Dissertation

Perceptual Visualization for Object Alignment in Mixed Reality

Alejandro Martin Gomez

Technische Universität München
Fakultät für Informatik
Lehrstuhl für Informatikanwendungen in der Medizin & Augmented Reality

Perceptual Visualization for Object Alignment in Mixed Reality

Alejandro Martin Gomez

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende:	Prof. Dr. Anne Brüggemann-Klein
Prüfer der Dissertation:	1. Prof. Dr. Nassir Navab 2. Prof. Dr. Christian Sandor

Die Dissertation wurde am 27.09.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 05.02.2022 angenommen.

Alejandro Martin Gomez
Perceptual Visualization for Object Alignment in Mixed Reality
Dissertation, Version 1.0

Technische Universität München
Fakultät für Informatik
Lehrstuhl für Informatikanwendungen in der Medizin & Augmented Reality
Boltzmannstraße 3
85748 and Garching bei München

Abstract

Over the past few decades, Mixed Reality (MR) has emerged as a technology that enriches human perception by generating virtual content that consistently co-exists and interacts with the real world. While this content can be delivered through any of the senses, vision-based MR has drawn special attention and represents the focal point of this dissertation. This MR modality has proven useful in assisting users during tasks that involve the manipulation and alignment of real and virtual objects, showing its value and flexibility in multiple domains, including academic, design, industrial, and medical applications. However, correctly estimating the virtual content's depth from the observer remains a challenge, frequently leading to inaccurate placement of the objects of interest in tasks that rely on the user's perception.

This dissertation investigates how designing visualization techniques that integrate visual perception concepts to represent the virtual content can provide relevant information to infer errors during interactive alignment in MR. These visualization techniques go beyond the simple representation of a virtual replica of the objects of interest. They aim to leverage the objects' geometry and appearance to provide virtual information useful to alleviate the estimation errors perceived during task performance.

This dissertation, divided into two main parts, focuses first on egocentric *single-view* scenarios. This part investigates the consequences of observing misleading occlusion when real and virtual objects overlap during alignment tasks in MR. In addition, it introduces the concept of COMPLEMENTARY TEXTURES as a novel approach that considers the textural, geometric, or semantic information of the objects of interest to generate virtual replicas that complement the visual information provided by their real counterparts. Furthermore, it presents a comparison of how two different display technologies, frequently used in commercial head-mounted displays, can influence depth estimation when the alignment task requires visualizing virtual content placed inside of real objects. Lastly, it proposes a structured decomposition of the visual properties of these techniques to design novel approaches that can lead to better depth estimation.

The second part of the dissertation explores the advantages of using egocentric *multi-view* approaches. Such approaches provide additional information that compensates for the errors observed in the user's view direction. This part explores how using external cameras and mirrors can support the user to improve the alignment. In addition, it introduces the AUGMENTED MIRRORS as a novel concept that uses a real mirror's surface to reflect the content of an MR environment dynamically. The additional viewpoint provided by the AUGMENTED MIRRORS allows the users to improve object alignment and can be used for additional applications such as exploration and scene understanding.

Zusammenfassung

In den letzten Jahrzehnten hat sich Mixed Reality (MR) als eine Technologie herauskristallisiert, die die menschliche Wahrnehmung bereichert, indem sie virtuelle Inhalte erzeugt, die mit der realen Welt koexistieren und interagieren. Während diese Inhalte über alle Sinne vermittelt werden können, hat die bildbasierte MR besondere Aufmerksamkeit auf sich gezogen und bildet den Schwerpunkt dieser Dissertation. Diese MR-Modalität hat sich als nützlich erwiesen, um Benutzer bei Aufgaben zu unterstützen, die die Manipulation und den Abgleich von realen und virtuellen Objekten beinhalten. Sie hat ihren Wert und ihre Flexibilität in verschiedenen Bereichen unter Beweis gestellt, darunter akademische, Design-, industrielle und medizinische Anwendungen. Die korrekte Einschätzung der Tiefe des virtuellen Inhalts durch den Betrachter stellt jedoch nach wie vor eine Herausforderung dar, die bei Aufgaben, die sich auf die Wahrnehmung des Benutzers stützen, häufig zu einer ungenauen Platzierung der betreffenden Objekte führt.

In dieser Dissertation wird untersucht, wie Visualisierungstechniken, die visuelle Wahrnehmungskonzepte zur Darstellung des virtuellen Inhalts integrieren, hilfreiche Zusatzinformationen zum Einschätzen von Positionierungsfehlern beim interaktiven Angleichen von realen und virtuellen Objektenliefern können. Diese Visualisierungstechniken gehen über die einfache Darstellung eines virtuellen Abbilds der interessierenden Objekte hinaus. Stattdessen zielen sie darauf ab, die Geometrie und das Aussehen der Objekte zu nutzen, um virtuelle Informationen zu liefern, die die während der Aufgabenausführung wahrgenommenen Schätzfehler reduzieren können.

Diese Dissertation, die in zwei Hauptteile gegliedert ist, konzentriert sich zunächst auf egozentrische *Single-View-Ansätzen*. In diesem Teil werden die Folgen der irreführenden Okklusion untersucht, wenn sich reale und virtuelle Objekte bei Ausrichtungsaufgaben in MR überschneiden. Darüber hinaus wird das Konzept von COMPLEMENTARY TEXTURES als neuartiger Ansatz vorgestellt, der die texturellen, geometrischen oder semantischen Informationen der interessierenden Objekte berücksichtigt, um virtuelle Nachbildungen zu erzeugen, die die visuellen Informationen ihrer realen Gegenstücke komplementieren. Darüber hinaus wird verglichen, wie zwei verschiedene Anzeigetechnologien, die häufig in kommerziellen Head-Mounted-Displays verwendet werden, die Tiefenschätzung beeinflussen können, wenn die Ausrichtungsaufgabe die Visualisierung virtueller Inhalte innerhalb realer Objekte erfordert. Schließlich wird eine strukturierte Zerlegung der visuellen Eigenschaften dieser Techniken vorgeschlagen, um neuartige Ansätze zu entwickeln, die zu einer besseren Tiefenabschätzung führen können.

Im zweiten Teil dieser Dissertation werden die Vorteile der Verwendung egozentrischer *Multiview-Ansätze* untersucht. Solche Ansätze liefern zusätzliche Informationen, die die Wahr-

nehmungsfehler in der Blickrichtung des Benutzers ausgleichen. In diesem Teil wird untersucht, wie der Einsatz von externen Kameras und Spiegeln den Benutzer bei der Verbesserung der Ausrichtung unterstützen kann. Darüber hinaus wird AUGMENTED MIRRORS (augmentierte Spiegel) als neuartiges Konzept vorgestellt, das die Oberfläche eines echten Spiegels nutzt, um den Inhalt einer MR-Umgebung dynamisch zu reflektieren. Der zusätzliche Blickwinkel, den AUGMENTED MIRRORS bieten, ermöglicht es dem Benutzer, die Ausrichtung von Objekten zu verbessern, und kann für zusätzliche Anwendungen wie Exploration und Szenenverständnis genutzt werden.

Acknowledgments

First, I would like to thank Prof. Dr. Anne Brüggemann-Klein, Prof. Christian Sandor, and Prof. Nassir Navab for being part of my committee. In particular, I want to express my deepest gratitude to Prof. Nassir Navab for sharing his knowledge and putting his trust in me during these past years. Thank you for granting me the freedom and supporting me in exploring novel concepts, and sharing that contagious positive attitude that drives your life.

In addition, I would like to express my gratitude to the Bayerische Forschungsstiftung (BFS) for all the support provided through the grant DOK-178-17. It would not have been possible for me to complete my studies without your help.

My most sincere appreciation to Dr. Ulrich Eck for his patience and constant support, for the precise and valuable feedback, and for sharing the vision with which you live your life in and out of academia. Of course, I don't want to miss the opportunity to acknowledge my colleagues in the Narvis lab. Thanks, Jakob Weiss, for the incredible friendship, the fantastic dinners, and being part of significant moments in my life. To Alex Winkler for the enjoyable discussions that led to exciting ideas and projects. To Matthias Grimm for the always exciting talks and the football nights. To Matthias Seibold for the time shared in Zurich and his always kind and attentive manner. To Felix Bork for inspiring me with his impeccable work ethic and organization. To Kevin Yu for sharing his ideas always full of creativity. To Tianyu Song for his always proactive attitude that invites to generate new ideas. To Michael Sommersperger for his friendship and for sharing the Hopkins experience with me. To Dr. Daniel Roth for his proactive attitude to carry out excellent projects and extend a helping hand whenever necessary. To Roger Soberanis for all his kind help and support. To my colleagues at CAMP for the wonderful times we shared during CAMPing and for pushing the boundaries of science. And last, but most certainly not least, to Martina Hilla, for all the help provided during my doctoral studies, and for being a fundamental element of the CAMP team.

To my friends and colleagues at Johns Hopkins University. Especially to Sing Chun Lee for his kind and sincere friendship and for sharing his life philosophy with me, to Javad Fotouhi for the wise advice and the fantastic coffee breaks, and to Prof. Mehran Armand for the sincere advice and unconditional support. I will always be thankful for having the chance to meet you. To Prof. Russ Taylor, Prof. Iulian Iordachita, Prof. Matthias Unberath, Dr. Greg Osgood, Dr. Alex Johnson, and all the researchers and students I have had the opportunity to collaborate including Wenhao Gu, Long Qian, Sasan Matinfar, Arian Mehrfard, Qiaochu Wang, Reza Roustaei, Nikola Budjakoski, Andreas Keller, Marc Fischer, Christoph Leuze, Caio Atayde, and Prof. Bruce Daniels.

To my parents, Altagracia and Juan, because your love for the sciences inspired me to follow this path. Especially to my mom, for her unconditional love and tireless support that allowed me to be here today. To my brother Jorge, for the adventures lived, the unforgettable childhood, and the experiences gained.

To my beautiful wife, Elsa, for her patience and unselfish love. Because all the sacrifices you have made for us are reflected in this achievement that is not mine but ours. For that fascinating soul of yours and for being my rock, peace, ally, best friend, and perfect complement. To her parents, Jorge and Rosy, for welcoming me into their family with open arms and making me feel part of it.

Finally, I would like to thank all the people I had the opportunity to meet in Munich. Particularly to Fabio Bracci for his wonderful friendship and help when I needed it the most, and to Herr Max and Frau Annette Schröener and their beautiful family for making me feel at home.

To all of you, because you all have contributed to my academic development and my personal growth, thanks.

Contents

Part I

Introduction

Introduction

1

Mixed Reality (MR) is an emerging technology capable of enriching a user's perception by seamlessly integrating interactive virtual content into the real world. While this technology can stimulate any of the senses to enhance the user's perception of the world, vision-based MR has drawn special attention from the research community, finding application in various fields, including academic, design, industrial, and medical scenarios. This modality represents one of the cornerstones of this dissertation. Hence, from now on, it will be referred to as MR for simplicity.

Within the numerous practical applications of MR, existing works have explored the benefits of using this technology to assist users in performing interactive alignment of real and virtual objects. A common approach in these scenarios is to present virtual replicas indicating the pose in which an object of interest must be placed. Hence, the information provided by the virtual content must be consistent with the visual stimuli perceived from the real world. Failing to generate consistent information between the real and virtual components may result in the observation of ambiguous information, hinder the correct interpretation of the environment, and complicate the performance of the alignment task. Although it has been shown that MR systems are capable of allocating virtual content in the real world accurately, the estimation of the virtual content's depth by human observers has proven challenging and remains an open problem. Multiple aspects, including human factors and technology limitations, frequently contribute to observing conflicting information when the real and virtual worlds are presented, hindering its adaptability for tasks that require high accuracy during the interactive alignment of real and virtual content.

This dissertation explores the benefits of designing novel visualization techniques for interactive alignment tasks in MR. These visualization techniques integrate fundamental concepts of visual perception with the properties of the objects of interest, such as their shape, texture, or even contextual information, to provide a meaningful virtual representation that facilitates the alignment task.

1.1 Motivation

The *extent of world knowledge* [113] is a concept that presents a linear continuum characterized by the information known about the shape and position of the real and virtual content of an MR environment. According to this concept, the amount of information available determines the possible operations a system can perform. One extreme of this continuum describes the *non-modeled world* as a scenario from which no knowledge of the object's shape and position is available. The other extreme describes the modeled world as a scenario where the object's shape and position and the observer's position and viewpoint are known. The relevance of this concept for this dissertation lies in the intermediate cases that require the accurate alignment of virtual and real objects from which only partial information regarding their shape and position is available.

In this context, assuming the desired position of an object to align is known, regardless if this object is real or virtual, it is common to find two alternatives to provide visual guidance to the users of MR applications:

1. If the up-to-date position of the object of interest is also known, it is possible to estimate the alignment error between the desired and current pose of the object. Thus, providing visual guidance to the users during the alignment task can be done in manifold ways. Examples of this include: using animations to indicate the action desired to be performed by the user and the expected pose of the object of interest, displaying the computed differences in the form of numerical errors or textual instructions, presenting visual signals such as arrows to indicate the desired pose, or implementing pseudo-chromatic representations encoding the alignment error.

2. Otherwise, if the up-to-date position of the object of interest is unknown, a common approach is to use virtual replicas of the real objects to assist users with the alignment task. When displayed as static representations, these replicas indicate the desired pose to allocate a real object (i.e., the case of real-to-virtual alignment). Alternatively, these replicas provide a visual reference of an intangible virtual object when it is aligned using a real object as reference (i.e., the case of virtual-to-real alignment).

The first case depicts a scenario where the alignment effectiveness depends largely on the accuracy of the system that computes the differences between the current and desired pose of the objects, and to a lesser extent, on the visual quality provided by the visualization techniques or the user's performance during the alignment process. These methods frequently require tracking systems that involve attaching passive components to the objects of interest–making them bulky and requiring a line of sight that allows uninterrupted observation of the objects of interest–, or the integration of active components by the addition of electronics. Alternative approaches use computer vision algorithms that allow determining the objects' pose. However, and although advances in this field promise good results in the future, nowadays are often unstable, imprecise, or object-specific, limiting their application in dynamic environments.

In contrast, the second case depicts a less restricted scenario in which the requirements involve knowing the desired pose of the object to align and its geometrical properties, for example, by having a computer-aided design (CAD) model of it. These conditions depict a scenario where the alignment accuracy depends to a higher degree on the user's skills and the quality of the information provided by the visualization techniques used to represent the virtual content. Although multiple studies have used this approach in the past, there seems to be no evidence of previous work investigating if the selection of visualization techniques used to represent the virtual content influences the accuracy achieved by the users during interactive alignment tasks in MR.

1.2 Objectives

This dissertation explores whether it is possible to design visualization techniques that provide meaningful information during the interactive alignment of objects in MR environments. To explore this, it focuses on studying the advantages of designing visualization techniques that consider the multiple cues that the human visual system uses to perceive depth and the physical properties of the objects involved in the alignment task, such as their geometry and texture. In addition, it distinguishes between exemplary visualization techniques that can be used when the observer has access to single egocentric viewpoints or when alternative viewpoints provided by external sensors or mirrors are available. Ultimately, it presents a general discussion on the current challenges and limitations of the proposed techniques and future directions that could support users during interactive alignment tasks in MR.

1.3 Contributions

The work presented in this manuscript initially focuses on the proposal of visualization techniques designed to support users during object alignment in egocentric environments where only one viewpoint is available. First, it presents an evaluation of visualization techniques traditionally used to represent virtual content during alignment tasks in MR. This evaluation explores how different levels of occlusion, observed when the objects overlap, can lead to conflicting visual information that influences the alignment accuracy and users' acceptance:

- ***Martin-Gomez, A.**, Eck, U., & Navab, N. (2019, March). Visualization techniques for precise alignment in VR: A comparative study. In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (pp. 735-741). IEEE.*

Results from this evaluation motivated the application and evaluation of these visualization techniques in medical settings requiring the alignment of anatomical structures in MR:

- *Fischer, M., Leuze, C., Perkins, S., Rosenberg, J., Daniel, B., & **Martin-Gomez, A**. (2020, November). Evaluation of Different Visualization Techniques for Perception-Based Alignment in Medical AR. In 2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) (pp. 45-50). IEEE.*

- *Leuze, C., Neves, C., **Martin-Gomez, A.**, Daniel, B. L., Navab, N., Blevins, N. H., ... & McNab, J. A. (2021). Augmented Reality Guided Retrosigmoid Approach. Journal of Neurological Surgery Part B: Skull Base, 82(S 02), S025.*

Moreover, additional studies, including a bachelor's thesis, explore considerations for designing techniques that require the visualization of virtual content within real objects and how the various technologies used to display this content may affect the depth perceived by the users:

- *Andreas Keller. Enhancing Depth Perception for Optical See-Through Head-Mounted Displays in Medical Applications. Advisor: **Martin-Gomez, A.**, Weiss J., & Navab N.*

- ***Martin-Gomez, A.****, Weiss, J.*, Keller, A., Eck, U., Roth, D., & Navab, N. (2021). The Impact of Focus and Context Visualization Techniques on Depth Perception in Optical See-Through Head-Mounted Displays. IEEE Transactions on Visualization and Computer Graphics.*

In addition, this dissertation introduces the COMPLEMENTARY TEXTURES as a novel concept that uses the physical properties of the objects of interest, such as their shape or texture, to present virtual replicas that provide meaningful information to the users performing interactive alignment tasks in MR.

Concerning those scenarios in which users have access to multiple views, this work presents a study comparing the benefits of providing alternative viewpoints using external cameras and mirrors during alignment tasks:

- ***Martin-Gomez, A.**, Fotouhi, J., Eck, U., & Navab, N. (2020, November). Gain A New Perspective: Towards Exploring Multi-View Alignment in Mixed Reality. In 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) (pp. 207-216). IEEE.*

When implemented in MR environments, these concepts provide alternative viewpoints that help mitigate the uncertainty observed in single-view scenarios:

- *Fotouhi, J., Song, T., Mehrfard, A., Taylor, G., Wang, Q., Xian, F., **Martin-Gomez, A.**, Fuerst, B., Armand, M., Unberath, M. and Navab, N. (2020). Reflective-ar display: An interaction methodology for virtual-to-real alignment in medical robotics. IEEE Robotics and Automation Letters, 5(2), 2722-2729.*

Another example of the benefits of using these alternative viewpoints is the AUGMENTED MIRRORS. A novel concept that exploits the physical properties of real mirrors to provide alternative dynamic viewpoints of the real and virtual content of an MR environment:

- ***Martin-Gomez, A.****, Winkler, A.*, Yu, K.*, Roth, D., Eck, U., & Navab, N. (2020, November). Augmented Mirrors. In 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) (pp. 217-226). IEEE.*

Although some of these concepts have been presented to the research community in the past, this manuscript unfolds the discussion on the implications they may have on tasks that require interactive object alignment in MR environments and, in some cases, extends the results obtained. Moreover, it introduces unpublished novel ideas that may contribute to inspire future ideas in this direction.

* The asterisk indicates an equal contribution from the corresponding authors.

1.4 Iconography

Along with this manuscript, two types of dialog boxes will be presented to the reader as follows:

Insights and further readings

This type of dialog box will provide useful insights into the topic discussed, including further readings and key elements that helped shape the ideas presented in this dissertation.

Findings and lessons learned

This type of dialog box will present interesting findings based on lessons learned during the performance of the experiments, observations derived from feedback provided by participants of the user studies, or discussions with participants and colleagues that were not included in the publications but are valuable for this dissertation.

2 On Perception and Mixed Reality. A General Overview

> *...because the human visual system is very robust and is able to work with only partial data, it is simple to create some sensation of depth. The difficulty lies in the creation of an accurate sense of depth.*
>
> — **David Drascic and Paul Milgram**
> (Perceptual Issues in Augmented Reality)

2.1 Fundamentals of Perception

The study of perception has captured the attention of philosophers, physiologists, physicians, and psychologists. According to formal definitions, *perception* can be described as the process of recognizing and interpreting information acquired through the senses, including how we use that information to interact with our environment. This concept differs from *sensations* as the former only comprehends the reception and transmission from external signals to the brain. Thus, perception considers creating an awareness of the external environment based on the signals generated from physical sensation. In this context, Gibson defines *sensations* as the raw material of human experience while perceptions represent the manufactured product [54]. According to Gibson, sensations represented by colors, sounds, touches, odors, and tastes such as a certain hue, a feeling of warmth, and a smell of smoke are not things in themselves. However, combining these specific sensations in *perception* may lead to assume there is something on fire.

The fundamentals of perception date back to ancient Greek philosophers and the branch of philosophy that studies the theory of knowledge and justified belief: *epistemology* [180]. This philosophical branch investigates the nature and conditions required to constitute knowledge, its potential sources, the structure of a body of it, and any possible claims that question its possibility. For example, it investigates whether the world perceived through our senses represents an accurate depiction of it. In this regard, Aristotle held that all knowledge is based on the senses and that sensory qualities alone do not reveal the essence of things. Instead, a higher cognitive faculty, the intellect, perceive these from observing changes over time [65].

While the philosophical approach explores the fundamental nature of knowledge and perception, physicians and physiologists have focused on investigating the physical means associated with the perception of the external world using scientific methods. In this regard, physicians and physiologists have tried to answer many of the questions raised by the philosophers. These questions include how perceptual systems sense the world and if the aspects of perception are learned through experience rather than being innate to the brain.

2.1.1 The Human Visual System

Although multiple theories exist about the number of perceptual systems the human being poses, these theories agree that we have at least five perceptual systems driven by the five basic human senses: touch, hearing, smell, taste, and sight–although some other perceptual systems can be considered, such as proprioception–. From these five, the visual system, and its associated visual perception, have captured special attention not only in the fields of philosophy and psychology but also in others such as robotics and computer vision.

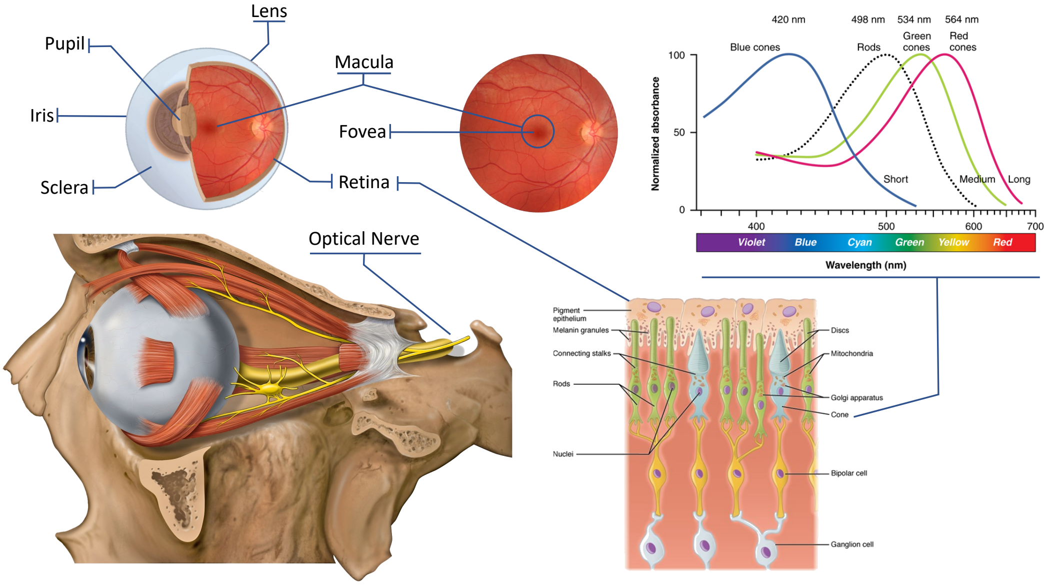


Fig. 2.1. The human visual system allows for generating a representation of our environment by converting a visual stimulus into electrochemical signals. In this process, the light that enters the eye hits the retina, a membrane that contains specialized photoreceptor cells that convert the light into electrochemical information. Such a signal is later transmitted to the brain through the optical nerve.

This system comprehends the physiological components in the human body that enable us to perceive our environment using the light in the visible spectrum reflected by the objects in the real world. This system occupies approximately 70 percent of the sensory receptors in humans [149] and can convert the light that enters the eye through the cornea into neuronal signals. This conversion process is performed by the retina, a photo-sensitive membrane located in the back of the eye. The retina contains two types of photoreceptor cells: rods and cones. These cells, named after their shape, are responsible for converting light into electrochemical signals. The rods, mostly found in the peripheral regions, enable night vision. The cones, concentrated in the central region of the retina–the macula, and more particularly the fovea–, are responsible for color vision and high acuity. The cones can be subcategorized into three types as they respond to green, blue, and red light. After this process, the light converted into electrochemical signals is transmitted to the brain through the optic nerve to the visual cortex. Here, the brain uses those signals to understand our environment (see Figure 2.1).

This process generates a two-dimensional representation of the three-dimensional world, like what is observed during the image formation when taking a picture using a camera. In this regard, Gibson [54] raised one question that has intrigued the researchers: how can the visual system depend on the pictures sensed by the eyes and still produce a scene that extends to the horizons? In other words, how the physical environment, which has three dimensions, is projected on a two-dimensional sensitive surface, but it is still perceived as a three-dimensional scene? How can perception restore the lost third dimension?

Certain theories of perception hypothesize that visual perception is created, in part, through the simultaneous action of feature detector neurons. These neurons can respond to color selectivity, speed, acuity, and contrast sensitivity at early visual areas; and to different aspects of vision such as form, color, movement, and stereopsis at higher stages. Moreover, the visual system has two main pathways for processing visual information. The ventral pathway analyzes color, texture, and shape. In contrast, the dorsal pathway analyzes motion and egocentric position.

2.1.2 Theories of Visual Perception

Over the last century, different theories have been proposed to explain how sensory inputs form the basis of perception. Each one of them has shown its difficulties in accounting for explaining how perception forms. A common approach from researchers in this field is not to accept one specific theory as final but rather to adhere to those theories with experimental foundations. Between these theories, two traditional and apparently contrasting approaches are currently predominant [121]: one consists of variants on the classical constructivist approach [171] (Helmholtzian), and the other of the ecological approach [53] (Gibsonian). Other theories include the Gestalt theory or modern sensory physiology.

The Constructivist Approach

This theory, considered a classical approach, hypothesizes that perception results from a process of unconscious inference about what the stimulus received by our sensors are most likely to be. In other words, it assumes that our mind makes mental adjustments to build a coherent picture of its experiences. According to Helmholtz [171], these inferences are supported by past experiences and learning and are not innate. Moreover, it assumes these are unconscious as the person is not aware of making them.

The *top-down processing theory* proposed by Gregory [58] follows the principles of the constructivist approach. It has its basis on the proposition that perception is *a continual series of simple hypotheses about the external world which are built up and selected by sensory experiences*. It also proposes that the information perceived by our senses is frequently ambiguous. Thus, higher cognitive information is required to make conclusions about what is perceived. This higher cognitive information can be obtained either from experiences or previous knowledge.

The Ecological Approach

Considered an ecological theory, also known as *bottom-up processing*, it presented a new perspective on perception. This theory hypothesizes that perception takes place in real-time and starts with a stimulus sensed from the environment. It assumes that the signal received from the sensors, transmitted from the retina to the visual cortex, becomes more complex on every successive stage in the visual pathway, providing more information and more complex input analysis. Moreover, it implies that perception involves innate mechanisms forged by evolution. Suggesting that no learning is required and that perception is evolutionary and not subject to hypothesis testing [55].

The Gestalt Theory

This theory of perception goes against the idea that perception can be subdivided into simpler components. Instead, it suggests that perception involves entire configurations or patterns and assumes that these entire configurations are more important than the sum of their parts [89]. It also suggests that perception is the result of learned mental associations between simple sensations. This approach differs from Helmholtz's theory as it does not consider the "*unconscious inference.*"

2.1.3 A Taxonomy on Visual Perception

Visual perception skills include recognizing and identifying shapes, objects, colors, and other qualities. These skills allow humans to make accurate judgments on the size, configuration, and spatial relationships of the objects in the environment. However, because visual perception has not been defined consistently in the literature, available resources use different terms when referring to these visual perception skills. This dissertation uses the taxonomy suggested by Schneck [149] to present a hierarchical organization of the fundamental visual skills and their functions (Table 2.1).

This taxonomy presents a classification of the visual perception skills based on the mental action -reception vs. cognition- and the different components that each one of these classes comprises. This subsection presents a general overview of these skills based on Schneck's taxonomy. However, further details regarding spatial perception, specifically depth perception, will be presented in Subsection 2.1.4.

Visual-Receptive Functions

Associated to the regions of the central nervous system that control the eye movements (oculomotor system), includes multiple components such as: **Visual Fixation.** It is the ability to focus and maintain the visual gaze on a single location or a stationary object. It represents a pre-requisite skill for other oculomotor responses, and it is a characteristic of animals that possess a fovea in the anatomy of the eye. **Pursuit Movements.** Also known as tracking involves continued fixation on a moving object to continuously maintain the image on the fovea. These are slow and smooth and can be performed independently by moving the eyes, the head, or both simultaneously. **Saccadic Movements.** Also known as scanning are associated with rapid changes in fixation from one point to another. These types of movements, voluntary or involuntary, are precise. However, it is normal to observe over- or under-shooting. **Other Components.** Despite the eye movements described before can be voluntary, additional components driven by the oculomotor system can react in response to the movements of the head or changes in position in the environment. These components are: *Acuity*. The capacity to discriminate fine details of objects observed in the visual field. *Accommodation*. The ability of the eye to change its focus to visualize objects at different distances. *Binocular fusion*. The process of combining the visual representations generated by both eyes to produce

Tab. 2.1. A taxonomy on visual perception as presented by Schneck [149]. This taxonomy distinguishes the multiple mental processes involved in visual perception based on their function, reception and cognition, and respective components. This taxonomy provides a hierarchical picture of the tasks involved during the visual-perceptual process.

A Taxonomy on Visual Perception			
Function	*Task and components*		
Visual-Receptive	Visual fixation		
	Pursuit movements		
	Saccadic movements		
	Other components	Acuity	
		Accommodation	
		Binocular fusion	
		Stereopsis	
		Convergence and divergence	
Visual-Cognitive	Visual attention	Alertness	
		Selective attention	
		Visual vigilance	
		Shared attention	
	Visual Memory		
	Visual Discrimination	Object (form) perception	Form constancy
			Visual closure
			Figure-ground recognition
		Spatial perception	Position in space
			Depth perception
			Topographic orientation
	Visual Imagery		

a single representation. *Stereopsis*. The perception of depth resulting from combining the visual representations produced by the eyes and their respective disparities or differences. *Convergence and Divergence*. The ability of the eyes to turn inward (convergence) and outward (divergence).

Visual-Cognitive Functions

The visual-cognitive components are the mental processes that interpret the visual stimulus sensed by the eyes. These functions include: **Visual Attention.** Involves the selection of relevant visual input and the ability to ignore irrelevant information. This function considers four components: alertness, selective attention, visual vigilance, and shared attention. **Visual Memory.** Involves the ability to relate visual information with previous experiences. Two main forms of visual memory are short- and long-term memory. **Visual Discrimination.** It is the ability to identify features and details in visual images for recognition, matching, and categorization. Recognition allows identifying key details and relating them with memory. Matching enables the identification of similar features. Categorization allows combining distinguishable features into discrete mental units treated as equivalent. This function also considers the distinction between identifying objects by their color, shape, or size (object vision) and their location in the environment (spatial vision). *Object vision* includes several components such as form constancy, visual closure, and figure-ground recognition. *Spatial vision* includes position in space, depth perception, topographic orientation. **Visual Imagery.** It is the ability to generate mental images even if the physical objects are not present.

Although this taxonomy provides a very general overview of the functions and properties involved in visual perception, it facilitates understanding the physiological and psychological components involved in this complex process.

2.1.4 Depth Perception

The ability to perceive the world in three dimensions (including the length, width, and depth) and judge how far an object is placed from an observer is known as depth perception. While the human visual system can generate a sensation of depth using only partial information, the real challenge lies in creating an accurate sense of it [29]. An example of this is the Necker cube, a collection of line segments perceived as a three-dimensional cube. However, the cube's orientation can be interpreted to have either the lower-left or the upper-right square as its front face (see Figure 2.2).

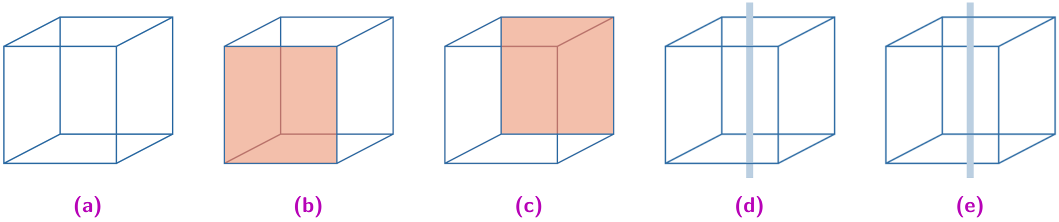


Fig. 2.2. *The Necker Cube* (a) is a bi-dimensional wireframe drawing of a cube that can be perceived as a three-dimensional object. Nevertheless, the visual information provided by the drawing can be interpreted to be facing towards the lower-left (b) or the upper-right (c). The integration of additional visual cues can assist the observer in solving these discrepancies (d),(e).

The human visual system uses several cues to improve depth estimation and solve conflicting information. These cues, depicted in Figure 2.3, can be classified into monocular and binocular cues depending on whether one or two eyes are required to perceive them. At the same time, they can be sub-classified into four types: pictorial, kinetic, physiological, and binocular disparities.

Pictorial Cues

Allow generating a sense of depth from bi-dimensional representations of the three-dimensional world (e.g., images or pictures). These include: **Interposition, overlapping, or occlusion** is a monocular cue observed when one object partially covers another. It is the strongest depth cue and allows to generate an idea of the relative depth order of the objects in the environment. **Linear perspective** can be defined as the effect observed when parallel lines appear to get closer or converge as a function of the depth observed. It is related to both relative size and texture gradient. **Texture gradient** is the effect observed when the texture of a surface gets finer and appears smoother as the surface gets farther away from the observer. This cue could be useful, for example, to determine the size of an object. **Aerial perspective**

or clearness is the result of light being scattered by particles in the air. This results in the loss of detail, clarity, coloration, and a tendency towards a bluish-grey color observed with very distant objects. **Relative brightness** is associated with the phenomenon observed when objects placed farther away from a light source appear darker than those closer. **Shadows** cast by the objects can play an important role in defining depth ordering and perceiving a form by giving the object a three-dimensional feel. This visual cue, under some conditions, can alter the interpretation we create of the environment. **Absolute size** is observed when the actual size of an object is unknown and there is only one object visible. Then, a smaller object seems further away than a large one presented at the same location. **Relative size** can provide information about the relative depth of two objects if their absolute size is unknown, but both objects are known to be the same size. **Familiar size** can be combined with previous knowledge of the object's size to determine its absolute depth. This notion originates under the assumption that an object projected into the retina decreases its visual angle as the distance increases.

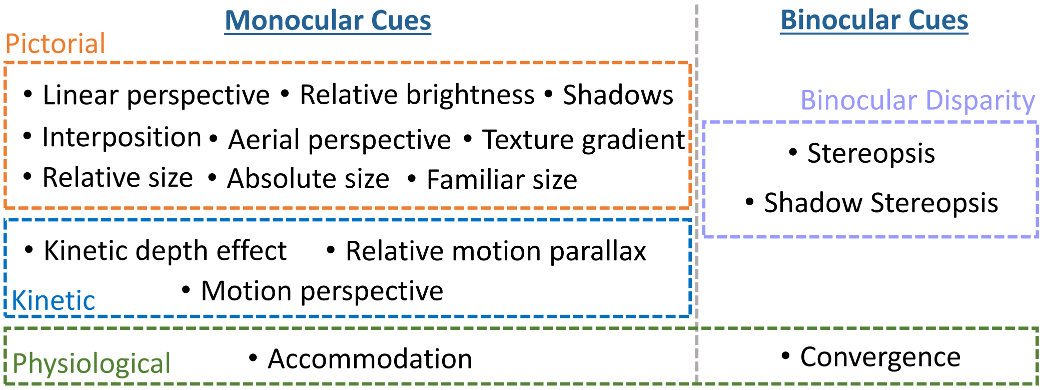

Monocular Cues
Binocular Cues
Pictorial
• Linear perspective • Relative brightness • Shadows
• Interposition • Aerial perspective • Texture gradient
• Relative size • Absolute size • Familiar size
Binocular Disparity
• Stereopsis
• Shadow Stereopsis
• Kinetic depth effect • Relative motion parallax
• Motion perspective
Kinetic
Physiological
• Accommodation
• Convergence

Fig. 2.3. The human visual system uses several visual cues to estimate the depth at which objects are observed.

Kinetic Cues

This type of cues requires the observer, or the objects observed in the environment, to be in motion. This motion allows providing depth information. These cues include: **Relative motion parallax** is a monocular depth cue that gives the impression that static objects follow an observer in movement. It can be defined as the apparent angular velocity of objects, which is inversely proportional to total distance and consequently permits a "safe conclusion" about distance [67]. **Motion perspective** is also a monocular depth cue. However, this visual cue refers to the relative speeds perceived when moving objects are closer or far away from the observer. This concept is similar to the effect observed in optical flow patterns. Whereas motion parallax is concerned with the relative movement of isolated objects, usually due to the observer's movement, motion perspective is concerned with whole gradients of motion that can occur whether the observer is moving or not [29]. The **Kinetic depth effect** allows perceiving the three-dimensional form of an object when two-dimensional representations of the object are in motion.

Physiological Cues

Are associated with the human visual system and how it adapts to changes in the objects observed at different depths, including: **Accommodation** is the eye's ability to adjust its optics (i.e., the shape of its lenses) to keep in focus objects located at different depths. It is driven by the ciliary muscles and enables retinal blur for objects out of focus. **Vergence** is the physical effect of turning inward (converge) or outward (diverge) our eyes, such that the viewed objects are projected to the central area of the retina. It is driven by the extraocular muscles and enables binocular disparity.

Binocular Disparity Cues

These cues rely on having a pair of images from different viewpoints to extract the depth information. This information is obtained by analyzing the differences between the pair of images resulting from the eyes' horizontal separation.

To better understand how these cues are used to perceive and understand the space and the depth and organization of the objects observed around us, Cutting and Vishton [24] investigated the relative efficacy of nine of these cues. These cues were selected based on three different criteria: i) the information provided inherently measured along with a particular scale type, ii) a set of assumptions about how light structures objects in the world, and most importantly, and iii) how the effectiveness of these visual cues vary at different distances. As a result of this, every cue is represented by its just-discriminable depth threshold as a function of the threshold ratio for judging two objects at different distances divided by the mean distance between these objects from the observer.

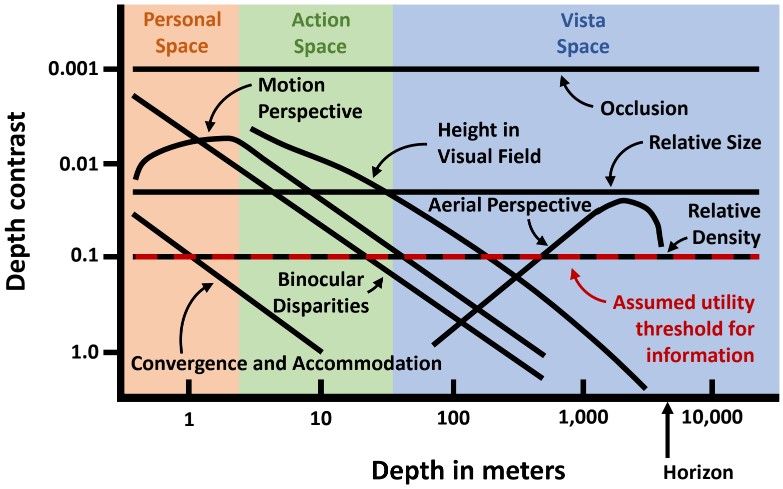


Fig. 2.4. Visual representation of the just-discriminable depth thresholds proposed by Cutting and Vishton [24].

The results from the just-discriminable depth thresholds, depicted in Figure 2.4, have led to the subdivision of the visual field into three egocentric regions defined in function of the actions that an observer can perform within a specific range. The personal space ranges from the observer's head and up to approximately 2 meters from them. This considers the area that can be reached by their arm and slightly beyond. The action space comprises distances between 2 and up to 30 meters where it is considered that the observer can interact and communicate publicly with a relative facility. Lastly, the vista space considers those distances above 30 meters from the observer.

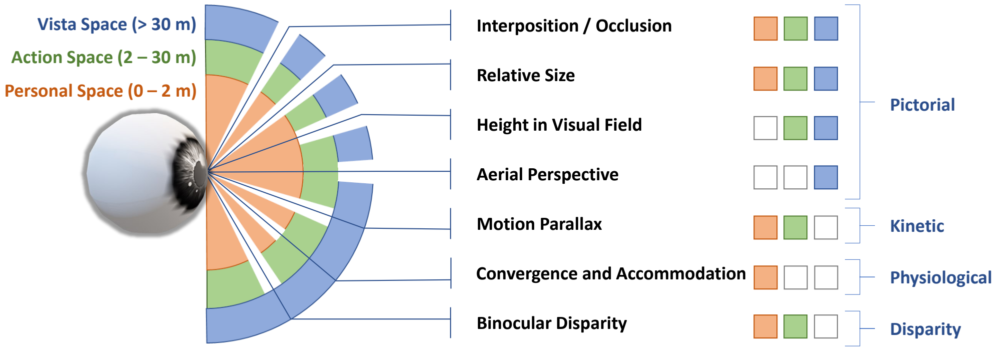


Fig. 2.5. An adaptation from the just-discriminable depth thresholds proposed by Cutting and Vishton [24] excluding *relative density* as proposed by Renner et al. [133]. This depth cue has been removed as it appears to be on the margin of the utility throughout the visible range.

On Visual Cues and Depth Perception

The evidence presented by Cutting and Vishton [24] demonstrates that the usefulness of some depth cues changes as a function of the distance. Thus, it is important to consider the subdivision of the visual space –and the influence of these visual cues– when designing visualization techniques that require the accurate estimation of the objects in the scene (see Figure 2.5).

2.2 Alternative Realities

> *...whereas virtual reality brashly aims to replace the real world, augmented reality respectfully supplements it.*
>
> — **Steven Feiner**
> (Augmented Reality: A New Way of Seeing)

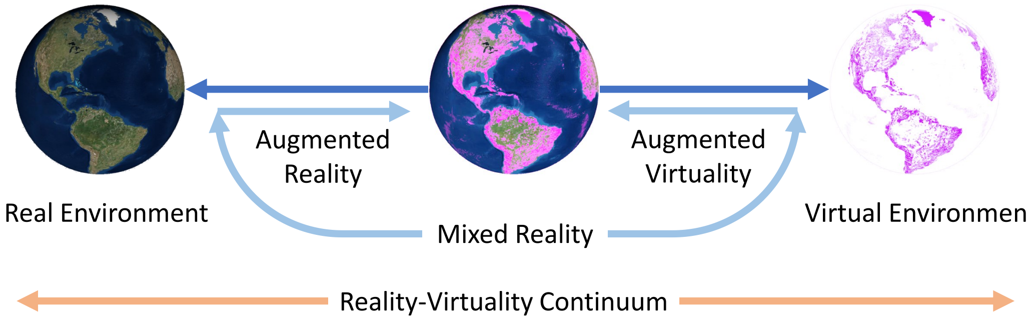


Fig. 2.6. An adaptation of the *Reality-Virtuality Continuum* presented by Milgram et al. [114]. This concept describes two worlds: i) the real one (*left*) that follows the physical laws of gravity, time, and material properties, and ii) the virtual one *right* that introduces a purely immersive and synthetic world, as completely opposites end of a continuum. Between these worlds, Mixed Reality environments combine the elements and properties of real and virtual worlds if an observer visualizes them simultaneously.

One could use the *reality-virtuality continuum* to understand better the relation between the real and virtual worlds and how they interact in MR applications [114]. This concept introduces the real and virtual worlds as the extreme opposites of a straight line. The real environment represents real objects visualized by direct observation or using any available display in this continuum. Correspondingly, the virtual environment acts in place of a construction where the totality of the content observed corresponds to virtual objects. All the other environments, excluded from the spectrum's extremes, in which virtual and real objects co-exist, belong to the MR. This subset of the reality-virtuality continuum includes Augmented Reality (AR) and Augmented Virtuality (AV) environments (Figure 2.6).

In the following, Subsection 2.2.1 provides a brief introduction of these environments. In addition, some alternative continuums that extent the initial definition of Milgram et al. [114] are described in Subsection 2.2.2.

2.2.1 Definitions

The reality-virtuality continuum introduced by Milgram et al. acknowledges the existence of multiple environments that combine virtual and real content to enrich the user's perception. Although it is hard to establish unbendable boundaries between these environments, one way to distinguish between them is to study the ratio of virtual and real content delivered to the user. Therefore, when situated at the left-hand side of the reality-virtuality continuum, no virtual content interacts with the observers in *real environments*. If small bits of virtual content consistently merge with the real world, *augmented reality (AR)* environments can enhance the user's perception and provide visual information otherwise inexistent in the physical world. It is continually moving towards the right-hand side of the reality-virtuality continuum that larger amounts of virtual content dominate the observation of real objects, characterizing the *augmented virtuality (AV)* environments. Lastly, *virtual reality (VR)* describes all those environments where computers completely generate the visual information perceived by the users and where no real objects are observed.

