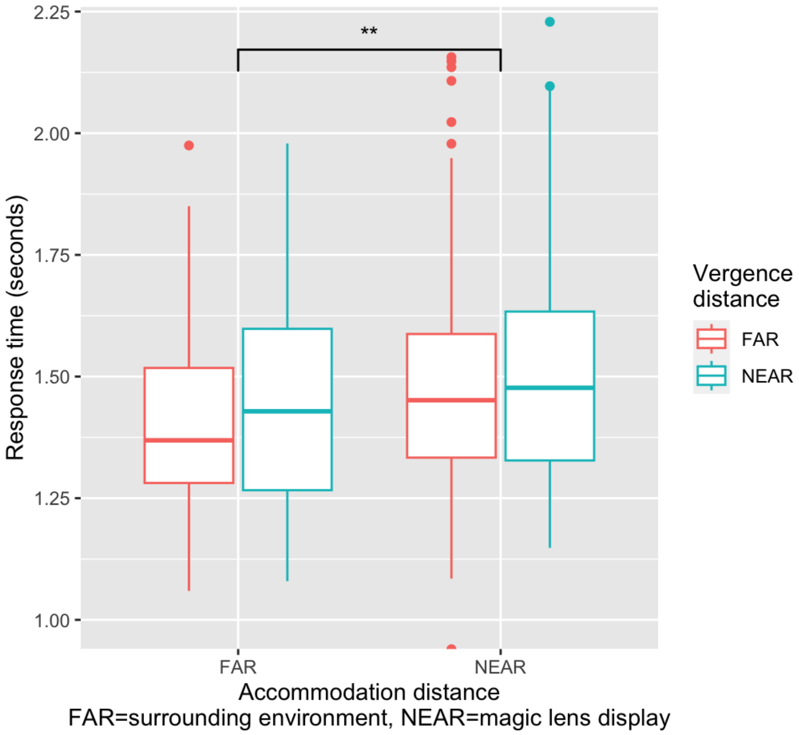


Figure 4.7: Study A performance results. The graph shows the average response time when shifting gaze between symbols on the surrounding (far plane) and a ML display (near plane) under four conditions of ML vergence distance and accommodation distance. Altering the accommodation distance of the ML display has a significant ($p < .01$) effect on response time, while vergence does not.

perspective of the user; Fig. 4.4), such that they were as close as possible to the participant's foveal vision.

The three C-symbols were displayed simultaneously for a duration that progressively decreased as the rounds advanced. In the pilot study (n=9), we established that all participants could achieve near 100% accuracy given > 0.5 s of symbol visualization time under any condition in our setup, with prior practice. When the visualization time was < 0.1 s, accuracy sharply declined and varied significantly from participant to participant, approaching the accuracy of the majority classifier. Therefore, the visualization duration for the C-symbols was set to start at 0.5 s, with 0.05 s decrements ending at 0.1 s, with 10 repetitions per visualization time, resulting in a total of 80 repetitions per condition. Participants were instructed to always focus on the *middle symbol*, and not to change their eye focus to the left or right symbols. During the study design, we discussed using an eye tracker to verify the participant's eye focus. However, the active shutter glasses prevented us from using screen-based eye-tracking technology. Wearing both shutter glasses and eye tracker glasses proved impractical, especially if additional correctional lenses were used. When the C-symbols are not visible, a question mark (?) was displayed in the middle symbol position so that the participants could focus on the accommodation and vergence distances of the current condition. In this study, we captured the symbol response correctness (three binary points per round).

4.6.3 Participants and procedure

We recruited 27 participants (18 males, 9 females) aged between 22 and 39 years, with a mean age of 28.1 years (SD=5.01). All participants were graduate university students, 3 of whom had in-depth knowledge of mixed reality, 6 had some experience, and 18 had no experience. Most participants (26) had previously wore shutter glasses to view a 3D movie and had no problems with depth-effect perception. We verified good visual acuity (with correction, if necessary) in 26 participants; one participant was excluded because of astigmatism.

This study consisted of three phases. In the first phase, participants were asked to fill out a demographic information questionnaire, and their eye acuity was discussed focusing on: any history of eye conditions, what is their prescription of glasses or contact lenses, and are they currently wearing corrections. Next, they were asked to sit on a chair in front of the experimental setup, and the intention of the study was explained. The participants were assured of three breaks between the next four rounds. After further clarification when requested, the room was dimmed.

In the second phase, the participants were asked to put on the stereo shutter glasses and put their chin on the head mount, with their forehead touching the top of the mount (see Fig. 4.4). The examiner then verified their ability to see stereoscopic depth by displaying a Landholt C-symbol on the ML display with zero disparity, and asked the participant to indicate how far the symbol appeared from them. The examiner then increased the disparity to simulate the far condition, and repeated the questions accordingly. The participants were familiarized with their tasks (section 4.6.2) followed by a practice round. During practice, the symbols were visible for 0.5 s and repeated 20 times, taking an average of 4 min.

The third (main) phase was repeated four times, once for each vergence and accommodation condition. First, the examiner sets up the ML display and the stereoscopic rendering according to the current conditions. After confirming that the participant was ready, a three-second countdown was displayed on the surrounding display, after which three C-symbols appeared in random configurations. After the viewing period elapsed, the symbols were substituted with question marks, prompting participants to input the directional openings of the three C-symbols using the arrow keys. This was repeated ten times, after which the viewing time period was decreased by one step. When it reached 0.1 s, the text "Finished" was displayed, and the condition round ended. Following each condition, participants were asked to take off their glasses, step outside the dimmed room to rest their eyes for 5 min, and complete the NASA TLX questionnaire. They were also asked to grade the condition between 1 and 4, based on how well they thought they performed, where 1 = worst and 4 = best.

4.6.4 Results

Fig. 4.8 (a) shows the proportion of correct responses per visibility duration in seconds, for all four combinations of conditions. The legend shows the vergence and accommodation distances for each display type; for example "Stereoscopic (NEAR–FAR)" is a display condition where accommodation is set to the near plane and vergence is set to the far plane. The vertical axis shows the accuracy of discerning the opening of C-symbols with a visible duration in seconds on the horizontal axis. As the visible duration decreased, the accuracy decreased as well. In Fig. 4.8 (b) this result is averaged over all visibility periods.

Owing to the non-normal distribution of accuracy data, we applied an Aligned Rank Transform followed by a two-way–repeated-measures ANOVA. The results showed that changing the eye vergence led to a statistically significant difference in accuracy ($F_{1,25} = 35.801$, $p < .001$). Moreover, changing the eye accommodation distance also led to a statistically significant difference in accuracy ($F_{1,25} = 6.10$, $p < .05$), but there was no significant interaction between vergence and accommodation.

Task load and questionnaire

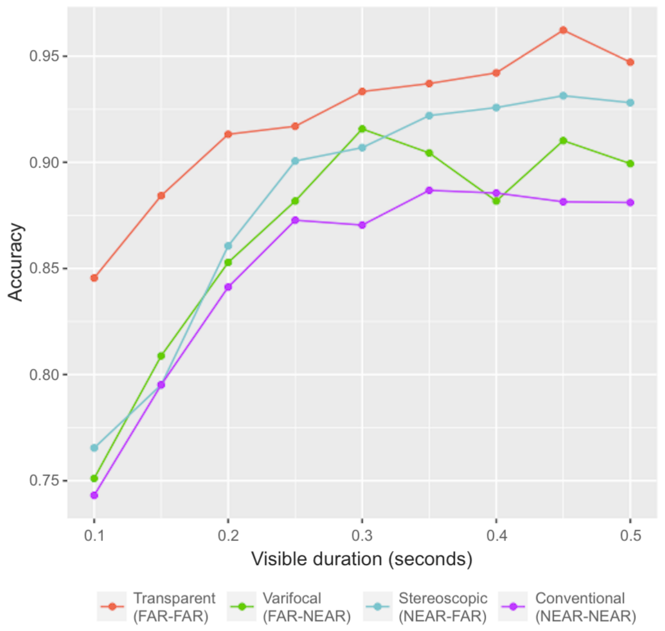
The NASA TLX questionnaire scores were averaged, as shown in Fig. 4.9. The display types had similar mean scores, with the transparent display having the lowest perceived task load (M = 8.82, SD = 4.73). The conventional (M = 11.26, SD = 4.87) and varifocal (M = 11.15, SD = 4.91) displays had similar higher values, whereas the stereoscopic display had a value in the middle (M = 10.69, SD = 5.01). We applied a two-way ANOVA, which showed that eye vergence had a significant impact on task load ($F_{1,99} = 7.396$, $p < .01$). In contrast, accommodation only showed weak evidence of influencing task load ($F_{1,99} = 2.973$, $p < .1$). The analysis revealed no significant interaction between vergence and accommodation.

Table 4.1 presents the participants' subjective grading of the four conditions, between 4 (best) and 1 (worst). The subjective grade was the highest (M = 3.58, SD = 0.86) for the transparent display and the lowest for the conventional display (M = 1.65, SD = 0.89).

Table 4.1: Subjective grading (1 to 4) of performance per condition and display type for Study B

Display type	Condition		Mean grade	SD
	Accommodation	Vergence		
Transparent	FAR	FAR	3.58	0.86
Varifocal	FAR	NEAR	2.15	0.92
Stereoscopic	NEAR	FAR	2.62	0.85
Conventional	NEAR	NEAR	1.65	0.89

(a)



(b)

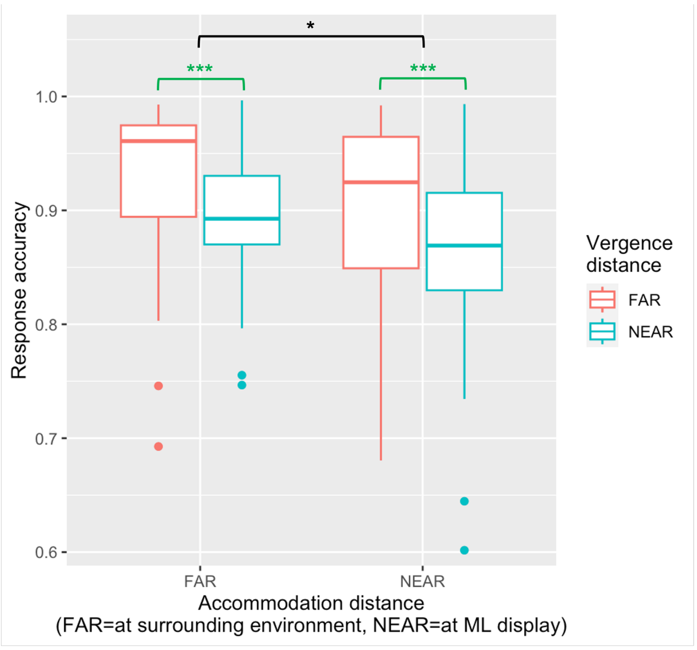


Figure 4.8: Study B results. (a) Accuracy results per display type for symbol visibility periods between 0.1 and 0.5 seconds. (b) Accuracy results averaged over all visibility periods. *** : $p < 0.001$, * : $p < 0.05$

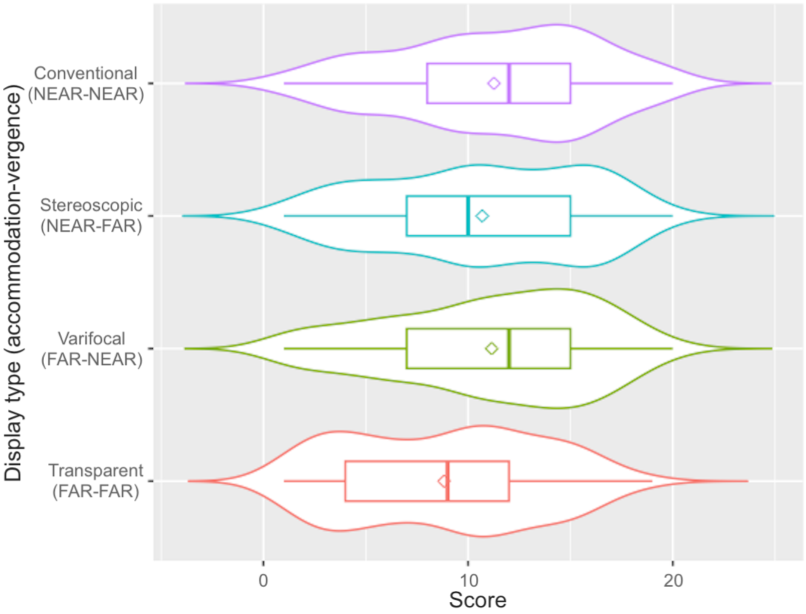


Figure 4.9: Results from the NASA TLX questionnaire on task load per condition averaged over all subscales for Study B. The outer edge of the violin plot shows the relative number of responses for each score.

4.7 Discussion

To interact with the (augmented) content on the ML display and its surroundings, we investigated two strategies: (1) rapid gaze shifting between the ML display and its surroundings and (2) fixed gaze on the ML display while viewing the display and its surroundings in a single context. To better understand how different conditions of vergence and accommodation distances affect user performance, we conducted two visual-acuity experiments, which are discussed hereafter.

4.7.1 Interacting with AR display and its surroundings by rapid gaze shifting

In our first hypothesis (H1), we predicted that when rapidly shifting the gaze, reducing the eye accommodation distance of the AR ML display in relation to its surroundings would result in a faster interaction between the AR content and its surroundings, more so than reducing the vergence distance. The results of Study A support this hypothesis: When the accommodation distance between the display and its surroundings was reduced, a significant reduction in symbol reading time was observed (on average 0.077 s). This decrease in reading time demonstrates a faster interaction between the AR content and its surroundings. However, there was no significant impact when the difference in the vergence distances was reduced. This is in line with the literature [108], where the accommodation response was found to be substantially slower than the vergence response when gaze fixation moved from near-to-far or far-to-near. This explains why accommodation was the prevailing human factor in this visual-acuity experiment. We furthermore observe that, in line with prior studies, focal distance switching caused reduced performance [111, 112] and switching both visual distances had the worst performance [110].

However, it is still interesting to note that when the eyes did not have to reaccommodate when shifting gaze, decreasing the vergence distance did not produce a significantly faster response time (Fig. 4.7: FAR-accommodation distance plots). It is possible that our experimental design did not have a sufficiently high resolution to capture small effects. Nevertheless, these results show that, in AR support tasks that require rapid shifting of gaze fixation between a display and its surroundings, matching the accommodation distance of the augmented content with that of its surroundings accelerates the task. For instance, a varifocal display (third image in Fig. 4.1) can help read augmented instructions on a surgery support display, where a surgeon often and rapidly shifts between AR instructions and its surroundings. While this type of display might induce discomfort owing to the vergence–accommodation conflict, we did not find evidence of its significant impact on cognitive-task load or task performance (i.e., no statistical analysis showed an interaction effect between vergence and accommodation). This suggests that the vergence-accommodation conflict is not a prevalent human factor in such scenarios.

4.7.2 Interacting with AR and its surroundings by viewing both as a single context

In our second hypothesis (H2), we predicted that when the gaze is fixed on a near-ML display, reducing eye vergence and accommodation distance in relation to its surroundings will result in a more accurate merging of AR content and its surroundings. This was based on the assumption that the resulting blurring and double vision from disparate vergence and accommodation, respectively, hinder accurate detection of the surroundings. The results of Study B provide evidence to support this hypothesis. The eye-acuity symbol identification accuracy was the highest when both eye vergence and accommodation distances were similar between the *near display* (ML display) and the far surroundings. For example, accuracy is best for a transparent display (Fig. 4.8, FAR-FAR condition), and worst when both distance differences are largest, as is the case when using a conventional ML display (NEAR-NEAR condition). This effect holds true over short and relatively long durations of an interaction switching task and increases as the task time shortens (the red line is far above other lines in Fig. 4.8 for short durations).

When the ML display acted like a stereoscopic display, rendering content at a vergence distance close to the surrounding distance (Fig. 4.8: accommodation=NEAR and vergence=FAR), also resulted in an improvement in accuracy over the conventional ML display. This suggests that simply using a display with stereoscopic capabilities and rendering the content close to the detected surroundings would allow for a more accurate user experience when the task involved requires the viewer to see both the surroundings and AR ML content simultaneously. Rendering content at a distance where eye accommodation is close to the surroundings also improves accuracy (FAR–NEAR), but to a lesser degree than vergence. This advantage of stereoscopic displays is beneficial because varifocal displays, which can change accommodation distances or have a specific set of focal planes, are less common than stereoscopic displays and often require large optics. Furthermore, stereoscopic displays are relatively affordable in contrast to varifocal displays. The results also allowed us to conclude that double vision was the prevalent human factor compared to blurred surroundings in our scenario, which was the case when users did not accommodate or verge at the correct distance.

4.7.3 Vergence-accommodation conflict

We also hypothesized (H3) that an AR ML display with mismatched vergence and accommodation distances would require higher physical and mental demands and would result in a less accurate interaction with AR content and its surroundings. Our results did not indicate a significant difference in the task load when vergence and accommodation were mismatched. On the varifocal and stereoscopic displays, the aggregated task load score (Fig. 4.9) was only slightly higher than the score for a transparent display where the conditions matched (FAR–FAR), and there was no significant difference from the conventional display (NEAR–NEAR). We observed an overall lower task load score in the transparent condition, as expected. The eye vergence and accommodation distances were identical for the *near* display and matched the surroundings. However, the scores were extremely dispersed under all conditions. This high standard deviation indicates that our method of measuring eye comfort is less predictable and difficult to generalize. However, a larger dispersion is expected in subjective data, and additional analysis is necessary. However, our results indicate that mismatching vergence and accommodation distance do not affect the cognitive-task load when users are required to merge their surroundings with an ML display. Furthermore, as already mentioned, no interaction effect of vergence and accommodation was detected in Studies A or B, again suggesting that vergence–accommodation conflict was not the key to accurate interaction with AR content and its surroundings.

4.7.4 User preference

Finally, when asked in Study B, under which condition users preferred to obtain the most correct answers, participants ranked the transparent display condition as the highest (Table 4.1). This was again expected, as the eye strain was the lowest, and the measured accuracy was also the highest under this condition. The second-most preferred display was stereoscopic, as in the NEAR–FAR condition. From the questionnaire responses, it seems that the participants had fewer problems with content being out of focus (as was the case in the non-matching accommodation distance condition) than with double vision. This binocular diplopia occurs when the eyes converge in front of or behind a focal plane, as is the case under non-matching vergence conditions. This further supports our recommendation for utilizing a stereoscopic ML display with a matching vergence distance, because when a strategy of fixing the gaze on the ML display is employed, the AR content on the display can be merged faster and more accurately with its surroundings.

4.8 Summary

In this chapter, we investigated eye vergence and accommodation distances in a typical scenario where users interacted with the content on a ML display at arm's length, as well as with its surroundings, in a short timeframe. We discussed the issues posed by these visual processes in contemporary mixed-reality displays and highlighted the lack of related materials concerning ML displays.

Two fundamental visual-acuity user studies were conducted in which both visual processes were compared by changing the display distances between near (arm's length) and far (1 m, similar to an office desk environment). In the first study, users interacted with the content on a near-ML display and its surroundings by rapidly shifting their gaze. We found that eye accommodation was bottlenecked and that reducing the distance of accommodation decreased the time needed to identify eye-acuity symbols. In the second user study, users focused on the ML display to view it and its surroundings as a merged, single context. We found that minimizing the eye's vergence distance discrepancy helps users most in accurately identifying eye-acuity symbols. Additionally, minimizing accommodation distance had a positive effect; however, the extent was less than that in the first study's results. These results coincided with the participants' subjective task performance and preferences. Thus, in a situation where the user has to frequently or rapidly compare content on a *near display*, such as a handheld device with AR support, with the physical surroundings, it is beneficial to reduce or equalize the stereoscopic vergence distance of that content, as well as its accommodation distance, in relation to the surroundings. Furthermore, if there are negative effects resulting from conflicting vergence–accommodation in a ML setup, they have no significant impact on the cognitive task load, nor are they detected as key to accurate interaction with the AR ML and its surroundings within a short timeframe.