Augmented Reality

One of the earliest and probably most accepted definitions of AR was presented by Azuma [4] in 1997. According to Azuma, AR systems are a variation of VR that enables the visualization of virtual objects that appear to co-exist in the real world. These systems must fulfill three basic requirements: i) to combine real and virtual content, ii) to allow the interaction with the content in real-time, and iii) to register the content in a three-dimensional space. Although multiple AR systems use Head-Mounted Displays (HMDs) to deliver the visual content to the observers, Azuma highlights the importance of not restricting this concept to such devices as some other devices, later discussed in this dissertation, can fulfill the requirements.

Augmented Virtuality

According to Milgram and Kishino [113], AV environments represent the converse case to AR. In this regard, a dominating virtual environment generated using computer graphics is enriched using real components that co-exist with the virtual content. These environments have been used, for example: to create virtual worlds augmented using video textures extracted from real objects such as textured windows, whiteboards, or virtual computer displays showing real content [154], in medical settings for the display and registration of real stereoscopic images from surgical microscopes with pre-operative virtual content [128], or to design serious gaming environments to develop hazard signal detection skills in construction settings [1]. The features presented by these environments give the users the flexibility to interact with the environment, relaxing the temporal, spatial, and physical constraints found in the real world [154], or avoiding the potential dangers of performing high-risk activities on-site [1, 128].

Virtual Reality

Contrary to the physical world, VR environments provide the user with fully computer-generated content. Even though AR, AV, and VR share multiple properties, VR environments produce a fully immersive experience that isolates users from the external physical world. This attribute has proven to be very valuable in multiple settings that require the user to be "disconnected" from the real world, for example, in medical scenarios for the treatment of phobias [13, 49, 103, 134, 139, 169], or as auxiliary systems in the treatment of pain [50, 73, 74, 75, 150, 152, 179]. Nevertheless, this same attribute may not be suitable or desired for other environments requiring full awareness and attention of the users, such as industrial settings.

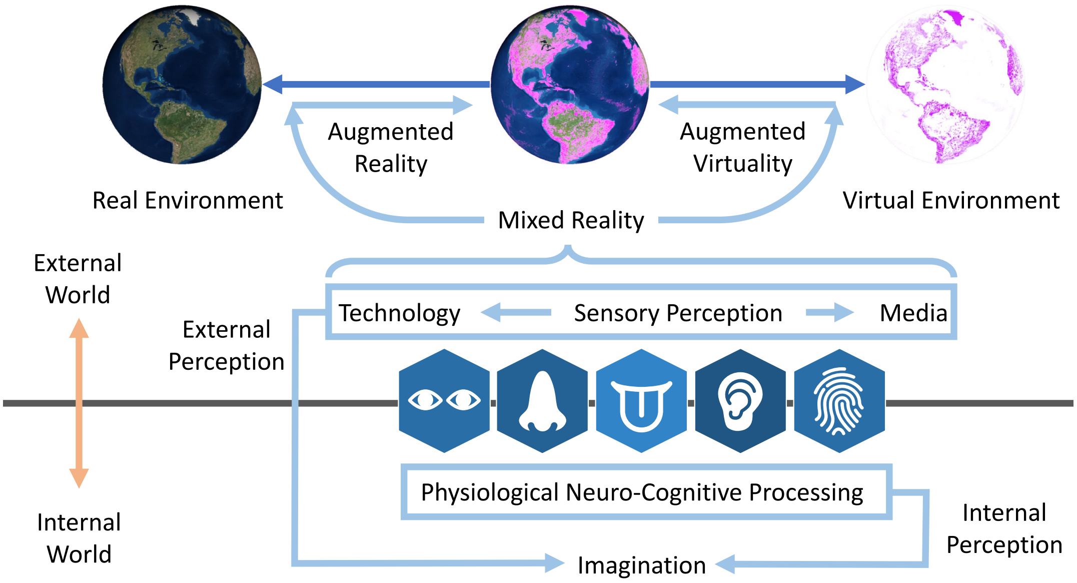


Fig. 2.7. The integration of the imagination can extend the virtuality continuum. In this concept, the real and virtual environments perceived by the senses create an external representation of the surroundings. Then, the imagination can produce its own internal perception to influence this external representation.

2.2.2 Other Continuums

The original definition of the reality-virtuality continuum can be extended from its linear form by adding a new dimension that recognizes the audience's imagination as an essential component of it [158]. The addition of this dimension allows considering ideas beyond the simple interpretation of external signals perceived by the senses (see Figure 2.7). This extension of the continuum acknowledges that the real and virtual worlds create an external perception that can be influenced by an internal perception produced by the imagination. Thus, it can convey the experience that the designer intends for the end-users.

Like the one proposed by Mann [104], alternative continuums extend the linear reality-virtuality continuum presented by Milgram et al. [114] by including another component: mediality. In addition to the considerations described by the MR environments, this concept considers broader aspects assuming that it is also possible to deliberately diminish or otherwise alter the perception of the real and virtual worlds. Therefore, all the environments described by the original continuum of Milgram form a subset of the mediated reality (see Figure 2.8). Such a concept involves those devices that, deliberately or accidentally, alter the user's perception of the environment.

In addition, Newman et al. [119] suggested extending the MR continuum by integrating ubiquitous computing environments. Although originally categorized as a roughly opposite extreme of VR by Weiser [174], Newman et al. recognized that ubiquitous computing could be seen as an orthogonal axis to an analogous continuum that they denominated the "Weiser's Continuum" (Figure 2.9). In this context, combining the Weiser and Milgram continuums would provide a general understanding of how systems, libraries, and frameworks could be used for developing MR and ubiquitous computing environments. Moreover, it would

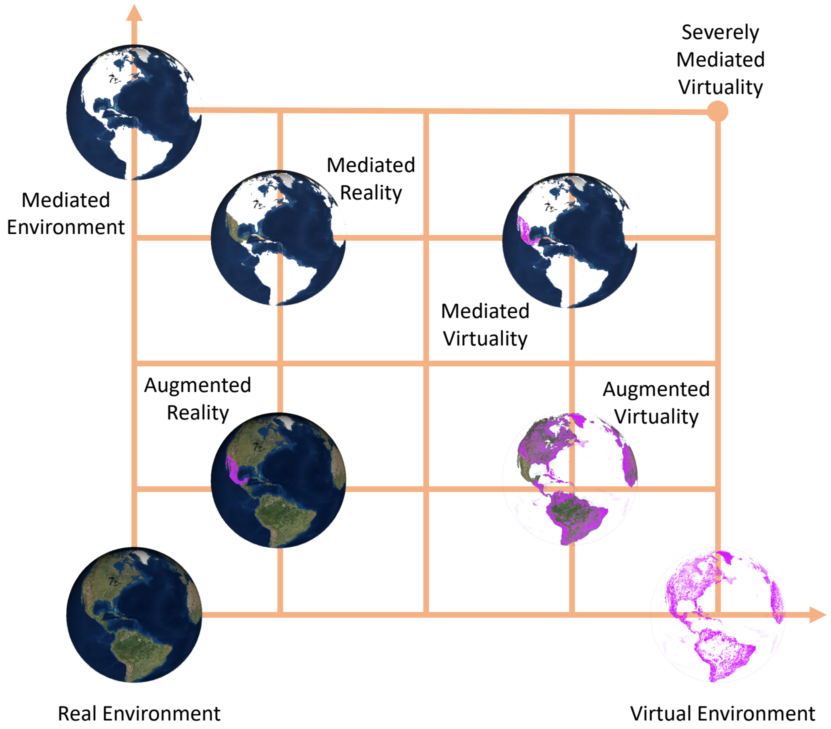


Fig. 2.8. The integration of mediality as a third aspect extends the reality-virtuality continuum. This concept acknowledges that the user's perception can be augmented with virtual content, but also diminished or altered in any form, leading to the concepts of *mediated reality* and *mediated virtuality*.

support developing generalized middleware and facilitate cross-disciplinary cooperation and collaboration between these two fields.

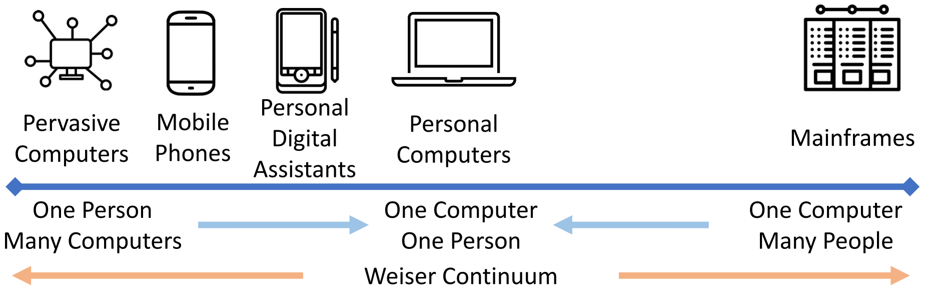


Fig. 2.9. The Weiser Continuum depicts a spectrum ranging from ubiquitous to monolithic computing.

On Mediated Reality and the Reality-Virtuality Continuum

For further details on this topic, please refer to the works of Steve Mann: 1) "Mediated reality with implementations for everyday life." Presence Connect 1 (2002); and 2) "Wearable, Tetherless, Computer-Mediated Reality (with possible future applications to the disabled)." Technical report 260 (1994).

2.3 Depth Estimation in Mixed Reality

Estimating the distribution, organization, and depth at which objects are observed requires the visual system to use several available and generally intersubstitutable information sources [24]. When the observer has plenty of these visual cues, perceiving and estimating distances can be done very accurately. However, this task becomes challenging when the available depth cues are limited or provide conflicting information.

In MR environments, some of the depth cues generated naturally in the real world cannot be faithfully reproduced by the virtual world. Multiple factors such as human factors and technology limitations contribute to the observation of conflicting visual cues when the real and virtual worlds are merged, frequently leading to errors in estimating the depth of the objects. Several studies have explored the accuracy achieved by users of MR applications at estimating the position of real and virtual objects: placed at different distances (i.e., personal, action, and vista spaces), in multiple environments (i.e., indoors vs. outdoors), using different judgment methods (e.g., open- and closed-loop methods), visualized using different technologies (i.e., head-mounted displays, hand-held devices or screens). To improve the general understanding and to provide a homogeneous language throughout this work, the following terminology will be used from now on when referring to:

> **Layout around the observer.** The spaces proposed by Cutting and Vishton [24] will be used when referring to the distance at which the objects are judged and perceived in real and virtual environments. Thus, three distances will be identified: i) the *personal space*, covering distances up to 2 meters from the observer., ii) the *action space*, for those distances between 2 and 30 meters from the observer, and iii) the *vista space* for those estimations made above 30 meters. In the literature, these same distances are also referred to as *near-*, *mid-*, and *far-*distances, respectively.
>
> **Distance estimation judgment.** Most of the studies found in the literature compute the error in the estimated distance using the formula: *error = judged distance – actual distance*. Therefore, any negative errors indicate observers judging the objects closer to themselves than the actual distance, indicating an *underestimation* of their perceived position. On the contrary, positive errors are indicative of an *overestimation* of the position of the observed objects.
>
> **Judgment methods.** Several methods have been used to judge the distance of the objects perceived by the observer. Between these methods, *verbal report* and *action-based* (open- and closed-loop) are used very frequently in the literature [161]. When a *verbal report* method is used, the observer directly communicates the perceived distance using familiar distance units (e.g., centimeters, meters, feet). *Closed-loop action-based* methods provide visual feedback that can be used as a reference to judge the perceived distance. An example of a widely used closed-loop action-based method for judging depth in personal space is *perceptual matching*. This method requires the observer to adjust the position of a physical indicator until, perceptually, it equals the target's depth. Although this method has been extensively used in studies investigating depth estimation in MR, it only provides a relative measure of the perceived distance. This relativity

results from the observer placing the indicator at a relative depth to the depth of the target distance. In contrast, *open-loop action-based* methods provide definite distance perception measurements. A good example of these methods used for depth estimation in personal space is *blind reaching*. This method requires the observer to reach the target distance using its hand but without being able to visualize it. Another example of a commonly used *open-loop action-based* method in action space is *blind walking*.

The following section introduces a literature review on depth estimation studies in MR environments. It also presents a general overview of the perceptual problems, technological limitations, and human factors associated with using this technology and how they affect depth perception and judgment.

2.3.1 Judging Depth in the Personal Space

Correctly estimating the depth of virtual content is one of the most challenging tasks in MR environments [23, 133]. Early work from Ellis and Menges [34] used optical see-through head-mounted displays to investigate how viewing conditions (monocular, binocular, and stereoscopic representations of the virtual content), as well as accommodation, age, or the position of real objects, affect the estimation of the object's depth. This study showed an overestimation of the distance when using the monocular condition (i.e., the object was perceived to be further away). Moreover, a close to correct estimation of depth was achieved using the binocular and stereoscopic conditions. Interestingly, additional experiments suggested that superpositioning virtual content over a real surface led to perceiving the virtual object moving closer to the observer.

Later work from Singh et al. [156] investigated the effects of highly salient occluders in estimating virtual objects' depth in personal space. More specifically, it used five different distances ranging from 34 to 50 centimeters. These experiments employed perceptual matching and blind reaching judgment methods to estimate the depth of virtual objects occluded by the highly salient real object using optical see-through head-mounted displays. Overall results showed an underestimation in the depth judgments, presenting less accuracy using the blind reaching method than perceptual matching. In addition, a constant underestimation error was reported without the presence of the occluder. Interestingly, although greater underestimation errors were reported at smaller distances whit the presence of the occluder, such errors decreased linearly as a function of the distance until they became almost equivalent to the results obtained without the occluder. A very interesting remark from the authors, aligned to the observations made by Ellis and Menges [34], is that virtual objects appear to be "*pushed*" towards the observer when a real object, initially placed behind the virtual object, is slowly moved in the direction of the observer. This visual effect later disappears, giving the illusion that the virtual object suddenly falls behind the real occluder. According to the authors, this effect provides "*a strong sense of transparency*" to the real object [156].

Posterior to the work of Singh and collaborators, Swan et al. [163] investigated the effects of judging not only virtual objects but also real ones using a similar study design. This work also used perceptual matching and blind reaching to estimate the distance of objects placed in the personal space at approximately 34 to 50 centimeters from the observer. Like the findings reported by Singh et al. [156], perceptual matching showed more accurate depth judgments than blind reaching. In addition, it showed that users were able to estimate the position of the real targets accurately. Nevertheless, without the presence of a physical occluder, the results presented by Swan and collaborators showed that observers systematically overestimated the distance of the virtual targets. This difference in the depth estimation accuracy between real and virtual objects was attributed to the display nature of the headset used to deliver the virtual content, causing the observer's eyes to rotate their vergence angle outward. Thus, raising the need to design headsets capable of providing adjustable focus to accurately estimate the position of the virtual content in the personal space.

More recent work presented by Singh et al. [155] explored the effects of focal distance, age, and brightness of the virtual content on depth estimation using perceptual matching in personal space (at approximately 33.3 to 50 centimeters). This work, divided into three experiments, supported the findings from Swan et al. [163] regarding: i) observers can estimate the position of real objects in personal space very accurately (even showing results that were not significantly different from the actual position of the objects), and ii) virtual observed show an overestimation when using collimated optics (that uses a focal distance at infinity). In addition, it extended these findings by providing evidence that adjusting the focal distance to be consistent with the position of the virtual objects contributes to mitigating this overestimation but presenting small underestimation results. Moreover, using a fixed focal distance placed in the middle of the effective depth range (approximately 40 centimeters) also contribute to mitigating these errors, suggesting that a fixed focal distance (set to the middle of the depth range) can provide as good results as optimizing the focal distance for every virtual object. These experiments have also shown that observers susceptible to suffer from age-related deductions in the accommodative capability of the eye did not affect their judgments compared with younger observers. Lastly, this study showed that the most accurate judgments are done for the virtual objects that more closely match the brightness of their real counterparts and that brighter objects are perceived to be closer by the observers.

As an alternative to collimated optics, Peillard et al. [129] compared the accuracy obtained by observers when estimating the distance of real and virtual objects using optical see-through head-mounted displays and retinal projection displays. The objects presented in this study were presented to the observers at distances ranging between 30 to 50 centimeters, and the depth was estimated using a blind reaching judgment method. This study reported an overall underestimation of the distances for real and virtual objects, probably due to the judgment method used. However, an interesting finding is that using optical see-through leads to overestimate the virtual objects' distance compared to the reported estimations for the real objects. In contrast, the use of retinal projection displays helps to mitigate this overestimation significantly. In addition, the estimation difficulty when using these two display technologies did not differ. These results may be an indication that "*the overestimation reported for OST devices is more likely to rely on a specific bias induced by accommodation, rather than being an effect of the vergence-accommodation conflict.*" Moreover, the authors hypothesize that it is better not to have an accommodation cue than an incorrect one.

2.3.2 Depth Judgments in the Action and Vista Spaces

Early work presented by Swan et al. [161], considered the first study to measuring depth judgments in the action and vista spaces, explored how accurately observers could estimate the distance of real and virtual objects, including estimating the distance of real objects, when observed using MR HMDs. This work, divided into two main experiments, used closed- and open-loop estimation methods for evaluating the objects' distance. This study led to a linear model that described the closed-loop distance estimation results of the virtual content, showing a change from under- to over-estimation of the virtual objects' distance at approximately 23 meters from the observer. In addition, using open-loop methods for estimating the object's distance in the action space showed very accurate results for the real-world objects and an underestimation of the virtual objects' distance. These results are consistent with the underestimation reported in VR environments. Nevertheless, the errors showed to be smaller.

Studies conducted by Jones et al. [83] compared egocentric depth perception of virtual and real objects in AR and VR conditions using the same MR device and investigated the influence of motion parallax when estimating distances in the action space. This study observed the commonly reported underestimation using the VR condition, although to a lesser degree than reported in previous studies. In addition, no underestimation was reported for the AR condition. Moreover, a very interesting effect regarding using motion parallax during the estimation of the objects was observed. Although the authors expected improvements when using motion parallax, as this would provide additional information from the scene, motion parallax did not assist observers in achieving more accurate judgments. On the contrary, the only effect observed was an interaction when estimating the distance of the real objects using the headset, making the depth judgments less accurate. Nonetheless, this effect could result from the influence that mass and inertia derived from wearing an HMD have over depth underestimation, as reported by Willemsen et al. in VR environments [178]. Further studies, also conducted by Jones et al. [82], aimed at exploring whether users gathered additional sources of implicit feedback from blind walking estimation methods, including proprioception and peripheral visual information, using a series of four different experiments. A first experiment replicated the setup presented in [83] using a between-subject instead of a within-subject experimental design. This study did not show significant estimation differences between both experimental designs. However, it served to identify that the distance judgments improved over time. These experiments also served to discriminate proprioception as a potential source of feedback and revealed that optical flow in an observer's periphery plays a determinant role in improving depth judgments when using directed walking techniques as judgment methods in VR and AR.

In addition, other studies have explored how accurately human observers can estimate the object's depth in MR environments using hand-held devices. The work presented by Swan et al. [162] explored how accurately users could estimate the depth of real and virtual persons in outdoor and indoor environments, placed at distances ranging between 15 and 30 meters under real and MR conditions. Although the environmental setting did not reveal any statistical significance, a very interesting finding from this study is that observers seem to overestimate the distance at 15 meters and underestimate it at 30 meters under the MR condition[1]. Further work from Liu et al. [100] used a similar setup to estimate the distance of persons standing between 10 and 40 meters from an observer visualizing an MR scene through a hand-held device. Like the earlier results reported by Swan et al. [162], participants of this study overestimated the virtual content when placed between 10 and 25 meters and underestimated it at 40 meters. Recent work conducted by Chakraborty et al. [22] replicated the study setup from Liu et al. [100] to explore the effects of using animated cues during depth estimation. This study showed similar results to those reported by Liu et al. [100]. Moreover, it implied that animated cues do not help get more accurate depth estimations but instead accentuate the virtual target's underestimation. Similar research presented by Gagnon et al. [48] used head-mounted displays to estimate the distance of real and virtual persons located at distances ranging from 10 to 35 meters from the observer in MR indoor environments. This work showed that the distance to the augmented targets is underestimated compared to real targets. However, this misestimation is less evident when performed in narrow spaces than in open spaces.

Alternative approaches have studied how observers perceive the depth of occluded scenes in MR. Early work from Furmanski et al. [44] explored the benefits of integrating depth cues into the design of specialized rendering techniques for visualizing virtual content located within real closed objects. This work showed promising results suggesting that using "cut-away" boxes, overlaid on top of the real content to visualize the objects of interest in-situ, would help mitigating estimation errors derived from misleading depth cues. In this context, Dey et al. [27] introduced visualization techniques to observe virtual content occluded by real objects in outdoor environments. These techniques allowed for the visualization of the virtual content by integrating ghosting techniques or occluding the real objects with virtual elements. In addition, integrated additional depth cues in virtual rulers to support the estimation of the virtual content's depth. Furthermore, a follow-up study evaluated depth perception in the vista space using such techniques and hand-held devices when visualizing virtual content placed between approximately 70 and 120 meters from the observer.

Despite the multiple studies exploring depth estimation in the action and vista spaces, very little work has been reported for estimating objects placed at distances larger than 30 meters. In this regard, more recent studies, like the one presented by Gagnon et al. [47], explored this phenomenon when estimating distances ranging from 25 and up to 500 meters using video see-through head-mounted displays in mediated MR environments. Such a system allowed the observer to visualize their own body while simultaneously observing a virtual environment. Results from this work differ from the underestimation errors reported by Dey et al. [27] and shown a relatively accurate distance estimation. However, a switch from underestimation to overestimation at approximately 200 meters was reported.

[1] In this study, observers used a bisection method to judge the distance between themselves and the target person. This task provides a relative distance perception instead of a direct one.

Part II

Single-View Alignment

As described in the previous sections, one of the core contributions of this dissertation is to investigate whether the selection of visualization techniques used to represent the virtual content influences the user's accuracy during interactive alignment tasks in MR. Thus, this part investigates how users benefit from observing virtual content represented using non-traditional visualization techniques during egocentric interactive alignment tasks where a single viewpoint of the scene is available. These visualization techniques exploit the physical properties of the real objects to align and consider the perceptual aspects discussed previously.

The first approach, discussed in Chapter 3, presents a user study that exposes how the observation of ambiguous information arising from misleading occlusion affects the user's performance during alignment tasks. This misleading occlusion originates from a common problem in MR applications in which virtual objects occlude their real counterparts when they overlap. The presented study, conducted using a VR environment, investigates how the occlusion ratio observed when the real and virtual objects overlap affects the user's accuracy and preference.

A second approach, presented in Chapter 4, introduces the COMPLEMENTARY TEXTURES. This novel class of visualization techniques provides meaningful representations of the virtual objects by exploiting the texture and geometry of their real counterparts. In addition, a subclass of COMPLEMENTARY TEXTURES aims at utilizing the user's familiarity with the real objects to provide semantic information during the alignment task.

In addition, Chapter 5 presents a study investigating the importance of providing appropriate visual cues for in-situ visualization (i.e., visualizing virtual content placed inside solid opaque real objects). This section provides a comparison between depth estimations when using video see-through (VST) and optical see-through (OST) head-mounted displays (HMDs). In addition, it discusses how the additive property of the display technology used in OST-HMDs hinders the generation of convincing visual cues to provide realistic occlusion using dark colors. Furthermore, it proposes a taxonomy for the decomposition of techniques used for in-situ visualization. This taxonomy aims at inspiring the design of alternative visualization techniques that can alleviate the limitations observed when using additive displays for in-situ visualization.

This part contains results from the works:

Martin-Gomez, Alejandro, Ulrich Eck, and Nassir Navab. "Visualization techniques for precise alignment in VR: A comparative study." In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 735-741. IEEE, 2019. © 2019 IEEE.

Martin-Gomez, Alejandro*, Jakob Weiss*, Andreas Keller, Ulrich Eck, Daniel Roth, and Nassir Navab. "The Impact of Focus and Context Visualization Techniques on Depth Perception in Optical See-Through Head-Mounted Displays." In IEEE Transactions on Visualization and Computer Graphics, 2021. © 2021 IEEE.

All authors have permitted the reproduction of this content.

* The asterisk indicates an equal contribution from the corresponding authors.

3 Misleading Occlusion

Several studies have explored the influence of multiple aspects on the estimation and perception of distances in MR. These aspects include if the objects observed are real or virtual [34, 156, 161], the modality or display technology used to visualize the objects [129], the methods used to judge the objects' depth [163], the distance at which the objects are presented [82, 83, 161], the environment and conditions where the estimation task is performed [44, 45, 46, 155, 156], or even the physiology of the observers [155]. However, there seems to be no evidence of previous work exploring whether the visualization technique used to present the virtual content influences how the observers perceive the objects and if this affects their performance while estimating the objects' depth.

Accurately estimating the virtual content's depth is particularly important for MR applications that use virtual replicas of real objects to assist users during alignment tasks [18, 37, 85, 122, 132, 135, 164, 165, 167, 175, 181]. As discussed in Section 1.1, a common factor in applications that use this approach is that the arrangement and sometimes the geometry of the real objects are unknown. This missing information hinders the generation of realistic occlusion, the strongest visual cue to determine depth ordering [29], and provides misleading depth cues. As a result of this, the virtual content consistently appears in front of the real objects, even when the virtual objects are behind their real counterparts. This limitation plays an important role during the interactive alignment of real and virtual content. Existing studies have shown that the observation of this misleading information results in the illusion of the virtual objects being pushed towards the observer [34, 156].

This problem motivates the question of whether observing misleading occlusion plays a role in the accuracy that the users of MR applications can achieve during alignment tasks. Especially in MR environments that cannot provide realistic occlusion when real and virtual objects overlap. As an initial effort towards answering this question, this section presents a user study conducted to compare the alignment accuracy, time to completion, and user's preference when utilizing classical visualization techniques to present the virtual objects during alignment tasks. These visualization techniques offer different ratios of occlusion when the objects overlap during the alignment process. Although implemented using a VR environment, this study aims at replicating the visual properties of the environments described here while allowing the collection of reliable data.

3.1 Visualization Techniques

This section presents a classification of classical visualization techniques used in MR applications to align real and virtual objects. These techniques are arranged considering the ratio of occlusion observed when the real and virtual objects are perfectly aligned.

Solid Replicas. One of the most common techniques to present virtual replicas to the users of MR technologies is by rendering them as textured or single-colored solid replicas of the objects of interest. This representation has been used in multiple alignment tasks involving user's guidance during the assembly of doors in automotive environments [132], the evaluation of the effectiveness of providing instructions using this modality in an assembly task [164], or for context-aware support systems in assembly tasks [85]. This type of visualization provides the highest ratio of occlusion when the objects overlap. Thus, they are frequently used in environments in which the evaluation of the alignment task uses discrete units as the scenarios involving the assembly of building blocks. In these applications, the alignment errors account for the number of times a user failed to place the block in the right place or counting how many units the blocks were misplaced.

Semitransparent Replicas. This visualization technique differs from the solid representation by changing the transparency value of the virtual objects. This effect is achieved by modifying the alpha channel of the virtual objects to any value bigger than zero and smaller than one. The rest of the color channels of the virtual objects remain the same regardless of if the replica presents texture or a homogeneous color. Existing work presented by Buchmann et al. [18] explored the effects of changing the transparency of the objects and the hands of the users of this technology. Results from this work suggest that users prefer alpha values between 0.6 and 0.8. An example of applications that have used this type of visualization to represent objects of interest is the work presented by Henderson et al. [68] to assist users during the performance of procedural tasks. Although this technique enables the visualization of the real objects during the alignment task, the occlusion ratio observed when the objects overlap equals the one observed when using solid replicas. Moreover, defining an optimal transparency level may depend on the type of display used to deliver the virtual content.

Wireframes. This technique represents one of the earliest visualization techniques to present virtual content in MR environments. This technique connects the vertices of the virtual models using single-colored solid lines to provide a mesh-like representation of the objects of interest. One of the advantages of this visualization technique is that it enables the visualization of the real object through the virtual replica. It also provides visual information that facilitates understanding the geometry of the object to align. Because of this, the occlusions ratio observed when overlap occurs strongly depends on the density of vertices the model contains. In this regard, when the vertices density is high, the occlusion ratio is high. Although reducing the number of model's vertices represents an alternative to decrease the mesh complexity and the occlusion ratio, this approach requires carefully selecting the decimation level to avoid compromising the visual quality of the virtual replica. A clear example of vertices reduction was used in [132], except that the virtual models were rendered as solid objects and used for instructional purposes.

Point Clouds. This type of visualization represents an alternative to the wireframes. This technique is commonly used to generate a virtual copy of a real object using depth cameras, three-dimensional scanners, or photogrammetry algorithms. However, it can be used to represents the vertices of a virtual object. Thus, allowing for the visualization of the three-dimensional shape of the object of interest while reducing the occlusion ratio. Nonetheless, this type of visualization provides visual information about the distribution of the vertices in the virtual model rather than providing information on the real object's geometry.

Besides the most common techniques used to represent virtual content during alignment tasks in MR, this work includes two additional visualization techniques that provide a low occlusion ratio when the objects overlap. In addition, these techniques provide useful information about the shape of the objects that can be used during the alignment task. These techniques are:

Fresnel Shaders. This type of visualization uses the object's surface normals and the observer's view direction to determine the reflectance observed over the surface of a virtual object. This technique uses the dot product of the direction vectors of these two components as a function of opacity. Therefore, the object's surface facing towards the observer is rendered as fully transparent, while the perpendicular sections are rendered opaque. Such property allows for the visualization of the real object when the objects overlap. Moreover, it still provides useful information about the object's shape.

Silhouette / Contours. The contours techniques are frequently used to attract the observer's attention in virtual environments. These methods provide an edge-like visualization of the virtual objects. Therefore, they allow visualizing the object's shape while providing an uncluttered view of the object of interest. Although these techniques do not provide larger amounts of visual information, the observer can still use the information to understand the shape and position of the objects.

3.2 User Study

A user study served to investigate the effects of observing conflicting visual cues during alignment tasks and if the occlusion ratio affects the users' accuracy. The user study was conducted using a virtual environment in which textured objects -–from now on, referred to as *interactable objects*— were aligned by human operators using the pose of a virtual replica as reference. The virtual replicas -–from now on referred to as *target objects*— were rendered using one of the classical visualization techniques presented in Section 3.1. The decision of implementing this study using a virtual environment was to ensure that the geometric properties of the objects to align would be identical. In addition, this would allow acquiring reliable measurements of position, orientation, and time.

3.2.1 Selection of Visualization Techniques

Four visualization techniques chosen from those presented in Section 3.1 were considered as part of the evaluation (see Figure 3.1). The target objects were rendered using single-colored materials to provide a fair comparison between the techniques.

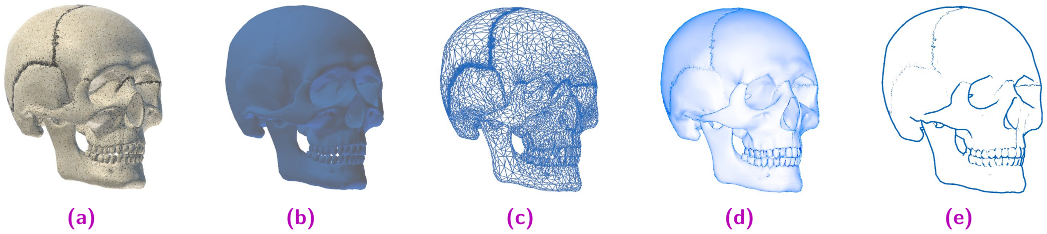


Fig. 3.1. Screenshots of the four visualization techniques employed to compare users' performance for precise object alignment in 6 DoF. (a) Interactable textured object and its four representations rendered with the visualization techniques –(a) *Semitransparent*, (b) *Wireframe*, (d) *Fresnel-Derivative*, and (e) *Silhouette*–.

In addition, the target objects were always rendered in front of the interactable objects regardless of their depth. Therefore, providing adequate occlusion is not possible. Furthermore, the interactable objects were not affected by shadows cast by the target objects. This design consideration allowed for the simulation of scenarios where no knowledge of the real environment's geometry is available. More details and design consideration of the individual techniques involved in the study are described as follows:

Semitransparent. This visualization technique allows rendering the target objects using a single color with an alpha value lower than 1.0. Modifying this alpha value allows the observation of the interactable object through the target counterpart, even in the presence of occlusion. Although results from Buchmann et al. [18] suggest that users prefer alpha values ranging between 0.6 and 0.8, these levels turned to be slightly high, impeding the observation of the objects. The alpha value used during this study was equal to 0.5.

Wireframe. This technique allows rendering the target objects as single-colored meshes connecting the edges of the model's faces using solid lines. The number of vertices of the virtual models was reduced when required to improve the objects' visibility during task performance.

Fresnel-Derivative. This variant of the Fresnel technique is composed of two core components. The first element, the Fresnel component, computes the inverse additive of the dot product between the normal surfaces of the object and the observer's view direction –the angle of incidence–. The resulting value modifies the object's transparency in the respective surface. Thus, allowing for the observation of the object's edges and surfaces with high angles of incidence. The second element, a derivative component, allows for observing the model surfaces with high curvature. This component results from comparing the gradient of the object's surface and a threshold d_t. A similar approach to this technique was presented by Bichlmeier et al. [8] to improve the perception of medical imaging data in MR. A mathematical representation of this model is presented in Equation 3.1.

$$\alpha = I_v * (d_f + d_c) \tag{3.1}$$

where α is the value of the object's alpha channel, I_v is a multiplicative intensity factor, while d_f and d_c are the Fresnel and derivative components, respectively. These components are computed using Equation 3.2 and Equation 3.3.

$$d_f = (1.0 - \delta)^{f_p} \tag{3.2}$$

where δ is the angle of incidence and f_p is the Fresnel factor power.

$$d_c = \begin{cases} 0 & if \left(\frac{\mathrm{d}}{\mathrm{d}x}SN + \frac{\mathrm{d}}{\mathrm{d}y}SN\right) \leq d_t \\ \frac{\mathrm{d}}{\mathrm{d}x}SN + \frac{\mathrm{d}}{\mathrm{d}y}SN & if \left(\frac{\mathrm{d}}{\mathrm{d}x}SN + \frac{\mathrm{d}}{\mathrm{d}y}SN\right) > d_t \end{cases} \tag{3.3}$$

where SN is the object surface normal, and d_t is a derivative threshold.

The parameters used during the user study were: $f_p = 17.5$ and $d_t = 0.215$.

Silhouette. This visualization technique involves the use of two rendering phases. The first phase extends the object's surfaces in the direction of their normals by a scale factor s_f. This first phase is rendered using a single solid color and producing an upscaled version of the original model. The second phase renders all the pixels of the model without growth as a fully transparent object. Combining these phases leads to visualizing the object's edges and some inner characteristic features that provide additional visual cues about the object's geometry. The value of s_f for this study was equal to 0.015 times the model size.

3.2.2 Selection of Models

Four models with different shapes, curvatures, and densities of vertices were used to diversify the objects that participants would manipulate. These models are shown in Figure 3.2. The *Mug* has a low density of faces and a smooth curved homogeneous surface. The handle provides visual cues that can be used during alignment, especially for orientation. The *Camera* combines curved and planar elements as well as a higher density of vertices than the *Mug*. The *Skull* is composed of smooth and curved surfaces with sharp edges and higher face density than the previous models. Lastly, the extremities of the *Balloon* are useful for precise alignment of the object's orientation, while its face density and curvature are similar to the *Skull*.

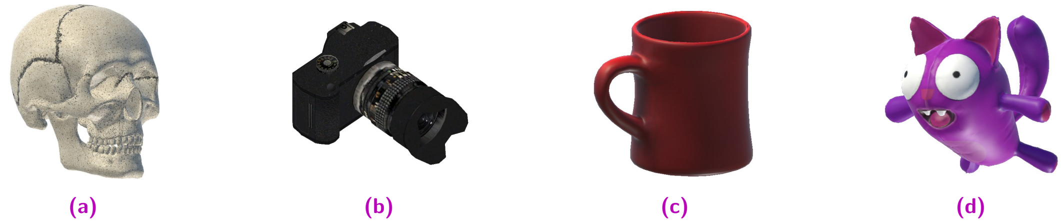


Fig. 3.2. Interactable objects utilized during the user study. (a) *Skull*, (b) *Camera*, (c) *Mug*, and (d) *Balloon*.

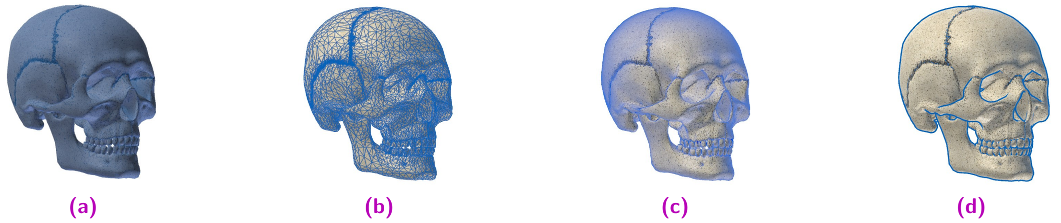


Fig. 3.3. Example of an *interactable object* aligned with each one of its *target objects*. (a) *Semitransparent*, (b) *Wireframe*, (c) *Fresnel-Derivative*, and (d) *Silhouette*.

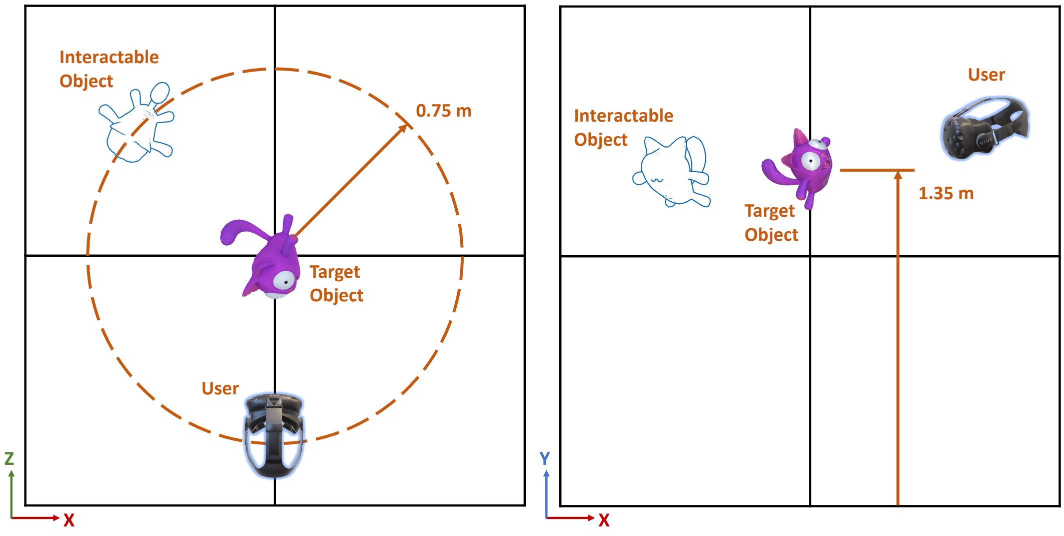
As shown in Figure 3.3, the density of the vertices affects the ratio of occlusion for the *Wireframe technique*. High curvature regions of the models generate higher levels of occlusion when using the *Fresnel-Derivative*, and sharp edges are more visible when using the *Silhouette technique*.

3.2.3 Participants

Thirty-two unpaid participants (11 female and 21 male), aged between 22 and 36 years old (mean age of 27.0 ±3.1 years), participated in the study. None of the participants had previous experience using the system. As a prerequisite, participants were asked to perform the standard Ishihara test [78] to detect color vision deficiency. After completing the Ishihara test, the headset was given to the participants, and they were asked to adjust the headset's interpupillary distance until they could read a welcome message. All participants wearing glasses to correct vision problems were allowed to use them during the experiment.

3.2.4 Experimental Setup

A preliminary pilot study helped to determine the parameters used to present the virtual objects to the participants. This preliminary study helped determine how the virtual objects would appear during the user study as depicted in Figure 3.4. The interactable and target objects appeared at 1.35 meters above floor level. The target objects were located pseudo-randomly over a circumference with a radius equal to 0.75 meters. The center of the circumference was used to place the interactable objects and served as the scene's center. The orientation of the target objects was also defined pseudo-randomly. The virtual environment presented an empty scene with a homogeneous, gray-colored background to avoid external distractions.



(a) Top-Down View

(b) Lateral View

Fig. 3.4. Placement of the virtual objects during the user study seen from (a) the top and (b) the side.

In addition, to mitigate learning effects, a 16 x 16 Latin square matrix was used to define the combination of *visualization technique* ⊗ *object type* presented to the participants. Lastly, two different alignment scenarios, with and without time constraints, were included in the study. In this regard, the pilot study helped establish the time limit for the constrained scenario to 30 seconds. This time corresponds to the time required for most participants to achieve an acceptable level of alignment during the preliminary study.

Interaction Modes

Two interaction modes were available for the manipulation of the interactable objects during the study. The *normal mode* enabled the participants to manipulate the interactable objects using a direct attachment to the interaction handle. The *precise mode*, enabled exclusively when the Euclidean distance between the target and interactable object was smaller than 10 centimeters, allowed for the manipulation of the interactable object's using a transformation ratio of one-tenth between the physical transformation and the transformation applied to the interactable object. Separate buttons controlled both modes. In addition, to avoid reporting false positives, a third button had to be pressed for two seconds by the participant to confirm the completion of the alignment task.

Tutorial

This session allowed participants to learn how to manipulate the interactable objects using the interaction methods designed for the study. Neither the object nor the visualization technique used in the tutorial was part of the experiments. The participants were allowed to interact freely with the object presented until they felt comfortable manipulating the object presented in this stage.

Object Alignment with No Time Constraints (NTC)

For the NTC stage, participants had to align as precisely as possible an interactable object using the pose indicated by a target object. Sixteen different scenarios resulting from the combinations of objects (Figure 3.2) and visualization techniques (Figure 3.1) were presented to the participants. Only one pair of corresponding objects was presented in the scene at a time. Participants were asked to align the objects as accurately as possible and were given as much time as needed to complete the alignment task. After user confirmation, a new pair of objects appeared in the scene until the completion of the sixteen alignment tasks. After completing the NTC stage, participants were allowed to take a rest.

Object Alignment with Time Constraints (TC)

This study stage considered a scenario in which the alignment task must be completed within a certain amount of time. The objects and techniques were the same as those used in the NTC stage; however, the objects' order of appearance was different. The participants were asked to align the objects as precisely as possible and were notified about the time constraints. Before triggering a countdown timer, always visible to the participants, they were allowed to inspect the scene. Any change in the pose of the interactable object triggered the countdown timer. Once this timer ran out of time, a new pair of corresponding objects was shown in the scene.

Usability and Mental Effort Questionnaires

After completing the NTC and TC scenarios, participants evaluated the visualization techniques using Brooke's system usability score (SUS) questionnaire [17]. They also reported the mental effort perceived (ME) using a nine-point Likert scale introduced by Paas [126].

3.2.5 Experimental Variables

Two independent variables: *model* and *visualization technique* –each with four different levels– were involved in both alignment scenarios –NTC and TC–. For the NTC scenario, three dependent variables were computed after confirmation: i) positional errors calculated as the Euclidean distance between the gravity centers of the interactable and target objects. ii) orientation errors computed as the axis-angle transformation between the interactable and target objects. iii) The time elapsed from the first change in the interactable object's pose until alignment confirmation. For the TC scenario, only the variables for position and orientation were considered.

3.2.6 Hypotheses

Motivated by the occlusion problems described in this section, the following hypotheses were proposed:

H1. Users perform better in terms of position, orientation, and time to completion under NTC conditions when visualization techniques with low occlusion rates are used compared to those that present high levels of occlusion.

H2. Users perform better in terms of position and orientation under TC conditions when visualization techniques with low occlusion rates are used compared to those that present high levels of occlusion.

H3. User's preference in terms of usability and mental effort is higher when visualization techniques with low occlusion rates are used compared to those that present high levels of occlusion.