4.8.1 Limitations

One limitation of our studies was the dependence on the size of the symbols and the accuracy of discerning them. The symbol sizes were determined in a pilot study (Sect. 4.3.3) and remained consistent throughout the experiments. However, it remains uncertain whether the accuracy is still affected by minimizing the eye accommodation and vergence distances when using larger-sized symbols. Furthermore, it is possible that visual tasks with many details, such as reading small texts, benefit more from minimizing the accommodation discrepancy, whereas depth-heavy tasks benefit more from minimizing the vergence discrepancy. This relationship should be investigated further in future studies.

Another limitation of our study design is the cognitive load of discriminating a symbol and matching it with the correct input key. It is easier for a user to input their answers when the eye acuity symbols are all equal or when two symbols (orientations) match. In future work, we hope to separate this difficulty factor.

Finally, although participants were given task instructions for shifting their gaze (Study A) or focusing on a single point (Study B), we used limited methods available to verify that these instructions were strictly followed. Although we did not find an uneven distribution of accuracy over the three symbol locations, employing eye tracking in future studies would be beneficial.

4.9 Future work

4.9.1 Diplopia

As highlighted in the discussion, stereoscopic disparity also causes diplopia, that is, double vision of objects or environments on which the eyes are not verging. In cases where the focus point in the surroundings is visible to one eye but occluded by the display for the other (Fig. 4.10), diplopia may hinder the ability to merge the two views. In future work, we plan to investigate the effects of creating a truly monoscopic ML (e.g., blackening the image on the ML for one eye or only for the border-case eye, as shown in Fig. 4.10, red line). It would be interesting

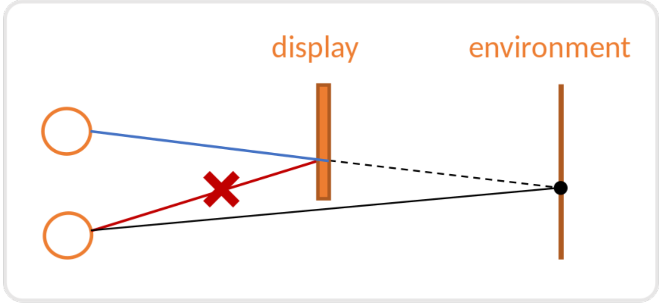


Figure 4.10: Diplopia (double vision) problem in ML displays. In edge cases, the surrounding environment is visible for one eye (black) but blocked by the display for the other eye (blue). Owing to stereoscopic disparity, focusing on the farther surroundings causes double vision (blue and red) of the closer display. Blackening the image of the display for one eye (red) or rendering it transparent may alleviate this problem.

to verify whether rendering content from a user's perspective has any impact on the performance shown in this study.

4.9.2 Depth perception

Vergence and accommodation are strong depth-cue providers. In this study, we did not focus on depth perception. Our results suggest that decreasing the difference between accommodation or vergence distances as displayed on an AR display and real surroundings will improve a viewer's task performance, but does not take into account the effect that this has on depth perception. Future work will need to verify the trade-offs between merging performance and depth-cue quality, based on the type of action performed.

4.9.3 Varying the far plane distance or context disparity

In the design of our experiments, we set the near plane at 35 cm from the user's eyes and the far plane at 105 cm. These distances are based on a common scenario in which a user interacts with their smartphone or tablet which augments content on a desktop in front of them. Whereas the near plane distance will have only slight variation (±5 cm) in common scenarios, the far plane is expected to have a much larger range. Cutting [117] defines the *near-field*—the distance slightly beyond arm's length at which the hands could still manipulate objects—to extend to about 1.5 m. Within the near-field, perceived depth distances are almost veridical. However, in scenarios where AR-supported tasks on a ML display are performed with haptic interactions (by hand or tool) without moving the ML display, the far plane distance is expected to be between 0.3 and 1.0 times arm's length, or approximately between 40 cm and 65 cm. We designed the far plane distance farther than this *personal workspace* to exaggerate the change needed by the accommodation-convergence reflex when shifting gaze between the two planes. This was to clearly show the effect of blur and double vision as a result of the disparate (70 cm) focal planes. We discuss the effect of shortening the distance between focal planes with both context integration tactics hereafter.

While shifting gaze

Our results indicate a round-trip (far-to-near then near-to-far) accommodation-convergence response time of 1510 ms, resulting in 21.57 ms/cm. If we can assume that the speed of the accommodation-convergence reflex scales linearly with distance (within converging or diverging conditions respectively), the response time for round-trip shifting gaze at personal workspace distance of 50 cm away would be 324 ms. In the optimal scenario (both accommodation and vergence distances of the ML display near to the far plane), the response time would lower to 289 ms under the conditions and environment described in Sect. 4.5. This improvement in response time may seem slight, however in scenarios where rapid gaze shifting happens frequently, it can cascade in a significant time gain.

While keeping a fixed gaze

Our results indicate an 11% accuracy improvement when integrating symbols on a near and far plane with 70 cm disparity, using a conventional ML display versus a transparent ML display. It is difficult to estimate whether or not this improvement is equal for all near-field distances. Future work should verify the experimental results of study B (Sect. 4.6) under varying far plane distances.

Influence on fusing disparity images

Both comparisons are speculative because the performance of fusing disparity images improves as we move the far plane (the workspace) closer to the near plane (the ML display), see green areas in Fig. 4.3. It is expected that shortening this distance would improve performance in both user studies in conditions where the vergence distance is set to the far plane, disproportionately to the performance gain from less blur and diplopia. However future research should confirm this.

5

Combined Study Analysis and Conclusion

In this chapter, we combine the results and discussions of the topics in Chapter 3 and Chapter 4 (hereafter: Study 1 and Study 2, respectively) to evaluate how they contribute to our overall goal of a true transparent magic lens. We discuss the implications of the integrated topics and their limitations. Then, we conclude this dissertation by summarizing its findings and contributions. We also propose the next steps for this research and explore future work.

5.1 Summary of Findings

5.1.1 Effects of perspective, haptic feedback, and visual context on AR spatial understanding

In Study 1 we created a user-perspective rendering ML display using off-the-shelf hardware. We investigated the accuracy of haptic interactions with virtual content and the physical world on an ML display. Our findings are as follows:

- Accurate visualization of the hand or tool on the ML display drastically improves haptic interaction accuracy.
- Cognitive awareness of the physical environment—through visual contextual cues and haptic interactions—improves spatial interaction accuracy on an ML display.
- User-perspective rendering on an ML display immediately improves spatial interaction accuracy, even without cognitive awareness of the physical

environment.

- Interactions with tactile feedback are more accurate than those without it, with user-perspective rendering providing a marginal improvement.

5.1.2 Effects of vergence and accommodation on AR context integration

In Study 2 we investigated visually interacting with content on an ML display and its physical surroundings, in a close-range scenario given a short timeframe. Specifically, we proposed two common integration strategies: rapid gaze shifting between the two contexts, and fixed gaze viewing them in a single context. Our findings:

- Eye accommodation is the prominent factor while shifting gaze between an ML display and its surrounding context. Reducing the distance of accommodation significantly improves visual acuity.

- Eye vergence is the prominent factor while viewing an ML display and its surroundings as a single merged context. Reducing the vergence distance improves visual acuity the most, whereas reducing accommodation distance has a lesser positive effect.

- Any effects from the vergence-accommodation conflict (VAC) have no impact on cognitive task load, nor play a role in accurate visual interaction with an ML display and its surroundings.

5.2 Discussion of the Combined Studies

Both studies considered the ML display and the physical environment as a single merged context. This came forth out of the desire for a true transparent ML display with seamless integration into the real environment. The results of the studies complement each other:

- User-perspective rendering provides the transparent effect of the ML display.

- Stereoscopic and varifocal visualization provides integration of the ML display in the depth direction.

Although we have not yet investigated an ML display that combines both functionalities simultaneously (see Sect. 5.4), we hypothesize that such an ML display will improve *depth perception* further. Depth cues from the accommodation-vergence system provided by the transparent ML would reinforce depth cues from perspective/motion parallax and coincide with natural depth cues from the physical surroundings of the ML display.

Both studies highlight the importance of accurate visualization for enhancing *interaction accuracy*. Study 1 focuses on haptic interactions, while Study 2 emphasizes visual interactions. In both cases, the accuracy of how the virtual content is rendered (whether hand/tool visualization, perspective rendering, or eye accommodation/vergence) plays a critical role in performance. Furthermore, Study 1 finds that user-perspective rendering improves spatial interaction accuracy. Although Study 2 does not directly address user-perspective rendering, its findings on eye accommodation and vergence can be seen as supporting the idea that rendering from the user's perspective (considering visual factors like accommodation and vergence) is crucial for accuracy.

5.2.1 Diverging factors

Study 1 focuses on haptic interactions, while Study 2 deals with visual interactions. This indicates that the type of interaction (haptic vs. visual) may require different considerations and optimizations in ML displays. For instance, when using an ML display in a visual comparison task, such as reading contextual AR instructions, the findings of Study 2 are of more import. When we consider a spatial task, such as moving an AR-annotated physical object to an AR-annotated position, the findings of Study 1 are more important. In spatial tasks, often haptic interactions play a larger role. However, in this last scenario of spatial awareness and hand-eye coordination, the true transparent ML—the result of the combined studies—is hypothesized to increase performance yet further.

5.2.2 Combined limitations

Physical constraints and user movement

The physical constraints of *stationary* ML displays (Sect. 2.1.4) restrict user movement and may lead to collisions, limiting the maneuverability required for high-precision tasks. This limitation underscores the need for more mobile (see Sect. 5.4 below) or ergonomically designed displays to facilitate better user interaction.