3.3 Results

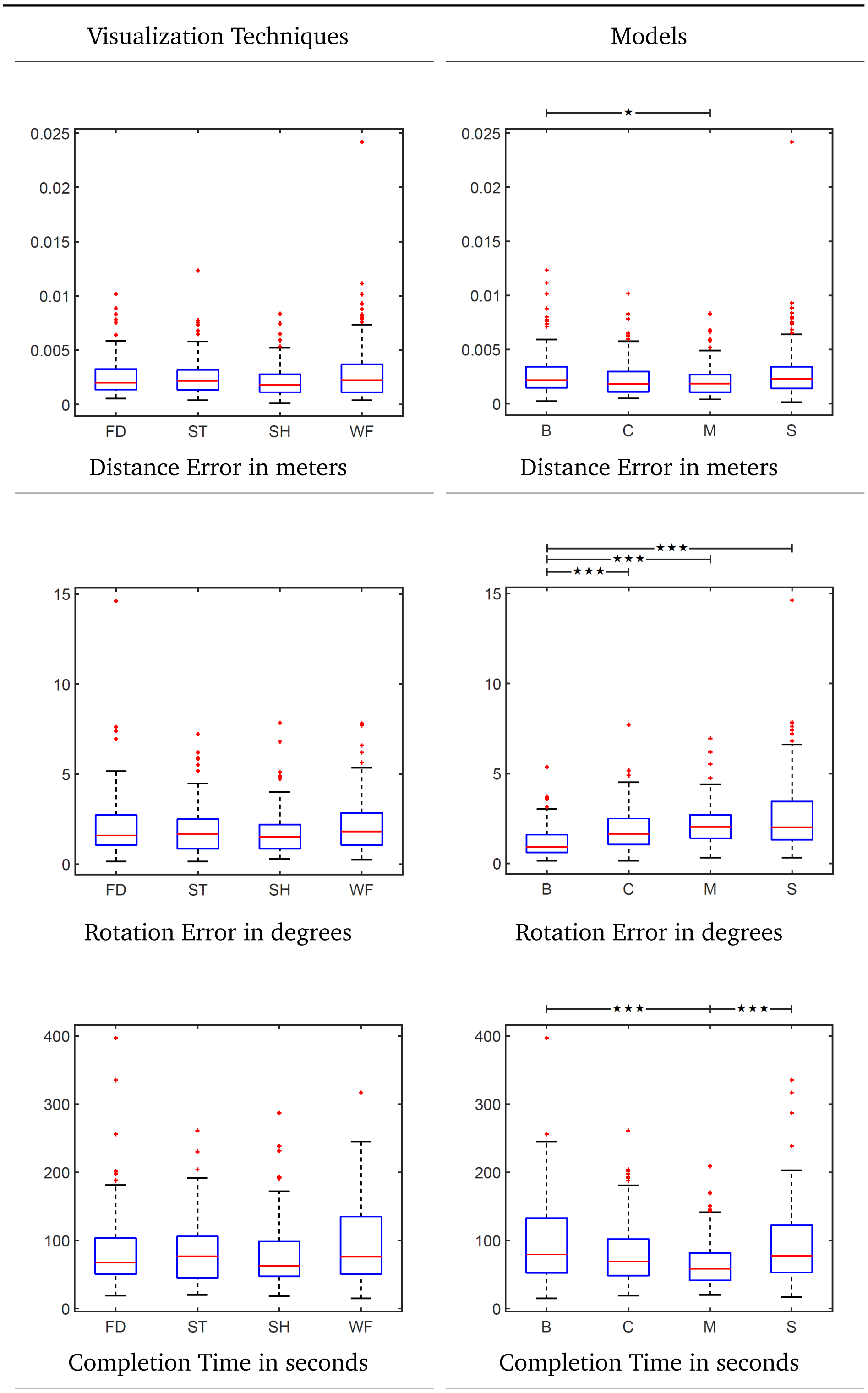
This section provides a statistical analysis of the results collected during the user study for the NTC and TC scenarios and the SUS and ME scores. Considering the data collected presents a non-normal distribution, Kruskal-Wallis tests with $\alpha = 0.05$ were used to compare the position and orientation scores for the NTC and TC scenarios and the time to completion in the NTC scenario. This test was also used to appraise the SUS and ME scores. Posterior Bonferroni tests were used to reveal significant differences between the variables based on the results obtained from the Kruskal-Wallis tests. Significant differences between variables are indicated as $\star\,(p < 0.05)$, $\star\star\,(p < 0.01)$, and $\star\star\star\,(p < 0.001)$ in Figure 3.5 (NTC), Figure 3.6 (TC), and Figure 3.7 (SUS and ME).

Overall, using the Silhouette yielded the best scores in position, orientation, time to completion for the NTC condition, rotation under the TC condition, and usability and mental effort. In addition, the Semitransparent technique scored the best in terms of translation for the TC scenario. Additional analysis regarding model behavior revealed the highest scores for position when using the Mug, and the Balloon obtained the best scores for orientation. Finally, during data analysis, one outlier with an orientation error of 169.52°, reported using the Skull model and rendered with the Wireframe technique, was replaced with the mean value of the sample.

3.3.1 NTC Statistical Analysis

The Silhouette visualization technique reported the lowest mean and standard deviation errors for translation of all visualization techniques, while the Wireframe reported the highest. However, Kruskal-Wallis tests did not reveal a significant interaction between these techniques. In contrast, statistically significant differences between models for errors in position ($\chi^2(3) = 13.01, p < 0.01$) were found. Participants were able to position the Mug significantly better than the Balloon ($p < 0.05$).

Participants reported the lowest mean errors using the Silhouette and the highest using the Wireframe in terms of rotation errors. Nevertheless, Kruskal-Wallis tests failed to reveal significant differences in the visualization techniques. Like the position scores, statistically significant differences between models were found for rotation ($\chi^2(3) = 79.21, p < 0.001$). Participants achieved significantly better orientation scores aligning the Balloon than the Camera ($p < 0.001$), the Mug ($p < 0.001$), and the Skull ($p < 0.001$).


Visualization Techniques

Models

Distance Error in meters

Distance Error in meters

Rotation Error in degrees

Rotation Error in degrees

Completion Time in seconds

Completion Time in seconds

Fig. 3.5. Scores obtained by users under NTC conditions. (Left) Results grouped by visualization techniques: *FD-Fresnel-Derivative, ST-Semitransparent, SH-Silhouette, WF-Wireframe*. (Right) Results grouped by model: *B-Balloon, C-Camera, M-Mug, S-Skull*. (The red line indicates the median, and the bottom and top edges of the blue box indicate the 25th and 75th percentiles).

In addition, and similarly to the position and orientation scores, the Silhouette technique obtained the best results in terms of time to completion, while the Wireframe performed the worst. Although, Kruskal-Wallis results revealed no significant difference between visualization techniques for the time scores. In terms of models, Kruskal-Wallis results revealed a statistically significant difference for the time scores ($\chi^2(3) = 23.74, p < 0.001$). Users completed significantly faster the alignment task when aligning the Mug compared to the Balloon ($p < 0.001$) and the Skull ($p < 0.001$).

These results are summarized in Figure 3.5 and Table 3.1.

Tab. 3.1. Mean and standard deviation scores for the position, orientation, and time to completion achieved by users under NTC conditions after grouping the results by visualization technique and model type.

	Distance (centimeters)		Rotation (degrees)		Time (seconds)	
Visualization Technique	Mean	SD	Mean	SD	Mean	SD
Fresnel-Derivative	0.26	0.17	2.07	1.76	84.98	58.60
Semitransparent	0.25	0.17	1.92	1.36	84.45	47.34
Silhouette	0.22	0.15	1.79	1.30	78.12	48.15
Wireframe	0.30	0.30	2.17	1.47	92.75	55.52
Model Type	Mean	SD	Mean	SD	Mean	SD
Balloon	0.28	0.21	1.20	0.86	96.22	58.81
Camera	0.23	0.17	1.92	1.21	83.64	50.19
Mug	0.22	0.15	2.20	1.18	67.80	38.15
Skull	0.30	0.28	2.62	2.06	92.76	57.19

3.3.2 TC Statistical Analysis

Similar to the NTC scenarios, Kruskal-Wallis and Bonferroni's post hoc tests were used to reveal statistical significance between the models and visualization techniques. In terms of errors in position, no significant difference between visualization techniques for the TC scores was revealed. However, significant differences between models ($\chi^2(3) = 41.3, p < 0.001$) were found. Participants were able to align more accurately the Mug than the Balloon ($p < 0.001$) and the Skull ($p < 0.01$), as well as the Camera compared to the Balloon ($p < 0.01$) and the Skull ($p < 0.05$).

In terms of rotation, participants achieved better scores using the Silhouette. However, Kruskal-Wallis results revealed no significant difference between visualization techniques. In addition, significant differences between models for the rotation scores ($\chi^2(3) = 61.6, p < 0.001$) were found. Users achieved lower orientation errors using the Balloon compared to the Camera ($p < 0.001$), the Mug ($p < 0.001$), and the Skull ($p < 0.001$); and using the Camera compared to the Skull ($p < 0.01$).

These results are summarized in Figure 3.6 and Table 3.2.

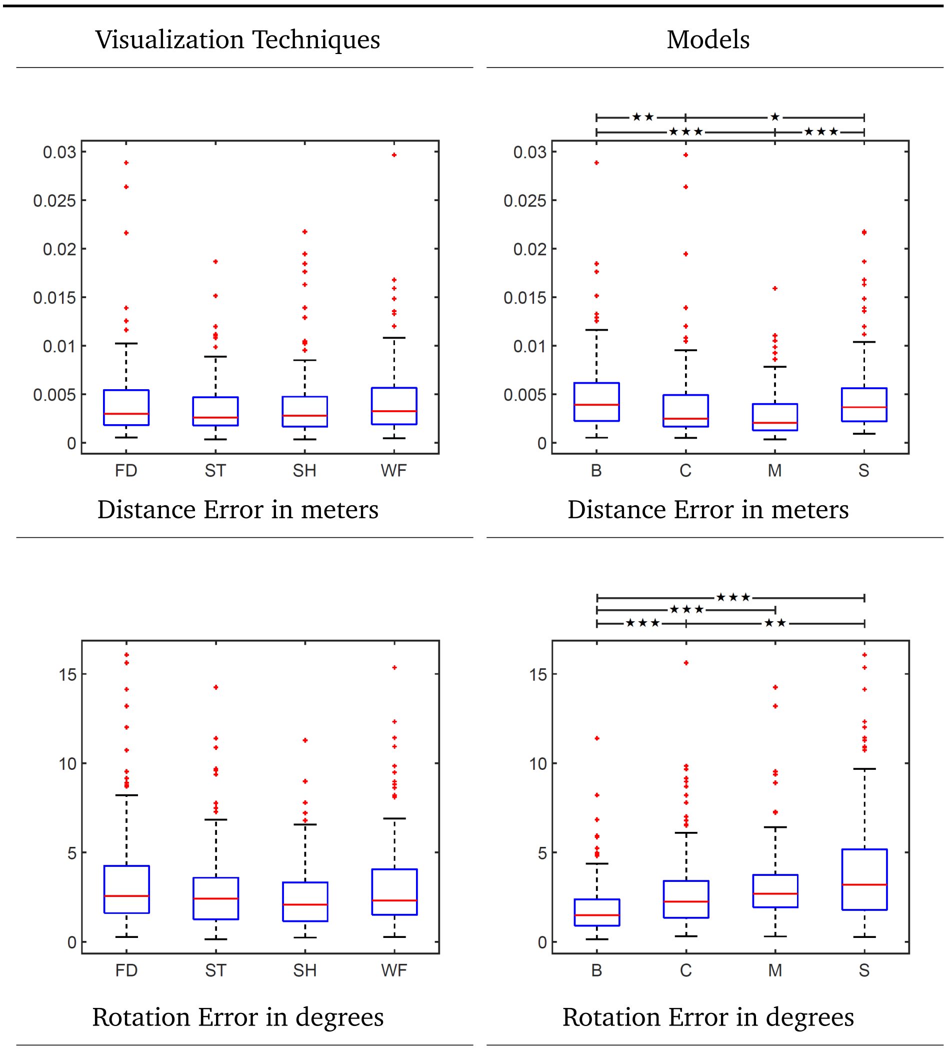


Fig. 3.6. Scores obtained by users under TC conditions. (Left) Results grouped by visualization techniques: *FD-Fresnel-Derivative, ST-Semitransparent, SH-Silhouette, WF-Wireframe*. (Right) Results grouped by model: *B-Balloon, C-Camera, M-Mug, S-Skull*. (The red line indicates the median, and the bottom and top edges of the blue box indicate the 25th and 75th percentiles).

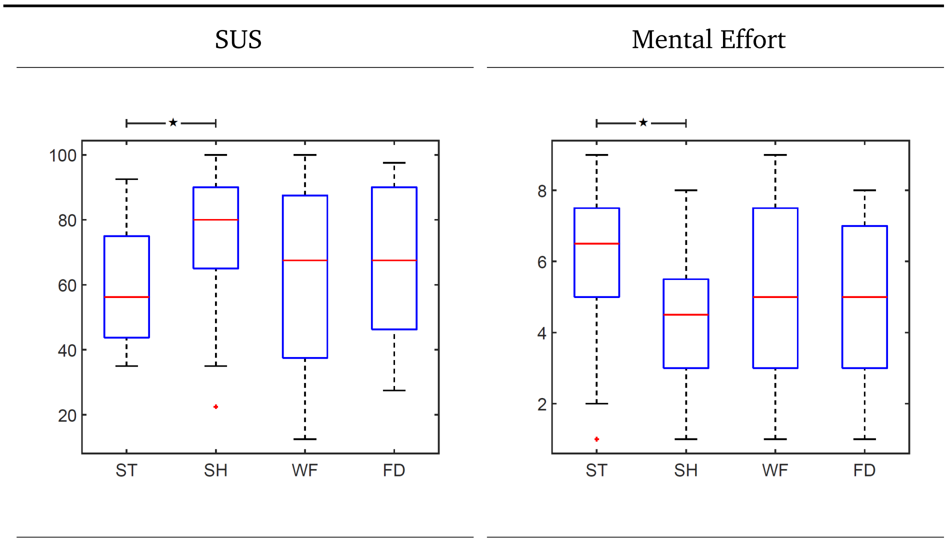
3.3.3 SUS and ME Statistical Analysis

The Silhouette and Fresnel-Derivative techniques obtained the highest scores regarding perceived usability and mental effort, followed by the Wireframe and Semitransparent. SUS scores between 51 and 68 are considered average results, scores between 68 and 80.3 are above average, and scores higher than 80.3 are considered in the top 10%. The SUS tests' mean and standard deviation results, summarized in Table 3.3, show that the Silhouette technique was the only visualization technique that scored above 68 but under 80.3. Kruskal-Wallis results revealed significant differences between visualization techniques for the SUS score $(\chi^2(3) = 10.26, p < .05)$. A Bonferroni post hoc test revealed that usability scores were significantly better for the Silhouette visualization technique than for the Semitransparent $(p < 0.05)$.

Tab. 3.2. Mean and standard deviation scores for the position, orientation, and time to completion achieved by users under TC conditions after grouping the results by visualization technique and model type.

	Distance (centimeters)		Rotation (degrees)	
Visualization Technique	Mean	SD	Mean	SD
Fresnel-Derivative	0.42	0.42	3.49	3.06
Semitransparent	0.38	0.30	2.98	2.50
Silhouette	0.41	0.40	2.61	1.91
Wireframe	0.44	0.40	3.22	2.69
Model Type	Mean	SD	Mean	SD
Balloon	0.49	0.40	1.98	1.66
Camera	0.39	0.43	2.91	2.38
Mug	0.29	0.25	3.24	2.24
Skull	0.48	0.40	4.18	3.32

Results for mental effort are similar to the SUS scores. The Silhouette and Fresnel-Derivative visualizations techniques yielded the lowest mental effort, while Wireframe and Semitransparent the highest. Results from the Kruskal-Wallis tests revealed a significant difference between visualization techniques for mental effort $(\chi^2(3) = 9.88, p < .05)$. Moreover, the post hoc test revealed that the Silhouette technique's mental effort was significantly lower than for the Semitransparent visualization technique $(p < 0.05)$. Summarized results of these scores are shown in Figure 3.7 and Table 3.3.


SUS

Mental Effort

Fig. 3.7. Scores for SUS and Mental Effort reported by users. *ST-Semitransparent, SH-Silhouette, WF-Wireframe, FD-Fresnel-Derivative*. (The red line indicates the median, and the bottom and top edges of the blue box indicate the 25th and 75th percentiles).

Tab. 3.3. Mean and standard deviation scores for usability and mental effort reported by users after grouping data results by visualization technique.

	Usability			
	SUS		ME	
Condition	Mean	SD	Mean	SD
Fresnel-Derivative	67.34	22.58	4.69	2.22
Semitransparent	59.45	17.61	5.97	2.02
Silhouette	75.78	18.53	4.28	1.89
Wireframe	63.44	25.84	5.00	2.59

3.4 Discussion

Based on the results obtained by comparing the visualization techniques, the Silhouette technique yielded better mean values and smaller standard deviations in 4 out of 5 accuracy scores: position (NTC), orientation (NTC and TC), and time to completion (NTC). These results align with the original reasoning presented in this section, suggesting that the selection of visualization techniques used to represent the virtual objects can influence the alignment accuracy achieved by the users. Moreover, it suggests that visualization techniques that provide useful visual cues and reduce the occlusion ratio during overlapping may assist the users in achieving better alignment scores. Although no statistical significance was observed after Kruskal-Wallis tests, future studies considering larger populations may be useful to confirm this hypothesis. Interestingly, and against the initial thoughts presented in this section, the Semitransparent technique performed better for position errors under the TC condition. Moreover, based on the observations after implementing the study, it seems plausible that the lack of differences in the results obtained, especially under NTC conditions, may have resulted from the amount of time available for the participants to complete the alignment tasks. However, the behavior was similar to the one observed in the TC scenario. These results may indicate that participants can achieve a relatively accurate alignment within a few seconds, using the remaining time to fine-tune the alignment. Considering the study results under the NTC condition, **H1** is partially supported by the results obtained with the Silhouette technique in terms of position, orientation, and time. In contrast, **H2** is not supported considering the results obtained by the Semitransparent technique. The scores of usability and mental effort achieved by the Silhouette technique were significantly better when compared to the Semitransparent technique and therefore support **H3**.

Regarding the geometry of the objects, two main tendencies can be observed. First, the study results show evidence that the Balloon model obtained the lowest rotation errors. These results can be related to the length of the object extremities that provide evident visual cues when misalignment in rotation between the interactable and target objects exist. A similar concept could explain why objects with relatively high curved surfaces and without long extremities, such as the Mug and the Skull, present the highest orientation errors. In contrast, the Mug appears to be the object with better position errors, while the Balloon shows the worst. This effect may lead to suppose there is a relation in how the geometrical properties of the objects lead users to concentrate their attention in aligning one of these two parameters –either position or rotation– underestimating the other. However, further studies would need to be conducted to validate this hypothesis.

Regarding the Objects' Geometry

Another interesting effect observed when grouping the results by model is that the position and orientation errors for the TC and NTC conditions showed a similar trend. Although the errors for the TC condition are higher than for the NTC, these differences seem to be the result of a consistent offset. This trend may indicate that the additional time available under NTC conditions helps to reduce the magnitude of the error. Still, the object's shape seems to introduce a bias towards observing specific errors either for position or orientation.

Additional research in this direction may consider the work presented by Singh et al. [155], which suggests that brightness is a component that influences depth perception in MR environments. Results from this study show that brighter objects appear closer than dimmer objects. Although it is not possible to derive conclusions from the study in this regard, further studies could explore if these findings contribute to explaining the result obtained by the study participants when performing alignment using semitransparent replicas.

3.5 Study Implications

The study presented in this section represents a necessary first step towards understanding the benefits of designing visualization techniques that exploit the properties of objects of interest to provide useful visual information during interactive alignment in MR. This work aims to inspire new concepts to create alternative visualization techniques that go beyond the simple representation of the object's shape or geometry.

Some initial benefits can be deducted from the results of this study as they suggest that the selection of visualization techniques used to present virtual content influences the accuracy that the users can achieve. Furthermore, the subjective measures collected during the study showed that visualization techniques that minimize occlusion when the objects overlap reduce mental effort and increase usability. Although no statistical differences were obtained for the objective measures, the observed trends show that this visualization technique leads to smaller position and orientation errors and reduced completion time.

This study has motivated additional work that used these visualization techniques to improve the interactive alignment of real and virtual content in medical settings. These works include exploring the feasibility of using MR applications for surgical procedures such as deep inferior epigastric perforators flap for breast reconstruction [40], and supporting surgeons performing the retrosigmoid approach in craniotomy [98].

Complementary Textures 4

> *com·ple·men·ta·ry*
> *Combining in such a way as to enhance or emphasize the qualities of each other or another.*
>
> — **Oxford Languages Dictionary**

Accurately estimating the spatial position of virtual content in MR scenarios has proven to be a challenging task and is still an open research topic. This misestimation problem becomes critical for tasks that rely on the user's ability to manipulate and align virtual content. In such environments, the alignment accuracy strongly depends on the relevance of the visual information presented to the observer. As presented in Chapter 3, traditional visualization techniques used for interactive alignment tasks normally present virtual replicas of the objects of interest in the form of solid [85, 132, 164], semitransparent [18, 122, 135, 175, 181], wireframe [37, 165], or even point cloud [41, 61] representations. However, these techniques can lead to incorrect estimation of depth or shape; and affect user's performance due to the perceptual ambiguities associated with this technology [112]. This outcome is not surprising as humans are not accustomed to visualizing real and virtual objects simultaneously and are not used to aligning virtual content using replicas of them. Moreover, the insights derived from the study presented in Chapter 3 suggest that the choice of visualization technique influences the user's performance in such alignment tasks [107].

This section explores the concept of Complementary Textures as a novel method that exploits the textural surface patterns of real objects to generate virtual replicas that provide rich visual cues during the performance of interactive alignment in MR. Although the visualization techniques presented in Chapter 3 can assist the observers to infer alignment errors and improve task performance, they are not designed to highlight potential misalignment but rather to indicate the desired target pose. In contrast, the Complementary Textures aim to modify the virtual objects' appearance (e.g., its color and texture) to provide highly salient visual cues when misalignment occurs. The introduction of this approach opens a set of concepts that go beyond the simple generation of contours, wireframes, or semitransparent visualizations and facilitate interactive alignment tasks in MR. More importantly, it represents an early attempt towards using the real objects' textural properties to optimize MR perceptual alignment tasks.

To further explore these ideas, a formal definition and three variations of Complementary Textures are introduced in Section 4.1. The first variation uses the textural properties of the real objects to provide photometric virtual complements. A second variation considers the geometric properties of the real objects to generate a virtual replica that complements the object's shape. The third variation acknowledges the observer's familiarity with the object to align to provide semantic visual information supporting the alignment task. The effects observed after using these textures for alignment tasks are presented in Section 4.2. In addition, some inspiring and promising proof-of-concept implementations of the COMPLEMENTARY TEXTURES and their counterparts using traditional approaches are introduced in Section 4.3. Furthermore, a discussion regarding the automatic generation of these textures and potential benefits and challenges for their implementation is presented in Section 4.4.

Please note that this new concept aims at taking advantage of specific visual, geometric, or semantic information of the object of interest. Therefore, it is expected that different classes of objects may require different methods for the automatic generation of their textures. The examples presented in this section represent the first set of COMPLEMENTARY TEXTURES and are not generated automatically. Moreover, the handful of solutions presented here only scratch the surface of this concept and can open the door to a new way of providing visual information during interactive alignment tasks in MR. For this reason, application designers could later propose specific methods for the automatic generation of most relevant COMPLEMENTARY TEXTURES or even lead to the discovery of new variations of them.

4.1 Definition

The complementary textures are defined as novel visualization techniques that exploit preexisting textural surface patterns on real objects to generate virtual replicas that can be used to improve object alignment in MR environments. This technique provides complementary visual cues to the user during real-to-virtual or virtual-to-real object alignment. Moreover, it aims at adopting perceptual properties like those associated with Gestalt psychology [96, 136] or constructivist approaches. These theories suggest that the parts of a geometrical shape, or even the tones of a melody, interact so that they produce a perceived whole that is distinct from the sum of its parts [136]. Thus, these components represent qualities of an experience that are not inherent in its components. The phenomenon of *amodal perception* that derives from *Gestalt psychology* indicates that it is possible to perceive spatial structure even when physical stimulation is absent, as described by Lehar [96]. This principle has been explored by Breckon [15] in the context of 3D computer vision to explore its application for volume completion.

Regarding alignment tasks in MR, such a concept could be applied to the combination of textures observed when the real and virtual objects overlap during the task performance. Achieving this would provide useful information that goes beyond the information conveyed by observing each object. Such a piece of complementary information can be delivered to the user, for example, by the generation of PHOTOMETRIC complements that lead to the observation of a homogeneous surface pattern when real and virtual objects are aligned. Furthermore, alternative modalities can provide GEOMETRIC elements drawn over the surface of a virtual replica to facilitate the alignment task. In addition, SEMANTIC augmentations relevant to the objects to align can be perceived as familiar by the users during the alignment task and therefore be used to provide interactive visual guidance.

4.1.1 Photometric Complements

This variant of complementary textures involves using pairs of interrelated color appearances that form a homogeneous single-colored object when proper alignment exists (4.1a-4.1e). These color appearances correspond to a real object's preexisting textural surface pattern, and a virtual texture generated using the real object's layout of shapes and patterns but inverting its colors.

In this context, a photometric complement consists of any alternative image that, when combined or added with its original counterpart, cancels out (i.e., loses hue) or creates a novel uniformly colored patch. Consequently, the photometric complementary textures generated for three-dimensional objects could result in a single or multi-color form composition when the real and virtual objects are aligned as depicted in Figures (4.1a-4.1e) and (4.1f-4.1j), respectively. Due to their complementary properties, the colors of the real and virtual textural surfaces create the strongest contrast when placed next to each other. In addition, when the alignment error between the objects is small, this type of complementary texture provides visual cues similar to the results obtained when using edge detection algorithms. These visual cues are observable even when the alignment errors occur in the user's viewing direction (e.g., errors in depth).

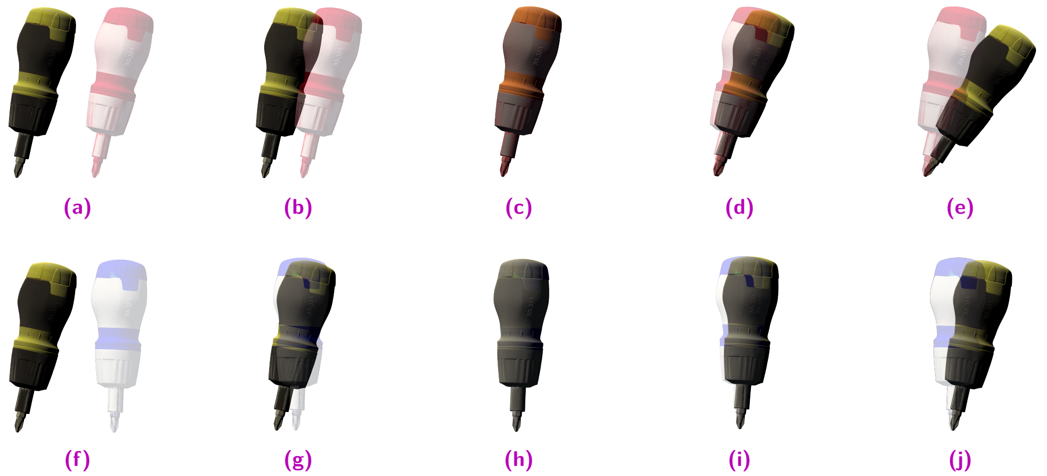


Fig. 4.1. COMPLEMENTARY TEXTURES for object alignment in Mixed Reality (MR). This concept utilizes the textural surface pattern of a real object to generate a virtual replica with a different but complementary texture to assist users during alignment tasks. The replica is designed to generate highly salient error visualization when the objects are not aligned in position (a),(b),(f),(g), or orientation (d),(e),(i),(j). In addition, it allows visualizing a new texture (c) or a homogeneous single-colored object (h) when proper alignment is achieved.

The display technology used to deliver the augmentations must also be considered when designing photometric complementary textures. For example, while using OST-HMDs provides an inherent additive aspect when the virtual and real textures overlap, the use of VST-HMDs would result in the virtual object fully occluding its real counterpart. In this regard, when using VST-HMDs, two approaches can be implemented to generate the photometric complements depending on the desired visual outcome. On the one hand, to observe a homogeneous pattern of the same color as the predominant real object's chromaticity, the complementary texture must inpaint those different pixels. In addition, the rest of the pixels should be rendered fully transparent. On the other hand, if the application requires a colorization of the patch generated after alignment, adjusting the transparency values, and blending the virtual texture with the background would produce a similar result to the one observed when using OST-HMDs.

4.1.2 Geometric Complements

The texture of the virtual replica can also be presented as a geometric complement, consisting of geometrical primitives including edges and diagonals, bisectors, arcs, inscribed and circumscribed circles, polygons, and more complex patterns such as fractals. These primitives can often be computed automatically from the geometry of the object of interest (see Figure 4.2).

The geometric complements provide multiple visual cues to improve the perception of alignment and help to reduce related errors. One of the advantages of this modality is that it provides visual feedback without leading to a visually occluded environment when the real and virtual objects overlap. Therefore, they can be used in scenarios where the alignment task requires the user to be aware of the real environment, as in several industrial and medical applications. This capability can be thought to be similar to the properties observed when using edge-based visualizations. However, it is important to distinguish that the geometric complements emphasize specific geometrical structures used during the alignment task instead of simply rendering the edges of the objects of interest. Thus, the cues provided by the geomet-

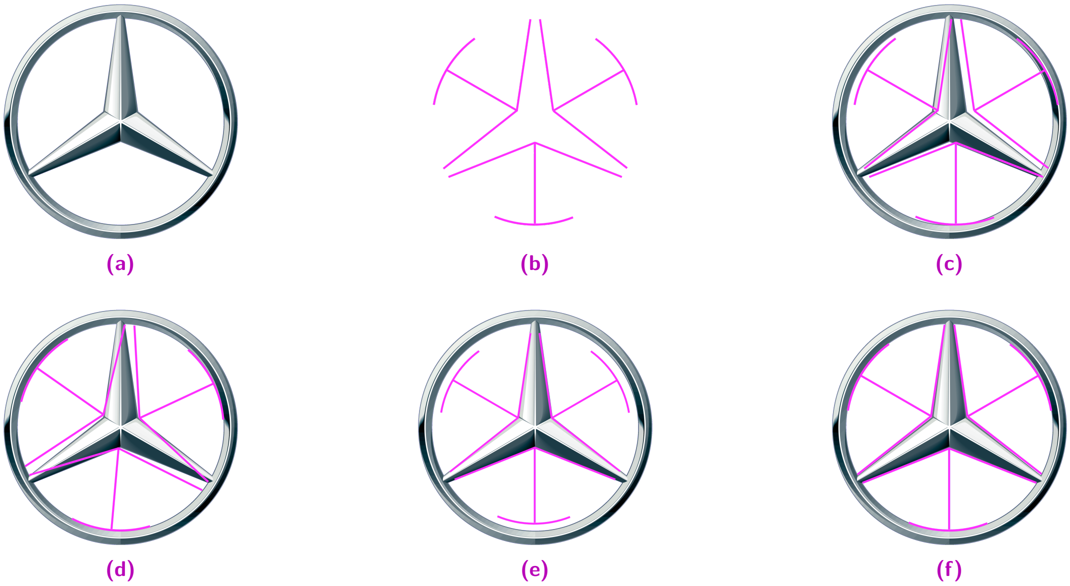
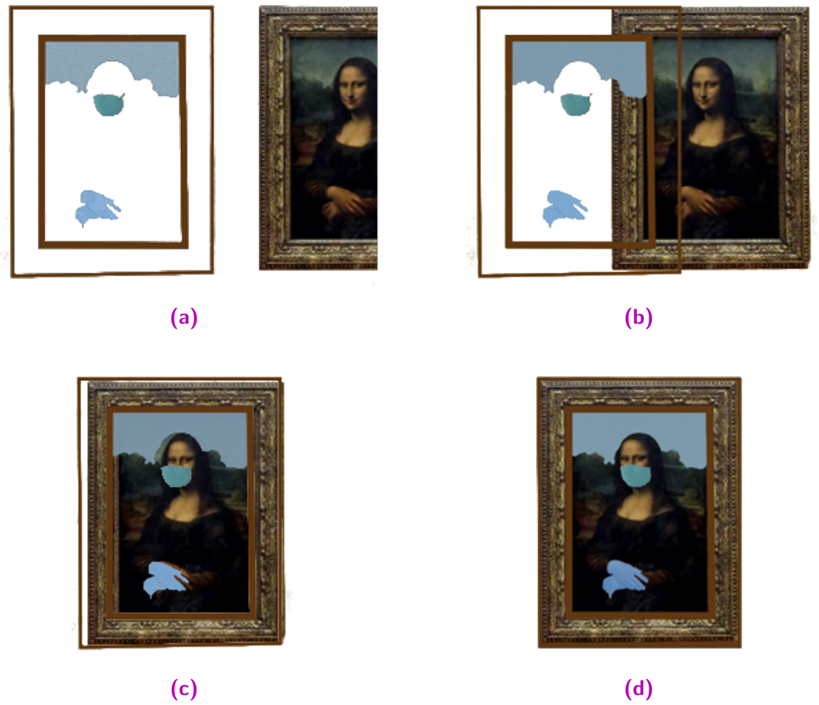


Fig. 4.2. **Geometric complementary textures** utilize geometrical primitives generated from the shape of a real object. These primitives are not limited to using edges, but diagonals, bisectors, inscribed circles, and more complex geometric patterns. In this example, the shape of a real object (a) is used to generate a geometric complementary texture applied to a virtual replica (b). These geometrical primitives allow identifying misalignment errors in position (c), orientation (d), and scale (e) (caused by misalignment in depth). When proper alignment is achieved, this type of texture leads to a seamless observation of the real and virtual objects as if they were part of one (f).

ric complements can assist the observer to infer errors during task performance and improve the alignment of the objects of interest. Note that, when using geometric complements, any misalignment between the real and virtual objects can be perceived globally and locally, even for textureless objects. Once proper alignment is achieved, the resulting overlap of the real and virtual content in the scene leads to the visualization of a harmonious scene in which the virtual and real content seem to belong to the same object. An exemplary implementation of this concept is depicted in Figure 4.2.

4.1.3 Semantic Complements

As an alternative to the previous modalities, the virtual replicas used during the alignment task can also be presented in the form of semantic complements (see Figure 4.3 and Figure 4.4). This type of texture extends from its geometric and photometric properties to include semantic content by presenting visual cues familiar to humans in the context of the objects to align. Thus, the virtual texture is designed to augment the corresponding real object so that the human observer perceives it as a natural addition of the object to align. Simple examples of this modality could include aligning a human face using virtual complements like sunglasses, hats, or masks. An exemplary image of this concept, illustrated in Figure 4.3, shows a semantic complement of the Mona Lisa (in the context of the COVID-19 pandemic) and a virtual complement consisting of a face mask and a pair of surgical gloves that are used to provide visual cues during an alignment task.



(a)

(b)

(c)

(d)

Fig. 4.3. The **semantic complementary textures** use semantic information derived from the object in the context of the alignment scenario. These semantic complements allow the identification of misalignment errors by visualizing augmented objects that are semantically familiar to the users. In this example, a face mask and a pair of gloves (a) are used to provide semantic information (Covid transmission protection) during the alignment of a frame to its targeted painting (b), (d).

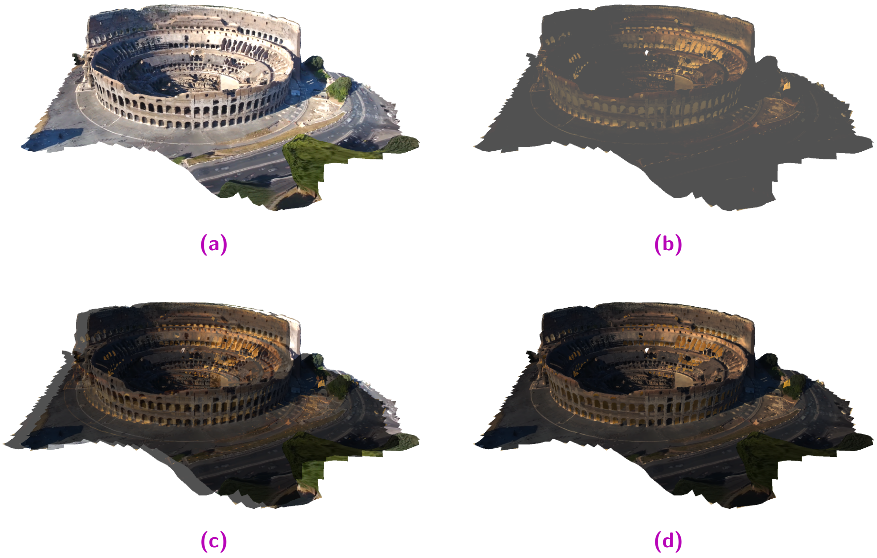


Fig. 4.4. The **semantic complementary textures** can be extended to three-dimensional objects. In this example, a model of the Colosseum acquired during the day (a) is used to generate a virtual replica (b) of the building at night. These textures provide contextual information that the users can understand during the alignment task (c)-(d). Similar to other modalities of the complementary textures, the virtual model allows identifying misalignment errors in the form of highly salient visual cues.

4.2 Effect

All the variants of complementary textures mentioned above share the property of presenting a single appearance that can be separated into two distinct components. Each component has a meaning of its own. Therefore, an observer can readily understand the two parts of the complementary textures when inspected individually. For example, a white triangle on a black background and a black triangle on a white background are complements, but even a single component forms a message for the observer. The same principle is true for an image of a landscape and the same image with its colors replaced by complementary colors. Consequently, both images evoke an understanding of the same landscape. Just as well, images of a face and a pair of sunglasses are both understandable individually. However, it is natural for a human observer to assume that the sunglasses must appear on the image's face.

Therefore, when these components are brought into proximity and misalignment occurs, the result becomes uncomfortable to look at. In this regard, additional edges that arise from the mixing process of the colors create visual noise when using photometric or geometric complements. At the same time, the semantic modality provides context that seems to be broken. This visual effect invokes a desire to move the components to remove the noise or the semantic absurdity so that the two parts form a new whole.

Such a need to bring together two related yet separated objects into alignment to provide a visually attractive image motivates the design and application of the complementary textures for object alignment in MR.

4.3 Proof-of-Concept

To explore the benefits of using complementary textures and to present a proof-of-concept, this section presents an exemplary implementation using scanned real objects to generate a set of complementary textures. These textures are then applied to virtual replicas of their real counterparts and used during alignment tasks. Visual salience maps computed during alignment tasks using complementary textures and traditional visualization techniques illustrate some potential benefits from using this concept in MR applications.

4.3.1 Texture Generation

As previously discussed in this dissertation, a common approach in MR applications that involve object alignment is to present a virtual replica of the object to be aligned, indicating the desired pose. These virtual models often represent existing 3D models of the object of interest, e.g., designed or scanned CAD models of industrial objects or segmented Computed Tomography or Magnetic Resonance Imaging of organs in medical applications. Moreover, depending on the complementary texture modality, the geometrical characteristics, the preexisting textural surface pattern of the object, or the semantic functionality or properties of the object can be used.

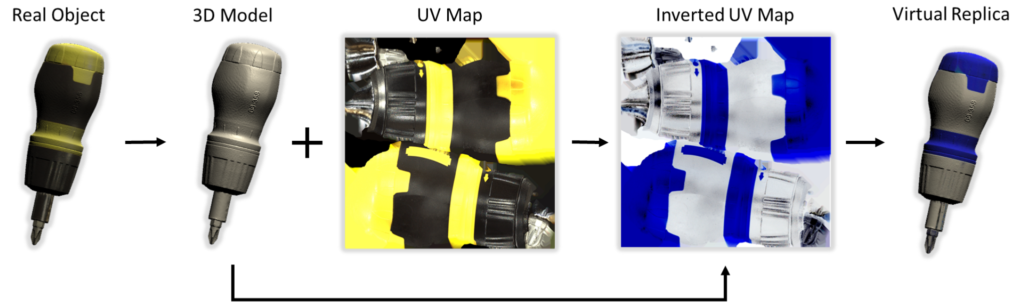


Fig. 4.5. A **photometric complementary texture** can be generated by inverting the pixel colors and brightness values of the textural surface pattern of a real object in the form of a UV map. This texture can be later applied to the virtual object that will be used during the alignment task.

To generate a homogeneous photometric complement, the corresponding pixel colors and brightness values of the preexisting surface pattern of the real object in the form of a UV map can be inverted and applied to the virtual object as shown in Figure (4.1h). On the other hand, when a non-homogeneous photometric complement is preferred, the corresponding alpha values or pixel colors of the original UV map in the region of interest can be modified and applied to the virtual replica, as depicted in Figure (4.1c). The overall procedure to produce a photometric complement is depicted in Figure 4.5. This specific texture, generated from inverting the colors of a real object's model acquired using 3D scanners[1], can be later be used to generate the specific modality of complementary texture. In addition, one may also need to take the properties of the AR display into account, in particular considering a color calibration of the display [80].

[1] The model used for Figure 4.1 and the related figures can be retrieved from https://www.artec3d.com/3d-models.

Alternatively, the geometric complements can be generated by selecting the geometrical structures that provide useful visual cues during the task performance, as depicted in Figure 4.2. Although currently generated manually, these visual cues could benefit from automatic algorithms involving mathematical models such as Delaunay triangulation [25], geometric interpolation, fitting curves, or even fractals. This type of automatically generated geometrical complements can be particularly interesting when dealing with the alignment of architectural elements, like the styles found in several Hindu and Gothic temples [140].

A similar approach can be used for the generation of semantic complements. Nevertheless, this procedure requires a more selective and less automatic design process. This variant of complementary textures must present understandable perceptual and semantic visual information to be useful during the alignment task. An example of this is presented in Figure 4.4, where a semantic complement designed to provide a night scene of the Colosseum is generated from a day scene. This section does not discuss the automatic generation of semantic complements, as they are highly dependent on particular scenes and their semantic interpretations. This modality of complementary texture could also represent a simple and attractive alternative for applications where very accurate alignment is not mandatory, as could be some architectural or entertainment applications.

4.3.2 Alignment Scenario

An exemplary alignment scenario implemented in a virtual environment is presented here to exemplify how the users could benefit from using the complementary textures. This scenario involves creating a virtual environment to represent the real and virtual objects using not only complementary textures but also traditional visualization techniques, the generation of the corresponding textures, and the computation of saliency maps using the representation of the virtual objects.

The model depicted in Figure 4.1, and its corresponding UV maps were used to represent the real objects for the generation and visualization of the objects of interest. In addition, the corresponding complementary texture was generated using the UV map of the textured object as described in Subsection 4.3.1. The alpha value of the virtual replica using the complementary texture was modified to exemplify the use case of MR applications that involve OST-HMDs. In addition, virtual replicas of the textured object were rendered using common visualization techniques frequently used for alignment tasks in MR [107]. These objects can be seen in Figure 4.6.

Visual salience maps of these objects during alignment were computed using the Quaternion [142] and Spectral Visual Saliency [147] algorithms to investigate the visual effects observed when using the complementary textures and the traditional visualization techniques. Furthermore, an Eigen-based Phase Quaternion Fourier Transform (Eigen-PQFT) approach was selected to compute the salience maps as this type of algorithm outperforms the low-level algorithms used to predict human gaze fixation [148].

The resulting salience maps obtained after applying the Eigen-PQFT algorithm to the virtual objects presented in Figure 4.6 can be seen in Figure 4.7. These salience maps show that using

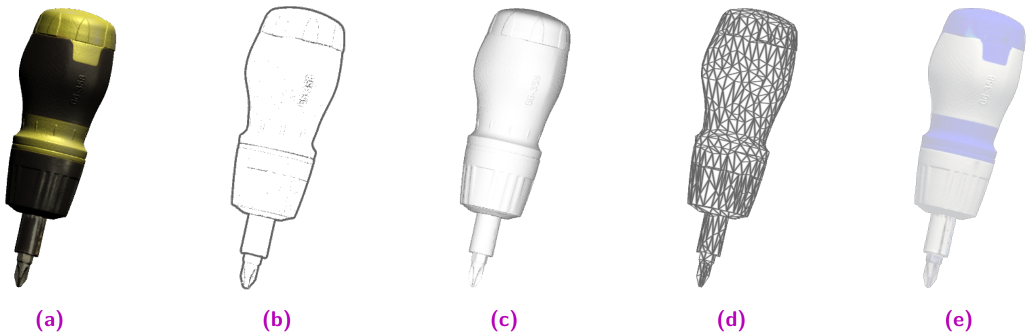


Fig. 4.6. Virtual objects used during the alignment scenario. The object with the real texture (a) is used as a reference to align virtual replicas using silhouette (b), fresnel (c), wireframe (d), and complementary textures (e).

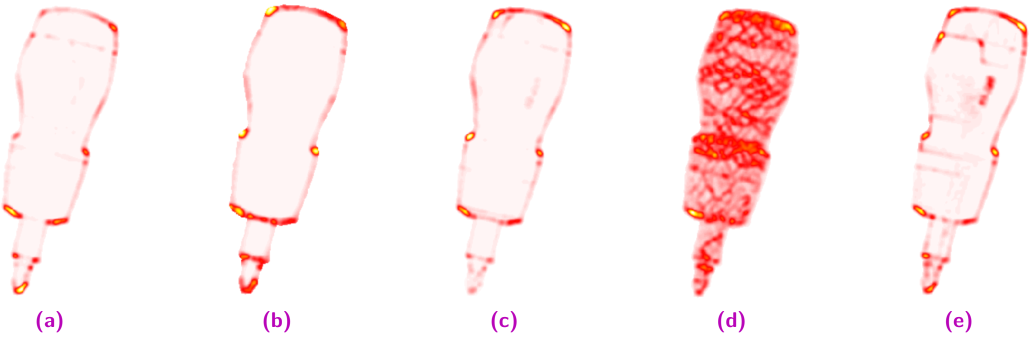


Fig. 4.7. Salience maps of the virtual objects used during the alignment scenario. The maps correspond to (a) the real, (b) silhouette, (c) fresnel, (d) wireframe, and (e) complementary texture.

complementary textures (4.7e) provides visual information from the outer and inner parts of the object of interest without leading to the observation of a cluttered environment. In contrast, existing approaches mostly provide visual information localized on the edges of the object to align (Figures 4.7b and 4.7c) or provide too many details that can end up saturating the visual field (4.7d).