Proximity and eye accommodation

Users' natural inclination to move closer to ML displays for detailed tasks impedes proper eye accommodation, affecting visualization. Both studies highlight the necessity for displays that support variable focal distances or enhanced optical systems to accommodate natural viewing behaviors and improve detail visualization.

Technical limitations of the hardware

Technical limitations, such as display resolution and the lack of advanced features like varifocal capabilities, present significant challenges. Addressing these constraints through improved hardware and software solutions could substantially enhance the effectiveness and usability of ML displays.

5.2.3 Design guidelines and recommendations

Our findings suggest several key design guidelines for AR ML systems:

General guidelines

1. Utilize front-facing cameras inside- or attached to ML displays to ensure clear and unobstructed tracking of the user's eyes, optionally enhanced with a fish-eye lens to improve the tracking field of view. This ensures stable eye tracking at close range as is required for UP rendering ML displays.

2. Capture and reconstruct a static (non-changing) environment in an offline step and visualize it during UP rendering. This improves performance as the ML device and/or additional hardware can focus on tracking and visualizing the interactive elements.

3. Ensure that depth cues are consistently presented across all visual contexts in both the ML display and the physical environment. This includes cues like shadowing, occlusion, and relative size.

4. Utilize 'X-ray visualizations' on the ML display for virtual objects inside physical geometry by rendering the physical geometry semi-transparent. This allows for parallax depth cues during UP rendering that improve depth perception.

Task-specific guidelines and User Perspective (UP)

5. Implement UP rendering for AR ML displays when the task involves *depth estimation* or *interaction at close range*. This ensures that users experience a more natural and intuitive depth perception, improving task accuracy.

6. Use UP rendering in AR-guided tasks where precise hand-eye coordination is required, such as in medical procedures or industrial applications, to reduce visual distortion and enhance spatial understanding.

7. UP rendering is preferable for *short* tasks or tasks requiring *immediate* spatial accuracy, such as in medical procedures where a trial-and-error approach is not desirable.

8. Use UP rendering in AR-guided tasks where the visual context of the physical environment is unavailable to maintain depth accuracy. Examples include virtual or simulated environments, remote assistance, and situations where the real environment is occluded by structures or geometry.

9. Use UP rendering in AR-guided tasks where haptic (tactile) feedback is unavailable, such as interactions with virtual objects or when augmented information is not anchored to physical structures.

Visual context and surrounding environment

10. Maintain a visible environment context whenever possible, as it enhances depth perception.

11. Maintain a consistent and visually rich background context when presenting AR content on ML displays. This helps users orient themselves in 3D space and improves their ability to judge distances accurately, especially with UP rendering providing parallax depth cues.

12. Display only the most relevant visual information on the ML display. Minimizing visual clutter in the surrounding environment reduces cognitive load and enhances the user's ability to focus on critical tasks.

Haptic feedback and interaction

13. Incorporate real-time haptic feedback that corresponds with the visual stimuli presented on the ML display. This helps users better estimate the position and depth of objects they interact with, especially in tasks that require precision, such as needle insertion.

14. Incorporate real-time visual representations of the hands or hand-guided tools. This enhances the user's ability to perform tasks accurately utilizing powerful occlusion depth cues, without relying solely on tactile feedback.

Designing for off-the-shelf devices

15. Leverage the existing hardware capabilities of consumer devices (e.g., ARKit, ARCore, front-facing TrueDepth cameras, and LiDAR cameras) to enable UP rendering without requiring additional hardware.

16. Prioritize larger display devices for use as *mounted* ML displays to provide a broader augmentable field of view.

17. Prioritize smaller display devices for use as *handheld* ML displays. This allows users to easily guide and aim the display, where maneuverability and ergonomics take precedence over field of view.

18. If mounting a ML in the environment, prioritize a moveable mount so that the user may move or aim the display as desired. This mitigates somewhat the issue of physical obstruction.

Using future technology

Several guidelines and recommendations depend on the technology that is available. Thus far, we discussed implementing a transparent ML display using off-the-shelf hardware, such as smartphones and tablets, available at the time of writing. In the future, it might well be that such consumer devices are fitted with stereoscopic capabilities or optics allowing varifocality. Furthermore, as camera technology is ever-improving, it is expected that most consumer handheld devices will be equipped with depth-sensing cameras. Without going into the specifics of implementation, we can pose several design recommendations for 'true' transparent ML displays making use of these capabilities:

1. Prioritize binocular depth cues over monocular ones. If binocular viewing is not feasible, enhance monocular cues with additional visual context and UP rendering to aid depth estimation.

2. Present content on the ML display with equal eye vergence and accommodation distances to maintain natural and comfortable viewing, especially for long tasks.

3. Interactively reconstruct physical objects and the environment using the available depth-camera technology on the fly, to allow visualization and augmentation when the environment is initially unknown or when preparation is not possible or feasible.

4. If utilizing a pure stereoscopic ML display, keep the disparity between the two eye images below 1/30 of the accommodation (eye-to-display) distance, to maintain a comfortable fusion of the disparity images and binocular depth perception.

5. Prioritize optically transparent displays, i.e. optical see-through ML displays, so that only virtual content needs to be rendered, reducing computa-

tion time and lag. Presented recommendations and guidelines should still be followed to improve performance and perception of the *virtual* content.

6. High-resolution displays are recommended especially when utilizing autostereoscopic capabilities, as such displays functionally reduce the *effective* resolution visible to the user.

5.3 Practical Implications

5.3.1 Dynamic accommodation and vergence adjustment

ML displays should dynamically adjust both accommodation and vergence distances based on the user's interaction context to optimize visual acuity and integration. The findings highlight a need for ML displays with varifocal capabilities. To our knowledge, only certain HMDs have optics to supply such capabilities. Therefore, it is not practical to recommend implementing the eye accommodation findings of Study 2 because the current state-of-the-art hardware cannot support them. However, stereoscopic ML displays do exist (see Sect. 2.1.4 and Sect. 5.4). Based on our findings it is therefore recommended to design AR applications with a single merged context in mind, that keeps a user's eyes focused on the display.

5.3.2 Dynamic perspective for haptic interactions

Our is focused on close-range tasks, at arm's length. This egocentric region is where haptic interactions naturally occur. Depth cues from both motion parallax and binocular disparity are less effective in the far periphery (see Sect. 2.2.1). To support multi-range tasks, ML displays should incorporate *dynamic switching* between device-perspective rendering and user-perspective rendering. For example, when the user's hand or tool is detected by the device camera, render the ML display from the perspective of the user to improve spatial interaction accuracy.

5.3.3 Practical applications and scenarios

The findings of the studies in this dissertation and the resulting guidelines and recommendations have implications for practical use-case scenarios of ML displays. For accurate visualization of physical objects or environments, from the perspective of the user, high-quality scene reconstruction is desired. Due to hardware limitations in the current consumer devices, real-time scene reconstruction is likely to result in low-resolution 3D structures. In the future, as discussed, consumer hardware is expected to be capable of real-time high-resolution scene reconstruction. Following are examples of practical use cases for the proposed system given the reconstruction quality.

Method 1: Pre-reconstruct and calibrate the physical environment

One method is to capture and reconstruct a high-quality detailed model of the physical environment in an offline phase, prior to the ML display to be used to present augmented information during the actual task performance. This method was employed before the experiments in Sect. 3.4. The *requirements* for this method are:

1. Knowledge about- and access to the physical environment before the AR task.

2. The physical environment remains unchanged during the AR task.

3. The ML display needs to be calibrated with the environment to determine its relative pose.

Prior reconstruction of the physical environment has several *benefits* for ML displays:

1. Able to provide high-resolution scenery necessary for precise interaction with UP rendering.

2. Low computational costs during the AR tasks.

Practical scenarios that benefit from this methodology are those that require sub-centimeter precision and haptic interactions:

- **Industrial assembly and maintenance.** In a factory, a technician uses the ML display to assemble or repair complex machinery. The machinery are known beforehand and its components have been pre-scanned and reconstructed in 3D. Detection of the machinery pose happens through markers that are attached to it, detected by the back-facing camera or through 3D model pose detection since its structure is known beforehand. During the assembly or repair, the system overlays augmented information, such as the location of internal components, wiring paths, and step-by-step instructions direction on the physical machinery.
- **Surgical preplanning and simulation.** In a medical setting, a surgeon uses the ML display to plan a complex surgery on a patient whose anatomy has been pre-scanned and reconstructed. Static surfaces of the surgery environment are pre-scanned as well to provide more visual context. The system overlays the patient's internal organs and structures, allowing the surgeon to simulate the procedure and make precise decisions before the actual surgery.

Method 2: Real-time reconstruction of the physical environment

For practical scenarios in which the physical environment or objects are unknown or unavailable in advance, they have to be scanned and reconstructed in real-time. The *requirements* for this method are:

1. Sophisticated depth sensor(s) on the ML display such as time-of-flight (LiDAR) or stereo cameras.
2. Scanning-related steps to be executed before AR-guided tasks can be performed, likely by the end user.
3. High computational costs during the AR tasks due to the continuous scene reconstruction process.

Real-time reconstruction of the environment has several *benefits*:

1. No prior knowledge or pre-processing of the physical environment or objects necessary.

2. Changes in the physical environment are acceptable during the AR task without requiring external hardware or pre-processing.

3. No calibration by the user is necessary.

Practical scenarios that benefit from this methodology are those in which the task environment is unknown or frequently changing while performing the AR-guided task:

1. **Remote assistance of physical tasks.** A remote expert guides a user to perform repair, assembly or operation training of complex machinery on location. In such scenarios, the physical environment and objects or their specific configuration might not be known beforehand. Thus, the user reconstructs the environment by scanning it from several viewpoints with the ML display. The remote expert can perform accurate AR annotations on the real environment to guide or train the user. UP rendering allows the remote expert to better understand the physical structure and place annotations more accurately.