A similar setup served to visualize the salience maps during the alignment process using a homogeneous photometric complementary texture. This texture provides visual areas that increase their salience when the objects are close to being aligned, even when the misalignment errors are small, and decrease once the objects are perfectly aligned. Compared to other visualization techniques such as the silhouette, these highly salient areas remain mostly constant during the alignment process (See Figure 4.8). Such a behavior is also depicted in Figure 4.9 and Figure 4.10, where the integral image and the maximum value of the salience maps computed during the observation of translation and orientation errors are presented. These values suggest that the photometric complement presents high integral values in translation and orientation that decrease when perfect alignment is achieved (Figures 4.9a and 4.9b). Moreover, when misalignment between the objects exists, the maximum salience values are observed using this type of visualization (Figures 4.10a and 4.10b).

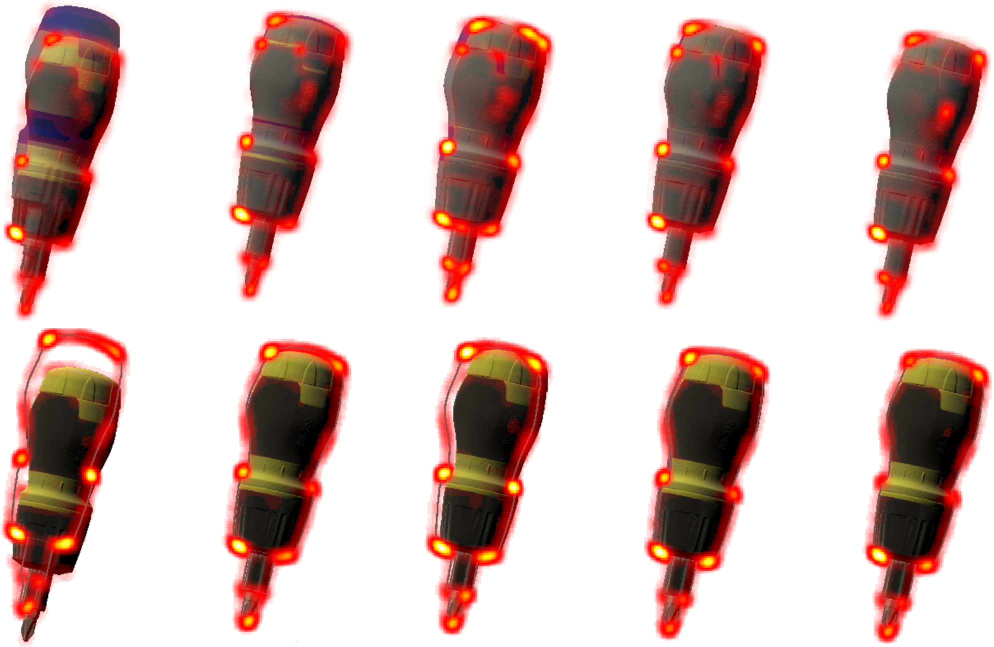


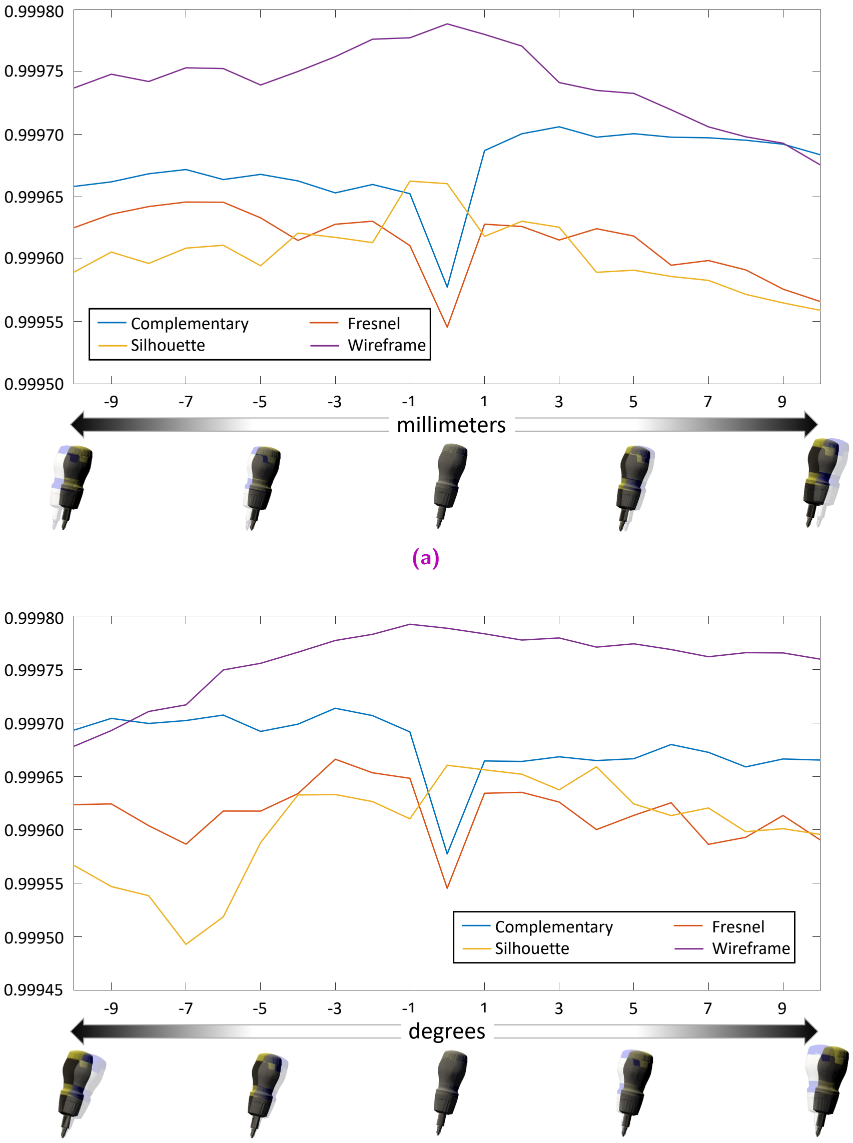
Fig. 4.8. Exemplary alignment of a screwdriver when translation errors are observed. The photometric complementary textures provide variable visual salience values that increase during the alignment process even when small misalignment errors are observed (top). In contrast, classical techniques such as silhouette visualizations provide relatively constant salience values during the alignment (bottom).

4.4 Challenges and Applications

The initial proof-of-concept of the complementary textures was designed and implemented using a non-automatic procedure. However, the nature of the photometric complements allows for generating the complements by simply inverting the pixel colors, adjusting the brightness values, and modifying the alpha value of the real object's texture. This procedure does not involve computationally complex algorithms.

Regarding the geometric complements, like those presented in this work, the geometric single-colored shapes are the easiest complements to generate. Exemplary automatic procedures for generating these textures could include the extraction of key features based on high curvature surfaces from which the corresponding complementary texture can be created. Alternatively, more complex mathematical models based on Delaunay triangulation, geometric interpolation, fitting curves, or fractals could lead to the generation of this variant of textures.

In contrast, creating automatic semantic complements presents a greater challenge as they should provide visual semantic information that the user can naturally understand during the alignment task. One can still imagine automatic methods for object classification that could contribute to the automatic generation of textures. For example, eyes that can be automatically detected using cameras may support the generation of matching glasses. Alternatively, automatic recognition systems could be used to detect words and generate complementary textures such that the alignment of two graphics would semantically make sense. One could even think



(a)

(b)

Fig. 4.9. Integral image of the salience maps computed during the observation of translation (a) and orientation (b) errors when using silhouette, fresnel, wireframe, and complementary textures.

that future deep learning approaches could support the generation of complementary textures based on the object geometry, texture, and even semantic properties.

The design of the complementary textures presented here assumes consistent lighting conditions under controlled environments. In this regard, it would be important to consider how the different lighting conditions observed in several real environments could influence how the complementary textures interact with their real counterparts. Still, it is expected that the adaptation of radiometric compensation techniques such as the one presented by Grundhofer and Bimber [60] could contribute to alleviating these effects. Using a geometric variation of complementary textures could represent an alternative for this type of environmental condition, especially when they would impede the use of radiometric compensation techniques.

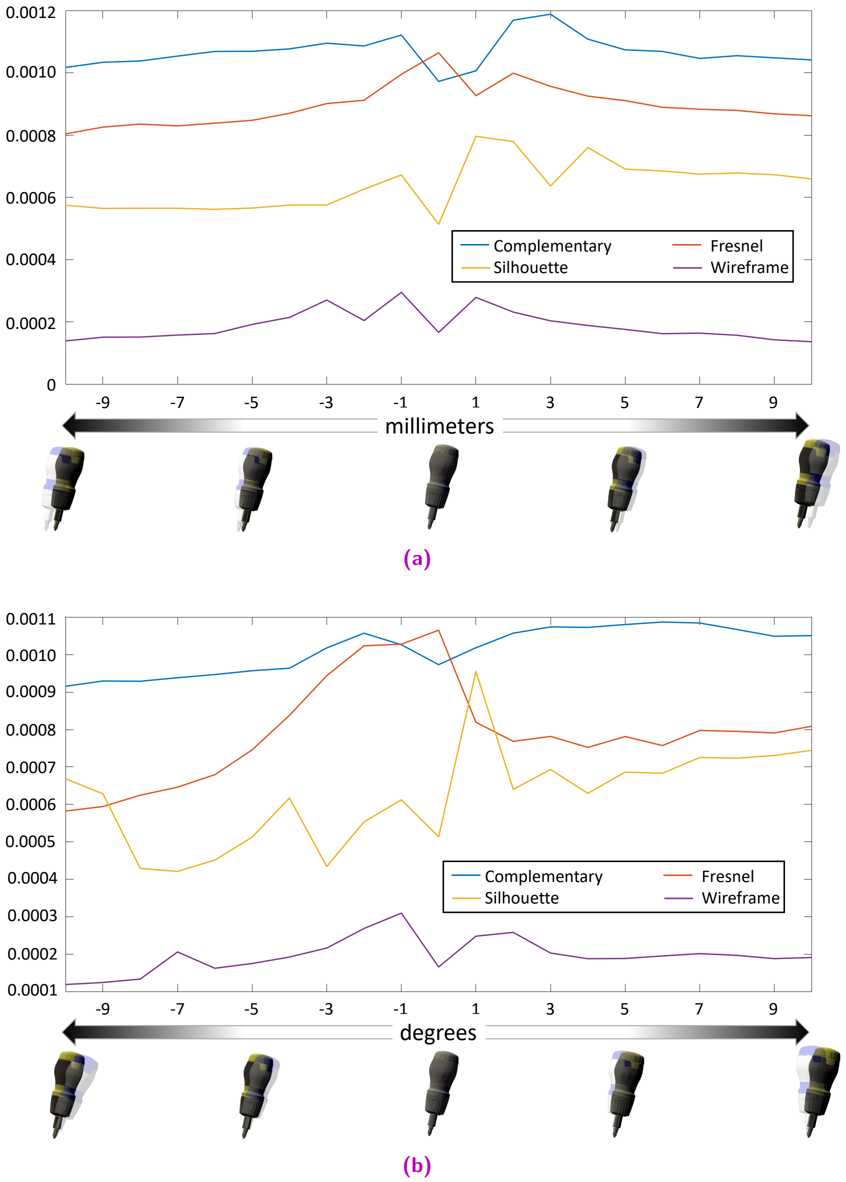


Fig. 4.10. Maximum value of the salience maps computed during the observation of translation (a) and orientation (b) errors when using silhouette, fresnel, wireframe, and complementary textures.

5 In-situ Visualization for Optical See-Through Head-Mounted Displays

> *If the hand be held between the discharge tube and the fluorescent screen, the darker shadow of the bones is seen within the slightly dark shadow-image of the hand itself. ... For brevity's sake I shall use the expression* ***rays*** *and to distinguish them from others of this name, I shall call them* ***X-rays****.*
>
> — **Wilhelm Röntgen**
> On the discovery of X-rays

Although MR applications have shown that it is possible to accurately place virtual content in the real world, another complex problem in these environments is estimating the position of virtual objects placed inside real opaque objects. This scenario raises new challenges as additional visual cues need to be considered to provide a realistic perception of the virtual content's location. One of these challenges involves the generation of an adequate sense of occlusion. As discussed in Section 2.1, this cue is particularly important as it represents the strongest depth cue and allows for generating an idea of the objects' relative depth order. Failing to provide consistent occlusion can create the illusion that the virtual content is floating on top of the real surface, even if the virtual object is presented in the correct position in the real world. This visual illusion has been reported in MR applications by Singh et al. [156]. According to the authors, when a virtual object appears in front of a real surface, and this surface moves towards the observer, the movement will provide the illusion that the virtual object is "pushed closer to the observer." This illusion will disappear at some point, and then the observer will perceive that the virtual object "falls back" behind the real surface. Despite Singh and collaborators observed this effect when using high salient real surfaces, Ellis and Menges reported similar observations [34].

This problem has been particularly interesting and widely explored in MR, for example, to observe anatomical structures within the patient's body in medical applications or to visualize, explore, and understand complex objects in industrial settings. Existing works in these domains have proposed multiple techniques to generate occlusion using virtual content displayed on the real object's surface. These techniques aid the observer in estimating the correct depth of the real and virtual objects in the scene as they provide a coherent sense of occlusion. Although previous work has used the term "X-ray vision" to describe these techniques, they will be referred to as "in-situ" visualization in this dissertation.

In the following, this section provides a general overview of existing techniques proposed for in-situ visualization, the design implications for implementing these techniques, and the challenges of using different display technologies such as OST- and VST-HMDs. In addition, it presents a study that compares the accuracy that users achieve when estimating the depth of virtual objects using traditional techniques for in-situ visualization displayed through optical and video see-through head-mounted displays. Moreover, a second study evaluates how users can estimate the depth of the objects using visualization techniques originally designed for VST-HMDs implemented using OST-HMDs.

5.1 Anatomical Illustrations and In-situ Visualization

From an anatomical standpoint, philosophers and physicians have dreamed of visualizing anatomical structures located inside the human body for centuries. In this context, classical Greek philosophers were intrigued to understand the anatomy of animals and the human body. Early studies from philosophers like Aristotle (384-332 BC) presented evidence of systematic dissections of animals. His studies presented detailed illustrations and descriptions of hundreds of mammals, birds, fish, reptiles, amphibians, cephalopods, insects, and other invertebrates. These studies aimed to investigate and compare the form, structure, physiology, and behavior to learn how and why these animals lived as they did [11]. In addition, Aristotle reported notes on human anatomy founded on his observations from the animal dissections, the reason why it is considered the father of comparative anatomy. Shortly after Aristotle's death, close to the end of the fourth century BC, two other philosophers appeared in the scene. Herophilus (335-280 BC), who may be called the founder of systematic anatomy, and Erasistratus (304-250 BC), the first scientific physiologist [28], performed routinary dissections of human cadavers and animals in their studies of the body. It is believed that thanks to these philosophers, the knowledge of human anatomy received major importance and attention. This search of knowledge and understanding of the human anatomy continued in ancient Greek with Galen (129-200 AC), who possessed the works from Herophilus and Erasistratus. This philosopher collected the ideas from his predecessors, combined them to create anatomical reports, and developed multiple theories. Even though some of them were incorrect in anatomy matters, such theories dominated Western medical theory and remained unchanged for centuries [14]. However, it was not until the 16th century during the Renaissance that Andreas Vesalius (1514-1564 AC) declared Galen's research valid only to animals and not humans. Vesalius –considered the founder of modern human anatomy– was a great anatomist and a wonderful illustrator. He created highly detailed illustrations of the human anatomy that allowed the observer to clearly understand and visualize the structure and function of the external and internal human anatomy. Similar illustrations can be found, for example, in Leonardo da Vinci's anatomical drawings.

This desire to visualize anatomical structures within the human body has motivated various mixed and virtual reality applications. An example of this is the early work from Edwards et al. [31, 32] that allowed the registration and visualization of pre-operative data accurately overlaid on top of a patient using a surgical microscope. This system enabled the visualization of critical structures (e.g., blood vessels and nerve fibers) when removing tumors in brain surgeries using a surgical microscope. In addition, it allowed delivering virtual content presented in the correct spatial position, enabling the visualization of anatomical data in-situ and eliminating the need for the surgeon to look away from the patient to focus on a screen. Later work presented by Grimson et al. [59] enabled the accurate registration and visualization of medical imaging (e.g., computed tomography or magnetic resonance imaging) overlaid on top of the patient in the context of neurosurgical procedures. This approach supported the interactive and non-intrusive visualization of intra- and extracranial anatomy in-situ. Additional work from Blackwell et al. [10] used image overlays of medical imaging positioned accurately within the patient's anatomy. This system required tracking the observer's head, the objects of interest, including tools and physical anatomical models, and the display's components to provide adequate visualization of the virtual content using semitransparent displays placed between the observer and the objects of interest. Although these systems enabled accurate registration of anatomical content and in-situ visualization of medical imaging within a patient's body, a common problem to all these approaches was observing inconsistent depth cues. These inconsistencies resulted from misleading information as the virtual content, placed at a greater depth, was never occluded by the real objects. This effect led to the assumption that the virtual content was placed on top of the real objects, hindering a more intuitive understanding of the patient's anatomy.

The capacity to convey visual information that facilitates an intuitive understanding of the internal anatomy of the human body depicted in Vesalius's illustrations becomes particularly relevant for mixed reality applications. These illustrations depict consistent visual cues that contribute to the understanding of the scene. In this regard, one of the first attempts to provide consistent visual cues during the observation of anatomical data in-situ was proposed by Bajura and collaborators [5]. This system used a video see-through head-mounted display to visualize ultrasound imagery in real-time. It allowed for the generation of a synthetic hole that gave the impression of a physical cut on the patent's skin, like those observed in Vesalius illustrations. The synthetic hole allowed for the visualization of ultrasound images acquired with a tracked probe according to the observer's viewpoint. The combination of the imagery and synthetic hole provided key visual cues such as the occlusion of the patient skin and the generation of shadows that enhanced the sense of depth inside the hole. This concept was later extended by Ohbuchi et al. [123] to visualize three-dimensional ultrasound imagery. Alternative approaches have used line contours projected over the surface of the real object to generate the impression of consistent occlusion combined with three-dimensional medical imagery [59] or using orthogonal bi-dimensional projections [170]. More recent work has used rendering techniques such as alpha blending to inpaint the real object's surface and generate a sensation of transparency, enabling the observer to focus on objects of interest located within the real objects [8, 90, 97]. These approaches have the characteristic to preserve the contextual information of the scene without creating visual clutter.

In addition to the medical field, similar visualization methods have been developed to visualize hidden content in office and industrial scenarios. An early example of these is the system presented by Furmanski et al. [44] that explored the benefits of integrating depth cues into rendering techniques for visualizing occluded content. Kalkofen et al. proposed the use of *Focus and Context* visualization techniques to understand complex objects and equipment by enabling the observation of their internal components [84]. Similar approaches were used by Schall et al. to assist workers with the visualization of structures placed underground [146] using hand-held devices. Alternative techniques have taken advantage of salient features from the occluding objects to provide contextual information while visualizing the hidden content.

These features include edges [3] or more complex components, including hue, luminosity, and motion [141].

On "X-Ray Vision"

Further information regarding in-situ visualization in MR can be found in the paper: **Pursuit of “X-ray vision” for Augmented Reality [101]**.

5.2 Depth Estimations for In-Situ Visualization with OST- and VST-HMDs

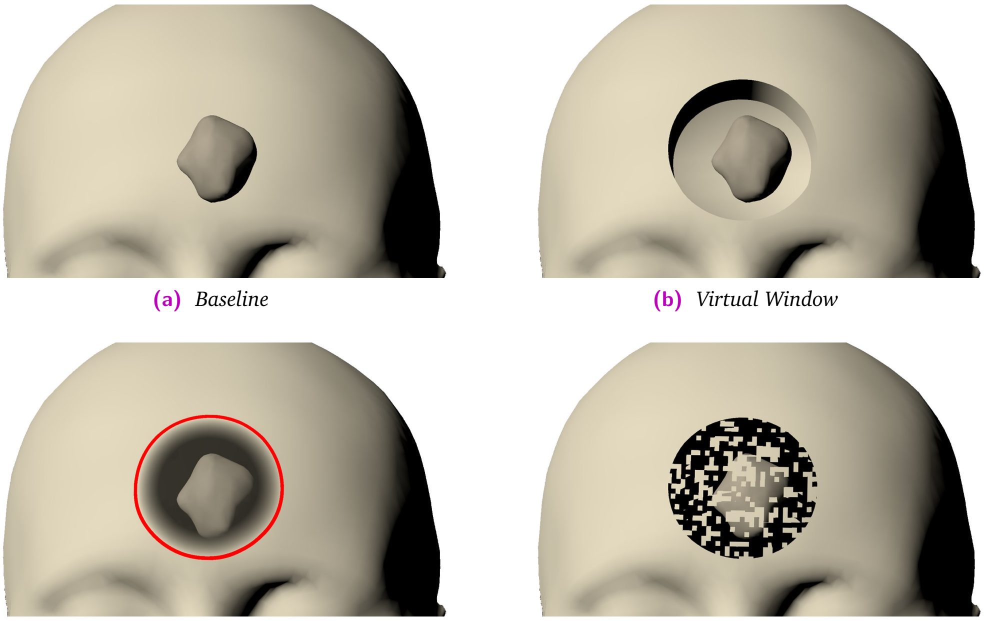
Although several approaches to provide focus and context visualization exist [5, 8, 43, 84, 97, 124, 141, 146, 153, 173], these approaches normally require hand-held devices or VST-HMDs as they inpaint the real object’s surface to provide a realistic sense of occlusion.

These techniques frequently rely on dark colors and shadows to allow the illusion of observing holes in the real object’s surface. While VST-HMDs can generate this range of colors, the additive nature of the display technology used in OST-HMDs impedes it. In this regard, it is possible to produce bright colors by delivering light wavelengths to the observer’s eyes using additive displays. Nevertheless, as black represents the total absence or complete absorption of light, additive displays will not generate any light wavelengths. Therefore, allowing for the visualization of the real object’s surface. For this reason, rendering dark colors using OST-HMDs will be perceived as changes in the opacity of the virtual objects.

Although existing works have shown the advantages of using focus and context visualization techniques to improve the perception of virtual content in-situ, there seems to be no evidence of any work comparing if these technological differences may affect the observer’s perception. Therefore, this section compares whether the observer’s perception is affected when presenting focus and context visualization techniques using video and optical see-through head-mounted displays.

5.2.1 Selection of Visualization Techniques

Within the multiple focus and context visualization techniques proposed over the past years, this study compares three approaches that have shown promising results when implemented on VST-HMDs and are suitable for adaptation in OST-HMDs. These techniques, in this dissertation, referred to as *virtual window* [5], *contextual anatomic mimesis* [8], and *virtual mask* [124] are depicted in Figure 5.1. These approaches, designed to provide useful visual cues in personal space, mainly take advantage of occlusion and motion parallax. In addition to rendering the virtual objects located inside the real object, they generate virtual content that emulates the effect of partially or fully transparent areas over the surface of the real object, helping to create a sense of proper depth ordering. This added virtual information does not conflict with the real surface depth as it merely "sits" on top. Such effect allows perceiving the virtual object



(a) *Baseline*

(b) *Virtual Window*

(c) *Contextual Anatomical Mimesis*

(d) *Virtual Mask*

Fig. 5.1. Base implementations of focus and context visualization techniques.

of interest as occluded by the real object's surface with correct depth ordering relative to the virtual transparent area while being visible to the observer. In addition, using motion parallax strengthens complements this illusion.

Virtual Window. The virtual window, first introduced by Bajura, Fuchs, and Ohbuchi [5], is the straightforward implementation of a virtual cutaway to visualize virtual anatomical content inside a patient. The original implementation of this virtual window, originally denominated "synthetic hole," was later adapted by Fuchs et al. [43] to partially cut into the tissue instead of giving the impression of a hollow body. The adaptation employed in this study shows cutaway edges of the virtual window to reveal the inner part of an otherwise hollow head, including a gray internal surface and diffuse shading.

Contextual Anatomical Mimesis. Originally introduced by Bichlmeier et al. [8], this visualization technique integrates additional subtle visual cues to the virtual window to retain some of the real object's surface features. This technique considers three factors to control the opacity of the real object's surface: i) a radial gradient increases the opacity towards the edges of the circular window, ii) areas with high curvature present higher opacity values, and iii) a modulation function that modifies the opacity of the object's surface as a function of the dot product between the view direction and surface's normal direction. In addition, a virtual circle indicates the limits of the augmentation. Combining these components creates the illusion of a semitransparent skin surface while the regions with high curvature or orthogonal to the view angle are preserved, providing contextual information. Similar techniques have been proposed for the visualization of hidden content in urban environments [97], or the visualization of contextual layers in industrial settings [84], and to observe anatomical content in minimally invasive surgery [173, 183]. Nevertheless, their image-based nature does not lend itself to straightforward application for OST.

Virtual Mask. Alternative approaches have been proposed to retain more contextual information by using less sparse windows. A clear example of this is the technique proposed by Otsuki et al. [124]. A virtual surface composed of a pattern of random dots controls the visibility of transparent patches of the window. This virtual mask occasions that parts of the real surface remain fully visible, occluding the virtual object, creating strong occlusion cues, and reinforcing the intended depth ordering when the observer uses motion parallax. Moreover, looking through these holes in the surface produces the effect described as *stereoscopic pseudo-transparency*. The technique implemented in this study is designed according to the original concept [125] using a dot size of $1/64$ and dot density of 50% as evaluated in [124].

5.2.2 Participants

Thirty-two paid participants recruited using mailing lists and campus announcements (12 female and 20 male), aged between 20 and 38 years old (mean age of 26.1 ± 4.6 years), took part in the study. To ensure correct or corrected-to-normal vision capabilities, a Landolt C-Test (EN ISO 8596) served to test for the participants' acuity; an Ishihara Color test was used for color blindness deficiencies [79]; and a Titmus test for stereo vision. All participants had normal or corrected-to-normal vision confirmed by these tests. Exclusion criteria included color vision deficiency, impaired stereopsis ($> 140°$ angle of stereopsis at 40cm), or a visual acuity below 63% (20/32). The participants' interpupillary distance (IPD), estimated using the OST HMD, ranged between 63 and $67.5mm$ (mean IPD of $64.74 \pm 1.48mm$). None of the participants had previous experience using the system. The study participation was voluntary, and participants could abort it at any time.

5.2.3 Experimental Setup

The main goal of this study was to investigate the impact of the focus and context visualization techniques on the perceived depth of a virtual object using different display technologies. A 2 *Display Type* $\otimes$ 4 *Visualization* $\otimes$ 3 *Target Position* within-subjects experiment with a perceptual matching protocol using a method of adjustments [35, 137] was conducted. Participants had to align a virtual object presented using four different visualization techniques with a physical reference object placed in three different positions within a 3D-printed real object. The visualization techniques correspond to those presented in Subsection 5.2.1 and a baseline condition (see Figure 5.1).

Apparatus

The physical setup used for this study (Figure 5.2) is inspired by neurosurgical guidance scenarios for tumor resection where frontal access to the brain is sometimes required. The apparatus consisted of two identical 3D-printed replicas of a human head with a hollow interior, a $10 \times 10cm$ optical marker, and an HTC Vive Tracker. The interior of the left head had a circular opening with a diameter of $4cm$ that allowed the visualization of the physical

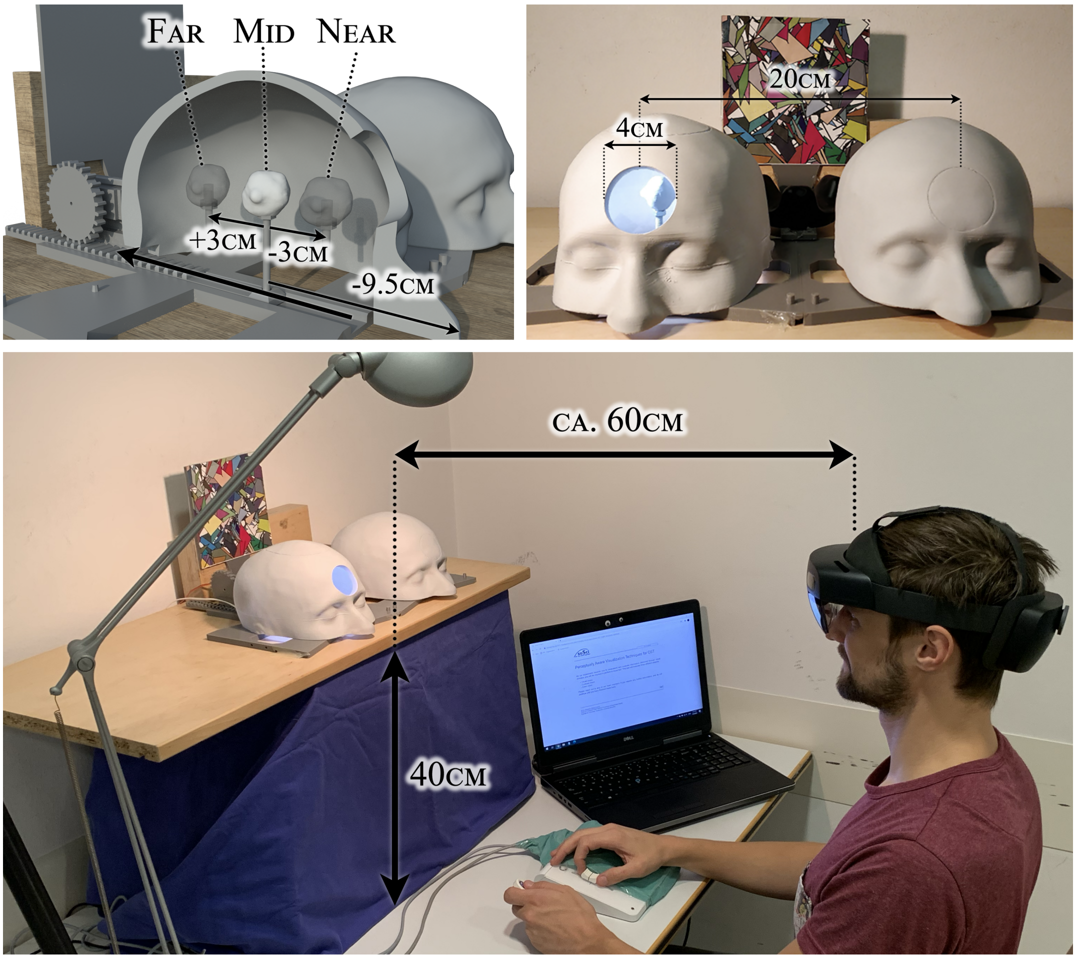


Fig. 5.2. **The experimental setup for Study 1.** Top: Two 3D printed phantoms of a head were placed in alignment with each other. The left head was used to present the physical target object. The right head was used to present the augmentations. Bottom: A participant executing the study by moving the augmentation on the right head to the perceived position of the physical object in the left head.

tumor and was illuminated using white LEDs. A servo motor coupled to a translational linear stage moved a 3D-printed physical tumor inside the left head into three defined positions. These positions were $3cm$ apart with a measured repositioning error smaller than $1mm$. The right head was completely closed to provide a low salient physical monochromatic occluder [156] and used to visualize a virtual replica of the physical tumor. All components were fixed to a wooden panel to avoid physical movement. In addition, a desk lamp was used to provide constant exterior lighting conditions (≈ 175 lx) throughout the study. The interior and exterior lights were aligned to match the light direction of the augmentations. A linear slider was used as an input device to position the virtual objects and a separate button to confirm alignment.

Participants were sitting on a chair at a fixed position, approximately $70cm$ away from the central tumor position, and allowed to move their heads from left-to-right and *front-to-back*, but not above the table. This constrained the participant's eye position to a range of approx. 55-75cm from the *Mid* tumor position. The spatial arrangement of the setup did not allow participants to visualize both tumors at all the alignment positions. This design consideration was intended to motivate participants to move their heads and provoke motion parallax.

The HMDs involved in the study were an HTC Vive Pro with a Stereolabs ZEDStereo camera for the VST condition and a Microsoft HoloLens 2 for the OST condition. The virtual environment was designed using Unity 3D. In addition, an optical marker using Vuforia tracking was used to calibrate the OST, whereas a SteamVR 2.0 optical tracking target was used in the VST condition.

System Error Assessment. To ensure there was no systematic bias and the accuracy of the setup was sufficient, a preliminary study involving $N = 10$ participants (3 female and 7 male, mean age of 26.9 ± 4.3 years) served to assess the potential error. This assessment was only used to assess the general error of the setup. It differs from the main study in that participants were asked to align the virtual tumor using the physical target in the left head as a reference. Thus, the virtual tumor was shown directly on top of the physical hole and rendered using the baseline technique. After a training stage involving six trials for each display modality, six repetitions per tumor position were collected for each HMD. This assessment revealed an average positioning error of $+0.10 \pm 2.61cm$ for the VST system and $+0.14 \pm 2.76cm$ for the OST system. These results indicate that participants placed the virtual object slightly farther away on average. There was no statistically significant difference between display technologies as determined by two-sample t-tests for alignment error ($t = 0.15$, $p = .879$), showing no indications for a systematic bias due to the tracking or display technology.

Just Noticeable Difference. In addition, the just noticeable difference (JND) [38, 51, 52] was calculated to ensure that the three different positions selected for the placement of the target were distinguishable by the participants during the alignment task. Data from the main study was used to calculate these values for each of the target positions and devices. The farthest position revealed the highest JND with $1.18cm$ (OST) and $1.22cm$ (VST), while the lowest JND was equal to $0.90cm$ (OST) at the nearest position, indicating that the $3cm$ spacing between positions was adequate.

Dependent Measures

Error and Time to Completion. The signed and absolute errors were assessed using the position of the virtual tumor relative to the physical reference as recorded by the Unity application when the users marked a trial as completed. The (signed) error was computed as the real minus the virtual tumor positions, using the tracked marker as the reference point. A very important note here is that due to the nature of the experiment (i.e., the participant placing the virtual object instead of indicating its position), this error has the opposite sign to the errors reported in the literature. Nevertheless, the meaning of over- and underestimation of depth will be presented consistently with these past studies. This means:

- An error $= 0$ is indicative of an accurate judgment.
- An error < 0 indicates an overestimation of the virtual tumor's depth (the virtual tumor was placed closer to the observer and thus perceived further away).
- An error > 0 means an underestimation in the perceived depth.

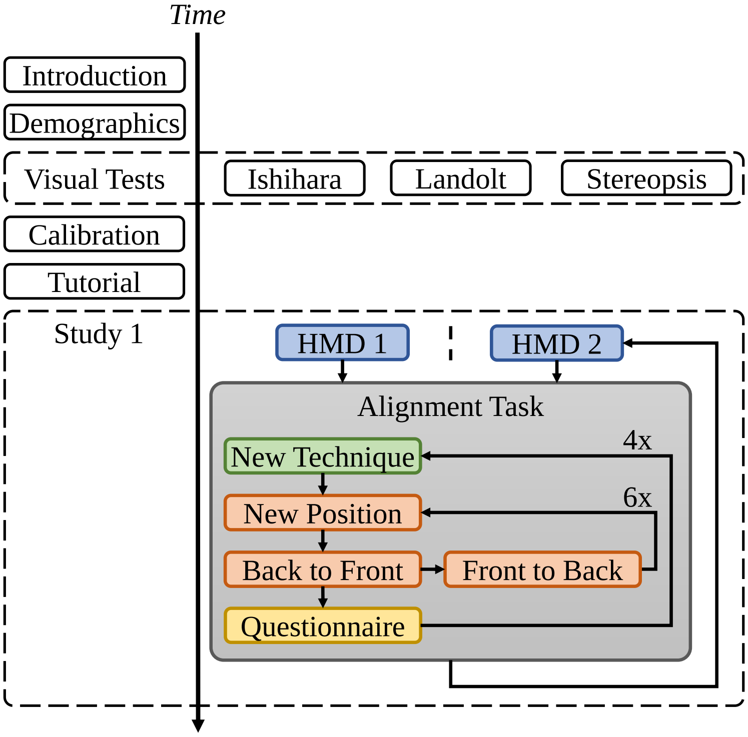


Fig. 5.3. The experimental procedure. Participants were randomly assigned to the order of conditions using a Latin-Square Matrix.

In addition to the alignment error, the time to completion elapsed from the presentation of the virtual tumor until trial completion was collected.

Subjective Measures. To derive assumptions related to the usability and task load, a NASA task load index questionnaire (TLX) [64] using a 21-point rating scale was used. It is important to mention that the individual subscales of this questionnaire were analyzed (Raw-TLX) [63] to avoid the introduction of additional sources of measurement errors [20]. Additionally, the single ease question (SEQ) [176] consisting of a 7-point Likert scale ranging from 1-(*Very easy*) to 7-(*Very difficult*) was assessed after every condition. Finally, participants completed a virtual reality sickness questionnaire (VRSQ) [86] composed of nine aspects, every one of them using a 4-Point rating scale: *None, Slight, Moderate*, and *Severe* to monitor for any negative effects resulting from using the HMDs.

Procedure

The study consisted of two main phases: i) an introduction and tutorial phase and ii) the evaluation phase. An overall picture of the experimental procedure is depicted in Figure 5.3.

Introduction and Tutorial. On arrival, participants were welcomed and informed about the study before presenting them with a consent form and performing the vision tests. After completing the tests, participants could familiarize themselves with the task and input device via a tutorial session. The tutorial involved the alignment of two virtual spheres presented using a desktop screen and the input device.

Evaluation. During this phase, participants completed one block of trials for every *Visualization* $\otimes$ *Display Type* combination. Each block consisted of twelve trials, and the target object was repositioned to one of three positions *near, mid, far* every trial (see Figure 5.2). The repositioning of the physical target was rapid (< 300ms to move from *Near* to *Far*). Hence, participants were not given special instructions during the repositioning phase. For each position, four alignment trials were performed. Half of these trials required the participants to move the virtual tumor from *back-to-front* until they were satisfied. For the other half, they moved it from *front-to-back*. The starting position of the virtual tumor was chosen randomly from a range of -5.5 ± 0.5 cm for *front-to-back* positioning and $7.5cm \pm 0.5$ cm for *back-to-front* positioning to prevent the participants from remembering the previous position of the slider. These distances were relative to the mid-position of the physical tumor. In addition, before the participants could begin a trial and move the virtual tumor, they had to reset the slider position to the end of its range (front or back depending on the trial mode).

A Latin squared matrix assigned the order of appearance of the visualization techniques and target object positions. Half of the participants started with the OST HMD before switching to the second HMD. The other half started with the VST HMD. Every time the participants were given a headset, they were assisted in wearing it comfortably and asked to perform a calibration routine to adjust their interpupillary distance (IPD). The participants were instructed to position the virtual tumor as precisely as possible, favoring precision over time. To be consistent with the *method of adjustments*, participants moved the tumor alternating between the two different directions (i.e., *front-to-back* or *back-to-front*) using the input device. In addition, participants were allowed to move the tumor only in one direction per trial to avoid very prolonged sessions observed during preliminary experiments. After completing every block, the participants were asked to fill out the subjective questionnaires. In addition, participants completed the VRSQ questionnaire after completing all blocks for every HMD.

A total of 96 positioning trials were performed for each participant, resulting from the possible combinations of display technology (2), visualization technique (4), tumor positions (3), approach directions (2), and approach repetitions (2). The total duration of the experiment was between 37 and 89 minutes.

5.2.4 Results

Three factorial (*Display Type* $\otimes$ *Visualization* $\otimes$ *Position*) repeated measures analysis of variance (ANOVA) was used to test for the objective assessments, using the mean of the four trials (2x *back-to-front* approach, 2x *front-to-back* approach) of each cell in the experiment.

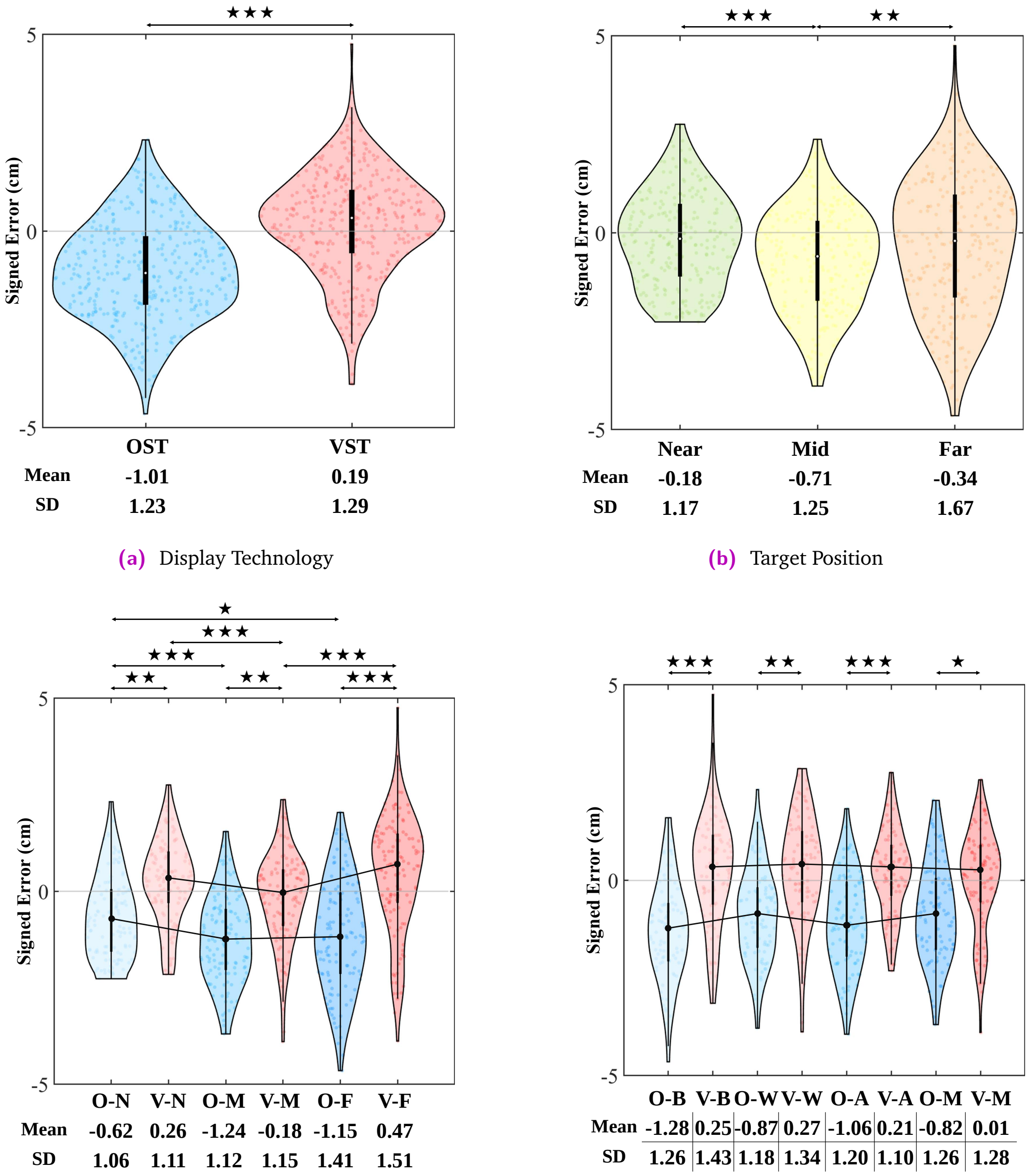
For the subjective measures, two factorial (*Display Type* $\otimes$ *Visualization*) ANOVAs were calculated. Significance was accepted at $\alpha = 0.05$. Greenhouse-Geisser corrected values were reported when the sphericity assumption was violated, assessed by Mauchly tests. Moreover, Bonferroni adjusted post hoc tests were used for all pairwise comparisons. Although all the interaction terms were considered in this study, only the significant ones are reported. Values greater than ± 3 standard deviations from the mean within each experimental condition (i.e., cell, factorial combination) were removed under the assumption that greater errors were not related to a misjudgment of depth but due to intermittent connectivity, tracking failure, or participant operating errors. A total of 21 samples for alignment (0.68 % of the sample points) and 64 samples for time to completion (2.08%) were removed and replaced with the new mean of the individual condition.

The figures presented in the following sections use $\star (p < 0.05)$, $\star\star\ (p < 0.01)$, and $\star\star\star (p < 0.001)$ to indicate levels of statistical significance. Plots presented are classic violin plots [69], indicating median with a white middle dot and the 25th and 75th percentiles with black bars. Reported descriptive values denote $mean \pm SD$ unless otherwise stated.

Alignment Error

Error. Significant interactions for *Display Type* $\otimes$ *Visualization*; $F(3, 93) = 3.59$, $p = .017$, $\eta_p^2 = .104$ were found (see Figure 5.4). Post hoc comparisons showed significant differences between the *Display Type* for all *Visualizations* ($ps \leq 0.023$), whereas the comparisons for the *Visualizations* when using the same *Display Type* were non-significant. However, the *Virtual Mask* showed the lowest alignment errors compared to the other visualizations using both OST and VST. Interestingly, the alignment errors reported for the *Virtual Window* showed the worst results when using the VST HMD and the second-best when using the OST, presenting values almost comparable to those obtained with the *Virtual Mask* (5.4d). In terms of *Display Type*, all OST conditions revealed that the virtual tumor was placed closer to the observer than the target position ($M = -1.01\ cm$). In contrast, all VST conditions showed the tumor was placed slightly further than the target position ($M = 0.19\ cm$). This effect was supported by a main effect for *Display Type*; $F(1, 31) = 16.88$, $p < .001$, $\eta_p^2 = .352$ (5.4a). A significant interaction for *Display Type* $\otimes$ *Target Position* was observed; $F(1.4, 41.8) = 20.54$, $p < .001$, $\eta_p^2 = .399$ (5.4c). Pairwise comparisons revealed that the *Mid* distance target position showed the strongest error in alignment (towards the participant) for both display types, corroborated by a main effect for *Target Position*; $F(2, 62) = 8.75$, $p \leq .001$, $\eta_p^2 = .220$ (5.4b). Comparisons revealed that the *Mid* target position showed a significant stronger error towards the observer (all $p \leq .004$) than *Near* and *Far*, whereas the *Near* and *Far* did not significantly differ. However, this error was mainly due to the OST, and for the OST, also strongly present for the *Far* position.

Absolute Error. The descriptives and pairwise comparisons for the absolute error are depicted in Figure 5.5. Results revealed a statistically significant interaction for *Display Technology* $\otimes$ *Position*; $F(1.90, 59.01) = 5.85$, $p = .005$, $\eta_p^2 = .159$. Pairwise comparisons revealed a significant difference between both displays for the *Mid* and position ($p = .008$), where the OST showed a higher absolute error compared to VST, but not for *Near* ($p = .523$) or *Far* ($p = .300$). Differences for *Target Position* were observed for the *OST Near* position, showing a higher accuracy than *Mid* and *Far* (all $p \leq .002$), see 5.5b. Further, the *VST Target Position* comparisons revealed that the *Far* position showed significantly higher errors than *Near* and *Mid* (all $p \leq .001$). These results were corroborated by a main effect for *Target Position*; $F(1.41, 43.73) = 26.6$, $p < .001$, $\eta_p^2 = .462$. Consecutive pairwise comparisons showed that the *Far* position evoked the largest absolute alignment error, significantly larger than the other positions ($ps < .001$).



(a) Display Technology

(b) Target Position

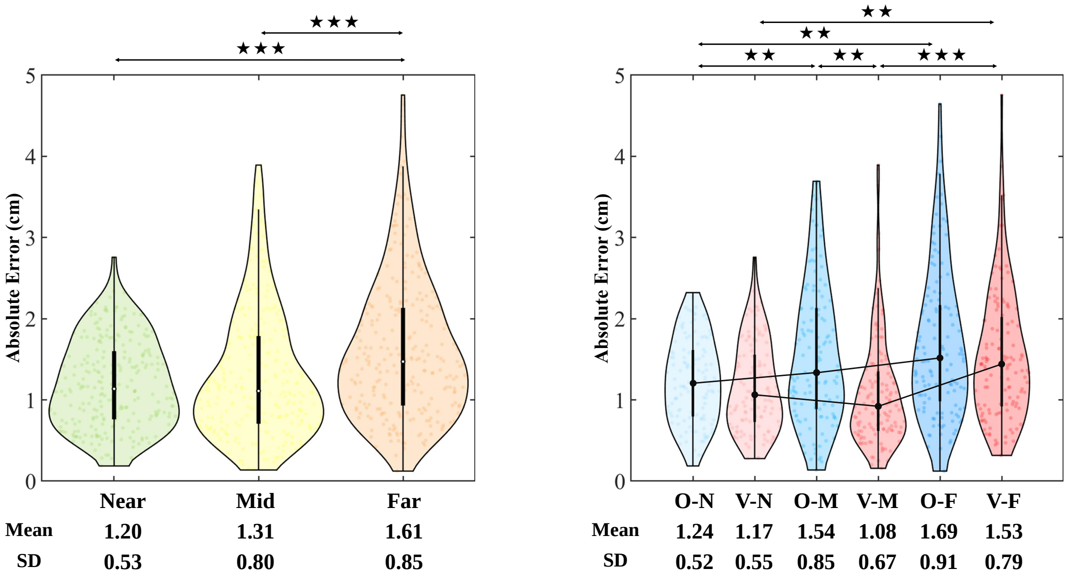
(c) Display Technology ⊗ Target Position

(d) Display Technology ⊗ Visualization Technique

Fig. 5.4. **Error (Study 1).** Violin plots depict distribution, median, and CI. Negative values describe a judgment of the virtual tumor closer to the participant compared to the actual target position of the real tumor. Results are depicted for Display Technology (O-*:*OST*, V-*:*VST*), Visualization (*-B:*Baseline*,*-W:*Virtual Window*,*-A:*Contextual Anatomical Mimesis*,*-M:*Virtual Mask*), and Target Position (*-N:*Near*, *-M:*Mid*, *-F:*Far*), as well as their interactions.

Time to Completion

The statistical analysis of the time to completion revealed a strongly significant main effect for *Target Position*; $F(1.37, 42.34) = 79.62$, $p < .001$, $\eta_p^2 = .720$. Pairwise comparisons showed that the *Far* position took the longest ($M = 13.04s$, $SD = 5.34$), followed by *Mid* ($M = 10.17s$, $SD = 3.80$) and *Near* ($M = 9.92s$, $SD = 3.83$; all $p < .001$). No further main or interaction effects were found.



	Near	Mid	Far
Mean	1.20	1.31	1.61
SD	0.53	0.80	0.85

(a) Target Position (Abs.)

	O-N	V-N	O-M	V-M	O-F	V-F
Mean	1.24	1.17	1.54	1.08	1.69	1.53
SD	0.52	0.55	0.85	0.67	0.91	0.79

(b) Display Technology ⊗ Target Position (Abs.)

Fig. 5.5. **Absolute Error (Study 1).** Results are depicted for Display Technology, (O-*:*OST*, V-*:*VST*), and Target Position (*-N:*Near*, *-M:*Mid*, *-F:*Far*), as well as their interactions. Values in tables depict means and standard deviations in centimeters.

Subjective Measures

NASA TLX. The descriptives of the Raw-TLX scores are reported in Table 5.1. There was a statistically significant interaction for *Display Technology* ⊗ *Visualization* for Mental Demand; $F(3,93) = 3.99$, $p = .010$, $\eta_p^2 = .114$. The analysis further revealed a main effect for *Visualization*; $F(3,93) = 20.20$, $p < .001$, $\eta_p^2 = .395$. Pairwise comparisons showed that using the *Virtual Mask* led to higher mental demand than the *Baseline*, *Virtual Window*, and *Anatomical Mimesis* (all $p < .001$). And further, when using the *Anatomical Mimesis* ($p = .010$) compared to the *Virtual Window*. A statistically significant interaction for *Display Technology* ⊗ *Visualization* for Effort was also found; $F(2.44, 75.5) = 4.79$, $p = .007$, $\eta_p^2 = .134$. A main effect for *Visualization*; $F(3,93) = 9.24$, $p < .001$, $\eta_p^2 = .230$ and pairwise comparisons showed that participants required more effort with the *Virtual Mask* than with the *Anatomical Mimesis* ($p = .005$) and *Virtual Window* ($p < .001$). Moreover, significant main effects for *Visualization* when analyzing the Physical Demand; $F(2.18, 67.42) = 3.52$, $p = .032$, $\eta_p^2 = .102$, the Overall Performance; $F(3,93) = 14.00$, $p < .001$, $\eta_p^2 = .311$, and the Frustration Level; $F(3,93) = 3.94$, $p = .011$, $\eta_p^2 = .113$ were found. Pairwise comparisons showed that the *Virtual Mask* required higher physical demand ($p = .013$) and led to lower overall perceived performance ($p < .001$), and higher frustration level ($p = .015$) than the *Virtual Window*, see Table 5.1.

Single Ease Question. The SEQ statistical analysis revealed significant interactions for *Display Type* ⊗ *Visualization*; $F(3,93) = 4.95$, $p = .003$, $\eta_p^2 = .138$. These results were corroborated by main effects for *Display Type*; $F(1,31) = 4.95$, $p = .034$, $\eta_p^2 = .138$, and *Visualization*; $F(3,93) = 24.67$, $p < .001$, $\eta_p^2 = .443$. Pairwise comparisons showed that the task was easier to preform when using the VST HMD ($p = .034$). Regarding the visualizations, the task was perceived to be significantly easier to perform when using the *Virtual Window* ($M = 3.27$,

$SD = 1.20$), followed by the *Anatomical Mimesis* ($M = 3.94$, $SD = 1.26$; $p = .002$), *Baseline* ($M = 4.11$, $SD = 1.39$; $p = .002$) and *Virtual Mask* ($M = 4.77$, $SD = 1.37$; $p < .001$).

Virtual Reality Sickness Questionnaire. The scores collected after using each display showed that participants reported significantly lower scores when using the OST; $F(1, 31) = 30.89$, $p < .001$, $\eta_p^2 = .499$, $M = 13.15$, $SD = 9.26$, than the VST ($M = 21.90$, $SD = 12.78$).

Tab. 5.1. Mean Raw-TLX scores reported by participants grouped by *Display Technology* and *Visualization*. MD:*Mental Demand*, PD:*Physical Demand*, TD:*Temporal Demand*, OP:*Overall Performance*, EF:*Effort*, FL:*Frustration Level*.

Disp. Tech. - Vis.	MD	PD	TD	OP	EF	FL
OST - Baseline	9.78	4.56	4.03	11.84	9.19	6.13
OST - Window	8.56	3.91	3.63	13.28	7.75	5.75
OST - Mimesis	11.19	5.09	4.09	10.38	9.81	6.78
OST - Mask	12.59	5.19	4.53	9.47	10.53	7.28
VST - Baseline	10.81	5.25	3.78	10.94	10.38	6.97
VST - Window	8.78	4.22	3.59	13.50	8.88	6.06
VST - Mimesis	9.81	4.56	3.50	11.94	8.00	5.97
VST - Mask	13.16	5.28	3.78	10.16	10.56	7.88

5.2.5 Discussion

This study showed that all visualization techniques achieved a lower alignment accuracy when participants performed the alignment task using the OST than the results obtained with the VST. In this regard, participants perceived the virtual tumor to be further away when using the OST HMD. Therefore, placing the virtual object closer to themselves than the real target. These results are consistent with the perceptual matching results reported by Swan et al. [163] for the virtual content. However, they differ from the underestimation reported by Singh et al. [156] when physical obstacles occlude the virtual objects. This difference may result from the highly salient texture used as the obstacle in such a study.

Conversely, participants slightly positioned the virtual tumor farther than the real in the VST condition. Nevertheless, the slight offset observed for the VST (M= 0.19cm) is similar to the results derived from the system error assessment for this condition (M= 0.14cm). Thus, the errors reported may not necessarily indicate misjudgments in depth caused by the in-situ visualization. Interestingly, these subtle differences, close to two millimeters, are similar to the results of the perceptual matching experiments reported by [163] for real content.

In terms of the target positions, users perceived the virtual tumor to be closer to themselves when positioned at the *Mid* position than in the *Near* and *Far* for both display technologies. In addition, the absolute positioning error was highest for the *Far* position for both display types. Although it may be intuitive to think that the absolute positioning error increases with the distance from the observer, the bias in the *Mid* position has no obvious explanation. One interpretation could be that at the *Near* and *Far* positions, the tumor may have been occluded

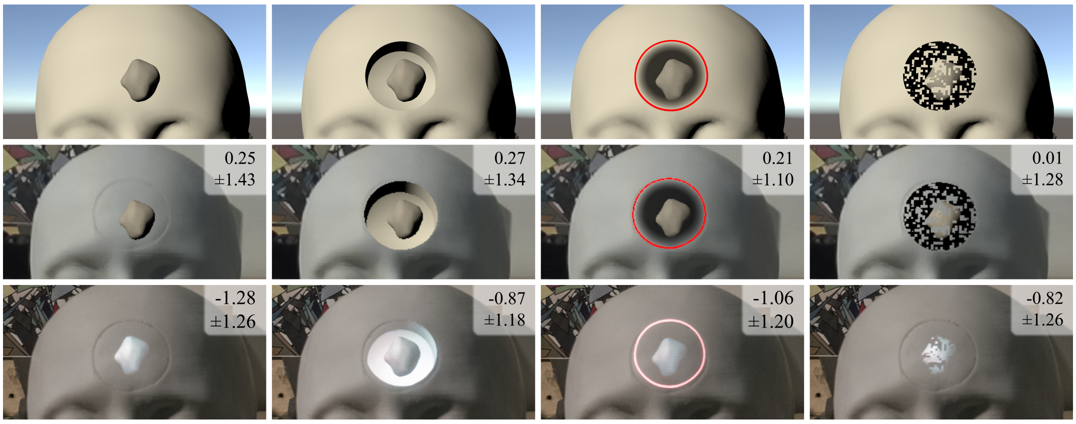


Fig. 5.6. **Base implementations of focus and context visualization techniques (*top*) and their appearance in video- (*mid*) and optical- (*bottom*) see-through head-mounted displays**. From left to right: *Baseline* overlay without contextual layer, *Virtual Window*, *Contextual Anatomical Mimesis*, and *Virtual Mask*. Mean, and standard deviation of corresponding alignment errors of study 1 are presented in centimeters. The OST images are captured using a smartphone camera placed at the eye position. Contrast and brightness have been adjusted for a faithful impression of the overlay as observed by the user.

by the contextual layer provided by the visualization techniques, allowing to derive additional information from occlusion. In contrast, this effect may be weaker at the *Mid* position, hindering the estimation of the tumor's position using only the information available. Nevertheless, this interpretation cannot be confirmed without performing further experiments.

According to findings from Otsuki and collaborators [124], the use of the virtual mask and cutouts, helps to improve depth judgments compared to the scenarios where *no masking* is used. These results correspond to the observation of virtual objects placed 0.1 to 2cm behind a surface. Although the distances estimated in this study required the estimation of larger distances, being the closest distance the *Near* position, the results suggest a similar trend when comparing the *Virtual Mask* and *Virtual Window* to the *Baseline* for the *OST*. Nevertheless, for the *VST* condition, this is only true for the *Virtual Mask* and not for the *Virtual Window* (Figure 5.6 and 5.4d).

Regarding the subjective measures, the *Virtual Window* consistently scored best while the *Virtual Mask* scored worst. The NASA TLX questionnaires show that participants perceive the alignment task to be more mentally and physically demanding, reporting higher effort and frustration levels and lower overall performance scores when comparing these visualizations. Interestingly, although the task performed was the same, observations collected from the study suggest that participants may have perceived the task to be more physically demanding using the *Virtual Mask* because it required using motion parallax to derive additional information. This assumption is in line with comments from participants who did not like the visual appearance of the virtual mask. Nevertheless, the alignment accuracy achieved using the *Virtual Mask* is comparable to one achieved with the *Virtual Window*. Although the virtual holes observed when using the *Virtual Mask* provide additional cues such as occlusion and parallax, the overall increased visual complexity and the lack of background in the OST condition made this visualization harder to understand and may explain the lower subjective scores.

5.3 Improving In-situ Visualization for OST-HMDs

The comparison between OST and VST display technologies presented in the previous section showed the need for more specialized visualizations to provide in-situ visualization when using OST-HMDs. This study revealed two important shortcomings that reduce the effectiveness of the existing techniques when using OST-HMDs. First, the inability to render opaque black colors in additive displays affects the visual outcome compared to the VST-HMDs. This aspect is particularly evident when using the *Contextual Anatomic Mimesis* and *Virtual Mask*. Second, the contrast observed with the additive displays reduces the visibility of shading details, for example, within the interior surface of *Virtual Window*.

To tackle these problems and motivated by the results of the previous study, a follow-up study presented in this section considers the design and evaluation of two techniques that have not been used for this purpose in MR applications. First, compressing the intensity range of the dark colors, mapping black to gray, can help mitigate dark regions' disappearance. These gray areas could help to provide a more noticeable visual boundary between non-augmented regions and originally augmented areas using black. In this regard, *chromatic shadows* could compensate for the lost contrast in luminance that this approach implies and introduce additional contrast through chrominance differences. This visualization technique has been used to enhance the contrast of shaded areas in scientific volume visualization [157]. Second, when using additive displays, regions with low luminance may appear translucent because the real world is visible behind the virtual content, whereas very bright overlay regions appear opaque. To exploit this aspect, *hatching* techniques commonly employed in illustrative rendering [93] may support the visualization of highly salient bright strokes that would occlude the real-world background, supporting the observation of the virtual content in place. These techniques also benefit from adding texture to an otherwise featureless surface, conveying the surface shape better than shading details on a low contrast OST. Furthermore, hatching techniques provide additional information derived from motion perspective, representing the third most significant depth cue in the personal space only after occlusion and binocular disparity [24]. These two techniques could also be combined to strengthen the illusion of observing consistent occlusion.

Therefore, to systematically evaluate the existing methods and assess the potential integration of the combinations proposed, this section presents a study where the visualization techniques are decomposed using three orthogonal properties: EXTERIOR *Visualization*, INTERIOR *Rendering* and SHADOW *Representation*. These properties characterize the previous methods and allow for a more natural integration of the chromatic shadows and hatching techniques.

Furthermore, this characterization supports a factorized analysis of the orthogonal properties and allows assessing their individual impact. Thus, enabling to describe the visualization techniques studied in Section 5.2 in terms of these properties as explained in Table 5.2 and shown in Figure 5.7. The orthogonal properties are defined as follows:

> **Exterior Visualization.** This property describes how the opacity of the real surface is occluded with the virtual content to generate the transparency effect and represents the most notorious difference between the methods evaluated in Section 5.2.

Interior Rendering. The interior rendering determines the internal appearance of the contextual virtual layer. It can be rendered in multiple styles, including traditional techniques such as Phong shading as observed on the interior surface of the *Virtual Window*. Additionally, non-traditional techniques such as *hatching* [131] can be used to create images with the appearance of pencil drawings. This study modified the hatches from their original form to provide bright streaks and create a stronger intensity contrast, adding additional visual detail in the background.

Shadow Representation. A minimum luminance even in fully darkened areas can be ensured using illustrative *chromatic shadows* [157] to avoid undesired effects resulting from rendering dark shadows using OST-HMDs. This method uses a *shadowiness* parameter S to compute a shadow color relative to the surface color. The shadow color is adapted by partially shifting the natural luminance from a gray-level to a chrominance contrast with a shadow color tone with the same perceptual distance as the original black shadow (measured in CIELAB color space).

$$C_{RGB} = (1 - S)C^O + SC_S \tag{5.1}$$

The implementation used during this study uses the clamped Lambert term $S = \max(n \cdot L, 0)$ as the *shadowiness* factor. Given that the interior surface color is pure white, the shadow color C_S does not depend on the surface position and can provide a consistent shadow color. Therefore, this study uses a blue-tinted shadow which performs well according to [157]. In addition, a grayscale shadow is used to compare whether the tinted shadow has a favorable effect.

Tab. 5.2. Possible values for the three visualization properties.

	EXTERIOR *Visualization*
Hole	A circular cutout hole with a hard edge.
Ghosted	Advanced opacity modulation based on curvature, normal, and Gaussian falloff as used in *Contextual Anatomical Mimesis*.
Mask	Circular cutout modulated by a binary random stencil texture as in *Virtual Mask*.
	INTERIOR *Rendering*
Constant	Constant background color.
Shaded	Diffuse shading applied to the interior.
Hatched	Illustrative hatching.
	SHADOW *Representation*
Black	A shadow color of $C_S^1 = (0, 0, 0)$, reduces Equation 5.1 to standard diffuse shading
Chromatic	A blue shadow of color $C_S^2 = (63, 89, 150)$
Bright	A gray value $C_S^3 = (89, 89, 89)$, chosen to have the same luminance as C_S^2

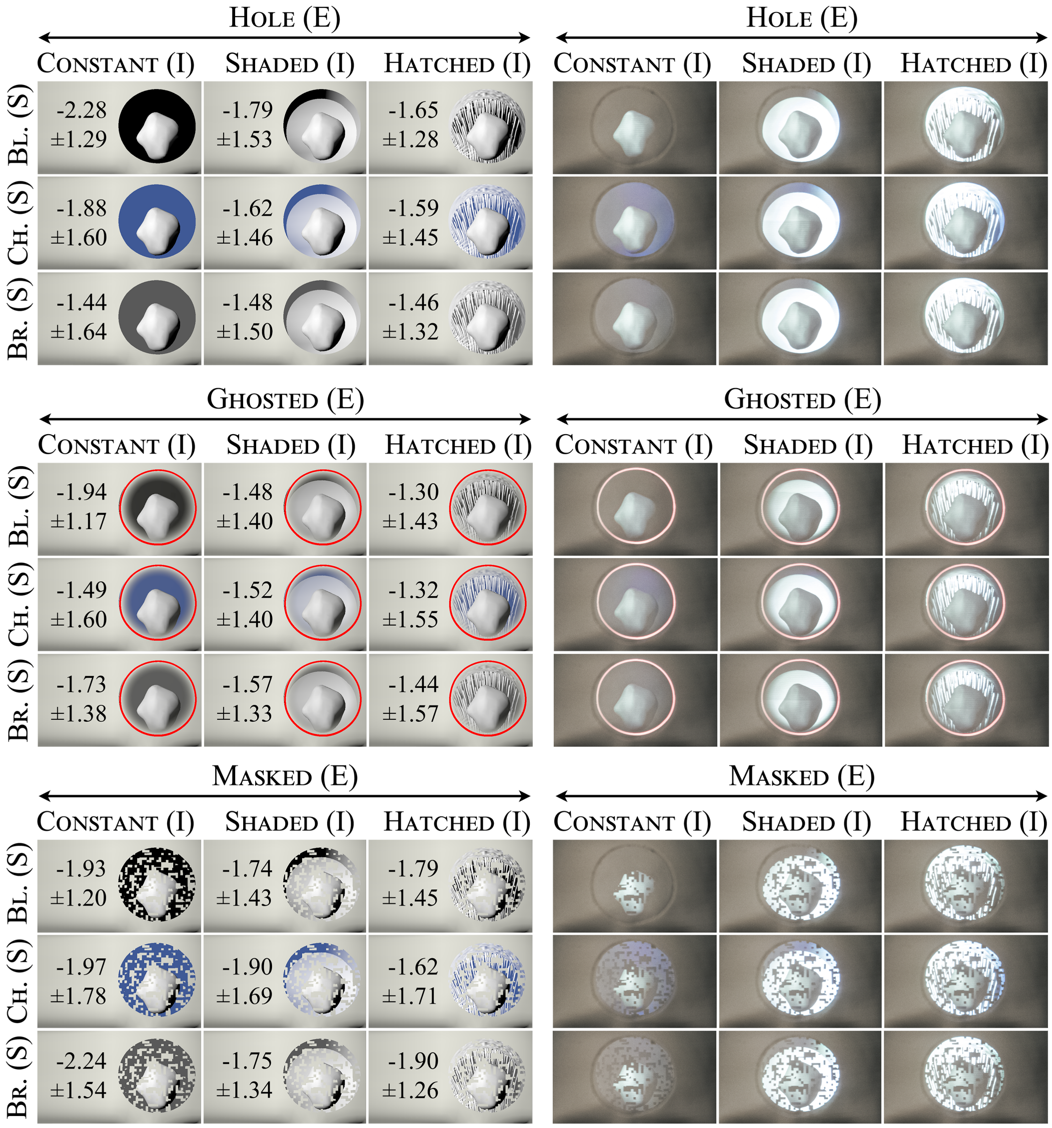


Fig. 5.7. All 27 visualization permutations tested in Study 2. Top: Base implementation in a virtual environment, with error $mean \pm SD$ in cm. Bottom: Corresponding appearance of the visualizations in the OST HMD. The images are captured with a smartphone camera placed at the eye position. Contrast and brightness have been adjusted to provide a faithful representation of the real view. (E) Exterior, (I) Interior, (S) Shadow.

Based on the findings from Section 5.2, and considering the perceptual advantages and visual cues provided by the techniques described in Section 5.3, this study aims to prove the following hypotheses: ***(H2.1)*** *Using visualization techniques that enhance the contrast between a virtual object and the background observed contribute to improve the depth estimation of the object and help to reduce the alignment error when using OST displays* ***(H2.2)*** *Using visualization techniques that enhance the contrast of shaded areas by ensuring a minimum luminance value contribute to improve the depth estimation of the virtual object and help to reduce the alignment error when using OST displays.*

5.3.1 Participants

Twenty-seven paid participants recruited using mailing lists and campus announcements (9 female and 18 male), aged between 22 and 34 years old (mean age of 26.2 ± 3.4 years), took part in this study. Like the previous study, a Landolt C-Test (EN ISO 8596), an Ishihara Color test, and a Titmus test for stereo vision were used to ensure correct or corrected-to-normal vision capabilities. The participants' interpupillary distance (IPD), estimated using the OST HMD, ranged between 59.3 and $67.7mm$ (mean IPD of $63.69 \pm 2.75mm$). None of the participants had previous experience using the system. The study participation was voluntary, and participants could abort it at any time.

5.3.2 Experimental Setup

A 3 (Exterior *Visualization*) ⊗ 3 (Interior *Rendering*) ⊗ 3 (Shadow *Representation*) ⊗ 3 (Target Position) within-subjects follow-up study using the apparatus as described in Section 5.2 was implemented using only an OST HMD. The factorial design aimed to decompose the aspects of the visualization techniques investigated previously, resulting in 27 different visualization variants (i.e., study cells) depicted in Figure 5.7.

Dependent Measures

The measures collected for this study correspond to those used in the previous study. The objective assessments were: (signed) error, absolute error, time to completion, and vision tests. Since this study focused on analyzing the visualization techniques' components, the NASA TLX questionnaire was excluded. Instead, a rating for visual attractiveness: "*Overall, I liked the appearance of the visual information provided by this technique*", answered using a 7-point Likert scale ranging from *1-Strongly agree* to *7-Strongly disagree* was included.

Procedure

The overall procedure of this study followed the structure of the study presented in Section 5.2. The number of repetitions was reduced to only two to compensate for the increased number of conditions: one *back-to-front* and one *front-to-back*. These new design considerations resulted in 162 trials (one block for each of the twenty-seven visualization combinations consisting of three target positions and two repetitions). The positioning and control of the virtual tumor were adapted to avoid any influence of the linear slider. The position and randomization range at which the tumor appeared increased to -9 ± 1cm (*front-to-back*) and $9 \pm 1cm$ (*back-to-front*). The movement scaling of the linear slider decreased by 45% to allow for more precise control. Participants filled the two subjective questions on a laptop while looking through the headset instead of folding up the HMD's display. This consideration streamlined the experiment and avoided the need for recalibration using the optical marker. Participants required 75 minutes, on average, to complete the study.

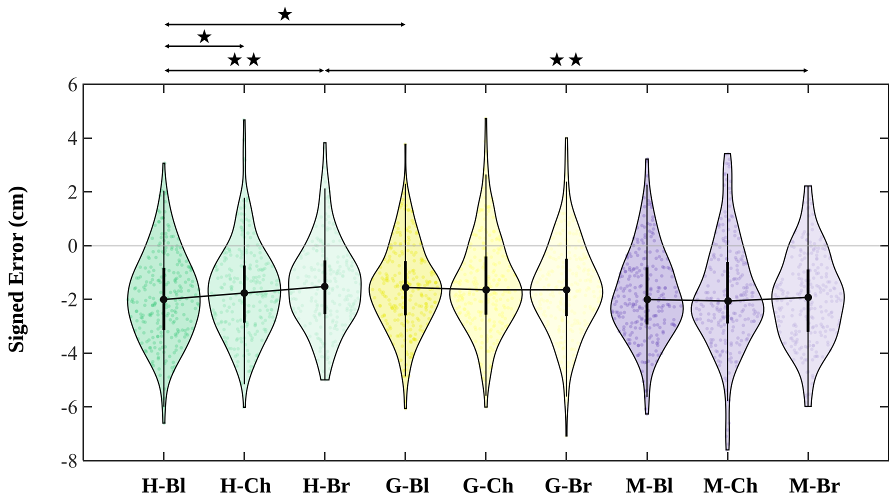
5.3.3 Results

The statistical analysis involved a four-way (EXTERIOR *Visualization* ⊗ INTERIOR *Rendering* ⊗ SHADOW *Representation* ⊗ *Target Position*) repeated measures analyses of variance (ANOVAs) for the objective assessments, aggregating the repetitive trials for each cell. For the subjective measures, three-way (EXTERIOR *Visualization* ⊗ INTERIOR *Rendering* ⊗ SHADOW *Representation*) ANOVAs were used. Significance was accepted at $\alpha = 0.05$. Greenhouse-Geisser corrected values are reported when the sphericity assumption was violated, assessed by Mauchly tests. Bonferroni adjusted post hoc tests for the pairwise comparisons are reported. The outlier correction was performed analogously to Section 5.2. This procedure removed 44 (1.01 %) of the values for the alignment accuracy and 85 (1.94 %) for time to completion. These values were replaced with the remaining mean of the specific cell to simplify the analysis and reporting.

Alignment Error

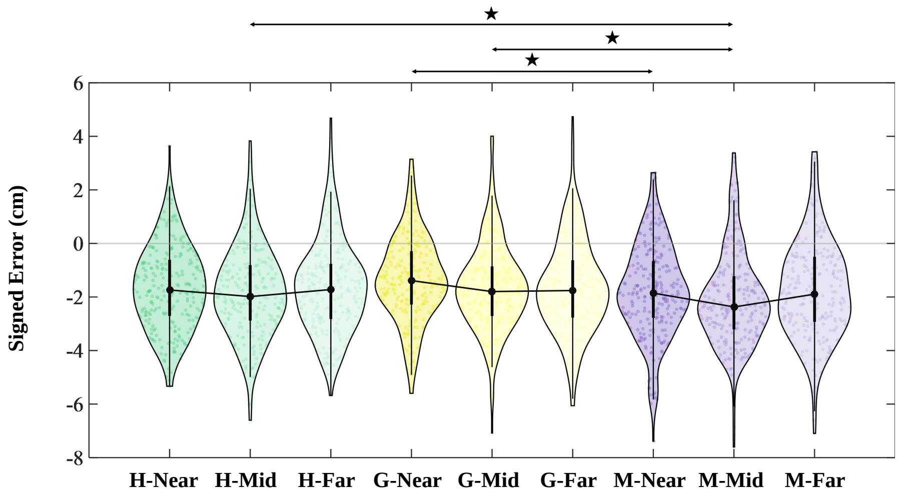
Error: In terms of signed error, significant interaction for EXTERIOR *Visualization*×SHADOW *Representation* was found; $F(2.71, 70.51) = 4.23$, $p = .010$, $\eta_p^2 = .140$. Errors were relatively similar for each shadow representation using the *Mask* exterior. The *Chromatic* shadow led to the smallest error using the *Mask* exterior, whereas the *Bright* shadow led to the smallest error with the synthetic *Hole* (Figure 5.8). Overall, the latter led to the lowest error in these combinations. In addition, significant interaction for EXTERIOR *Visualization*×*Target Position* was found; $F(4, 104) = 3.38$, $p = .012$, $\eta_p^2 = .115$. Pairwise comparisons showed that participants performed better using the *Ghosted* than the *Mask* exterior at *Near* ($p = 0.013$) and *Mid* ($p = 0.025$) positions. Moreover, they performed better using the *Hole* than the *Mask* at the *Mid* position ($p = 0.027$). These results are presented in Figure 5.9. The relatively strong impact of the EXTERIOR *Visualization* was corroborated by a main effect; $F(1.26, 32.84) = 4.71$, $p = .030$, $\eta_p^2 = .153$. Pairwise comparisons showed that, overall, the *Mask* evoked the largest error, significantly larger than the *Hole* ($p = .007$). Additionally, a main effect for INTERIOR *Rendering* was found; $F(2, 52) = 6.46$, $p = .003$, $\eta_p^2 = .199$. Pairwise comparisons revealed that participants scored the largest negative errors using a *Constant* rendering, significantly larger than with the *Hatched* ($p = .015$). These results are summarized in Figures 5.10a and 5.10b. Overall, the *ghosted black* visualization combined with the *hatching* technique presented the smallest error ($M = -1.30$, $SD = 1.43$) compared to the other combinations (Figure 5.7).

Absolute Error: Statistical analysis for absolute error revealed significant interaction for EXTERIOR *Visualization*×*Target Position*; $F(4, 104) = 2.93$, $p = .024$, $\eta_p^2 = .101$. A main effect for EXTERIOR *Visualization*; $F(1.44, 37.55) = 7.79$, $p = .004$, $\eta_p^2 = .230$ was found. Pairwise comparisons showed that the *Mask* led the largest absolute error, significantly larger than the *Ghosted* ($p = .012$) and the *Hole* ($p = .003$), as depicted in Figure 5.11a. In addition, the statistical analysis revealed a main effect for INTERIOR *Rendering*; $F(1.50, 39.09) = 7.25$, $p = .004$, $\eta_p^2 = .218$. Absolute errors for *Hatched* were significantly lower compared to *Constant* ($p = .014$) as depicted in Figure 5.11b.



	H-Bl	H-Ch	H-Br	G-Bl	G-Ch	G-Br	M-Bl	M-Ch	M-Br
Mean	-1.91	-1.70	-1.46	-1.57	-1.44	-1.58	-1.82	-1.83	-1.96
SD	1.63	1.66	1.63	1.53	1.73	1.67	1.64	1.96	1.64

Fig. 5.8. **Error for Exterior×Shadow.** Lines connect median values. Results are depicted for Exterior, (H-*:*Hole*, G-*:*Ghosted*, M-*:*Mask*), and Shadow Representation (*-Bl:*Black*, *-Ch:*Chromatic*, *-Br:*Bright*), as well as their interactions. Values in tables depict means and standard deviations in centimeters.

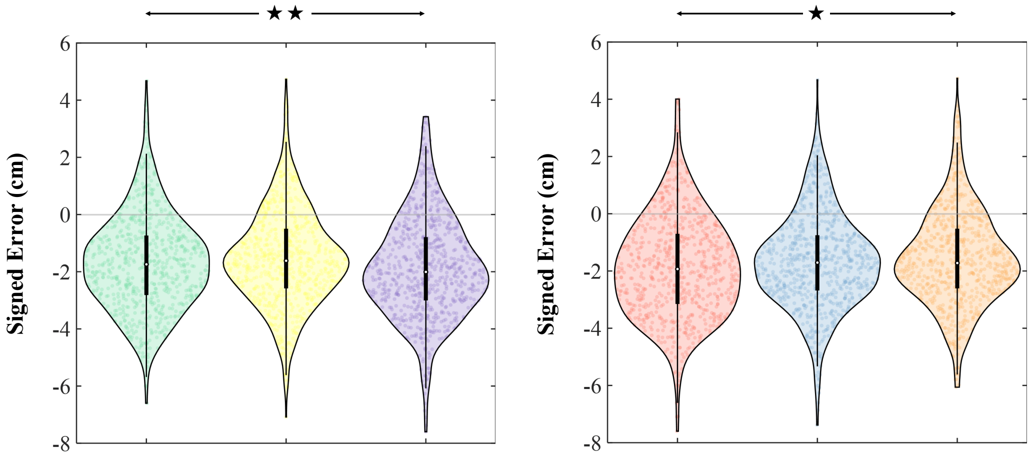


	H-Near	H-Mid	H-Far	G-Near	G-Mid	G-Far	M-Near	M-Mid	M-Far
Mean	-1.64	-1.81	-1.61	-1.34	-1.63	-1.62	-1.78	-2.08	-1.75
SD	1.55	1.66	1.72	1.57	1.62	1.73	1.72	1.74	1.79

Fig. 5.9. **Error for Exterior×Position.** Lines connect median values. Results are depicted for Exterior, (H-*:*Hole*, G-*:*Ghosted*, M-*:*Mask*), and Position (*-Near, *-Mid, *-Far), as well as their interactions. Values in tables depict means and standard deviations in centimeters.

Time to Completion

During the analysis of the time to completion, a strong significant main effect for *Target Position* was found; $F(2,52) = 25.35$, $p < .001$, $\eta_p^2 = .494$. Pairwise comparisons showed that the *Far* position took the longest ($M = 11.89$, $SD = 7.54$), followed by *Mid* ($M = 10.77$, $SD = 6.62$; $p < .001$) and *Near* ($M = 10.61$, $SD = 6.49$; $p < .001$). No other main or interaction effects were found.



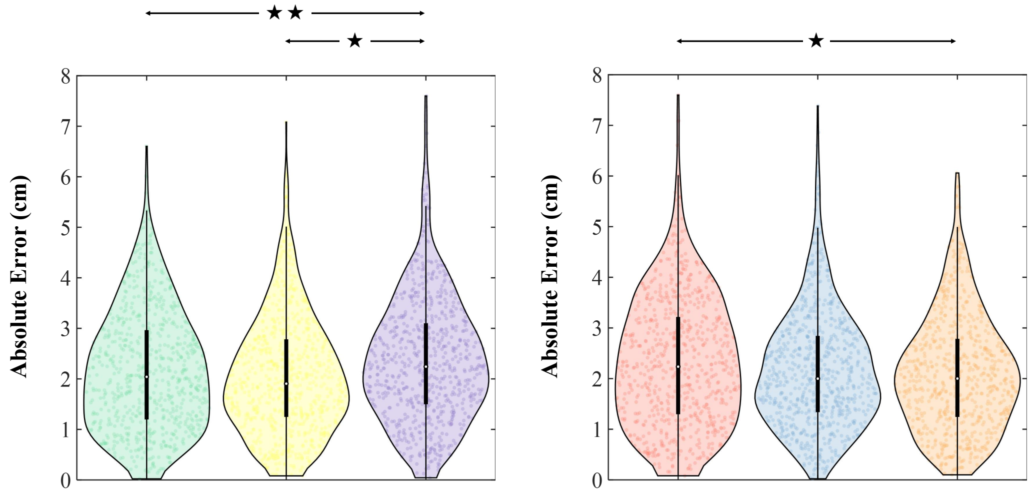
	Hole	Ghosted	Mask
Mean	**-1.69**	**-1.53**	**-1.87**
SD	**1.65**	**1.64**	**1.75**

(a) EXTERIOR *Visualization*

	Constant	Shaded	Hatched
Mean	**-1.88**	**-1.65**	**-1.56**
SD	**1.75**	**1.65**	**1.64**

(b) INTERIOR *Rendering*

Fig. 5.10. **Error (Study 2).** Results are depicted for (a) EXTERIOR *Visualization* and (b) INTERIOR *Rendering*. Values in tables depict means and standard deviations in centimeters.



	Hole	Ghosted	Mask
Mean	**2.17**	**2.07**	**2.37**
SD	**1.18**	**1.15**	**1.23**

(a) EXTERIOR *Visualization*

	Constant	Shaded	Hatched
Mean	**2.35**	**2.16**	**2.10**
SD	**1.27**	**1.15**	**1.14**

(b) INTERIOR *Rendering*

Fig. 5.11. **Absolute Error (Study 2).** Results are depicted for (a) EXTERIOR *Visualization*, and (b) INTERIOR *Rendering*. Values in tables depict means and standard deviations in centimeters.

Subjective Measures

Single Ease Question: Results for the SEQ (Table 5.3) revealed significant interactions between EXTERIOR *Visualization*×SHADOW *Representation*; $F(4, 104) = 4.04$, $p = .004$, $\eta_p^2 = .135$. These results were corroborated by main effects for EXTERIOR *Visualization*; $F(1.26, 32.83) = 24.72$, $p < .001$, $\eta_p^2 = .487$. Pairwise comparisons showed that participants

Tab. 5.3. **SEQ Results for Study 2**, reported as $mean \pm SD$ in terms of EXTERIOR Visualization (*columns*), INTERIOR rendering (*Constant*, *Shaded*, *Hatched*), and SHADOW representation (*Black*, *Chromatic*, *Bright*).

	Hole	Ghosted	Mask
Constant - Black	3.67 ±1.44	4.44 ±1.62	4.85 ±1.58
Constant - Chromatic	2.74 ±1.14	4.04 ±1.60	3.89 ±1.75
Constant - Bright	3.63 ±1.47	4.41 ±1.28	4.74 ±1.69
Shaded - Black	2.59 ±1.28	3.11 ±1.26	3.81 ±1.33
Shaded - Chromatic	2.67 ±1.12	3.26 ±1.17	3.07 ±1.02
Shaded - Bright	2.74 ±1.26	3.67 ±1.31	3.33 ±1.33
Hatched - Black	2.52 ±1.00	3.07 ±1.15	4.00 ±1.49
Hatched - Chromatic	2.37 ±0.99	3.30 ±1.24	3.44 ±1.40
Hatched - Bright	2.52 ±1.17	3.33 ±1.39	3.59 ±1.28

perceived the task easier to complete using the *Hole* ($M = 2.83$, $SD = 1.30$; $p < .001$) than the *Ghosted* ($M = 3.63$, $SD = 1.44$) and the *Mask* ($M = 3.86$, $SD = 1.56$). Additionally, a significant main effect for SHADOW *Representation* was found; $F(2, 52) = 14.051$, $p < .001$, $\eta_p^2 = .351$. Posterior pairwise comparisons revealed that users find the alignment task easier to achieve using *Chromatic* SHADOWS ($M = 3.20$, $SD = 1.40$) than with the *Black* ($M = 3.56$, $SD = 1.56$; $p < .001$) and *Bright* ($M = 3.55$, $SD = 1.52$; $p = .001$) representations.

Furthermore, significant interactions for INTERIOR *Rendering*×SHADOW *Representation*; $F(4, 104) = 3.87$, $p = .006$, $\eta_p^2 = .130$ were found. Main effects for INTERIOR *Rendering*; $F(2, 52) = 19.33$, $p < .001$, $\eta_p^2 = .351$, revealed that users found easier to complete the task using the *Hatched* ($M = 3.13$, $SD = 1.35$) and *Shaded* ($M = 3.14$, $SD = 1.30$) rendering compared with the *Constant* ($M = 4.05$, $SD = 1.64$; $p < .001$).

Visual Attractiveness: Results obtained after analyzing the participants' opinion regarding visual attractiveness showed significant interaction for INTERIOR *Rendering*×SHADOW *Representation*; $F(4, 104) = 4.46$, $p = .002$, $\eta_p^2 = .146$. A main effect for INTERIOR *Rendering* revealed that participants found the *Shaded* ($M = 3.35$, $SD = 1.54$; $p < .001$) and the *Hatched* ($M = 3.45$, $SD = 1.50$; $p = .004$) renderings more visually appealing than the *Constant* ($M = 4.17$, $SD = 1.63$).

Even more, the statistical test revealed a strong significant main effect for EXTERIOR *Visualization*; $F(2, 52) = 27.06$, $p < .001$, $\eta_p^2 = .510$. Posterior pairwise comparisons showed that the *Mask* ($M = 4.30$, $SD = 1.59$) representation was less visually appealing than the *Ghosted* ($M = 3.73$, $SD = 1.49$; $p = .005$) and the *Hole* ($M = 2.95$, $SD = 1.41$; $p < .001$), as well as the *Ghosted* than the *Hole* ($p < .001$).

5.3.4 Discussion

In terms of EXTERIOR *Visualization*, results from this study show that the *Mask* performs badly for accuracy and usability-related metrics. Even combinations that use a brighter background cannot improve the estimation error, and the subjective metrics show similar behavior. The *Hole* and *Ghosted* perform similarly for error metrics. However, participants' preference seems to benefit the *Hole* visualizations over the *Ghosted* techniques. This is an interesting finding considering the *Hole* is one of the earliest works introduced for in-situ visualization. Even more, the *Ghosted* and *Mask* visualization techniques correspond to more recent methods and are based on careful arguments and considerations on visual perception.

The analysis of the INTERIOR *Rendering* revealed that the participants prefer the addition of interior geometry (*Shaded* or *Hatched*). In addition, the *Hatched* modifications outperformed the *Constant* shading and reduced the alignment error, supporting ***H2***.

In terms of the SHADOW representation, results from this study suggest that this component interacts with the EXTERIOR. However, the descriptives presented in Figure 5.7 suggest that *Bright* shadows work best using the *Hole* and *Ghosted* exterior. However, the use of *Chromatic* shadows does not improve the results when using the *Mask*. These results partially support ***H3*** as the visualizations with *Chromatic* shadows only helped improve the alignment error in some cases. Additionally, using *Chromatic* shadows showed a positive effect for SEQ. Therefore, it would be recommendable to utilize this technique in circumstances where usability is of decisive importance.

Lastly, the descriptives presented in Figure 5.7 can be used to derive two general recommendations for the estimation error, yet not easily shown through statistical analysis. First, the HOLE exterior seems to benefit from brightening dark colors, with the INTERIOR rendering playing a minor role. Second, the HATCHED interior seems to play a more determinant role than the shadow color when using a GHOSTED exterior.