2. **Custom close-range assembly.** The ML display can display augmented information or annotations made by the user while they construct or assemble structures at arm's length. Examples include soldering components, building LEGO constructions, or figurine shaping and painting. In such cases, stages of the physical objects are not fixed or known beforehand but can be continuously reconstructed by the depth sensors on the ML display. Subsequently, UP rendering allows the user to more accurately place AR annotations (e.g. how to cut the clay, where to attach the LEDs) and interact with the physical object while viewing it through the ML display.

Real-time high-resolution reconstruction of the interaction medium

All previous example practical scenarios require that the interaction medium, i.e. the user's hands or tools, are accurately tracked and visualized on the ML display through external hardware. This is the case for the experiments described in Sect. 3.4 where a haptic pen tool provides real-time pose information to a

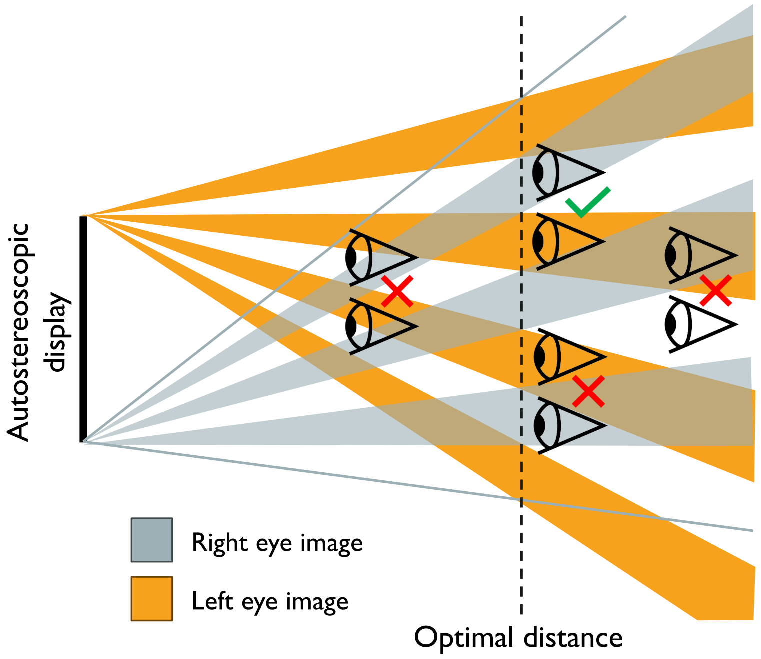


Figure 5.1: Conventional autostereoscopic displays using lenticular lens or parallax barrier technology have an optimal distance at which 3D images can be perceived.

pre-processed 3D reconstruction of the physical tool. This is a hard restriction for using UP rendering on a consumer ML display to support haptic interactions. In the future, when depth cameras and software on consumer devices gain performance in resolution, speed and field-of-view, the interaction medium could be visualized in real-time with sufficient precision, without the necessity of external hardware.

5.4 Future Directions

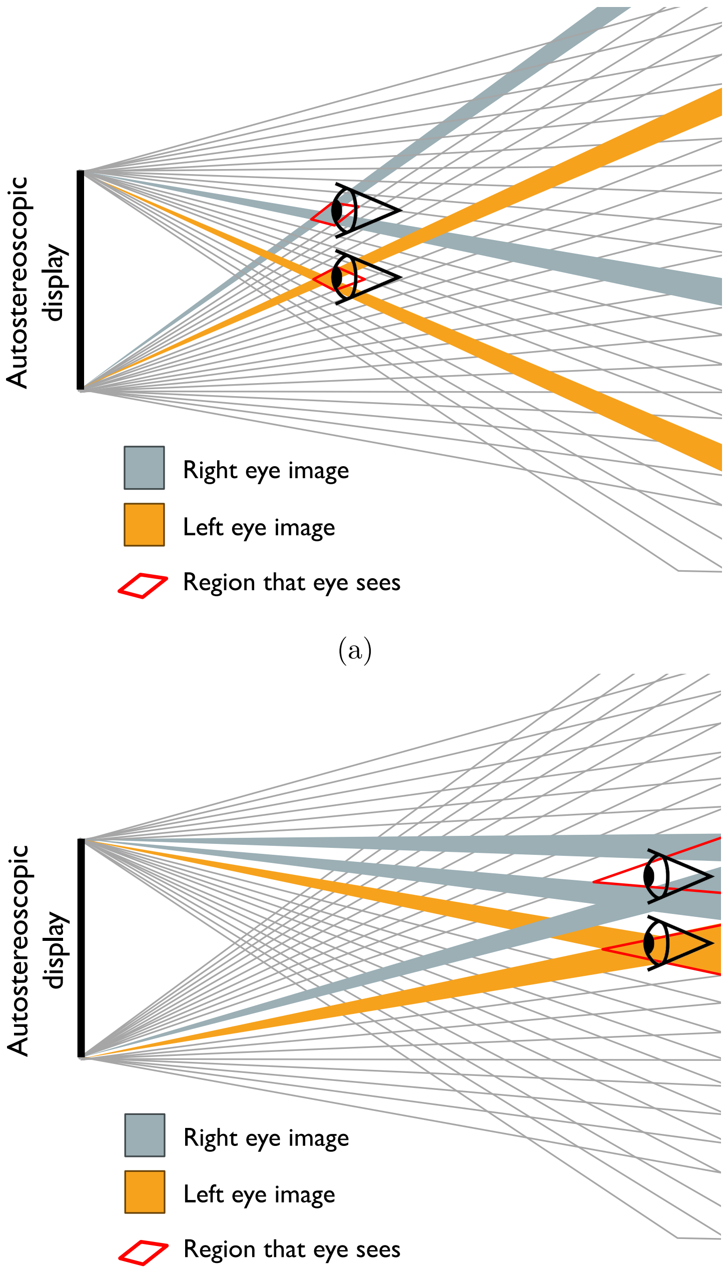
The integration of the two prototype ML displays discussed in our studies (Chapter 3 and Chapter 4) to form a true transparent ML, using *consumer devices*, remains a future work. Technical proposals towards this goal are made in this final section.

5.4.1 Autostereoscopic ML display

Arguably the most impactful improvement on the ML display presented in this dissertation is binocular vision, through the use of a mobile or handheld ML display. In Study 1 we utilized a 3D projector in combination with shutter glasses. As a next step, we want to implement the proposed user-perspective rendering on an autostereoscopic display such as Leia [118], improving practicality and mobility. These displays usually use parallax barriers or lenticular lenses to present each eye with a different disparity image, creating a 3D effect without the need for glasses. As discussed in Sect. 3.9.2, our method of face tracking can keep track of both eye positions of the user. With relative ease, we can produce two user-perspective images, one for each eye. Image processing can then interlace the images, as seen in Fig. 5.1, providing each eye with a disparate binocular image.

Issues and eye-aware rendering

A caveat of autostereoscopic display technology using a parallax barrier or lenticular lens is that it requires the viewer to be at a certain distance to allow the stereoscopic image to be perceived optimally, as shown in Fig. 5.1. When the viewer's eyes are horizontally offset it can result in crosstalk (both right/left eye images are partially visible to one eye) or incorrect depth perception (due to inversion of the right/left eye images). Similarly, when the viewer's eyes are too close to, or too far from, the display, it results in blurry vision due to the blending- or absence of the right/left eye images (see erroneous eye positions in Fig. 5.1). Increasing the resolution of the autostereoscopic display and eye-aware rendering (such as UP rendering) can mitigate these issues. First, as the technology of consumer devices advances it is expected that the resolution of 3D displays will increase. Such high-resolution autostereoscopic displays have an increased density of 'sweet spots' in which 3D viewing is optimal (see Fig. 5.2). Second, knowing the position of each eye allows the system to dynamically select the viewing area per eye and render the correct right/left image, as shown in Fig. 5.2a). This has the added benefit of increased performance because a smaller number of viewing areas have to be rendered simultaneously. While this approach is expected to improve stereoscopic ML experiences at close range (such as at arm's length, see



(a)

(b)

Figure 5.2: Increasing the resolution of the autostereoscopic display and eye-aware dynamic view rendering allows the viewer to (a) perceive the 3D image from variable distances. (b) 3D viewing performance worsens as the distance from the display increases.

Sect. 2.1.4), at larger viewing distances blurry vision as a result of incorrect or missing right/left images are more likely (see Fig. 5.2b). In such scenarios, it is recommended to employ alternative methods to present stereoscopic imagery, such as light fields [36, 88, 94], volumetric displays [119, 120, 121], or 3D projector technology in combination with shutter glasses.

5.4.2 Optical see-through ML display

An intriguing future direction is to replicate the studies proposed in this dissertation on an OST ML display. The scene reconstruction as described in Sect. 3.3.1 would not be necessary, saving on computational power. It would further reduce the need for eye re-accommodation when shifting focus from the physical environment to the ML display, to only virtual content.

Acknowledgements

First of all, I would like to thank Professor Hirokazu Kato, my supervisor. My interest in, and journey of augmented reality research started many years ago with Professor Kato allowing me to do an internship in his lab. I have always felt honored to work with such a famous and influential researcher in the mixed-reality field. Thank you, Professor Kato, for the ideas and discussions during our meetings, and your patience with a student that continually proposed new research topics throughout the years.

Secondly, I would like to thank the other professors of the Interactive Media Design lab at NAIST. Professor Masayuki Kanbara, for ever-valuable feedback during progress reports and lab presentations. And Asst. Prof. Taishi Sawabe, for the positive energy he brings to the lab. Thank you Max, for always making time to discuss interesting research topics and brainstorming new ideas together. And lastly, Asst. Prof. Yuichiro Fujimoto, thank you for your insightful feedback on my research and your help with scientific writing and data analysis.

I would also like to thank professors Klen Čopič Pucihar and Matjaž Kljun, for interesting and engaging discussions and research ideas and for hosting me in their lab in Koper, Slovenia, during our collaborative research.

Lastly, I thank my girlfriend Hao, for occasionally dragging me away from work and the computer, to live. I look forward to the next chapter in our life.