5.4 Limitations

The visualization techniques presented in these studies do not consider how the augmented objects of interest are rendered, as this would increase the complexity of the factorial design. Techniques such as depth-encoding outlines [62] or pseudo chromadepth [138] have been shown to aid perception effectively [66]. In this regard, future experiments, including combinations of these techniques with the hatching techniques, might provide interesting results. Further interesting aspects can include different cutaway geometries [99], other illustrative surface shading techniques [93], or even animated surface visualizations. Although these techniques may also lead to interesting results, the studies presented in this section were constrained to the techniques proposed to avoid very extensive experimental sessions and participant fatigue.

In addition, it is important to mention that the studies presented here considered adjusting the brightness to ensure a minimum luminance and enhance the visibility of the virtual content when using the OST displays. Recent studies have shown that the brightness of the virtual content influences the accuracy achieved in near-field depth matching tasks [155]. However, these studies involved the judgment of virtual content that was not presented inside real objects.

Moreover, the interpretation of the results presented here is limited to the specific characteristics of the headsets used. In this regard, different headsets may cover a wide range of intrinsic parameters such as different FoV, focal planes, display resolutions, brightness, and other characteristics that can influence depth perception. To name an example, the comparably low angular resolution of the VST device used in the first study might have negatively impacted how the *Virtual Mask* performs, as the mask cutouts caused some aliasing at the pixel boundaries. Furthermore, it is expected that additive displays with different overall brightness will likely benefit from adapting the luminance of the shadow color C_s of the SHADOW representations accordingly. Although the visualizations investigated here do not strongly rely on the display's color accuracy, an interesting point for future research would be to investigate whether spatial perception is affected by displays with limited color uniformity like the HoloLens2 display. In addition, VST and OST differ fundamentally concerning accommodation, making a direct comparison between the two technologies relatively complex.

In terms of statistical power, the size of the cohorts recruited for both studies was driven by considering the balanced randomization of the experiments through a Latin-Square Matrix in a repeated measure fashion. Considering the complexity of the models and the multiple levels investigated in the studies, future studies should include larger samples, focusing on combinations and factor interactions.

Lastly, the experimental setup is limited to near-field depth judgments that represent a prototypical distance for medical scenarios. As a result of this, the studies did not investigate depth estimation at mid-to-far distances, a larger separation between the target positions, or larger object geometries.

5.5 Study Implications

The first study presented in this section has provided a direct comparison of how VST and OST-HMDs can influence depth estimation of objects shown below a real object's surface when using F+C visualization techniques. This comparison has demonstrated that depth judgments using F+C techniques for in-situ visualization are more accurate with VST-HMDs. This results from the VST technology providing more consistent visual cues such as a believable occlusion of the real objects using virtual content.

The second study enabled investigating how the proposed techniques can impact estimating the objects' depth when using OST-HMDs. The decomposition of the proposed techniques in this study showed that using *interior hatching* can provide useful cues to improve accuracy in the estimation of the object's depth, as confirmed by both signed and absolute errors. In addition, the use of *chromatic shadows* showed significant improvements for the subjective scores. In this regard, the novel visual components proposed in this work contribute to reducing the perceived complexity of the task. Moreover, they increase the visual attractiveness of the content presented without increasing, and in many cases decreasing, the estimation errors during perceptual matching tasks. These results suggest that *interior hatching* and *chromatic shadows* can effectively improve in-situ visualization with OST HMDs.

Regarding the original visualization techniques, the studies presented in this section provide evidence that the masking method proposed by Otsuki et al. [125] presents adequate depth cues and produce similar alignment errors as the other methods. However, users do not find the visualization technique visually appealing, as indicated by the subjective scores from both studies. In addition, the participants seem to perceive the perceptual matching task to be harder when using this technique. For this reason, EXTERIOR rendering approaches such as the *Hole* or *Ghosted* should be preferred for the presentation of in-situ content at near-field distances in MR.

Overall, results from these studies showed that virtual objects tend to be consistently perceived further away from the observer when using OST-HMDs. These results differ from the findings reported in studies that explored how physical occluders affect the perceived depth of virtual content at near-field distances in MR [156]. Nonetheless, it is important to mention that such studies used a highly salient textured occluder. In contrast, the experimental setup presented in this section used a monochromatic 3D printed head with only a few salient features.

Regarding the Occluder's Texture

The differences between the results presented in this section and those reported by Singh et al. [156] may be an indication that the occluder's texture influences the user's perception of the virtual content. However, the experimental setup presented in this section was designed to provide contextual information when using the visualization techniques proposed, and not to explore the effects of the occluder's texture over the estimation of depth. Therefore, further investigations need to be conducted in this regard.

The conceptual and explicit decomposition of $F+C$ visualization techniques presented in this work aims to provide useful insights for developing novel approaches for in-situ visualization using OST- and VST-HMDs. The structured analysis used to develop and evaluate the visualizations techniques presented in this work could serve as a model for future extensions and contribute to forming the basis for future investigations of in-situ visualization techniques. In this regard, users would benefit from the adaptation of visualization techniques designed to provide optimal visual cues based on the demands of specific tasks.

Part III

Multi-View Alignment

In addition to the visualization techniques presented in Part II, other approaches can aid the users of MR by providing alternative viewpoints of the scene. These alternative viewpoints can help resolve the ambiguous information observed along the user's line of sight in single-view approaches. Therefore, contributing to mitigating the alignment errors. This part explores the advantages of using external camera views and mirrors to present the additional information. Moreover, it investigates how the users take advantage of the supporting information during task performance. In addition, it proposes a new class of MR mirrors that provides alternative viewpoints of the real and virtual worlds using real mirrors.

At first, results from a study comparing the alignment accuracy achieved using additional views generated from virtual cameras and mirrors are presented and discussed in Chapter 6. In addition to the clear advantage of providing an alternative view of the scene, these virtual helpers present multiple benefits that can be used during alignment tasks, such as recurring to motion parallax to generate dynamic viewpoints when using mirrors.

The idea of providing these additional viewpoints, originally validated in a virtual environment, inspired the concept of the AUGMENTED MIRRORS. This concept introduces a new class of MR mirrors that can reflect the real and virtual content of an MR environment in real-time, allowing the observer to change its viewpoint dynamically. This concept, presented in Chapter 7, can also be used for exploration, scene understanding, or multi-modal data visualization.

This part contains results from the works:

> **Martin-Gomez, Alejandro**, Javad Fotouhi, Ulrich Eck, and Nassir Navab. "Gain A New Perspective: Towards Exploring Multi-View Alignment in Mixed Reality." In International Symposium on Mixed and Augmented Reality (ISMAR), pp. 207-216. © 2020 IEEE.
>
> **Martin-Gomez, Alejandro***, Alexander Winkler*, Kevin Yu*, Daniel Roth, Ulrich Eck, and Nassir Navab. "Augmented Mirrors." In International Symposium on Mixed and Augmented Reality (ISMAR), pp. 217-226. © 2020 IEEE.

All authors have permitted the reproduction of this content.

* The asterisk indicates an equal contribution from the corresponding authors.

6 Alternative Viewpoints

The advantages of using multiple viewpoints to enrich what users perceive in mixed and virtual reality is not a new concept and have largely been explored in the past. One of the initial efforts towards achieving this is the *worlds in miniature* metaphor introduced by Stoakley et al. [159]. This metaphor enabled the visualization of multiple virtual views using a miniature replica of the virtual environment. This property allowed users to manipulate objects while changing their viewpoints simultaneously. Posterior work presented by Kiyokawa and Takemura [87] proposed the *tunnel window* concept. The tunnel window possessed the capacity of providing multiple viewpoints of a virtual environment, allowing users to navigate and interact with the content via single or multiple scenes. Alternative approaches as the one presented by Nakamura et al. [117] have used these additional viewpoints to generate third-person views for their application in VR gaming setups. This concept used a video sequence acquired from an external camera and presented it to the users using an HMD, showing that users can adapt relatively easily to these views. Nevertheless, a very interesting remark from the authors was that cognitive difficulties were observed when the users were facing the external camera. These cognitive difficulties result from the observation of mirrored views. The visualization of multiple viewpoints has also been explored in AR. In this regard, the *augmented viewport* presented by Hoang and Thomas [71, 72] enriched the egocentric viewpoint with a picture-in-picture video sequence of an external camera in outdoor applications. The viewport, used for manipulating and aligning virtual objects placed on top of buildings, showed improved precision during task performance, requiring less time and effort. In addition, Sukan et al. [160] presented a concept to align pairs of real objects using pre-acquired photographs from different viewpoints. This system allowed users to achieve faster alignment when compared to scenarios where users had to walk to acquire additional information from the scene.

Furthermore, the use of multiple views has also been used for collaborative environments. The *Dollhouse* system introduced by Ibayashi et al. [77] used VR environments in architectural settings to design living or working spaces. The dollhouse presented a top-down view of a floorplan allowing non-immersed users to provide instructions to users immersed in the virtual world. The system enabled the communication between the users by integrating pointing targets and allowing the immersed user to see through the ceiling of the virtual environment. Additional work, presented by Grandi et al. [56], utilized hand-held devices to enable simultaneous collaborative manipulation of virtual objects in AR. In addition, more recent work presented by Kunnert et al. [91] explored the advantages of using egocentric and allocentric views to provide navigation capabilities during exploration tasks.

As an alternative to using external cameras to generate the additional viewpoints, other solutions have used mirrored-like views to improve users' experience in MR and VR applications. In this regard, Bichlmeier et al. [7] proposed the use of *virtual mirrors* as an interaction paradigm. This concept was used in a medical setting to augment the direct view of physicians in laparoscopic applications, providing additional views from volumetric medical data. Alternatively, the *AR-reflective displays* introduced by Fotouhi et al. [42] can be used to improve the workflow of medical procedures involving robots. This work explores the advantages of using *AR-reflective displays* when setting up robotic arms used for trocar placement in minimally invasive surgery. This concept required the pre-acquisition of images from multiple viewpoints of the scene. These images were later inverted over its horizontal axis to provide a quasi-mirrored view from their respective viewpoints. Although this concept showed improvement in registering a virtual replica of the robot with its real counterpart, the *AR-reflective displays* cannot reflect changes in the real content after the images are acquired.

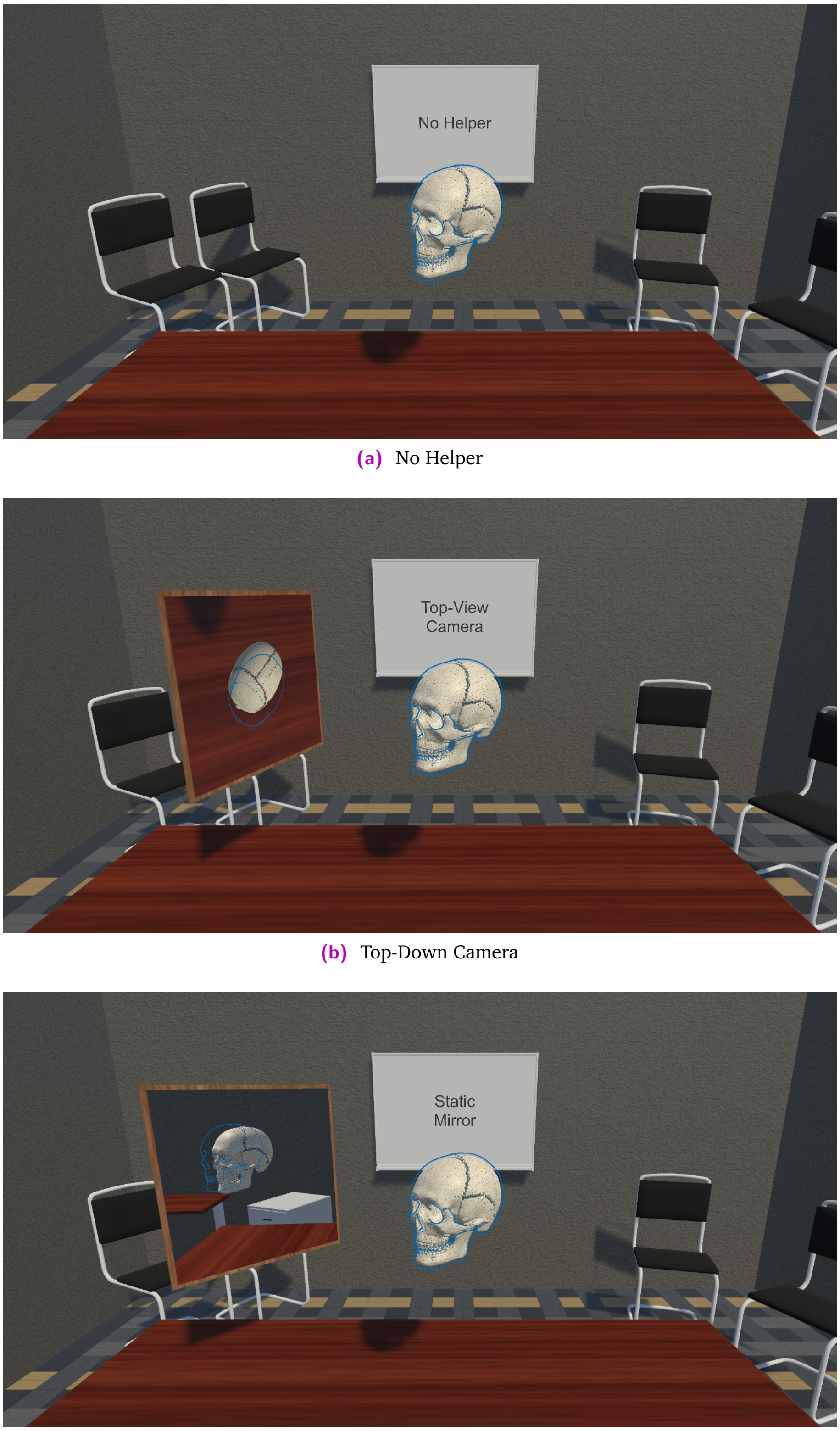
Until now, multiple studies have shown the undeniable advantages of using multiple viewpoints in MR and VR environments. These advantages include their application for manipulation and navigation tasks [70], improving perception in medical applications [7], reduce user displacement [160], or assisting users during the setup of robotic arms [42]. In contrast to existing works, this section thoroughly evaluates the user's performance when using multiple views for alignment tasks.

6.1 Virtual Helpers

This section introduces two types of helpers that provide meaningful information to the users by presenting additional viewpoints of the alignment scenario (Figure 6.1). These helpers provide different perspectives, mainly using external cameras and mirrors as follows:

> **Top-Down Camera Views.** This type of helper provides an additional viewpoint using a virtual camera placed on top of the alignment area. This top-down camera view is relatively common in surgical settings where cameras and scanners provide useful images inside the operating room. Users familiar with using design and medical imaging software may accustom relatively easily to this type of representation. This type of software frequently uses a set of orthogonal views to visualize three-dimensional objects using bi-dimensional displays.
>
> **Mirrors.** Compared to the images generated when using top-down views from external cameras, using mirrors provides an additional advantage: any changes in the user's viewpoint will lead to the observation of different projections over the mirror's surface, allowing to explore further the environment. Although the observed image would present the inherent property of being flipped horizontally, the users' familiarity with these devices may facilitate the understanding and interaction when utilizing these helpers. In fact, it is common to use mirrors in medical fields such as in dentistry, in industrial scenarios to perform maintenance and repair tasks, or when we use them while parallel parking a car.

The additional information provided by these helpers can support resolving ambiguities during object alignment and overcome the physical restrictions observed in cluttered environments that limit the user's movements. Moreover, as these additional views are familiar to humans, it is expected that the user could take advantage of the information relatively easily.



(a) No Helper

(b) Top-Down Camera

(c) Mirror

Fig. 6.1. Examples of the virtual helpers provided to assist users during alignment tasks. Apparent correct alignment can be perceived (a) when no virtual helper is present. The use of virtual helpers such as (b) camera views and (c) virtual mirrors provide additional information to identify misalignment errors.

6.2 User Study

A study conducted in a controlled virtual environment helped evaluate the benefits of using the helpers listed in the previous section. Analogous to the study presented in Chapter 3, this virtual environment allowed to design a controlled environment where the geometrical properties of the objects involved in the alignment tasks were known and identical. The concept of evaluating MR concepts using VR environments has been presented in the past by Lee et al. [94]. A study reported in this work showed that VR simulations of AR environments could be a viable alternative to performing controlled experiments.

This design consideration enabled collecting reliable data and diminishing external factors, including system-related errors commonly associated with AR HMDs. Some examples of these factors: are imprecise eye to display calibration, reduced field of view that can lead to split attention problems as the user would not be able to visualize the object of interest and the virtual helper simultaneously, lack of sufficiently accurate interaction methods such as hand gestures and air taps, and inaccurate tracking results or drift effects observed when using commercially available AR OST HMDs.

On Diminishing External Factors

Based on the experience gained from the study presented in Chapter 3, the alignment errors achieved by the users could be smaller than those observed when using commercially available AR OST HMDs.

A preliminary pilot study served to determine the optimal settings for the experiment. All virtual objects were presented at 1.35 meters above the floor. The target objects appeared in the center of the scene, while the interactable objects appeared pseudo-randomly oriented at 0.5 meters to the right from the target. The virtual helpers were positioned at 0.5 meters with an angle of $150°$ between the Interactable-to-Target and Target-to-Helper vectors, as illustrated in Figure 6.2.

The system incorporated direct manipulation techniques for the interaction during task performance, following the results from the studies conducted by Mine et al. [115] that explored how body-relative interactions affect user performance during object docking tasks. During the study, participants aligned textured virtual objects –referred to as *interactable objects*– using the pose defined by their virtual replica –referred to as *target objects*– as reference. To mitigate potential learning effects, a 16 x 16 Latin square matrix defined the order of appearance of the objects and helpers.

6.2.1 Virtual Helpers

To provide additional and meaningful information to the users and overcome the limitations associated with scenarios where physical restrictions in the environment limit the user movements, and considering the helpers described in Section 6.1 the following virtual helpers were evaluated during the user study:

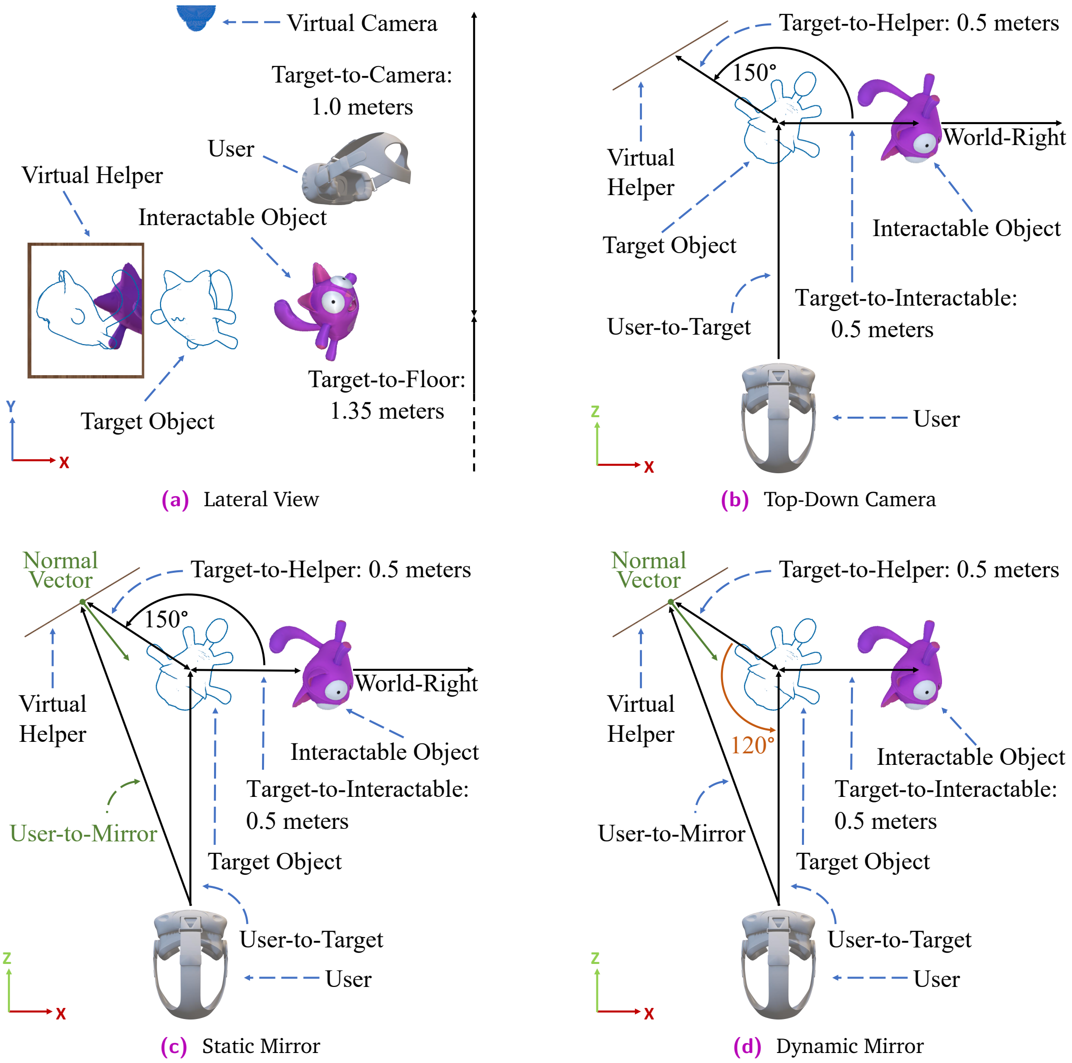


Fig. 6.2. Distribution and placement of the virtual objects and helpers on the scene. The lateral view (a) is the same for all the virtual helpers. The virtual helper is static in position and orientation for the *Top-Down Camera* (b), static in position and dynamic in orientation for the *Static Mirror* (c), and dynamic in position and orientation for the *Dynamic Mirror* (d).

Top-Down Camera. This helper presented an image generated by a virtual camera placed on top of the *target object*. The virtual camera used for the study had a field of view of 30° and was placed 1 meter over the target. The image generated from the virtual camera was presented at 0.5 meters from the target object to mitigate undesired split attention artifacts when the participants tried to visualize the virtual helper and the object to align. The virtual helper was positioned forming an angle of 150° between the World-Right and Target-to-Helper vectors as depicted in Figure 6.2a and Figure 6.2b. Figure 6.1b depicts an example of the image rendered using this virtual helper.

Static Mirror. Conversely to the *Top-Down Camera*, the *Static Mirror* simulates the reflection observed when looking at a real mirror. Thus, any changes in the user's viewpoint will produce changes in the texture observed over the mirror's surface, enabling the acquisition of additional information from the scene using motion parallax. The position

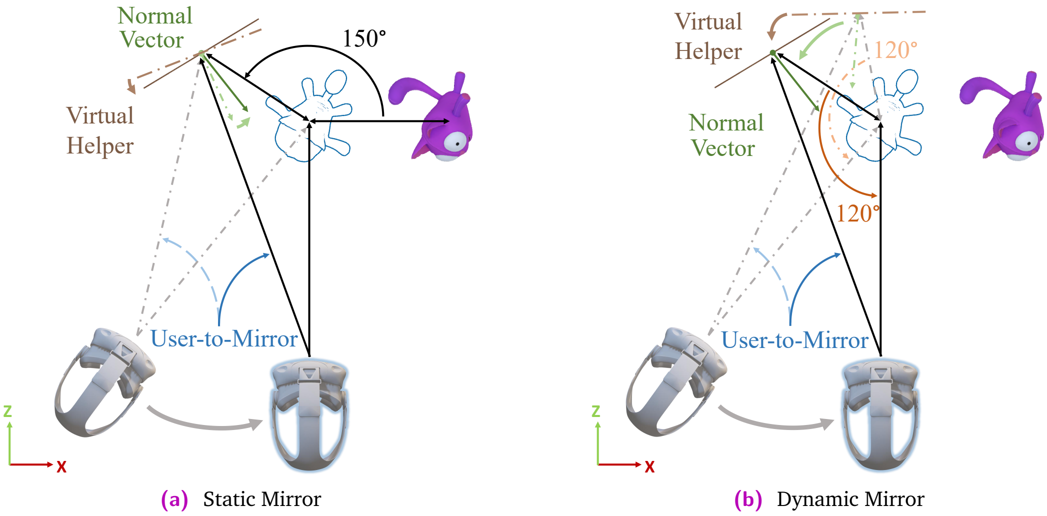


Fig. 6.3. Distribution and placement of the virtual objects and helpers when the user moves during the alignment trial. The *Static Mirror* modifies its orientation to ensure that the target object is always visible (a). At the same time, the *Dynamic Mirror* additionally changes its position to ensure an angle of 120° from the user (b).

of the mirror in the virtual environment matches the one used for the top-down camera view. However, the orientation of the static mirror was updated continuously to ensure the visibility of the target object during task performance. This design consideration aims to observe an image comparable to the one presented by the top-down view. To achieve this, the mirror's normal vector was always facing towards the middle angle defined by the Mirror-to-User and Mirror-to-Target vectors (see Figure 6.2c and Figure 6.3a). Results from the images generated using this virtual helper are shown in Figure 6.1c.

Dynamic Mirror. This mirror differs from the static version because both position and orientation are dynamically updated based on the user's viewpoint. In this regard, the orientation is computed the same way as for the static version. In contrast, its position is always at 0.5 meters from the Target object, forming a 120° angle between the User-to-Target and Target-to-Helper vectors (as depicted in Figure 6.2d and Figure 6.3b).

6.2.2 Objects Representation

To obtain data comparable with existing studies presented in this dissertation and attain a baseline for this comparison, the models used during this study correspond to those introduced in Subsection 3.2.2. These models are shown in Figure 3.2.

Similarly, the visualization technique selected to represent the *target objects* corresponds to the silhouette visualization technique presented in Subsection 3.2.1. This representation reported the best overall scores during the study reported in Section 3.2. This technique highlights the model's borders by generating an outline based on the view direction of the users and the surface normal of the model. As the outline visualization grows the model surface outwards, the *interactable object* can fit inside of its respective *target object* without generating significant occlusion when the objects overlap (see Figure 3.1e and Figure 3.3d).

6.2.3 Participants

Thirty-two unpaid volunteers (9 female and 23 male), aged between 21 and 46 years old (mean age of 26.8 ± 4.6 years), gave their consent to participate in the study.

6.2.4 Experimental Setup

A user study served to evaluate the effects of augmenting the environment with additional viewpoints during alignment tasks. The experience was designed in a virtual environment using the HTC Vive Pro Eye headset and Unity 3D. A pilot study served to define, validate, and refine the initial design considerations of the system proposed, according to the description presented in Subsection 6.2.1. The subsequent user study was conducted in a private office, where external distractions were reduced.

The virtual environment comprised two stages: a tutorial and an evaluation scenario. Data regarding accuracy in position and orientation, time to completion, user interaction, visual fixation, distance traveled, and average head velocity was collected during the evaluation scenario. After completing the experiments, users answered subjective questionnaires for usability and mental effort. The total duration of the experiment was between 25 and 40 minutes.

Interaction Modes

Users had two interaction modes to manipulate the object in the scene. The *normal mode* was used to manipulate the *interactable object* by directly attaching it to a controller. The *precise mode* reduced the translation and orientation changes of the interactable object by one-tenth based on the changes observed in the controller's pose. The latter mode was enabled exclusively when the distance between the objects was smaller than 10 centimeters. Separate buttons were used to activate these modes. In addition, a third button served to confirm the alignment of the objects. After pressing down the confirmation button, any interaction was terminated, and the pose of the interactable object was fixed. Moreover, to avoid registering false positives, the confirmation button must be continuously pressed for two seconds.

Tutorial

The purpose of this stage was to train the participants on how to manipulate virtual objects. In this tutorial, users interacted with the objects and the virtual helpers. The objects used for the tutorial were not part of the formal study. The three virtual helpers presented in Subsection 6.2.1, and a scene without any helper, were introduced in this stage. Participants had the freedom to interact with the objects and all the helpers until they felt comfortable.

Alignment Task

For the formal study, participants had to align as best as possible an interactable object using the pose of a target object as a visual reference. Participants were allowed to take as much time as needed to complete the alignment task. Sixteen different alignment trials composed this stage. This number results from the possible combinations of objects and helpers, including a baseline scenario without helpers. One pair of the interactable and target objects was presented every trial. These objects were replaced with a new pair of objects after the user had confirmed their alignment.

Visual Fixation

The built-in eye-tracking system of the HTC Vive was used to collect information regarding the user's visual fixation during task performance. The three-dimensional points where the users focused while looking at the interactable objects and the virtual helpers were computed and recorded by intersecting the scene geometries with the gaze direction.

System Usability Score and Mental Effort

Participants rated the virtual helpers using the system usability score (SUS) questionnaire [17] and reported the perceived mental effort (ME) [126] after completing all the alignment tasks.

6.2.5 Experimental Variables

Two independent variables: *virtual helper* and *model type* –each with four different levels– were involved were considered in this study. The poses of the controller, user's head, interactable and target objects, and the task duration, visual fixation, and subjective measures were used to compute nine dependent variables. (i) Position errors computed as the Euclidean distance between the gravity centers of the interactable and target objects. (ii) Orientation errors between the target and interactable objects in degrees using an axis-angle representation. (iii) Time to completion calculated as the time elapsed since the first change in the interactable object's pose until alignment confirmation. (iv) User interaction computed as the number of normal and precise manipulations used during the alignment trials. (v) Visual fixation determined as the time spent by the participants looking at the helpers and interactable objects. (vi,vii) Distance traveled and average head velocity using the head pose and time to completion. (viii,ix) Usability and mental effort.

6.2.6 Hypotheses

The general assumption was that the additional views provided by the virtual helpers would improve the user's performance, increase usability, and reduce the mental effort. Especially in scenarios where the environment restricts the user's movements (e.g., presence of obstacles such as furniture or equipment). Therefore, the following hypotheses were proposed:

H1. Users perform significantly better in terms of position and orientation when virtual helpers are present in the scene.

H2. Users complete the alignment task significantly faster when using virtual helpers.

H3. The presence of virtual helpers significantly reduces the distance traveled and the average head velocity of the users.

H4. The presence of virtual helpers does not significantly affect the user interaction during task performance.

H5. The presence of virtual helpers significantly increases usability and reduces the mental effort perceived by users.

6.3 Results

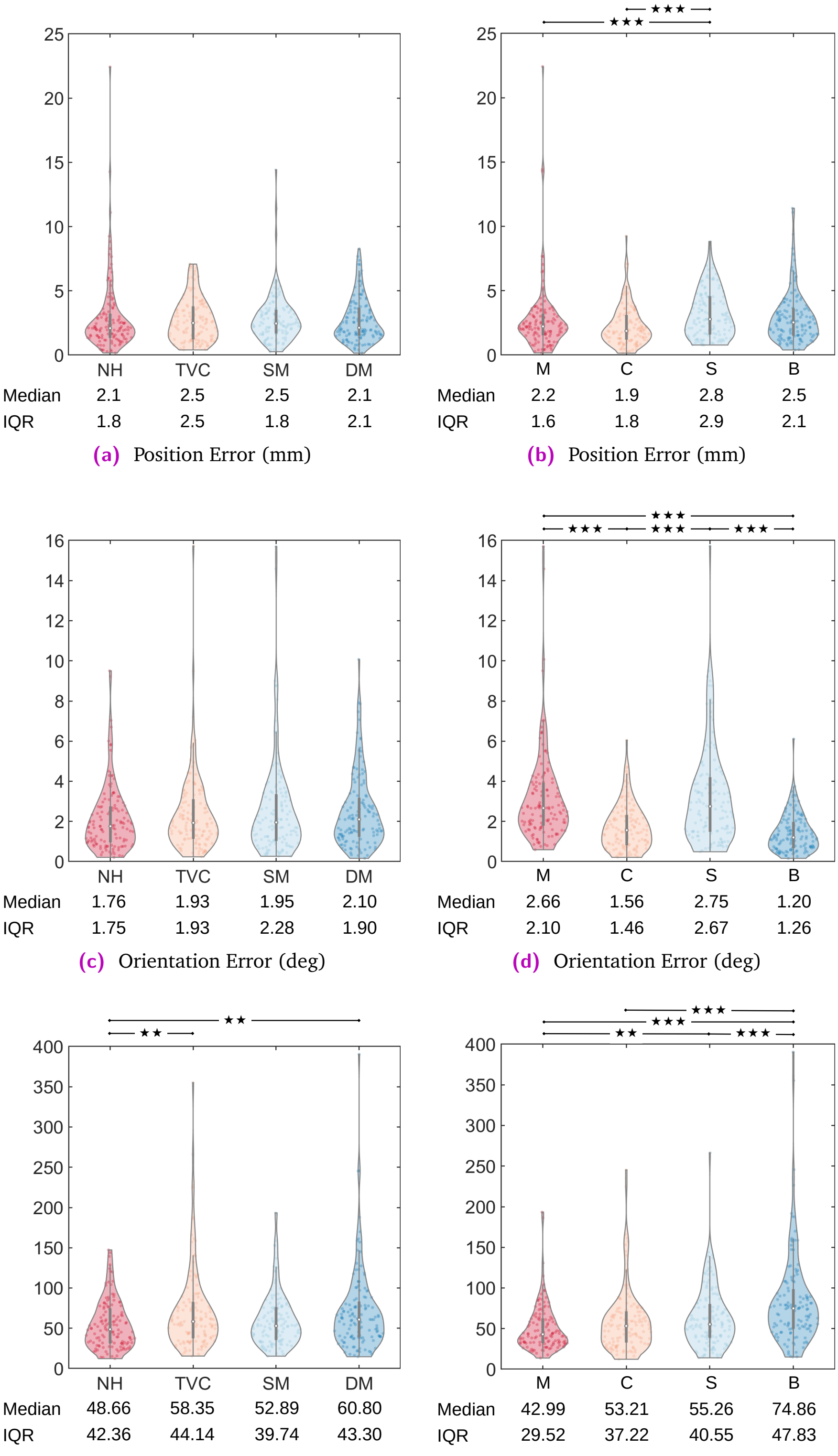
After retrieving the data collected from the study, normality tests revealed a non-normal distribution. In addition, due to the within-subject design of the study, Friedman tests with $\alpha = 0.05$ were used to analyze the experimental variables. Posterior Dunn-Sidák tests were used to reveal statistical significance between the variables. These differences are represented as $\star\,(p < 0.05)$, $\star\star\,(p < 0.01)$, and $\star\star\star\,(p < 0.001)$ in the violin plots [69] (e.g., Figure 6.4). The dots in these plots represent sample points, and the area surrounding them illustrates the data distribution. The white dots in the middle indicate the median, while the bottom and top edges of the gray boxes indicate the 25th and 75th percentiles.

Although this study aimed to explore the feasibility of using the virtual helpers during alignment tasks, rather than understanding how the model selection influences users' performance, this section discusses this information as it may result of special interest to the reader.

6.3.1 Position, Orientation, and Time to Completion

To results collected for alignment accuracy and time to completion were grouped by the type of virtual helper used during alignment. Interestingly, the scenarios without virtual helpers reported the lowest median values for position, orientation, and time to completion. Friedman tests did not reveal statistical significance for position nor orientation. However, statistically significant differences for time to completion ($\chi^2(3) = 16.19$, $p < 0.01$) were found. Participants aligned the objects without using virtual helpers ($Mdn = 48.66sec$, $IQR = 42.36$) significantly faster when compared to the Top-View Camera ($Mdn = 58.35sec$, $IQR = 44.14$, $p < 0.01$) and the Dynamic Mirror ($Mdn = 60.80sec$, $IQR = 43.30$, $p < 0.01$). These results are summarized in Figure 6.4.

On the contrary, after grouping the scores by models, Friedman tests revealed significant interaction for position ($\chi^2(3) = 29.33$, $p < 0.001$), orientation ($\chi^2(3) = 118.04$, $p < 0.001$), and time to completion ($\chi^2(3) = 63.7$, $p < 0.001$). Participants achieved significantly better position scores using the Mug ($Mdn = 2.2mm$, $IQR = 1.6$) than using the Skull ($Mdn = 2.8mm$, $IQR = 2.9$, $p < 0.001$), and when aligning the Camera ($Mdn = 1.9mm$, $IQR = 1.8$) compared to the Skull ($p < 0.001$). Dunn-Sidák tests for orientation error showed strongest statistical significance when aligning both, the Balloon ($Mdn = 1.20°$, $IQR = 1.26$) and the Camera ($Mdn = 1.56°$, $IQR = 1.46$), compared to the Skull ($Mdn = 2.75°$, $IQR = 2.67$, $p < 0.001$) and the Mug ($Mdn = 2.66°$, $IQR = 2.10$, $p < 0.001$). Similarly, users where



	NH	TVC	SM	DM
Median	2.1	2.5	2.5	2.1
IQR	1.8	2.5	1.8	2.1

(a) Position Error (mm)

	M	C	S	B
Median	2.2	1.9	2.8	2.5
IQR	1.6	1.8	2.9	2.1

(b) Position Error (mm)

	NH	TVC	SM	DM
Median	1.76	1.93	1.95	2.10
IQR	1.75	1.93	2.28	1.90

(c) Orientation Error (deg)

	M	C	S	B
Median	2.66	1.56	2.75	1.20
IQR	2.10	1.46	2.67	1.26

(d) Orientation Error (deg)

	NH	TVC	SM	DM
Median	48.66	58.35	52.89	60.80
IQR	42.36	44.14	39.74	43.30

(e) Time to Completion (sec)

	M	C	S	B
Median	42.99	53.21	55.26	74.86
IQR	29.52	37.22	40.55	47.83

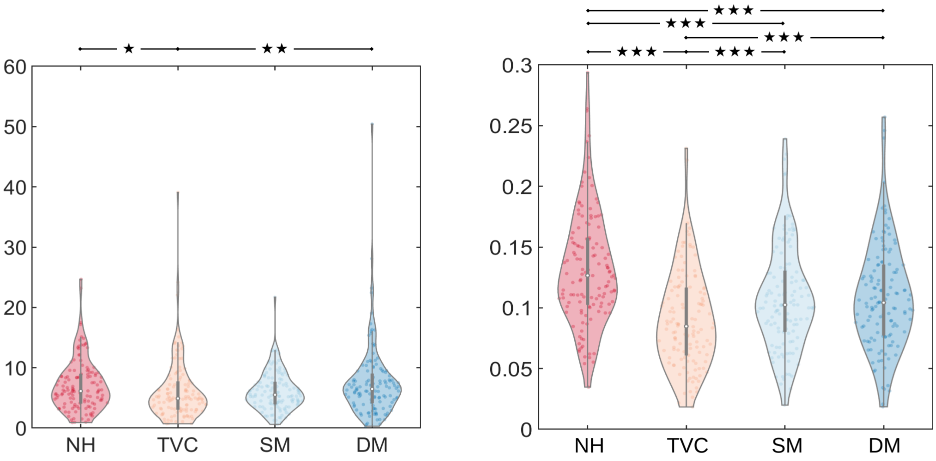
(f) Time to Completion (sec)

Fig. 6.4. Scores obtained by users under test conditions. Results grouped by virtual helpers (*left*): NH-*No Helpers*, TVC-*Top-View Camera*, SM-*Static Mirror*, DM-*Dynamic Mirror*; and model type (*right*): M-*Mug*, C-*Camera*, S-*Skull*, B-*Balloon*.

able to align the Mug ($Mdn = 42.99sec$, $IQR = 29.52$) significantly faster than the Skull ($Mdn = 55.26sec$, $IQR = 40.55$, $p < 0.01$) and the Balloon ($Mdn = 74.86sec$, $IQR = 47.83$, $p < 0.001$); the Camera ($Mdn = 53.21sec$, $IQR = 37.22$) when compared to the Balloon ($p < 0.001$); and the Skull than the Balloon ($p < 0.001$). These results are summarized in Figure 6.4.

6.3.2 Normalized Distance Traveled and Head Velocity

To evaluate if using the virtual helpers contributes to reducing the distance traveled, resulting from using motion parallax to generate additional information, the total distance traveled by the users for every alignment trial was computed. This information was grouped by helper and model type. In this regard, Friedman's results revealed statistical significance ($\chi^2(3) = 15.27$, $p < 0.01$) between virtual helpers. Posterior Dunn-Sidák tests revealed that users significantly traveled lower distances using the Top-View Camera ($Mdn = 4.91m$, $IQR = 4.53$) when compared to the scenarios without virtual helpers ($p < 0.05$, $Mdn = 6.13m$, $IQR = 4.77$) and with the dynamic mirror ($p < 0.01$, $Mdn = 6.47m$, $IQR = 4.75$). These results are shown in Figure 6.5. In terms of model type, the Friedman tests also revealed statistical significance ($\chi^2(3) = 35.99$, $p < 0.001$). The distance traveled while aligning the Mug ($Mdn = 5.20m$, $IQR = 3.92$) and the Camera ($Mdn = 5.01m$, $IQR = 3.67$) was significantly shorter than the one required to align the Balloon ($Mdn = 7.19m$, $IQR = 5.01$, $p < 0.001$ for both), as well as when aligning the Camera when compared to the Skull ($Mdn = 6.12m$, $IQR = 5.12$, $p < 0.05$) and the Balloon ($p < 0.001$). These results are shown in Figure 6.6.



	NH	TVC	SM	DM
Median	6.13	4.91	5.51	6.47
IQR	4.77	4.53	3.57	4.75

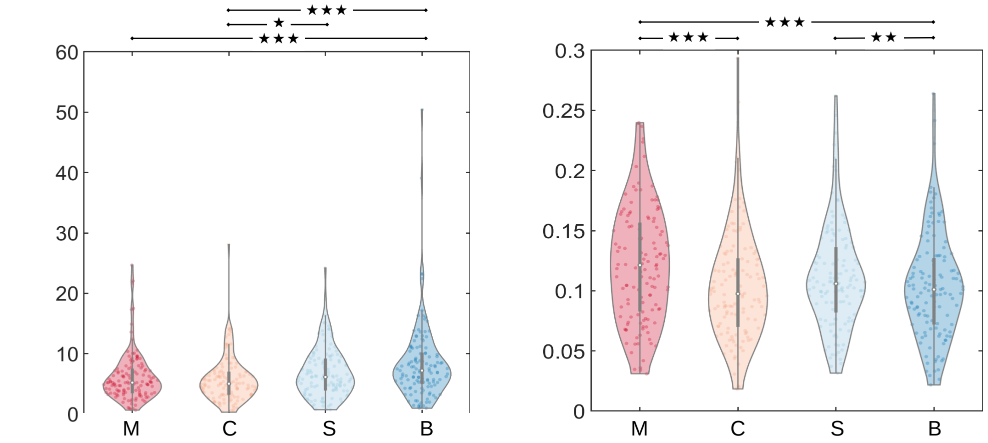
(a) Distance Traveled (m)

	NH	TVC	SM	DM
Median	0.13	0.08	0.10	0.10
IQR	0.05	0.06	0.05	0.06

(b) Head Velocity (m/s)

Fig. 6.5. Scores obtained by users under test conditions. Results grouped by virtual helpers: NH-*No Helpers*, TVC-*Top-View Camera*, SM-*Static Mirror*, DM-*Dynamic Mirror*.

In additionally, the average velocity through computed as the ratio of total distance traveled and trial duration was analyzed. The statistical analysis revealed significant differences between virtual helpers ($\chi^2(3) = 98.73$, $p < 0.001$) presenting a higher velocity the scenarios without virtual helpers ($Mdn = 0.13m/s$, $IQR = 0.05$) when compared to all the virtual helpers ($p < 0.001$). The Top-View Camera ($Mdn = 0.08m/s$, $IQR = 0.06$) registered the slowest velocity when compared to both proposed mirrors ($p < 0.001$). These results are shown in Figure 6.5. On the other hand, statistical significance was found between models ($\chi^2(3) = 37.3$, $p < 0.001$) presenting the Camera ($Mdn = 0.10m/s$, $IQR = 0.06$) and Balloon ($Mdn = 0.10m/s$, $IQR = 0.05$) significantly slower velocities than the Mug ($Mdn = 0.12m/s$, $IQR = 0.07$, $p < 0.001$), and the Balloon compared to the Skull ($Mdn = 0.11m/s$, $IQR = 0.05$, $p < 0.01$) (see Figure 6.6).



	M	C	S	B
Median	5.20	5.01	6.12	7.19
IQR	3.92	3.67	5.12	5.01

(a) Distance Traveled (m)

	M	C	S	B
Median	0.12	0.10	0.11	0.10
IQR	0.07	0.06	0.05	0.05

(b) Head Velocity (m/s)

Fig. 6.6. Scores obtained by users under test conditions. Results grouped by model type: M-*Mug*, C-*Camera*, S-*Skull*, B-*Balloon*.

6.3.3 User Interaction

The number of normal and precise manipulation events observed across every alignment trial served to investigate if the model or virtual helper influenced how users interacted with the virtual objects. In this regard, a manipulation event corresponds to the press-release action over any of the two interaction buttons used to modify the pose of the interactable object. These manipulation events do not consider the duration of the action. Figure 6.7 depicts an example of these manipulation events where four precise manipulation events are identified.

After collecting the normal and precise manipulation events, these were grouped by model and virtual helper. Friedman tests did not reveal statistical significance for the manipulation events when grouped by virtual helpers (see Figure 6.8). However, statistical significance was found when the interaction events were grouped by model for normal ($\chi^2(3) = 27.71$, $p < .001$) and precise ($\chi^2(3) = 29.52$, $p < .001$) modes (see Figure 6.9). Users required a significantly higher amount of normal manipulation events during the alignment of the Balloon ($Mdn = 6.0$, $IQR = 7.0$) when compared to the Mug ($Mdn = 5.0$, $IQR = 4.0$, $p < 0.001$), the Camera ($Mdn = 5.0$, $IQR = 5.0$, $p < 0.001$), and the Skull ($Mdn = 5.0$, $IQR = 5.0$, $p < 0.001$), and a significantly higher number of precise manipulations to align the Balloon ($Mdn = 15.0$, $IQR = 20.0$) when compared to the Mug ($Mdn = 8.0$, $IQR = 11.5$, $p < 0.001$), the Camera ($Mdn = 10.5$, $IQR = 16.0$, $p < 0.05$), and the Skull ($Mdn = 11.0$, $IQR = 18.5$, $p < 0.05$).

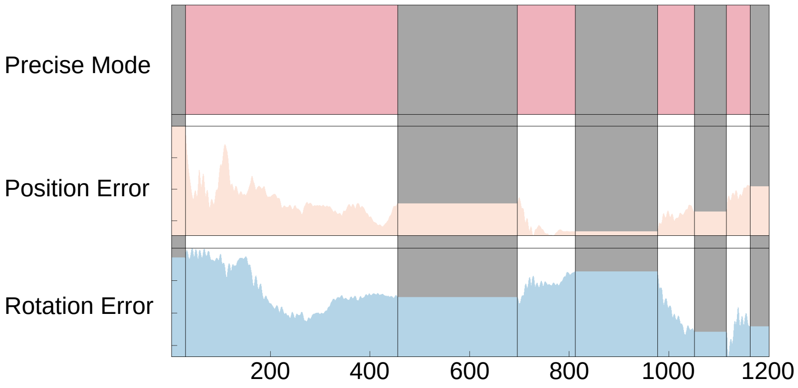
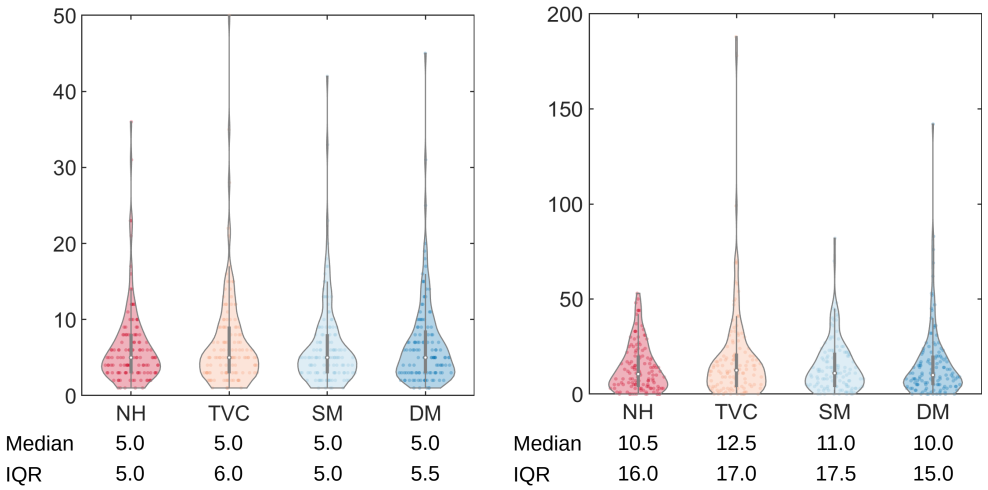


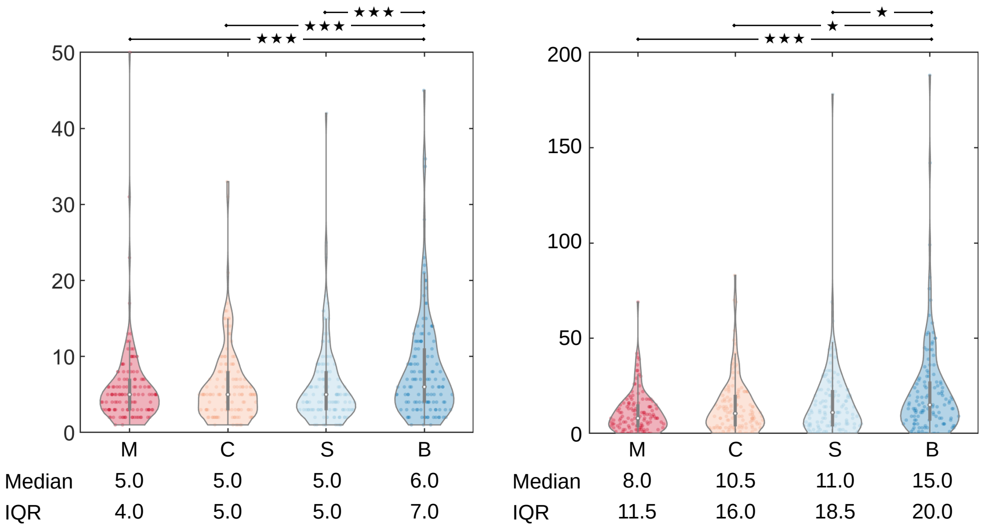
Fig. 6.7. Data sequence corresponding to *User2* while aligning the *Camera model* using the *Dynamic Mirror*. This sequence illustrates 1200 samples collected corresponding to: a) Precise mode manipulation event (*top*). b) Euclidean distance from target position (*middle*). c) Angular error from desired orientation (*bottom*). The gray shaded areas indicate no interaction; thus, the respective errors remain unchanged.



(a) Normal Mode **(b)** Precise Mode

Fig. 6.8. Normal and precise manipulation events grouped by virtual helper. NH-*No Helpers*, TVC-*Top-View Camera*, SM-*Static Mirror*, DM-*Dynamic Mirror*.

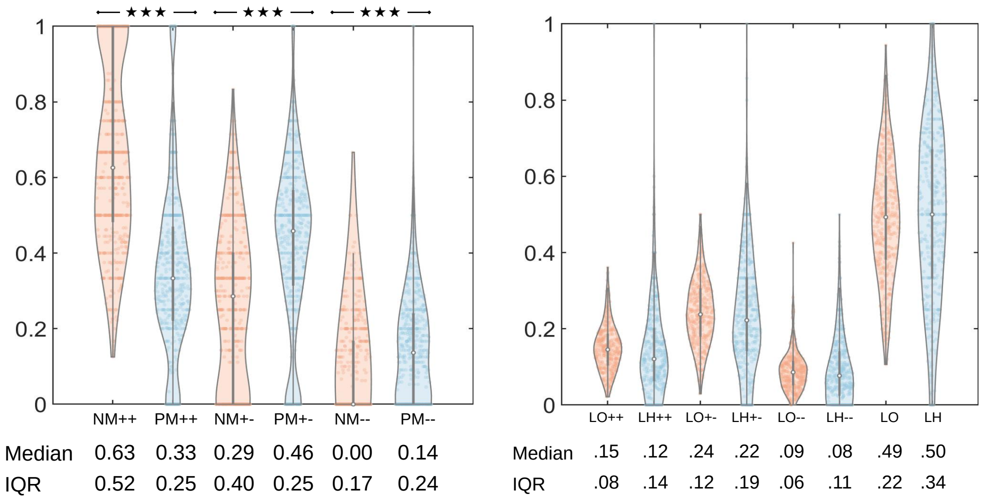
In addition, the manipulation events for every trial were classified using three categories to evaluate if the use of the interaction modes supported the users to improve the alignment. Therefore, any normal manipulation event –*NM*– leading to improved position and orientation was classified as *NM++*. If the manipulation event resulted in an improvement in position but negatively affected the orientation or vice versa, it was labeled as *NM+-*. Finally, if the interaction event negatively affected both position and orientation, it was classified as *NM–*. In addition, the individual scores for these subsets were divided by the total amount of normal manipulation events, helping normalize the data sample based on the specific alignment trial. The same classification procedure was used to compute the scores for the precise mode. An example of this classification is presented in Figure 6.7. In this example, the first manipulation event classifies as a PM++, the second and third as PM+-, and the fourth as PM–.



(a) Normal Mode (b) Precise Mode

Fig. 6.9. Normal and precise manipulation events grouped by model type: M-*Mug*, C-*Camera*, S-*Skull*, B-*Balloon*.

Friedman tests revealed significant differences between the improvement observed using the interaction modes ($\chi^2(5) = 1196.97$, $p < .001$). Users significantly improved the position and orientation of the objects more frequently using the normal mode ($Mdn = 0.63$, $IQR = 0.52$) than the precise mode ($Mdn = 0.33$, $IQR = 0.25$, $p < 0.001$). In addition, users worsen the position and orientation of the objects significantly less frequently using the normal mode ($Mdn = 0.00$, $IQR = 0.17$) than the precise mode ($Mdn = 0.14$, $IQR = 0.24$, $p < 0.001$). Although statistical significance between all six groups depicted in Figure 6.10a was found, only those corresponding to the same type of improvement classification are shown.



(a) Interaction Modes (b) Visual Fixation

Fig. 6.10. Overall improvement achieved by users during alignment. (a) NM-*Normal Mode*, PM-*Precise Mode*, '++'-*Improved translation and rotation*, '+-'-*Improved translation or rotation*, '–'-*Worsen translation and rotation*. (b) LO-*User looking at interactable object*, LH-*User looking at virtual helper*.

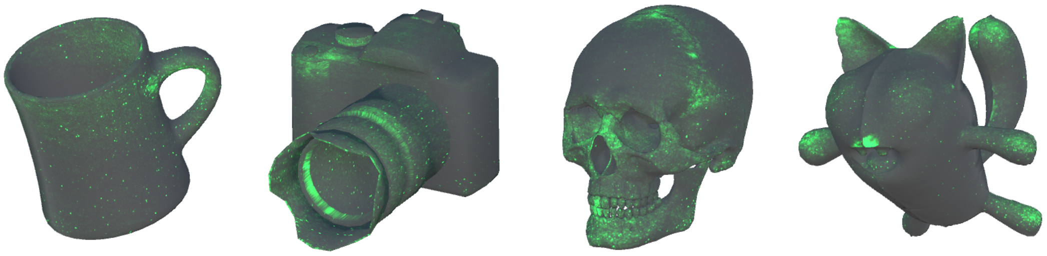


Fig. 6.11. Model-based UV maps generated from users' attention during alignment tasks. This data is retrieved from eye-tracking values collected during task performance.

6.3.4 Visual Fixation

The eye-tracking data collected during the study served to identify the time frame during which the users looked at the objects and virtual helpers, supporting investigating how the latter may influence the user's attention. Like the interaction modes, the alignment errors of the interactable objects were analyzed to examine if alignment improvement was observed during this action. These results were classified using the three categories used for interactions (i.e., "++," "+-," and "–"). In addition, a fourth class indicated if the users were looking at the object (LO) or the helper (LH) without manipulation. Friedman's test did not reveal statistical significance between any of these classes when compared against its simile.

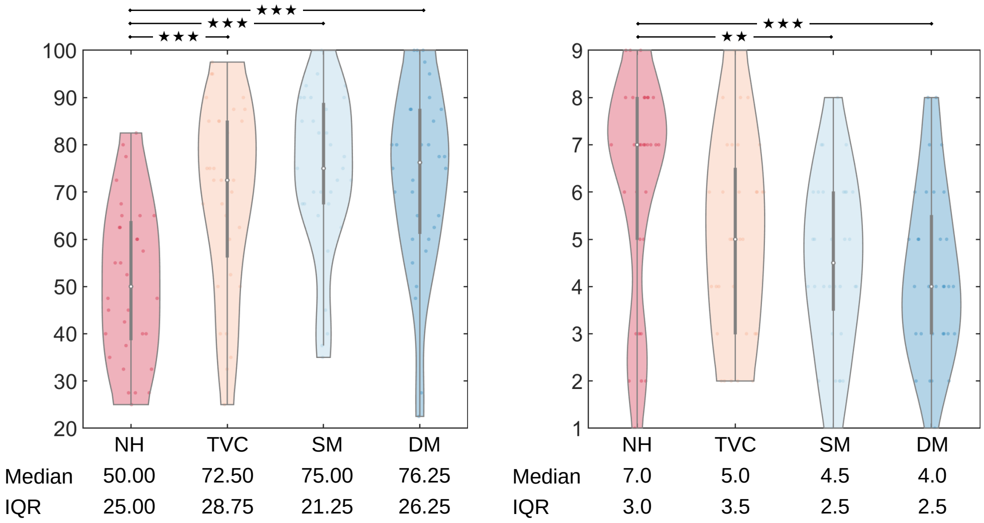
Additionally, the data collected from the eye-tracking helped to identify the structures of the interactable objects where the users focus their attention during the alignment. The three-dimensional points of the object's surface where the users focused their attention during alignment were collected and used to visualize this data. These points were later mapped to their respective UV Map and used to generate a heat map. A bi-dimensional normal distribution centered in the points collected was used to increase the visibility of the samples and mitigate possible inaccuracies in the values acquired from the tracking system. The cumulative UV Maps generated from this analysis are shown in Figure 6.11.

6.3.5 Usability and Mental Effort

Results collected for usability (SUS) and mental effort (ME) perceived by users show that users prefer the environments where virtual helpers are present. As depicted in Figure 6.12, the Dynamic Mirror obtained the best results for both SUS and ME, followed by the Static Mirror and the Top-View Camera. The three virtual helpers obtained SUS scores considered above the average, while the scenario without helpers scored below the average[1]. Friedman results revealed significant differences between virtual helpers for SUS ($\chi^2(3) = 35.16$, $p < .001$). A Dunn-Sidák post hoc test revealed that SUS scores were significantly higher for the Top-Camera View ($Mdn = 72.50$, $IQR = 28.75$), Static Mirror ($Mdn = 75.00$, $IQR = 21.25$), and Dynamic Mirror ($Mdn = 76.25$, $IQR = 26.25$) compared to the scenes without virtual helpers ($Mdn = 50.00$, $IQR = 25.00$, $p < 0.001$). Similar to the SUS scores, Dunn-Sidák results for ME revealed statistical significance between virtual helpers ($\chi^2(3) = 16.97$, $p < .001$). A post

[1] Scores between 68 and 80.3 are considered above average, while scores between 51 and 68 are considered average.

hoc test revealed that the mental effort using the Static ($Mdn = 4.5$, $IQR = 2.5$, $p < 0.01$), and Dynamic ($Mdn = 4.0$, $IQR = 2.5$, $p < 0.001$) mirrors was significantly lower when compared to the scenario without helpers ($Mdn = 7.0$, $IQR = 3.0$).



(a) System Usability Scale

(b) Mental Effort

Fig. 6.12. Scores for SUS and ME reported by users. NH-*No Helpers*, TVC-*Top-View Camera*, SM-*Static Mirror*, DM-*Dynamic Mirror*.

6.4 Discussion

Overall, results from the position showed average total errors no larger than 2.5 mm for all the scenarios. Nevertheless, the Top-View Camera and Dynamic Mirror showed the most consistent values. Users were able to achieve alignment errors under 10 mm for all the alignment trials using these helpers. In addition, participants registered overall orientation scores with median error values under $2.5°$ and $10°$ for most of the individual trials. Even more, results from the study revealed statistical significance for time to completion scores. The scenarios without virtual helpers registered lower values than the Top-View Camera and Dynamic Mirror. Comparing the results with those reported in Section 3.1, this study showed a similarly strong influence of the model type on the values for position, orientation, and time to completion. Nevertheless, it is worth noting that the values for time to completion reported in this study are lower than those observed in Section 3.1.

Although it was expected that users would achieve better alignment **–H1–** significantly faster when using the virtual helpers **–H2–**, the results of the study did not show such a trend. However, these results could be attributed to several factors. In this regard, according to Buhrmann [19], the performance of complex tasks requires the coordination of multiple sensorimotor patterns. Such patterns involve actions and perceptual systems that are enhanced based on the user's experience. As a result of this, the lack of familiarity using the additional views for alignment tasks may have influenced task performance. This hypothesis could also explain why the time to completion without helpers and with the static mirror are comparable, as humans are more or less accustomed to their daily use.

Another factor to consider is the divided attention nature of the task, as described by Wickens et al. [177]. This concept explains our limited ability to perform multiple tasks simultaneously. In this respect, while an optimal scenario would require users to process visual input from the direct and alternative viewpoints in parallel, physiological aspects such as the size of the visual area that the human eye can cover limit our ability to achieve this.

Concerning the distance traveled, the Top-View Camera achieved comparable scores to the Static Mirror but significantly lower values than the Dynamic Mirror and the scenes without virtual helpers. These results suggest that users might benefit from camera views or static mirrors during alignment tasks in environments with physical constraints. The increased distance traveled in scenarios without helpers may be caused by induced motion parallax necessary to generate meaningful information for alignment. Conversely, the increase in scores achieved with the Dynamic Mirror may indicate that users were moving around the object to generate multiple dynamic viewpoints. The results from analyzing the average head velocity illustrate the effects observed from the distance traveled. The Top-View Camera showed lower velocity values than all other scenarios, while the no-helper scenario showed significantly higher values than the mirrors. These results support **–H3–**, hence the recommendation to use virtual cameras and static mirrors for scenarios with space constraints.

In terms of interaction, results from this study suggest that the geometry of the object to align plays a stronger role in the number of normal or precise manipulation events required to achieve alignment. On the contrary, virtual helpers do not have a significant impact on this metric, supporting **–H4–**. Results from the study also suggest that users are more likely to simultaneously improve the position and orientation of the interactable object when using normal mode. However, this may result from the precise mode being activated only when the virtual objects are closer than 10 centimeters. In addition, the statistical results inferred from the user's visual fixation showed comparable performance improvement when looking at the interactable object or the virtual helpers.

Lastly, this study revealed significantly higher scores for usability and lower scores for mental effort when using virtual helpers, supporting **–H5–**. Interestingly, whereas no statistical significance was found for the alignment scores, and although users aligned objects significantly faster without the helpers, the SUS and ME scores reflect a clear preference for the presence and use of the virtual helpers. In addition, an interesting phenomenon is observed for the reported ME scores without helpers. These results show a clearly divided opinion regarding the mental effort perceived.

On Visual Fixation and the Object's Geometry

Based on the observation of the heat maps created using the visual fixation data, it appears that users focus their attention on surface areas with strong curvatures, such as the camera's lens and shutter button; the nose, jaw, and bones near the eyes of the skull; or the balloon's mouth. These results could suggest that the distinctive geometric structures of the objects are used during the alignment process.

6.5 Study Implications

The results derived from this study motivated the idea of evaluating such visualization techniques using real mirrors in MR environments. A literature review on this topic showed no evidence of MR mirrors capable of dynamically reflecting the real and virtual content of an MR environment. Therefore, influencing the design of the Augmented Mirrors introduced in the following section.

7 Reflection

> *The importance of mirror-reflection symmetry to our perception and aesthetic appreciation, to the mathematical theory of symmetries, to the laws of physics, and to science in general, cannot be overemphasized...*
>
> — **Mario Livio**
> (The Equation That Couldn't Be Solved: How Mathematical Genius Discovered the Language of Symmetry)

As discussed in the previous section, providing alternative viewpoints to the users of MR applications can be particularly useful during alignment tasks in MR environments. This hypothesis inspired the idea of introducing physical mirrors into MR environments to better understand their potential use during alignment tasks. The concept of replicating the physical properties of the mirrors in mixed and virtual reality applications has been a topic of interest during the past decades, leading to the conception of multiple classes of virtual mirrors. These classes include the creation of virtual mirrors that replicate the properties of their real counterparts [7, 118], the use of external cameras and screens [12, 33, 39, 92, 102, 111, 168], or the integration of half-silvered mirrors and displays [81, 127, 144]. Alternative classes have used pre-acquired pictures of the scene to generate multiple viewpoints [160] later modified to replicate the image observed from a mirror [42].

This section introduces a new class of mirrors for their use in MR applications: the Augmented Mirrors (AM). This new class of mirrors provides a simple yet effective manner to reflect the real and virtual content of an MR application, offering interactive and dynamic viewpoints of the scene while preserving the perceptual benefits of using mirrors and facilitating the exploration and understanding of the environment.

The AMs only require the user and mirror's poses to generate spatially consistent augmentations observed over the mirror's surface. This feature provides the flexibility to implement this concept regardless of the method used to estimate these poses or the display device to visualize the MR scene. In addition, as the AMs obey the same optical laws as real mirrors, all visual cues provided by the latter apply to the former.

In this regard, the AMs provide additional perspectives when the subject or the object changes their position or orientation in the environment, allowing to deduce valuable spatial information. This new perspective provides valuable depth information as the three-dimensional composition of the environment can be better understood by using two views. Furthermore, using motion parallax when interacting with the mirror provides stronger depth information. In addition, the combination of the direct and mirrored viewpoints can help resolve ambiguous information frequently observed in MR, as discussed in Chapter 3. These features become even more important as interposition and motion parallax are considered two of the most important depth cues in personal space [24]. Lastly, the information presented by the mirror provides proprioception input that can be used for the decision-making process for motor movements [143].

Mirrors in Mind

For interesting information related to mirrors and how they have influenced mankind, society, arts, and science, it is highly recommended to read the book "Mirrors in mind" by Richard Gregory (1998) [57].

7.1 A Brief History of Mirrors

It is an accepted belief that some of the first manufactured mirrors date back from prehistoric times. The first of them, made of polished obsidian, were recovered from graves in Anatolia and dated approximately 6000 to 5900 BC [36]. Other mirrors made of polished copper have been discovered in Mesopotamia and ancient Egypt between 4000 and 3000 BC. It is believed that Egyptians may have used these objects to direct the sunlight into dark places and may even be considered sacred [57].

Other cultures, such as the Chinese civilization, attributed mythical and magical powers to these objects and even understood their optics before the Greeks around the fourth century BC. According to the *Mo Ching*, a technical document from the fourth century BC, the Chinese described the optics of plane, concave, and convex mirrors. This knowledge enabled them to use concave mirrors as searchlights to illuminate distant places, make fires, or even capture the Moon's light [57]. These mirrors served as inspiration in the design of Japanese mirrors. Greeks' mirrors, also dating from the fourth century BC, seem to be influenced by the Egyptian designs and motivated more poetic, philosophical, and mythological questions and myths such as those of Narcissus or Perseus.

A common denominator observed along with the different civilizations is the mythical powers associated with these objects. In this regard, Mesoamerican religions believed that Tezcatlipoca, one of their deities, drew his powers from an obsidian mirror. Similar properties have been attributed to mirrors in medieval Europe as they were associated with witchcraft. However, the introduction of the scientific method has helped to detach their physical properties from the mythical powers associated with them.

7.2 Benefits and Challenges of Using Mirrors

Contrary to the intuition that the reversal of axes observed when using mirrors may confuse the users, existing works suggest that users adapt to these images after an adaptation period. The degree to which users benefit from additional viewpoints is related to the users' familiarity with the systems. This statement can be explained using the theory of sensorimotor contingencies that describes the sensitivity for the link between action and its consequences. According to Buhrmann [19], performing complex tasks requires multiple sensorimotor coordination patterns that involve different actions and perceptual systems. Such patterns, evaluated based on experience as preferred for achieving a specific task, are regularly executed by every individual.

An example of this is the work presented by Dunnican et al. [30] for training in laparoscopic scenarios that present the problem of reverse alignment visualization (*mirror image*). This problem occurs when the camera ends up facing the surgeon, leading to the visualization of the tools' movements in reverse order, hindering the performance of the task. In this context, Dunnican et al. suggested intentionally training a subset of users adopting a reverse alignment configuration to overcome this problem. This study showed that users without this training improved their performance only when a regular view was shown. On the other hand, the users trained with the mirrored view improved their performance in both situations, the regular and the mirrored. This phenomenon becomes more evident in other medical fields, such as dentistry. In this field, mirrors represent an essential tool as they enable the exploration of reduced environments, the visualization of content that otherwise would be hidden to the observer, and even facilitate the illumination of the mouth by reflecting the exterior light.

Another advantage derived from using mirrors is that they provide proprioception input beneficial for the decision-making process for motor movements. In this regard, while humans use the visual information to plan the trajectory and kinematics involved in reaching movements, the proprioception is critical in transforming this plan into the motor commands sent to the muscles of the arm [143].

The fact that users can get used to mirrored images, in addition to their perceptual benefits presented in Chapter 6, leads to the hypothesis that *Augmented Mirrors* can support users performing exploration, scene understanding, and alignment tasks in MR. Moreover, considering that the *Augmented Mirrors* represent an extension of the world observed when using mirrors, by the addition of the virtual content, it is expected that users will easily and naturally transfer their knowledge to interact with this new class of MR mirrors.

Mirror Reversal

A common misconception about mirrors is the belief that a left-right reversal is inherent to the image of a flat mirror. However, this reversal is a front-back effect "...*caused by the light rays going forward toward the mirror and then reflecting back from it*" [151].

A Detailed Explanation on Mirror Reversal

A very descriptive and highly recommended explanation of this problem can be found in "Chapter 4: Puzzles of Images" of the book *Mirrors in Mind* by Richard Gregory [57].

7.3 A Taxonomy of Mirrors in Mixed Reality

Several works have implemented multiple architectures that reproduce the physical properties of physical mirrors for their integration into MR environments to a greater or lesser degree. These works can be distinguished based on the architecture used to reproduce the mirror's properties and depending on the functionality of the proposed methods. In this regard, Portales et al. [130] have proposed a useful classification that discusses the integration of the mirror paradigm in these environments. Although this classification presents an extensive examination of this paradigm in MR, it mainly focuses on architectures composed of cameras and displays or half-silvered mirrors, excluding less conventional approaches that have been proposed more recently. Therefore, this section aims at presenting an extended taxonomy of the mirror paradigm in MR environments.

7.3.1 Cameras and Displays

One of the initial efforts towards integrating the mirror paradigm into MR applications was the *Magic Mirror*. Introduced by Maes et al. [102], it used a video camera facing towards the user to acquire a video sequence that was horizontally flipped, enriched with virtual content, and displayed on a large screen presented in front of the users. In addition, this system used body tracking algorithms to enable users to interact with the virtual content. This type of architecture has been used for multiple applications, including entertainment [39, 168], marketing [33], anatomical education [12], and visualization of medical data such as brain signals in real-time [111].

Although these approaches provide a mirror-like view of the scene, one limitation of these systems is that any changes in the users' viewpoint will not produce perspective changes on the image observed. This is due to the position and orientation of the camera used to capture the scene is typically fixed. Therefore, limiting the possibility of changing viewpoints and failing to replicate the properties of the physical mirrors realistically.

7.3.2 Half-Silvered Mirrors and Displays

An alternative class of AR mirrors combines semi-transparent mirrors normally attached in front of a computer display. This class of mirrors has been used by Pardhy et al. [127] to augment the side and rearview mirrors of vehicles with virtual lane boundaries and road information using geospatial data. Sato et al. [144] used this class of mirrors, together with an array of video cameras and marker-based body tracking, to present interactive AR objects observed over the reflected image. Further work presented by Jang et al. [81] replaced the markers used in [144] with a body-tracking algorithm using depth cameras. More recently, Lee et al. [95] introduced a system that integrated three-dimensional displays to replace regular displays. Results from implementing this system suggest that this type of display can be used to improve depth perception.

One of the advantages of this class of mirrors is that they can reflect the virtual and real content of the MR environment and provide the additional benefits of using real mirrors.

However, the virtual content observed over the mirror's surface is perspectively correct only to the user facing the display. In addition, the virtual objects can be seen only as a reflection over the mirror's surface, but not in the observer's direct view.

7.3.3 Virtual Mirrors

Another class of MR includes those approaches that replicate the mirror paradigm using fully virtual content. An example of these is the *virtual mirror* paradigm introduced by Navab et al. [118]. This type of mirror can reflect the virtual content of an MR scene without the need to use physical mirrors. This paradigm was used as an interactive visualization tool to improve depth perception in monocular systems. Further developments of this concept, later extended by Bichlmeier et al. [7]. Results from these studies have shown that this class of mirrors facilitates the visualization of physically restricted areas, the improvement of navigation tasks, the understanding of complex models, and improved depth perception. However, a noticeable limitation observed when using these mirrors is that they are not capable of reflecting the real objects over the surface of the virtual mirror.

7.3.4 Reflective Displays

More recently, Fotouhi et al. [42] introduced the *Reflective-AR displays*. This concept can acquire pre-acquired images from multiple viewpoints of interest in a real scene. Such images are mirrored over the horizontal axis and used to overlay any virtual objects observed in an MR scene from their respective viewpoints, allowing for the simultaneous visualization of the multiple viewpoints.

The generated displays can effectively reflect any changes in the virtual content while showing the real scene. However, because it uses pictures of the viewpoints of interest to generate the reflective displays, this approach requires the real scene to be static. In addition, as it does not use any reflective surface, any changes in the observer's viewpoint will not produce changes in the reflection of the environment.

7.3.5 Virtual Reflections over Static Mirrors

A conceptual idea of presenting virtual content over the surface of a mirror was suggested by Zimmer et al. [182]. This concept introduced the idea of increasing immersion in MR applications by reflecting virtual content over the surface of a static mirror using HMDs. In this regard, the position of a physical mirror would be used to define the boundaries of a virtual mirror surface. This surface would later serve to render the reflections of the virtual world and be employed to explore user interaction techniques.

Using this class of mirrors would enable the user to change their viewpoint to explore the scene and gain additional information from the scene. However, the implementation details provided by the authors suggest that the pose of the real mirror cannot be changed after the mirror's pose has been calibrated. Therefore, limiting the interaction between the user and the mirror.

7.3.6 Augmented Mirrors

As previously discussed in this section, the Augmented Mirrors represent a new class of MR mirrors. This class of mirrors allows for the interactive, dynamic, and simultaneous visualization of the real and virtual content of an MR environment using the surface of a real mirror. This concept is simple yet effective as it only requires tracking the poses of the observer and the mirror. Due to its simplicity, an AM can be implemented regardless of the tracking technology available or the display device used to deliver the MR experience. This flexibility facilitates its integration in MR environments requiring alignment, exploration, spatial understanding, or selective contextual visualization of real and virtual content.

7.4 Methods

As mentioned in Subsection 7.3.6, to generate a virtual reflection that would enable the generation of an *Augmented Mirror*, two fundamental poses are required to be known: the viewpoint of the observer and the mirror. If the poses are known with sufficient accuracy, generating an accurate virtual reflection can be produced and overlaid onto the plane of the physical mirror. In addition to these two poses, any other object of interest can be integrated into the MR environment by simply estimating its pose. The simplicity of this concept allows the observer to use any MR-enable device, including cameras and monitors, handheld devices, or HMDs. Even more, any preferred tracking technology can be used to determine the poses of the mirror and the observer. Therefore, leaving the selection of these two components, visualization and tracking, open to the requirements that a specific application would demand.

Once the poses of the basic components of the *Augmented Mirror* are known, it is possible to estimate the pose of a virtual camera that generates a virtual image comparable to the one created by a real mirror (Figure 7.1). This camera enables the creation of the virtual reflection that is later overlaid over the surface of the physical mirror.

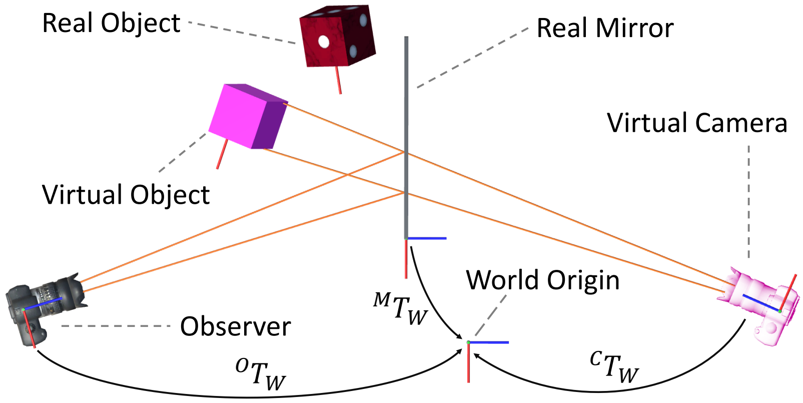


Fig. 7.1. The *Augmented Mirror* concept. The poses of the observer O, physical mirror M, and virtual object are assumed to be known relative to the world coordinate system W. The pose of the virtual camera C is computed from the observer's pose and the mirror to generate the virtual reflections.

Suppose the normal to the physical mirror is defined as the vector $(0, 0, 1)$. In that case, the virtual camera's pose in mirror coordinates is equal to the multiplication of its original pose in the local mirror coordinate system with the matrix:

$$M_{reflect} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \text{-1} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

This transformation corresponds to scaling the z-direction by -1 as the reversal in a mirror corresponds to a front-back reversal that results from the light rays being reversed on the mirror surface and reflecting back from it [151].

To find the world pose of the virtual camera C, it is necessary to first transform the observer O from world into the mirror local coordinates M. A follow step performs the image mirroring, and transforms back the virtual camera pose from mirror into world coordinates, formalized as[1]:

$$C =^M T_W \cdot M_{reflect} \cdot^W T_M \cdot O$$

Once these transformations are known, the virtual reflection can be rendered over the surface of the real mirror as follows:

1. Duplicate the observer's camera as a virtual camera.
2. Place the virtual camera in the position of the virtual observer.
3. Set up the mirror plane as the near plane of the virtual camera to avoid rendering any virtual content placed behind the mirror plane.
4. Render the scene from the virtual camera's viewpoint into a texture.
5. Apply the texture to the real mirror in screen space when rendering the scene using the main camera.

The transformations required for the *Augmented Mirror* following this approach are shown in Figure 7.2. In addition, the first implementation of this concept is depicted in Figure 7.3.

Special attention must be taken when using an off-axis projection matrix or rendering stereoscopic viewpoints, as in the case of HMDs. In addition, the texture used to represent the virtual mirror should be excluded from the rendering when using the mirror camera. Failing to do this may result in only visualizing the back of the virtual mirror. Furthermore, to present more advanced visualizations, a set of culling masks can facilitate selecting the components in the scene that will be rendered in both the direct view and the reflection of the *Augmented Mirror*. Further details about this topic are presented in Section 7.5.

[1] ${}^{A}T_{B}$ describes a transformation from the coordinate system A to B.

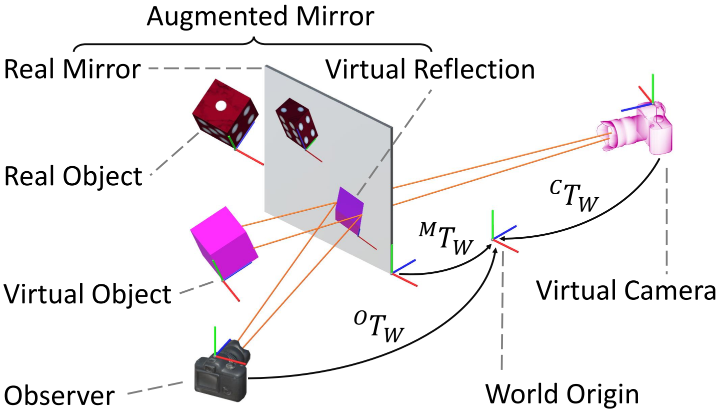


Fig. 7.2. Theoretical representation of real and virtual objects reflected over the surface of an *Augmented Mirror*.

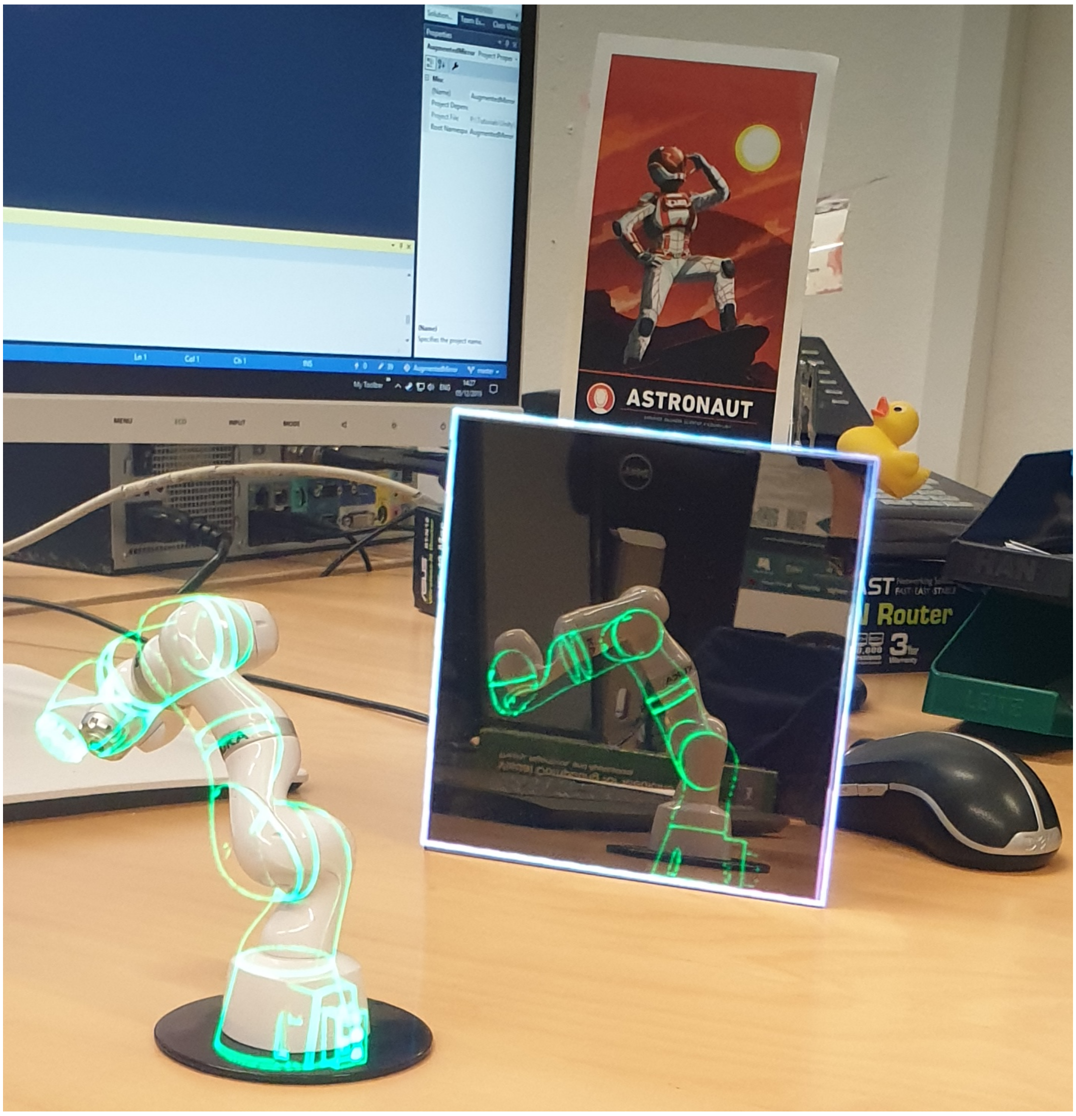


Fig. 7.3. Early implementation of an Augmented Mirror using Vuforia optical markers and an OST HMD.

7.4.1 Theoretical Influence of Tracking Errors

A very important thing to consider when using Augmented Mirrors is the tracking technology used to estimate the poses of the observer and the physical mirror. As previously discussed in Subsection 7.3.6, any tracking technology can be used to estimate these poses. However, every tracking technologies exhibit different fidelity and tracking accuracy.

In this regard, it is possible to imagine a scenario in which the system accurately tracks the observer's viewpoint and mirror pose. In addition, there is a real object of interest from which its virtual reflection is of interest for the application. Then, if the estimated pose of this object is inaccurate, the virtual reflection observed over the surface of the real mirror will correspond to the pose estimated by the tracking system, leading to the observation of visual discrepancy between the real and virtual views in the reflection.

A similar situation occurs when the spatial relationship between the mirror and observer's poses is estimated inaccurately. In this specific case, the incident rays that collide with the physical mirror will not be equivalent to the rays estimated for the virtual mirror. Hence, the view provided by the real mirror will not correspond to the image rendered by the *Augmented Mirror*. Although this problem is not notorious when the reflected objects are closer to the mirror, this problem becomes more evident when this distance increases due to the longer lever arms.

This section illustrates the theoretical influence of error propagation derived from tracking errors in the translation and rotation of the mirror's pose. The perceived error Δx_t introduced by a translation z of a reflected object in the mirror and the perceived error Δx_r due to the error in rotation β both depend on: i) the distance between the observer and the mirror $\|\vec{OM}\|$, ii) the distance between the augmented object and the mirror $\|\vec{PM}\|$, and iii) the angle of incidence θ. These spatial relations are depicted in Figure 7.4 and Figure 7.5.

Translation Error

It is important to distinguish between the errors observed within the different axis when talking about translation errors. In this context, tracking errors in the left-right (x-axis) and up-down (y-axis) directions of the mirror will not influence the perceived position of the virtual reflection concerning the real world. This is because both mirror planes, the real and virtual, remain coplanar. Therefore, the angles of incidence for the real and virtual mirrors remain the same, and the poses of the real mirror's virtual observer and the virtual camera are consistent. Nonetheless, undesired visual artifacts may arise if the distance between the virtual content and the real mirror's edge is smaller than the tracking error magnitude in the respective axis. This scenario results in visualizing a cropped image of the virtual content. As an analogy, one could visualize an object through a physical mirror and then translate the latter in the vertical or horizontal axis until the real object appears to be cropped by the mirror.

A completely different scenario is observed when the errors exist along the normal of the mirror surface (i.e., forward-backward). These errors affect the perceived position of the augmented content as the physical and virtual mirror planes are not coplanar anymore (see

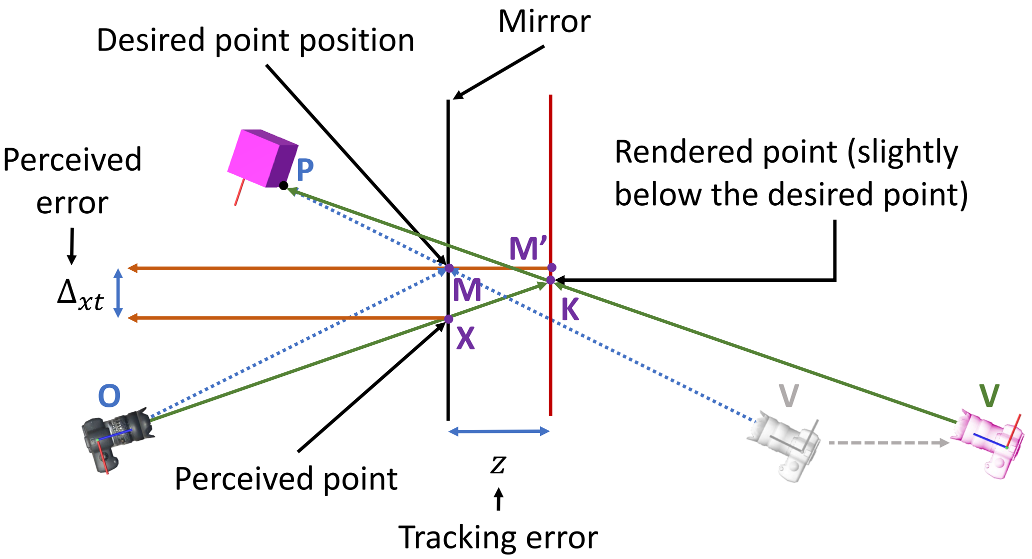


Fig. 7.4. Graphical representation of the error propagation associated with tracking errors in translation.

Figure 7.4). These errors would result in observing changes in both the apparent size and horizontal position of the virtual content reflected in the mirror. To obtain the perceived error Δx_t, the distances from the observer to the translated mirror $\|\vec{OM'}\|$, and from the augmented object to the translated mirror $\|\vec{PM'}\|$ are required. These distances can be derived using the trigonometric law of cosines:

$$\|\vec{OM'}\| = \sqrt{\|\vec{OM}\|^2 + \|\vec{z}\|^2 + 2\|\vec{OM}\|\|\vec{z}\| \cdot cos(2\theta)}$$

$$\|\vec{PM'}\| = \sqrt{\|\vec{PM}\|^2 + \|\vec{z}\|^2 + 2\|\vec{PM}\|\|\vec{z}\| \cdot cos(2\theta)}$$

As $\|\vec{OM'}\| = \|\vec{VM'}\|$, the segments $\|\vec{OM'}\|$ and $\|\vec{PM'}\|$ can be used to estimate the angle $\angle PM'V$ using the law of sines, and the segment $\|\vec{VP}\|$ using the law of cosines.

At the same time, this segment can be used to obtain $\angle PVM'$ and $\angle KVM'$, as well as $\angle VKM'$. The law of sines serves to obtain the segment $\|\vec{KM'}\|$.