Bibliography

[1] ZealAR, "Progressive applications of augmented reality in healthcare," 2022, accessed on June 28th, 2024. [Online]. Available: https://zealar.com.au/augmented-reality-in-healthcare-industry/

[2] Niteesh Yadav, "Understand display techniques in augmented reality," 2018, accessed on July 9th, 2024. [Online]. Available: https://niteeshyadav.com/blog/understanding-display-techniques-in-augmented-reality-7485/

[3] G. Kramida, "Resolving the vergence-accommodation conflict in head-mounted displays," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 7, pp. 1912–1931, 2015.

[4] LG, "Vehicle windshield hud," 2021, accessed on July 9th, 2024. [Online]. Available: https://www.lg.com/global/newsroom/lg-story/beyond-news/enhancing-the-driving-experience-with-augmented-reality/

[5] D. Andersen, V. Popescu, M. E. Cabrera, A. Shanghavi, B. Mullis, S. Marley, G. Gomez, and J. P. Wachs, "An augmented reality-based approach for surgical telementoring in austere environments," *Military medicine*, vol. 182, no. suppl_1, pp. 310–315, 2017.

[6] K. Čopič Pucihar, P. Coulton, and J. Alexander, "Evaluating dual-view perceptual issues in handheld augmented reality: device vs. user perspective rendering," in *Proceedings of the 15th ACM on International conference on multimodal interaction*, 2013, pp. 381–388.

[7] O. Bimber and R. Raskar, “Modern approaches to augmented reality,” in *Acm siggraph 2006 courses*, 2006, pp. 1–es.

[8] H. Altinpulluk, “Current trends in augmented reality and forecasts about the future,” in *ICERI2017 Proceedings*. IATED, 2017, pp. 3649–3655.

[9] L. Jensen and F. Konradsen, “A review of the use of virtual reality head-mounted displays in education and training,” *Education and Information Technologies*, vol. 23, pp. 1515–1529, 2018.

[10] L. P. Berg and J. M. Vance, “Industry use of virtual reality in product design and manufacturing: a survey,” *Virtual reality*, vol. 21, pp. 1–17, 2017.

[11] P. D. Schlosser, B. Matthews, and P. M. Sanderson, “Head-worn displays for healthcare and industry workers: A review of applications and design,” *International Journal of Human-Computer Studies*, vol. 154, p. 102628, 2021.

[12] R. Pausch, D. Proffitt, and G. Williams, “Quantifying immersion in virtual reality,” in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, 1997, pp. 13–18.

[13] M. A. Livingston, A. Dey, C. Sandor, and B. H. Thomas, *Pursuit of “X-ray vision” for augmented reality.* Springer, 2013.

[14] Y. F. Cheng, H. Yin, Y. Yan, J. Gugenheimer, and D. Lindlbauer, “Towards understanding diminished reality,” in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–16.

[15] E. Chang, H. T. Kim, and B. Yoo, “Virtual reality sickness: a review of causes and measurements,” *International Journal of Human–Computer Interaction*, vol. 36, no. 17, pp. 1658–1682, 2020.

[16] C. Moro, C. Phelps, P. Redmond, and Z. Stromberga, “Hololens and mobile augmented reality in medical and health science education: A randomised controlled trial,” *British Journal of Educational Technology*, vol. 52, no. 2, pp. 680–694, 2021.

[17] M. Kaufeld, M. Mundt, S. Forst, and H. Hecht, “Optical see-through augmented reality can induce severe motion sickness,” *Displays*, vol. 74, p. 102283, 2022.

[18] H. Tan, T. Nie, and E. S. Rosenberg, “Invisible mesh: Effects of x-ray vision metaphors on depth perception in optical-see-through augmented reality,” in *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2024, pp. 376–386.

[19] M. Lambooij, W. IJsselsteijn, M. Fortuin, I. Heynderickx *et al.*, “Visual discomfort and visual fatigue of stereoscopic displays: a review,” *Journal of imaging science and technology*, vol. 53, no. 3, pp. 30 201–1, 2009.

[20] V. Biener, S. Kalamkar, N. Nouri, E. Ofek, M. Pahud, J. J. Dudley, J. Hu, P. O. Kristensson, M. Weerasinghe, K. Čopič Pucihar, M. Kljun, S. Streuber, and J. Grubert, “Quantifying the effects of working in vr for one week,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 11, pp. 3810–3820, 2022.

[21] J. Kim, D. Kane, and M. S. Banks, “The rate of change of vergence–accommodation conflict affects visual discomfort,” *Vision research*, vol. 105, pp. 159–165, 2014.

[22] J. Park, S. Lee, and A. C. Bovik, “3d visual discomfort prediction: vergence, foveation, and the physiological optics of accommodation,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 3, pp. 415–427, 2014.

[23] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, “Vergence–accommodation conflicts hinder visual performance and cause visual fatigue,” *Journal of vision*, vol. 8, no. 3, pp. 33–33, 2008.

[24] R. Zabels, K. Osmanis, M. Narels, U. Gertners, A. Ozols, K. Rūtenbergs, and I. Osmanis, “Ar displays: Next-generation technologies to solve the vergence–accommodation conflict,” *Applied Sciences*, vol. 9, no. 15, p. 3147, 2019.

[25] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks, “Visual discomfort with stereo displays: effects of viewing distance and direction of vergence-

accommodation conflict," in *Stereoscopic Displays and Applications XXII*, vol. 7863. SPIE, 2011, pp. 222–230.

[26] Y. Zhou, J. Zhang, and F. Fang, "Vergence-accommodation conflict in optical see-through display: Review and prospect," *Results in Optics*, vol. 5, p. 100160, 2021.

[27] J. Guo, D. Weng, H. B.-L. Duh, Y. Liu, and Y. Wang, "Effects of using hmds on visual fatigue in virtual environments," in *2017 IEEE Virtual Reality (VR)*. IEEE, 2017, pp. 249–250.

[28] G. A. Koulieris, K. Akşit, M. Stengel, R. K. Mantiuk, K. Mania, and C. Richardt, "Near-eye display and tracking technologies for virtual and augmented reality," *Computer Graphics Forum*, vol. 38, no. 2, pp. 493–519, 2019.

[29] G.-A. Koulieris, B. Bui, M. S. Banks, and G. Drettakis, "Accommodation and comfort in head-mounted displays," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, pp. 1–11, 2017.

[30] D. S. Andersen, M. E. Cabrera, E. J. Rojas-Muñoz, V. S. Popescu, G. T. Gonzalez, B. Mullis, S. Marley, B. L. Zarzaur, and J. P. Wachs, "Augmented reality future step visualization for robust surgical telementoring," *Simulation in Healthcare*, vol. 14, no. 1, pp. 59–66, 2019.

[31] M. Kersten-Oertel, P. Jannin, and D. L. Collins, "The state of the art of visualization in mixed reality image guided surgery," *Computerized Medical Imaging and Graphics*, vol. 37, no. 2, pp. 98–112, 2013.

[32] C. Alves and J. Luís Reis, "The intention to use e-commerce using augmented reality-the case of ikea place," in *Information Technology and Systems: Proceedings of ICITS 2020*. Springer, 2020, pp. 114–123.

[33] S. Ozturkcan, "Service innovation: Using augmented reality in the ikea place app," *Journal of Information Technology Teaching Cases*, vol. 11, no. 1, pp. 8–13, 2021.

[34] D. Curtis, D. Mizell, P. Gruenbaum, and A. Janin, “Several devils in the details: making an ar application work in the airplane factory,” in *Proc. Int’l Workshop Augmented Reality*, 1999, pp. 47–60.

[35] S. J. Henderson and S. Feiner, “Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret,” in *2009 8th IEEE International Symposium on Mixed and Augmented Reality.* IEEE, 2009, pp. 135–144.

[36] J.-h. Lee, I. Yanusik, Y. Choi, B. Kang, C. Hwang, J. Park, D. Nam, and S. Hong, “Automotive augmented reality 3d head-up display based on light-field rendering with eye-tracking,” *Optics Express*, vol. 28, no. 20, pp. 29 788–29 804, 2020.

[37] H. Kim, X. Wu, J. L. Gabbard, and N. F. Polys, “Exploring head-up augmented reality interfaces for crash warning systems,” in *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2013, pp. 224–227.

[38] K. Čopič Pucihar, M. Kljun, and P. Coulton, “Using a mobile phone as a 2d virtual tracing tool: static peephole vs. magic lens,” in *Technology and Intimacy: Choice or Coercion: 12th IFIP TC 9 International Conference on Human Choice and Computers, HCC12 2016, Salford, UK, September 7-9, 2016, Proceedings 12.* Springer, 2016, pp. 277–288.

[39] J. Franz, M. Alnusayri, J. Malloch, and D. Reilly, “A comparative evaluation of techniques for sharing ar experiences in museums,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, pp. 1–20, 2019.

[40] E. A. Bier, M. C. Stone, K. Pier, W. Buxton, and T. D. DeRose, “Toolglass and magic lenses: the see-through interface,” *Proc. of ACM SIGGRAPH’93, Anaheim, CA*, pp. 73–80, 1993.

[41] A. Olwal, C. Lindfors, J. Gustafsson, T. Kjellberg, and L. Mattsson, “Astor: An autostereoscopic optical see-through augmented reality system,” in *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR’05).* IEEE, 2005, pp. 24–27.

[42] M. Berning, D. Kleinert, T. Riedel, and M. Beigl, "A study of depth perception in hand-held augmented reality using autostereoscopic displays," in *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2014, pp. 93–98.

[43] X.-L. Ma, R.-Y. Yuan, L.-B. Zhang, M.-Y. He, H.-L. Zhang, Y. Xing, and Q.-H. Wang, "Augmented reality autostereoscopic 3d display based on sparse reflection array," *Optics Communications*, vol. 510, p. 127913, 2022.

[44] H. Urey, K. V. Chellappan, E. Erden, and P. Surman, "State of the art in stereoscopic and autostereoscopic displays," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 540–555, 2011.

[45] D. Baričević, C. Lee, M. Turk, T. Höllerer, and D. A. Bowman, "A hand-held ar magic lens with user-perspective rendering," in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2012, pp. 197–206.

[46] P. Mohr, M. Tatzgern, J. Grubert, D. Schmalstieg, and D. Kalkofen, "Adaptive user perspective rendering for handheld augmented reality," in *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, 2017, pp. 176–181.

[47] S. Hueber, J. Wilhelm, R. Schäfer, S. Voelker, and J. Borchers, "User-aware rendering: Merging the strengths of device-and user-perspective rendering in handheld ar," *Proceedings of the ACM on Human-Computer Interaction*, vol. 7, no. MHCI, pp. 1–18, 2023.

[48] K. Čopič Pucihar, P. Coulton, and J. Alexander, "The use of surrounding visual context in handheld ar: device vs. user perspective rendering," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 197–206.

[49] J. E. Cutting and P. M. Vishton, "Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth," in *Perception of space and motion.* Elsevier, 1995, pp. 69–117.

[50] J. P. Rolland, R. L. Holloway, and H. Fuchs, "Comparison of optical and video see-through, head-mounted displays," in *Telemanipulator and Telepresence Technologies*, vol. 2351. SPIE, 1995, pp. 293–307.

[51] D. Medeiros, M. Sousa, D. Mendes, A. Raposo, and J. Jorge, "Perceiving depth: optical versus video see-through," in *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, 2016, pp. 237–240.

[52] D. Baričević, T. Höllerer, P. Sen, and M. Turk, "User-perspective augmented reality magic lens from gradients," in *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, 2014, pp. 87–96.

[53] D. Baričević, C. Lee, M. Turk, T. Höllerer, and D. A. Bowman, "A hand-held ar magic lens with user-perspective rendering," in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2012, pp. 197–206.

[54] R. M. Kitchin, "Cognitive maps: What are they and why study them?" *Journal of environmental psychology*, vol. 14, no. 1, pp. 1–19, 1994.

[55] D. Valkov, A. Giesler, and K. Hinrichs, "Evaluation of depth perception for touch interaction with stereoscopic rendered objects," in *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*, 2012, pp. 21–30.

[56] L. Gao, Z. Liu, Z. Chen, J. S. Pan, and M. Yu, "Targeted reaching with monocular depth information and haptic feedback: Comparing between monocular patients and normally sighted observers," *Vision Research*, vol. 211, p. 108274, 2023.

[57] D. Schmalstieg and T. Hollerer, *Augmented reality: principles and practice*. Addison-Wesley Professional, 2016.

[58] D. Baricevic, C. Lee, M. Turk, and T. Hollerer, "User-perspective augmented reality magic lens," *GSWC 2012*, p. 7, 2012.

[59] L. Gombač, K. Čopič Pucihar, M. Kljun, P. Coulton, and J. Grbac, "3d virtual tracing and depth perception problem on mobile ar," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2016, pp. 1849–1856.

[60] R. Nomura, T. Komuro, S. Yamamoto, and N. Tsumura, “Object manipulation for perceiving a sense of material using user-perspective mobile augmented reality,” *ITE Transactions on Media Technology and Applications*, vol. 8, no. 4, pp. 245–251, 2020.

[61] L. Schütz, C. Brendle, J. Esteban, S. M. Krieg, U. Eck, and N. Navab, “Usability of graphical visualizations on a tool-mounted interface for spine surgery,” *Journal of Imaging*, vol. 7, no. 8, p. 159, 2021.

[62] K. Kassil and A. J. Stewart, “Evaluation of a tool-mounted guidance display for computer-assisted surgery,” in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2009, pp. 1275–1278.

[63] J. E. Swan, G. Singh, and S. R. Ellis, “Matching and reaching depth judgments with real and augmented reality targets,” *IEEE transactions on visualization and computer graphics*, vol. 21, no. 11, pp. 1289–1298, 2015.

[64] K. Čopič Pucihar, P. Coulton, and J. Alexander, “Creating a stereoscopic magic-lens to improve depth perception in handheld augmented reality,” in *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*, 2013, pp. 448–451.

[65] F. Kerber, P. Lessel, M. Mauderer, F. Daiber, A. Oulasvirta, and A. Krüger, “Is autostereoscopy useful for handheld ar?” in *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia*, 2013, pp. 1–4.

[66] G. Lugtenberg, K. Čopič Pucihar, M. Kljun, T. Sawabe, Y. Fujimoto, M. Kanbara, and H. Kato, “Effects of eye vergence and accommodation on interactions with content on an ar magic-lens display and its surroundings,” *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–11, 2024.

[67] R. Hecht, M. Li, Q. M. de Ruiter, W. F. Pritchard, X. Li, V. Krishnasamy, W. Saad, J. W. Karanian, and B. J. Wood, “Smartphone augmented reality ct-based platform for needle insertion guidance: a phantom study,” *Cardiovascular and interventional radiology*, vol. 43, pp. 756–764, 2020.

[68] M. Li, R. Seifabadi, D. Long, Q. De Ruiter, N. Varble, R. Hecht, A. H. Negussie, V. Krishnasamy, S. Xu, and B. J. Wood, "Smartphone-versus smartglasses-based augmented reality (ar) for percutaneous needle interventions: system accuracy and feasibility study," *International journal of computer assisted radiology and surgery*, vol. 15, pp. 1921–1930, 2020.

[69] F. El Jamiy and R. Marsh, "Survey on depth perception in head mounted displays: distance estimation in virtual reality, augmented reality, and mixed reality," *IET Image Processing*, vol. 13, no. 5, pp. 707–712, 2019.

[70] J. A. Jones, J. E. Swan, G. Singh, E. Kolstad, and S. R. Ellis, "The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception," in *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, 2008, pp. 9–14.

[71] A. P. Mapp, H. Ono, and R. Barbeito, "What does the dominant eye dominate? a brief and somewhat contentious review," *Perception & psychophysics*, vol. 65, no. 2, pp. 310–317, 2003.

[72] L. Nissen, J. Hübner, J. Klinker, M. Kapsecker, A. Leube, M. Schneckenburger, and S. M. Jonas, "Towards preventing gaps in health care systems through smartphone use: Analysis of arkit for accurate measurement of facial distances in different angles," *Sensors*, vol. 23, no. 9, p. 4486, 2023.

[73] 3D Systems, "Touch," 2015, accessed on May 31st, 2024. [Online]. Available: https://www.3dsystems.com/haptics-devices/touch

[74] I. M. Vogels, "Detection of temporal delays in visual-haptic interfaces," *Human Factors*, vol. 46, no. 1, pp. 118–134, 2004.

[75] S. Okamoto, M. Konyo, S. Saga, and S. Tadokoro, "Detectability and perceptual consequences of delayed feedback in a vibrotactile texture display," *IEEE Transactions on Haptics*, vol. 2, no. 2, pp. 73–84, 2009.

[76] A. J. Doxon, D. E. Johnson, H. Z. Tan, and W. R. Provancher, "Human detection and discrimination of tactile repeatability, mechanical backlash, and temporal delay in a combined tactile-kinesthetic haptic display system," *IEEE Transactions on Haptics*, vol. 6, no. 4, pp. 453–463, 2013.

[77] W. R. Miles, "Ocular dominance in human adults," *The journal of general psychology*, vol. 3, no. 3, pp. 412–430, 1930.

[78] H. Kang and G. Shin, "Effects of touch target location on performance and physical demands of computer touchscreen use," *Applied ergonomics*, vol. 61, pp. 159–167, 2017.

[79] T.-C. Lin, A. U. Krishnan, and Z. Li, "Comparison of haptic and augmented reality visual cues for assisting tele-manipulation," in *2022 International Conference on Robotics and Automation (ICRA).* IEEE, 2022, pp. 9309–9316.

[80] O. Ariza, G. Bruder, N. Katzakis, and F. Steinicke, "Analysis of proximity-based multimodal feedback for 3d selection in immersive virtual environments," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).* IEEE, 2018, pp. 327–334.

[81] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev, *3D user interfaces: theory and practice.* Addison-Wesley Professional, 2017.

[82] L. Ma, T. Huang, J. Wang, and H. Liao, "Visualization, registration and tracking techniques for augmented reality guided surgery: a review," *Physics in Medicine & Biology*, vol. 68, no. 4, p. 04TR02, 2023.

[83] N. Rosa, W. Hürst, P. Werkhoven, and R. Veltkamp, "Visuotactile integration for depth perception in augmented reality," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, 2016, pp. 45–52.

[84] J. Troccaz, G. Dagnino, and G.-Z. Yang, "Frontiers of medical robotics: From concept to systems to clinical translation," *Annual Review of Biomedical Engineering*, vol. 21, no. Volume 21, 2019, pp. 193–218, 2019. [Online]. Available: https://www.annualreviews.org/content/journals/10.1146/annurev-bioeng-060418-052502

[85] J. L. Bacca Acosta, S. M. Baldiris Navarro, R. Fabregat Gesa, S. Graf *et al.*, "Augmented reality trends in education: a systematic review of research and applications," *Journal of Educational Technology and Society, 2014, vol. 17, núm. 4, p. 133-149*, 2014.

[86] M. Dunleavy, C. Dede, and R. Mitchell, “Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning,” *Journal of science Education and Technology*, vol. 18, pp. 7–22, 2009.

[87] J. Maxwell, J. Tong, and C. M. Schor, “Short-term adaptation of accommodation, accommodative vergence and disparity vergence facility,” *Vision Research*, vol. 62, pp. 93–101, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0042698912000867

[88] R. Swannack and O. D. A. Prima, “An assessment of human depth understanding in handheld light-field displays,” *International Journal on Advances in Life Sciences*, vol. 14, no. 3-4, pp. 100–109, 2022.

[89] B. A. Holden, T. R. Fricke, D. A. Wilson, M. Jong, K. S. Naidoo, P. Sankaridurg, T. Y. Wong, T. J. Naduvilath, and S. Resnikoff, “Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050,” *Ophthalmology*, vol. 123, no. 5, pp. 1036–1042, 2016.

[90] K. Kaur, B. Gurnani, S. Nayak, N. Deori, S. Kaur, J. Jethani, D. Singh, S. Agarkar, J. R. Hussaindeen, J. Sukhija *et al.*, “Digital eye strain-a comprehensive review,” *Ophthalmology and therapy*, vol. 11, no. 5, pp. 1655–1680, 2022.

[91] S. Jaiswal, L. Asper, J. Long, A. Lee, K. Harrison, and B. Golebiowski, “Ocular and visual discomfort associated with smartphones, tablets and computers: what we do and do not know,” *Clinical and Experimental Optometry*, vol. 102, no. 5, pp. 463–477, 2019.

[92] S. Gowrisankaran and J. E. Sheedy, “Computer vision syndrome: A review,” *Work*, vol. 52, no. 2, pp. 303–314, 2015.

[93] K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks, “A stereo display prototype with multiple focal distances,” *ACM transactions on graphics (TOG)*, vol. 23, no. 3, pp. 804–813, 2004.

[94] F.-C. Huang, D. P. Luebke, and G. Wetzstein, “The light field stereoscope.” in *SIGGRAPH emerging technologies*, 2015, pp. 24–1.

[95] D. Lanman and D. Luebke, "Near-eye light field displays," *ACM transactions on graphics (TOG)*, vol. 32, no. 6, pp. 1–10, 2013.

[96] G. Westheimer, "The maxwellian view," *Vision research*, vol. 6, no. 11-12, pp. 669–682, 1966.

[97] J. Lin, D. Cheng, C. Yao, and Y. Wang, "Retinal projection head-mounted display," *Frontiers of Optoelectronics*, vol. 10, pp. 1–8, 2017.

[98] A. Rastogi, "Design of an interface for teleoperation in unstructured environments using augmented reality displays." University of Toronto, 1996, master's Thesis.

[99] D. Drascic and P. Milgram, "Perceptual issues in augmented reality," in *Stereoscopic displays and virtual reality systems III*, vol. 2653. Spie, 1996, pp. 123–134.

[100] O. Cakmakci and J. Rolland, "Head-worn displays: a review," *Journal of display technology*, vol. 2, no. 3, pp. 199–216, 2006.

[101] N. Kim, A.-H. Phan, M.-U. Erdenebat, A. Alam, K.-C. Kwon, M.-L. Piao, and J.-H. Lee, "3d display technology," *Disp. Imag*, vol. 1, no. 1, pp. 73–95, 2014.

[102] D. M. Krum, E. A. Suma, and M. Bolas, "Augmented reality using personal projection and retroreflection," *Personal and Ubiquitous Computing*, vol. 16, pp. 17–26, 2012.

[103] O. Bimber, G. Wetzstein, A. Emmerling, and C. Nitschke, "Enabling view-dependent stereoscopic projection in real environments," in *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)*. IEEE, 2005, pp. 14–23.

[104] H. Mizushina, J. Nakamura, Y. Takaki, and H. Ando, "Super multi-view 3d displays reduce conflict between accommodative and vergence responses," *Journal of the Society for Information Display*, vol. 24, no. 12, pp. 747–756, 2016.

[105] S. Yano, S. Ide, T. Mitsuhashi, and H. Thwaites, "A study of visual fatigue and visual comfort for 3d hdtv/hdtv images," *Displays*, vol. 23, no. 4, pp. 191–201, 2002.

[106] G. Heron, W. N. Charman, C. M. Schor *et al.*, "Age changes in the interactions between the accommodation and vergence systems," *Optometry and vision science*, vol. 78, no. 10, pp. 754–762, 2001.

[107] B. Lijka, S. Toor, and G. Arblaster, "The impact of diplopia on reading," *The British and Irish Orthoptic Journal*, vol. 15, no. 1, p. 8, 2019.

[108] A. Khurana, A. Khurana, and B. Khurana, *Theory and Practice of Optics and Refraction*, ser. Modern system of ophthalmology (MSO) series. Elsevier India, 2017. [Online]. Available: https://books.google.co.jp/books?id=tnX7zQEACAAJ

[109] R. L. Newman, *Head-up displays: Designing the way ahead.* Routledge, 2017.

[110] A. Eiberger, P. O. Kristensson, S. Mayr, M. Kranz, and J. Grubert, "Effects of depth layer switching between an optical see-through head-mounted display and a body-proximate display," in *Symposium on Spatial User Interaction*, 2019, pp. 1–9.

[111] J. L. Gabbard, D. G. Mehra, and J. E. Swan, "Effects of ar display context switching and focal distance switching on human performance," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 6, pp. 2228–2241, 2018.

[112] M. S. Arefin, N. Phillips, A. Plopski, J. L. Gabbard, and J. E. Swan, "The effect of context switching, focal switching distance, binocular and monocular viewing, and transient focal blur on human performance in optical see-through augmented reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 5, pp. 2014–2025, 2022.

[113] Y. Bababekova, M. Rosenfield, J. E. Hue, and R. R. Huang, "Font size and viewing distance of handheld smart phones," *Optometry and Vision Science*, vol. 88, no. 7, pp. 795–797, 2011.

[114] L. Lipton, *Foundations of the stereoscopic cinema: a study in depth.* Van Nostrand Reinhold Company, 1982.

[115] P. T. de Jong, "A history of visual acuity testing and optotypes," *Eye*, pp. 1–12, 2022.

[116] B. Rassow and Y. Wang, "Correlation of letter optotypes with landholt ring for different degrees of visual acuity," *Klinische Monatsblatter fur Augenheilkunde*, vol. 215, no. 2, pp. 119–126, 1999.

[117] J. E. Cutting, "How the eye measures reality and virtual reality," *Behavior Research Methods, Instruments, & Computers*, vol. 29, no. 1, pp. 27–36, 1997.

[118] Leia, "Leia lume pad," 2022, accessed on July 10th, 2024. [Online]. Available: https://www.leiainc.com/

[119] K.-J. Kim, B.-S. Park, J.-K. Kim, D.-W. Kim, and Y.-H. Seo, "Holographic augmented reality based on three-dimensional volumetric imaging for a photorealistic scene," *Optics Express*, vol. 28, no. 24, pp. 35 972–35 985, 2020.

[120] L. Lisle, K. Tanous, H. Kim, J. L. Gabbard, and D. A. Bowman, "Effect of volumetric displays on depth perception in augmented reality," in *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2018, pp. 155–163.

[121] L. Lisle, C. Merenda, K. Tanous, H. Kim, J. L. Gabbard, and D. A. Bowman, "Effects of volumetric augmented reality displays on human depth judgments: Implications for heads-up displays in transportation," *International Journal of Mobile Human Computer Interaction (IJMHCI)*, vol. 11, no. 2, pp. 1–18, 2019.

Publication List

Journal Articles

1. **G. Lugtenberg**, K. Č. Pucihar, M. Kljun, T. Sawabe, Y. Fujimoto, M. Kanbara, and H. Kato, "Effects of eye vergence and accommodation on interactions with content on an AR magic-lens display and its surroundings," in *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, pp. 1-11, IEEE, 2024.
2. **G. Lugtenberg**, T. Sawabe, Y. Fujimoto, M. Kanbara, and H. Kato, "Effects of user perspective, visual context, and feedback on depth perception of AR content on Magic-lens displays during haptic interactions," (under review).

Peer-Reviewed Conference Publications

1. C.T. Chu, **G. Lugtenberg**, G. Yamamoto, T. Taketomi, J. Miyazaki, and H. Kato. "Augmented Reality Agent as Desktop Assistant," In *Proceedings of First International Symposium on Socially and Technically Symbiotic Systems. Okayama*, pp. 29-31. 2012.
2. **G. Lugtenberg**, C. Sandor, W. Hürst, A. Plopski, T. Taketomi, and H. Kato, "Changing perception of physical properties using multimodal augmented reality: position paper," in *Proceedings of the 2016 workshop on Multimodal Virtual and Augmented Reality*, pp. 1–4, 2016.

3. **G. Lugtenberg**, W. Hürst, N. Rosa, C. Sandor, A. Plopski, T. Taketomi, and H. Kato, "Multimodal augmented reality–augmenting auditory-tactile feedback to change the perception of thickness," in *MultiMedia Modeling: 24th International Conference, MMM 2018, Bangkok, Thailand, February 5-7, 2018, Proceedings, Part I 24*, pp. 369–380, Spritnger, 2018.

4. B. Ye, Y. Fujimoto, T. Sawabe, M. Kanbara, **G. Lugtenberg**, and H. Kato, "A rendering method of microdisplay image to expand pupil movable region without artifacts for lenslet array near-eye displays.," in *ICAT-EGVE*, pp. 131–138, 2022.

5. **G. Lugtenberg**, K. Mori, Y. Matoba, T. Teo, and M. Billinghurst, "The magicbook revisited," in *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 801–806, IEEE, 2023.

Other Conference or Workshop Presentations and Demos

1. J. Cambpell, E. Barnes, J. D. Fraser, B. Twynham, X. T. Pham, N. T. Hien, **G. Lugtenberg**, N. Yoshinari, S. Al Akkad, A. G. Taylor, et al., "Rockembot boxing: Facilitating long-distance real-time collaborative interactions with limited hand tracking volumes," in *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2020.

2. **G. Lugtenberg**, Y. Fujimoto, T. Sawabe, M. Kanbara, H. Kato, "The effect of parallax depth cues on AR supported hand tasks using a non-wearable display," *The 14th Asia-Pacific Workshop on Mixed and Augmented Reality (APMAR2022)*, Japan, Yokohama, 3 Dec. 2022.

3. **G. Lugtenberg**, T. Teo, H. Kato, M. Billinghurst, "The MagicBook (2000): A Transitional MR Experience," Demonstration at *SIGGRAPH '23: Special Interest Group on Computer Graphics and Interactive Techniques Conference*, Los Angeles, USA, August 6 - 10, 2023.