The Pythagorean theorem and the law of cosines make possible to estimate the segments $\|\vec{MK}\|$ and $\|\vec{OK}\|$, respectively. These segments are at the same time used to estimate $\angle MOX$ and $\angle OXM$. Lastly, after using the law of sines, we obtain Δx_t:

$$\Delta x_t = \frac{\|\vec{OM}\| sin(\angle MOX)}{sin(\angle OXM)}$$

Orientation Error

Analogously to the translation errors, it is also important to distinguish between tracking errors observed along the different axis for orientation. In this case, contrary to the translational case, any orientation errors observed along the normal of the mirror surface (z-axis) will not influence the perceived position of the virtual reflection concerning the real world. Again, this is the result of the virtual and real mirror planes remaining coplanar.

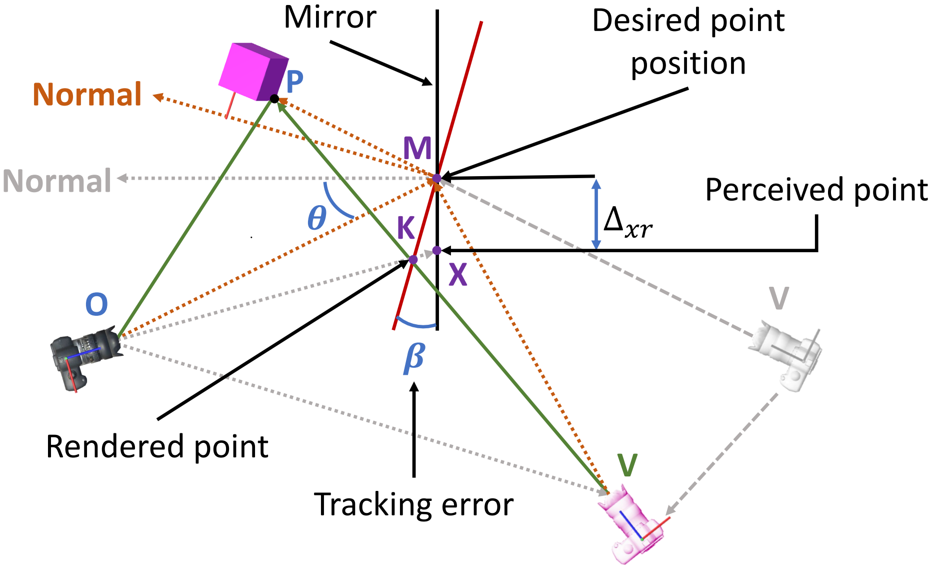


Fig. 7.5. Graphical representation of the error propagation associated with tracking errors in orientation.

On the contrary, any rotation errors along the x- and y-axis will derive in rendering the virtual reflection using a plane that does not matches the real mirror plane (see Figure 7.5). In this case, the distance between the reflected observer and the augmented object $\|\vec{VP}\|$ is calculated using the trigonometric law of cosines:

$$\|\vec{VP}\| = \sqrt{\|\vec{OM}\|^2 + \|\vec{PM}\|^2 + 2\|\vec{OM}\|\|\vec{PM}\| \cdot cos(2\beta)}$$

Then the angle between the object, the mirrored observer, and the mirror, α, follows with the law of sines:

$$\alpha = sin^{-1}\left(\frac{\|\vec{PM}\|sin(2\beta)}{\|\vec{VP}\|}\right)$$

Finally, using the law of sines and triangle postulate multiple times:

$$\Delta x_r = \frac{\|\vec{PM}\|sin(2\beta - \alpha)}{sin(90° - \alpha + \theta)}$$

Tracking Errors and Perceived Augmentations

A numeric analysis using the models presented in this section and considering two different tracking technologies can be found in the paper: "Augmented Mirrors." [110]

7.4.2 Alternative Architectures

Another form to implement the Augmented Mirrors is to replicate the virtual content behind the mirror plane. Therefore, this approach would avoid the necessity of a secondary camera to render the reflected content but would require flipping all the objects along their z-axis. In addition, the area defined by the real mirror size would play the role of a virtual window that would enable the observation of the content behind it. This concept is, in principle, comparable to the *tunnel window* paradigm introduced by Kiyokawa and Takemura [87]. Although this approach allows reducing the number of cameras in the virtual scene, the complexity of the virtual environment and the quantity of the virtual objects may increase the scene's complexity. An exemplary implementation of this concept is depicted in Figure 7.6

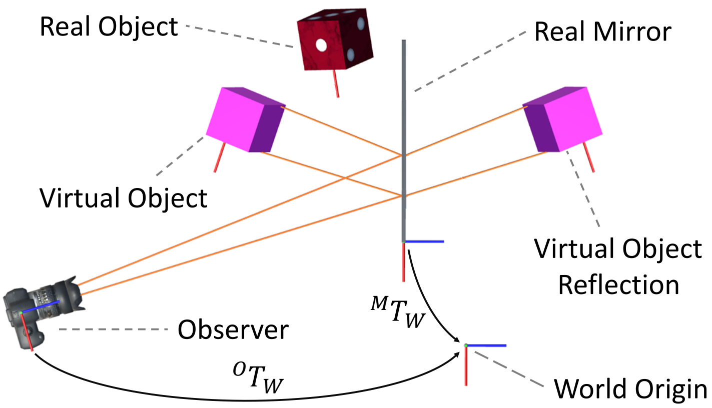


Fig. 7.6. An alternative implementation of the Augmented Mirrors. This approach duplicates the virtual content of the virtual scene to provide the mirrored image.

7.5 Application Domains

The Augmented Mirrors were initially motivated by interactive alignment scenarios in which an additional viewpoint would enhance the observer's perception helping to mitigate depth estimation errors in MR. However, the visual properties of the Augmented Mirrors provide extra benefits that can be useful for other tasks. For example, the additional viewpoint provided by an Augmented Mirror could assist the user in visualizing content hidden from its direct view. In addition, the virtual reflections provided by the Augmented Mirror could be presented using different visualization techniques than those applied to the direct view (e.g., using different colors, rendering techniques, or data representations). This section discusses some of those application domains and how the users of MR technologies could benefit from using them.

7.5.1 Alignment

As discussed throughout this dissertation, a major challenge in egocentric MR applications is accurately estimating the depth of the virtual content. Although some MR technologies such as HMDs provide stereovision, the relatively small baseline between the eyes limits depth perception. In this context, the *Augmented Mirrors* can help mitigate this perceptual problem as the alternative viewpoint can be used to retrieve the unperceivable depth information from an egocentric viewpoint.

Figure 7.7 depicts an industrial scenario that requires aligning a drill using a planned trajectory as a visual reference, exemplifying how the *Augmented Mirrors* can assist users during alignment tasks. Although this example may resemble a very naive scenario, even a simple trajectory can be perceived as correctly aligned using a single view when it is not (c.f., Figure 7.7a). Such scenarios require the user to repeatedly re-position their viewpoint while trying to preserve the pose of the tool, frequently leading to an iterative process that does not guarantee the desired outcome. In contrast, the integration of the *Augmented Mirror* in the scene enables the user to perceive the alignment error between the pose of the drill and the planned trajectory (c.f., Figure 7.7b).

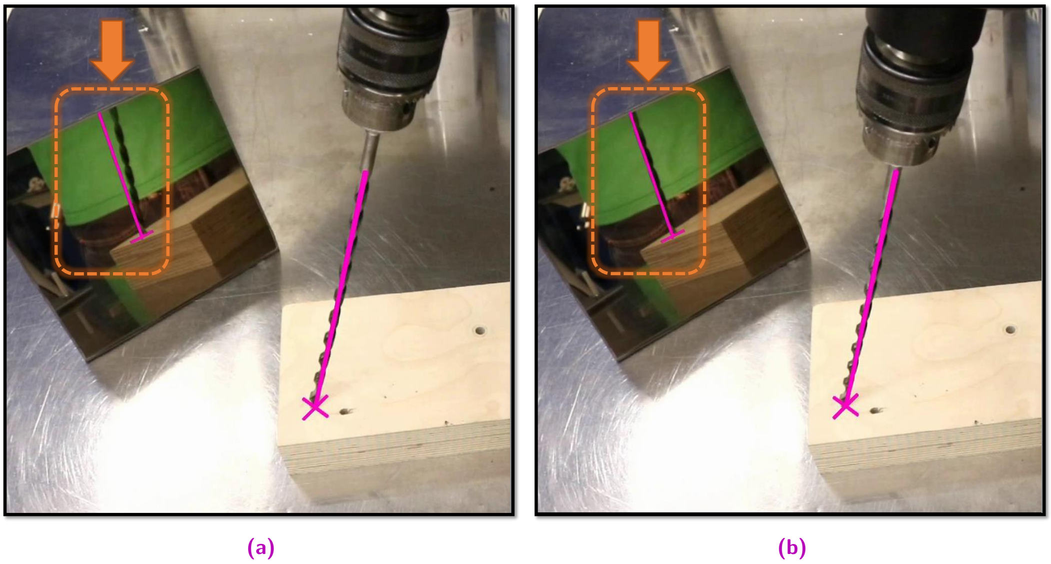


Fig. 7.7. *Augmented Mirrors* for alignment in industrial applications. Seemingly adequate alignment can be perceived when using direct view approaches. However, by integrating the *Augmented Mirror*, alignment errors are emphasized in the reflection (a). By aligning the drill in two views, the drill is ensured to align with the path (b).

Moreover, several tasks performed by humans involve this type of activity and demand a high level of accuracy. Examples of these are medical scenarios that require the insertion of needles, trocars, or pedicle screws during spine procedures and bone drilling or Kirschner wire placement in orthopedics. The relevance of using MR applications in this field has motivated efforts in bringing this technology into the operation rooms [2, 21, 26, 116]. In this context, achieving accurate alignment becomes particularly important for the procedure's outcome and the potential improvement of the patient's quality of life. Therefore, a simulated environment depicting the potential deployment of the Augmented Mirrors in medical environments is presented in Figure 7.8. This example depicts the use of this concept in vertebroplasty procedures that require the alignment of a trocar.

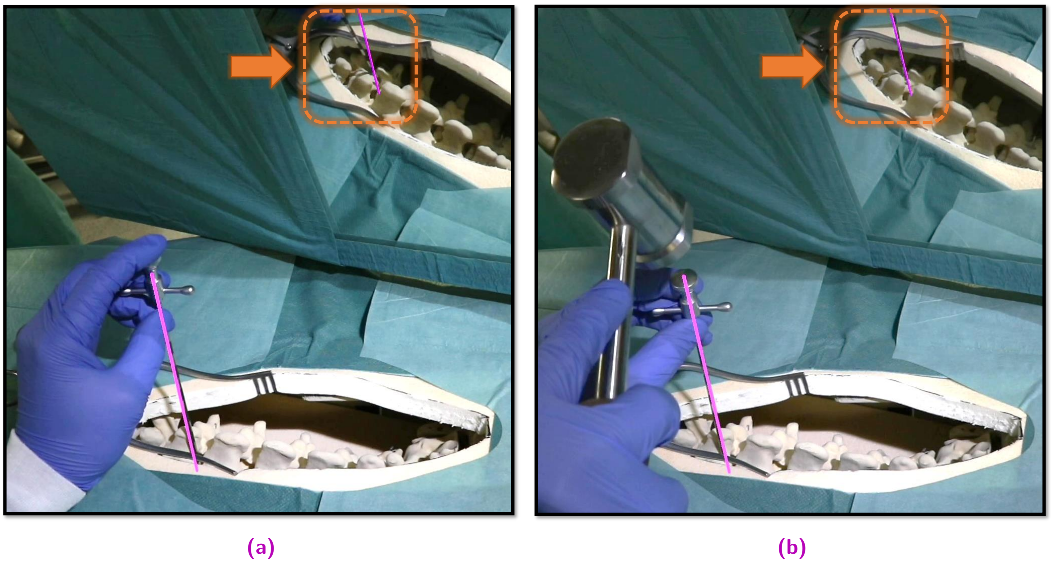
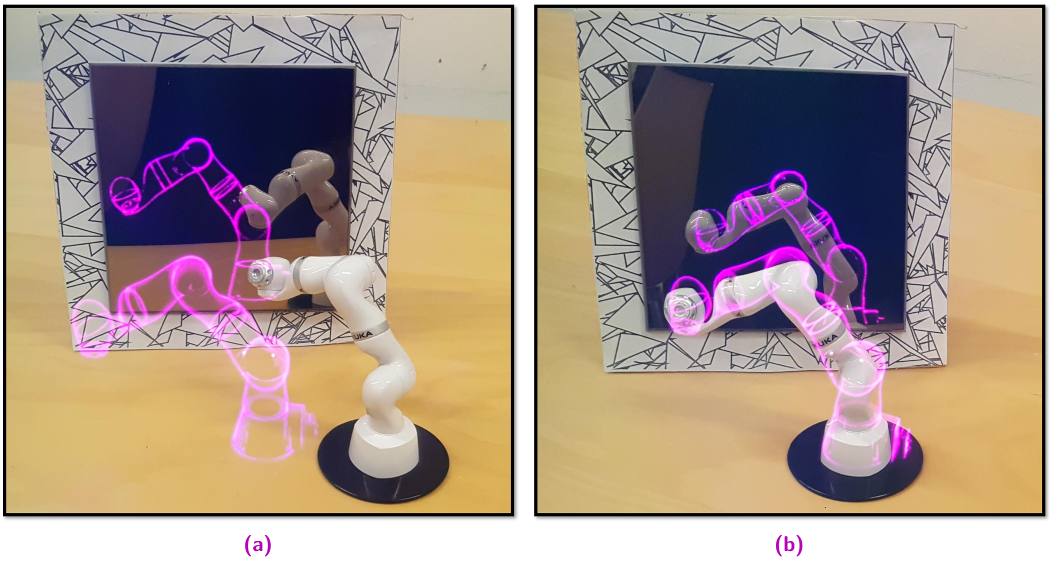


Fig. 7.8. *Augmented Mirrors* for spine procedures. The procedure's outcome may be jeopardized by inserting the trocar in the direction depicted in (a). In contrast, the *Augmented Mirrors* can assist the surgeon in achieving a suitable alignment of the trocar (b).

Furthermore, the potential of *Augmented Mirrors* for alignment tasks is manifold and reaches farther than just trajectories. Existing work has shown that applications such as the setup of robotic arms benefit from using MR approaches [42]. Inspired by this scenario, an *Augmented Mirror* can also assist during the alignment of articulated robotic arms with multiple joints. The example depicted in Figure 7.9 shows how the Augmented Mirrors assist users during the spatial alignment of a miniature replica of a robot with multiple degrees of freedom using the pose of an augmented replica as a reference frame. Even though this scenario depicts an environment using miniature-sized, the selection of a miniature was to exemplify the use case. Therefore, the implementation of the Augmented Mirrors is not limited by this fact and can be easily translated to full-sized devices. Like the alignment of trajectories, the additional view provided by the *Augmented Mirror* allows for the visualization of misalignment between the objects. Observing these errors is possible even without the user's need to physically move within the scene, presenting important advantages for several scenarios such as medical and industrial where the reduced or cluttered environments restrict the user's movement.

7.5.2 Exploration and Spatial Understanding

Another challenging task, not only in MR environments, involves the understanding of complex geometries. Multiple procedures in the medical field demand the correct spatial understanding of anatomical structures before planning or starting a procedure. A special case is orthopedic surgeries involving comminuted fractures in which a bone breaks in multiple parts. These procedures may represent a real challenge for the surgeon when visualizing and understanding the lesson during surgery. Although multiple procedures require pre-planning the steps to treat the fracture, better understanding the spatial arrangement of the bone fragments can


(a)

(b)

Fig. 7.9. *Augmented Mirrors* for the alignment of complex objects. The virtual replica of a robotic arm represents the desired pose (a), which needs to be achieved in the setup of a robotic arm (b).

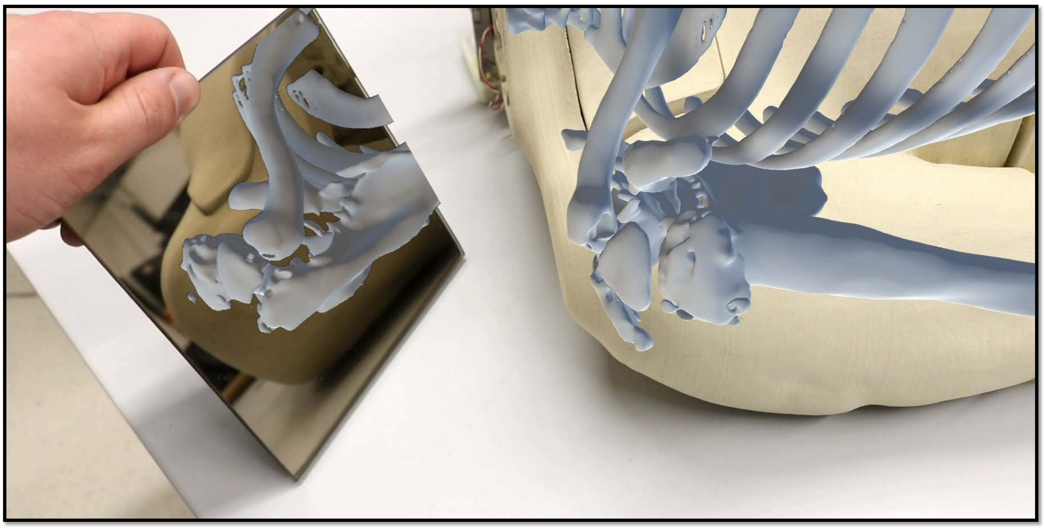


Fig. 7.10. *Augmented Mirrors* in Orthopedics. The use of the *Augmented Mirror* enhances a surgeon's spatial understanding of the anatomy of a shoulder. This way, the surgeon is assisted in finding bone splinters that would not be visible directly.

expedite the decision-making process during surgery. In this context, using an *Augmented Mirror* could assist the surgeon during the localization, identification, and re-assembly of the broken bone. Even more, this would facilitate the observation of the fragments by using the physical mirror as an interaction tool, reducing the need to move the surgeon's head or the patient's body. This exemplary use case is shown in Figure 7.10).

A more natural scenario in which users could benefit from the Augmented Mirrors is dentistry. This medical field already uses mirrors as an essential part of the workflow, potentially facilitating the seamless integration of this concept. For example, in this scenario, one could think about using the *Augmented Mirror* to visualize optimal drilling trajectories for the placement of an implant (see Figure 7.11a) or to visualize and understand anatomical information (Figure 7.11b). Even though this image depicts a mock-up scenario, the reconstruction and tracking accuracy to implement this concept can be achieved using dental navigation systems.[2]

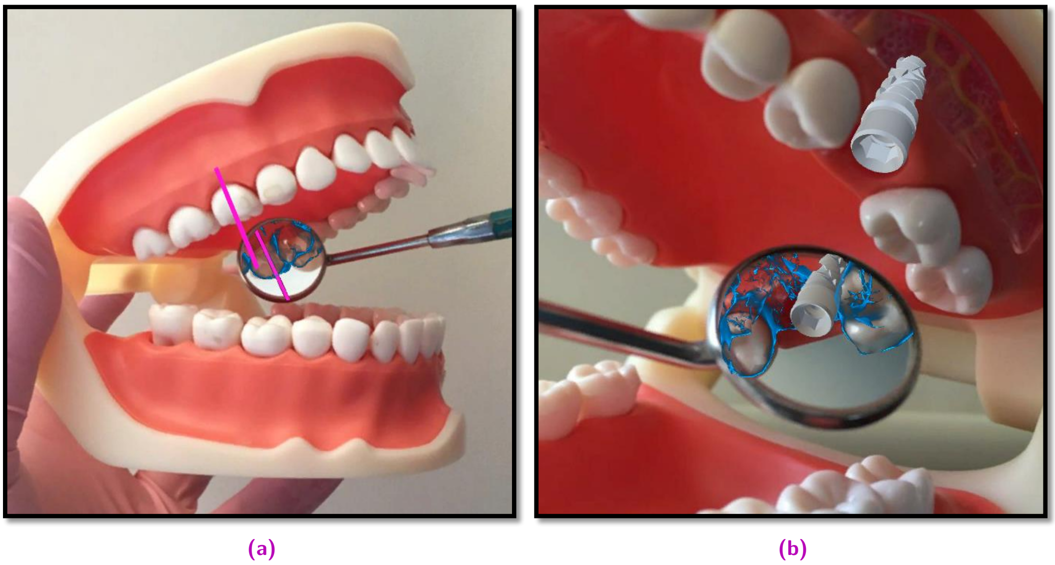


Fig. 7.11. *Augmented Mirrors* in Dentistry can show the optimal trajectory for drilling during dental implant procedures (a). Alternatively, it can visualize the fitting of the planned implants (b).

This concept could also be used during maintenance tasks in industrial scenarios in which the *Augmented Mirror* would enable exploring narrow spaces. Alternative use cases are visualizing areas of interest, guiding users while repairing electronic devices, highlighting objects of interest such as screws and crucial mechanical connections in car maintenance, or aiding operators in setting up cable connections.

7.5.3 Selective Content Visualization

Another particularity of the Augmented Mirrors is that they can use any visualization technique to present the virtual content to the observers, enabling them to utilize multiple visualization methods such as ray-casting, non-photorealistic, or physically-based rendering. This advantage can be used to exploit specific visual aspects of the MR environment, even contributing to understanding the scene better. For example, in the medical context, one could envision using ray-casting volume rendering techniques to visualize X-rays, Computed Tomography (CT) scans, or Magnetic Resonance Imaging (MRI) in the mirror. At the same time, a direct view of the scene would show a surface rendering of the anatomy of interest. An exemplary implementation of this concept is shown in Figure 7.11. In this use case, a virtual trajectory and an implant are visible in the direct and mirror views; however, the three-dimensional model of the patient's teeth can only be seen using the mirror.

[2] These images are generated by manual pose estimation of the model and mirror and only depict a conceptual idea.

7.5.4 Multiple Mirrors

Another advantage of this new class of MR mirrors is that they allow for the integration of multiple instances. In this regard, the number of *Augmented Mirrors* that can be used simultaneously is only limited by the number of objects that the tracking system can detect and by the computational power available to render the augmentations. An example of this feature is presented in Figure 7.12, in which a medical scenario with two entities of *Augmented Mirrors* serves to visualize CT and MRI volumes simultaneously. [3] Although this figure shows multiple instances of Augmented Mirrors in a medical scenario, their application is by no means restricted to them. Therefore, they can also be used, for example, to emulate the orthogonal views frequently observed in Computer-Aided Design software.

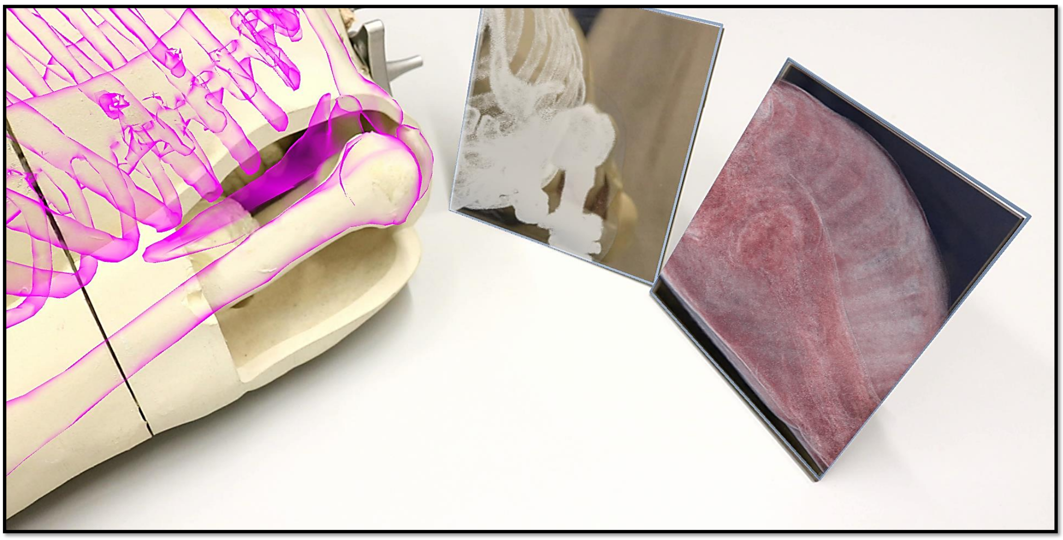


Fig. 7.12. Multiple *Augmented Mirrors* visualize a CT and an MRI scan of a patient using selective visualization.

7.6 Future Directions

Despite the multiple benefits derived from using the Augmented Mirrors in MR applications, the concept presented in this dissertation illustrates an initial implementation, leaving the door open to integrate numerous additional features. For this very same reason, the current state of this MR mirror class presents some limitations.

The current state of the augmented mirrors cannot handle the occlusion of real and virtual objects. This limitation may potentially lead to the observation of ambiguous information derived from misleading occlusion, hindering the interaction with this device as previously discussed in Chapter 3. A second limitation occurs in scenarios involving multiple instances of the *Augmented Mirrors*. In this regard, the current state of this concept does not account for multiple reflections for the virtual content as it would happen when using real mirrors. On the other hand, additional features could be integrated into this concept to account for the magnification of the real and virtual content, which could be highly beneficial in exploration

[3] This figure also exemplifies another use of the selective content visualization described in Subsection 7.5.3

tasks. Even more, a new class of MR mirrors could involve the use of external cameras that could replace the physical mirrors but replicate their optical properties.

7.6.1 Occlusion Handling

Similar to the case of many MR displays, the current implementation of the *Augmented Mirror* is not capable of providing adequate occlusion when the real and virtual objects overlap, presenting the virtual content on top of the real objects. This effect is observed from both the direct and reflected views. In this regard, a real-world object placed between the observer and mirror, or between the mirror and a virtual object, would obstruct the line of sight fully or partially to the virtual object from the observer's viewpoint to provide a correct visualization.

Ongoing research on occlusion handling in MR environments has suggested model-based approaches [16], contour tracking for occlusion masking [6, 88], or the generation of occlusion meshes from depth maps [76, 166], or machine-synthesized three-dimensional geometry from color images [172]. In the context of the Augmented Mirrors, spatial information of the scene from the observer and mirror's viewpoints is necessary to generate consistent occlusions. Therefore, the following cases may apply to generate realistic occlusion when using this concept:

> *Object tracking.* If we suppose it is possible to track the occluders, then model-based approaches will allow for the generation of accurate occlusions using pre-computed three-dimensional models of the real-world occluders [145]. This technique would allow for implementing the Augmented Mirrors without any further requirements or alteration of the system.
>
> *Depth-cameras.* Now, suppose it is impossible to track the occluders, then it is necessary to generate some form of representation of the real environment. In this regard, several forms exist to reconstruct the real environment and generate virtual meshes from it. A first approach involves using RGB-D cameras, as those integrated into commercially available OST HMDs, to generate three-dimensional spatial information to reconstruct the environment. However, because these cameras are aligned with the user's view direction, the reconstructed meshes would generate realistic occlusion only when facing the observer, leaving holes in the reconstructed meshes in all the hidden areas (e.g., the sides and back of the objects). Thus, to integrate this approach for the *Augmented Mirrors*, generating an occlusion mesh of the scene from the mirror's viewpoint is necessary. In this regard, the missing information can be recovered using a depth camera and the reflected depth image from the mirror [120].
>
> *Sparse reconstruction.* Another alternative is to use sparse SLAM reconstruction algorithms for the generation of occlusion meshes [76]. Nevertheless, it is important to consider that this method could cause a delayed response observed in the occlusion mesh when the real environment changes. This delay is the result of the temporal filters used to reduce noise in the input data. Therefore, using these algorithms might be prohibitive in dynamic environments.

Contour-based segmentation. In addition, contour-based methods could provide an alternative option by segmenting the foreground from bi-dimensional images [9], avoiding three-dimensional methods. This segmentation can be used as a mask to generate the occlusion and is well suited for the nature of the mirror images of the *Augmented Mirror* if the camera and observer's view directions are aligned.

7.6.2 Multiple Reflections

A special case and another feature that would further improve the similitudes between augmented and real mirrors is providing multiple reflections when using multiple instances. In general, if we assume two adjacent plane mirrors, the total number of reflections that can be seen will be determined by the angle formed between the planes of the mirrors and using the equation:

$$n = \frac{360^\circ}{\alpha} - 1 \tag{7.1}$$

where α is the angle between the mirrors and n is the number of reflections observed.

A peculiarity of this mirror configuration happens when both mirrors form an angle of 90 degrees, leading to a right-angle, non-reversing, or true mirror. In this case, the number of reflections observed is equal to 3, and the reflection observed in the middle will present the same size and will be seen at the same distance from the mirror as the reflected object. In addition, the reflection presented will be an upright image.

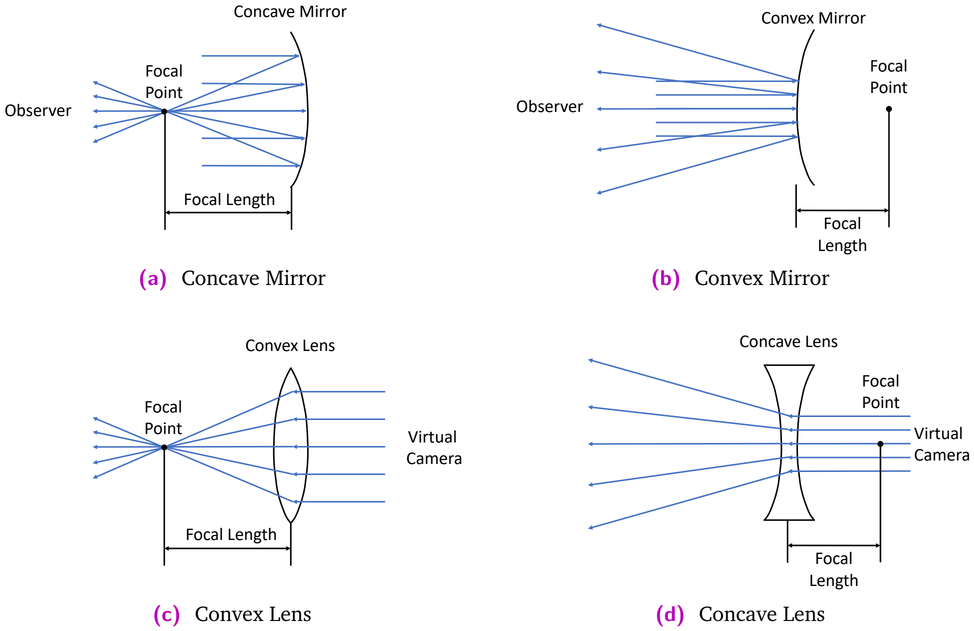
One aspect to consider when using this configuration is that the number of reflections increases as the angle between the mirrors decreases. While this visual effect may be interesting for the observer, it may also reduce the system's applicability as this will, at the same time, reduce the space in which the objects can be manipulated.

A potential approach to integrating this functionality as part of the Augmented Mirrors would require using Equation 7.1 to compute the number of reflected objects. This number could be later used to duplicate the virtual objects and position them in the scene using a single camera to visualize the scene as described in Subsection 7.4.2.

An even more complex case is the event in which more than two adjacent mirrors exist in the scene. In this case, the number of reflections increases, and what the observer perceives highly depends on its position in the scene. Moreover, the number of complete and partial reflections changes depending on the angles formed between the mirrors and the observer's viewpoint.

7.6.3 Non-planar Mirrors

In addition, it is possible to integrate non-planar mirrors as part of the new class of MR mirrors discussed in this dissertation. In this regard, two types of non-planar mirrors exist: convex and concave.



(a) Concave Mirror (b) Convex Mirror

(c) Convex Lens (d) Concave Lens

Fig. 7.13. The optical properties of concave and convex mirrors can be modeled using convex and concave lenses, respectively. This potentially enables the design and implementation of Non-planar Augmented Mirrors.

Convex mirrors present the characteristic to be curved outwards and enable the observation of a wider field of view. However, they reduce the apparent size of the objects observed over their surface. This type of mirror is commonly used in applications that require a wider field of view, such as in the car's side mirrors, on walls, ceilings, or hallways to allow for the visualization of incoming persons or potential obstructions, as well as in some streets and alleys for the visualization of incoming cars.

Concave mirrors are curved inwards and can be used to focus light. Unlike the convex mirrors, this type of mirror can invert the image observed if the distance from the object to the mirror is larger than its focal length. On the contrary, if this distance is smaller, they provide a magnified image without any reversal. Therefore, they are frequently used in dental applications because of their capacity to magnify images and focus light.

Regarding integrating these types of mirrors into the concept presented in this section, it is important to mention that mirrors have similar properties to lenses. In general terms, they both can be modeled using the Gaussian mirror equation:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f} \tag{7.2}$$

where d_o represents the object distance, d_i the image distance, and f the focal length.

In this regard, a concave mirror presents similar optical properties to a convex lens and vice versa (see Figure 7.13). Therefore, it would not be hard to imagine modeling a concave augmented mirror using a virtual camera with a convex lens replicating the same optical properties as the mirror.

7.6.4 Magnification

An additional feature that can be integrated into the Augmented Mirrors without the necessity of modifying its current state is magnification. As the position of the mirror concerning the observer is known, the real image reflected over the mirror's surface, visible to the camera of an MR-enabled device, could be cropped on a certain area of interest and magnified. The magnification ratio used for the real image would then be applied to the virtual content using the virtual camera placed behind the mirror. Nevertheless, it is important to mention that this method could lead to a major drawback. The real image acquired from the MR device will suffer from distortion due to the change in scale associated with the magnification.

Part IV

Conclusion

This dissertation explored whether the design of visualization techniques used to represent virtual content could provide relevant information to infer errors during interactive alignment in MR.

Early work presented in this dissertation compared traditional visualization techniques, commonly used during alignment tasks in egocentric MR scenarios, against alternative techniques that provided low levels of occlusion once the real and virtual content overlapped during task performance. Results from this comparison showed that reducing occlusion while providing meaningful visual information leads to more accurate alignment. In addition, the observation of lower levels of occlusion reduces the mental effort required to complete the task and increases the user's preference for the visualization technique. These results support the idea that the appearance of the virtual content influences the user's performance during the alignment task. Therefore, highlighting the importance of designing novel visualization techniques that provide relevant visual cues for interactive alignment tasks in MR. Additional work in this direction motivated the question of whether the appearance of the real objects could be used to generate virtual replicas capable of providing meaningful information when misalignment was observed? This principle motivated the conception of the COMPLEMENTARY TEXTURES. A new class of visualization techniques that use the textural, geometric, or even semantic properties of the real objects to generate complementary virtual replicas of them. These replicas provide meaningful visual information when misalignment occurs and, therefore, can be used to provide visual information during interactive alignment tasks in MR. Furthermore, this dissertation presented a direct comparison between traditional visualization techniques for in-situ visualization using OST and VST HMDs. This work investigated how to design visualization techniques that could enable in-situ visualization of virtual content placed inside real objects despite the optical limitations of commercially available OST HMDs. Even more, it introduced a taxonomy for the decomposition of these techniques to allow and motivate the generation and exploration of alternative techniques.

In addition, this dissertation explored the benefits of providing additional non-egocentric viewpoints during interactive alignment in MR environments. An initial study compared the advantages of using external cameras and mirrors to provide additional information to the observers during task performance. The study showed that having external cameras and static mirrors reduces motion parallax, traveled distance, and average head velocity. Therefore, they could be used in scenarios where the user's movements are restricted by the environment, such as industrial and medical setups. In addition, the results suggested that using dynamic mirrors motivates the exploration of the object's geometry and the environment. These results motivated the conceptualization and design of the AUGMENTED MIRRORS, a simple yet effective concept capable to dynamically provide alternative viewpoints of an MR scene using real mirrors. The use of this new class of MR mirrors showed additional benefits other than their simple use for alignment tasks. In this regard, they showed to be useful for exploration tasks, scene understanding, and multimodal visualization.

Overall, the concepts presented in this work serve as a step towards highlighting the importance of designing non-traditional visualization techniques to display virtual content during interactive alignment tasks in MR. These concepts aim at providing meaningful visual information when misalignment occurs. In this context, novel visualization techniques must integrate the fundamental concepts of visual perception and the depth cues used by the human visual system to infer these errors. In addition, the consideration of the environment and the physical properties of the objects involved in the alignment task, such as their textural patterns, geometry, and even their semantic information, can provide useful information to improve the alignment. Therefore, they must be considered when designing novel visualization techniques for alignment purposes.

Part V

Appendix

List of Publications

A

2021

[109] **Martin-Gomez, Alejandro***, Jakob Weiss*, Andreas Keller, Ulrich Eck, Daniel Roth, and Nassir Navab. "The Impact of Focus and Context Visualization Techniques on Depth Perception in Optical See-Through Head-Mounted Displays." *IEEE Transactions on Visualization and Computer Graphics (2021).*

[98] Leuze, Christoph, Caio Neves, **Alejandro Martin-Gomez**, Bruce L. Daniel, Nassir Navab, Nikolas H. Blevins, Vaisbuch Yona, and Jennifer A. McNab. "Augmented Reality Guided Retrosigmoid Approach." *Journal of Neurological Surgery Part B: Skull Base 82, no. S 02 (2021): S025.*

2020

[40] Fischer, Marc, Christoph Leuze, Stephanie Perkins, Jarrett Rosenberg, Bruce Daniel, and **Alejandro Martin-Gomez**. "Evaluation of Different Visualization Techniques for Perception-Based Alignment in Medical AR." *In 2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 45-50. IEEE, 2020.*

[108] **Martin-Gomez, Alejandro**, Javad Fotouhi, Ulrich Eck, and Nassir Navab. "Gain A New Perspective: Towards Exploring Multi-View Alignment in Mixed Reality." *In 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 207-216. IEEE, 2020.*

[110] **Martin-Gomez, Alejandro***, Alexander Winkler*, Kevin Yu*, Daniel Roth, Ulrich Eck, and Nassir Navab. "Augmented Mirrors." *In 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 217-226. IEEE, 2020.*

[105] **Martin-Gomez, Alejandro***, Colin Hill*, Hui-Yun Lin, Javad Fotouhi, Sarah Han-Oh, Ken Kang-Hsin Wang, Nassir Navab, and Amol Kumar Narang. "Towards Exploring the Benefits of Augmented Reality for Patient Support During Radiation Oncology Interventions." *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization (2020): 1-8.*

[106] **Martin-Gomez, Alejandro**, Ulrich Eck, Javad Fotouhi, and Nassir Navab. "Looking Also From Another Perspective: Exploring the Benefits of Alternative Views for Alignment Tasks." *In 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp. 811-812. IEEE, 2020.*

[42] Fotouhi, Javad, Tianyu Song, Arian Mehrfard, Giacomo Taylor, Qiaochu Wang, Fengfan Xian, **Alejandro Martin-Gomez** et al. "Reflective-ar display: An interaction methodology for virtual-to-real alignment in medical robotics." *IEEE Robotics and Automation Letters 5, no. 2 (2020): 2722-2729.*

2019

[107] **Martin-Gomez, Alejandro**, Ulrich Eck, and Nassir Navab. "Visualization techniques for precise alignment in VR: A comparative study." *In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 735-741. IEEE, 2019.*

*The asterisk indicates an equal contribution from the corresponding authors.

B Abstracts of Publications not Discussed in this Thesis

Augmented Reality Guided Retrosigmoid Approach

Christoph Leuze, Caio Neves, **Alejandro Martin-Gomez**, Bruce L. Daniel,
Nassir Navab, Nikolas H. Blevins, Vaisbuch Yona, Jennifer A. McNab

Abstract. While medical imaging data has traditionally been viewed on 2D displays, medical augmented reality (AR) allows physicians to project the medical imaging data on patient's bodies. An important application of medical AR is intra-operative surgical guidance by providing the physician with the ability to "see through" the patient's skin and locate important anatomy.

We present a medical AR application to support the retrosigmoid approach, an important approach to access the internal auditory canal. During the retrosigmoid approach, the craniotomy window is placed directly behind the sigmoid sinus, a large venous blood drainage running inside the posterior cranial fossa adjacent to the skull. The current standard is to use a surgical navigation system and anatomical surface landmarks to guide the surgeon during the procedure. However, surgical navigation systems require a long setup time and lack intuitiveness in presenting reformatted oblique planes to the surgeon while surface landmarks lack anatomical accuracy. As a simple and accurate alternative, we propose the use of an AR application that augments the surgeon's vision to guide the targeting procedure.

Journal of Neurological Surgery Part B: Skull Base (2021)

Evaluation of Different Visualization Techniques for Perception-Based Alignment in Medical AR

Marc Fischer, Christoph Leuze, Stephanie Perkins, Jarrett Rosenberg,
Bruce Daniel, **Alejandro Martin-Gomez**

Abstract. Many Augmented Reality (AR) applications require the alignment of virtual objects to the real world; this is particularly important in medical AR scenarios where medical imaging information may be displayed directly on a patient and is used to identify the exact locations of specific anatomical structures within the body. For optical see-through AR, alignment accuracy depends both on the optical parameters of the AR display as well as the visualization parameters of the virtual model. In this paper, we explore how different static visualization techniques influence users' ability to perform perception-based alignment in AR for breast reconstruction surgery, where surgeons must accurately identify the locations of several perforator blood vessels while planning the procedure. We conducted a pilot study in which four subjects used four different visualization techniques with varying degrees of opaqueness and brightness as well as outline contrast to align virtual replicas of the relevant anatomy to their 3D-printed counterparts. We collected quantitative scores on spatial alignment accuracy using an external tracking system and qualitative scores on user preference and perceived performance. Results indicate that the highest source of alignment error was along the depth dimension, with users consistently overestimating depth when aligning the virtual renderings. The majority of subjects preferred visualization techniques rendered with lower levels of opaqueness and brightness as well as higher outline contrast, which were also found to support more accurate alignment.

IEEE International Symposium on Mixed and Augmented Reality Adjunct (2020)

Towards Exploring the Benefits of Augmented Reality for Patient Support During Radiation Oncology Interventions

Martin-Gomez, Alejandro*, Colin Hill*, Hui-Yun Lin, Javad Fotouhi, Sarah Han-Oh, Ken Kang-Hsin Wang, Nassir Navab, and Amol Kumar Narang

Abstract. Traditionally, patient education has been limited to verbal exchanges between providers and patients, along with paper handouts that summarise relevant information. While such exchanges are a natural step in educating patients, they are limited for several reasons, including the lack of time that provider teams are afforded, and the inherent challenge of communicating nuanced concepts related to complex medical procedures. A clear example of this is radiation oncology, in which traditional routes of patient education may not satisfy the patient's needs. Although existing work has demonstrated the ability of audio-visual systems to improve patient engagement during medical procedures, the integration of emerging technologies such as Augmented Reality (AR) remains largely untapped. In this work, we propose an innovative proof-of-concept AR system as a first step towards exploring the benefits of using this technology during radiotherapy sessions. Our concept uses an AR headset to provide visual feedback of the patient's respiratory trace presented using two different forms: (i) a bi-dimensional graph and (ii) a game-based user interface. Moreover, we explore how interactive environments have the potential to contribute to better user experience and improve engagement, and discuss different challenges that must be addressed to deploy this technology to radiation treatment sessions.

Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization (2020)

*These authors contributed equally to this work.

Reflective-AR Display: An Interaction Methodology for Virtual-to-Real Alignment in Medical Robotics

Javad Fotouhi, Tianyu Song, Arian Mehrfard, Giacomo Taylor,
Qiaochu Wang, Fengfan Xian, **Alejandro Martin-Gomez**, Bernhard Fuerst,
Mehran Armand, Mathias Unberath, Nassir Navab

Abstract. Robot-assisted minimally invasive surgery has shown to improve patient outcomes, as well as reduce complications and recovery time for several clinical applications. While increasingly configurable robotic arms can maximize reach and avoid collisions in cluttered environments, positioning them appropriately during surgery is complicated because safety regulations prevent automatic driving. We propose a head-mounted display (HMD) based augmented reality (AR) system designed to guide optimal surgical arm set up. The staff equipped with HMD aligns the robot with its planned virtual counterpart. In this user-centric setting, the main challenge is the perspective ambiguities hindering such collaborative robotic solution. To overcome this challenge, we introduce a novel registration concept for intuitive alignment of AR content to its physical counterpart by providing a multi-view AR experience via reflective-AR displays that simultaneously show the augmentations from multiple viewpoints. Using this system, users can visualize different perspectives while actively adjusting the pose to determine the registration transformation that most closely superimposes the virtual onto the real. The experimental results demonstrate improvement in the interactive alignment a virtual and real robot when using a reflective-AR display. We also present measurements from configuring a robotic manipulator in a simulated trocar placement surgery using the AR guidance methodology.

IEEE Robotics and Automation Letters (2020)

C Acronyms and Abbreviations

– A –	
AM	Augmented Mirrors
AR	Augmented Reality
AV	Augmented Virtuality
– C –	
CT	Computed Tomography
– H –	
HMD	Head-Mounted Display
– J –	
JND	Just Noticeable Difference
– M –	
MR	Mixed Reality
MRI	Magnetic Resonance Imaging
– O –	
OST	Optical See-Through
– S –	
SEQ	Single Ease Question
SUS	System Usability Scale
– T –	
TLX	Task Load Index
– V –	
VR	Virtual Reality
VST	Video See-Through

Bibliography

[1] A. Albert, M. R. Hallowell, B. Kleiner, A. Chen, and M. Golparvar-Fard. "Enhancing construction hazard recognition with high-fidelity augmented virtuality". In: *Journal of Construction Engineering and Management* 140.7 (2014), p. 04014024 (cit. on p. 19).

[2] "Augmented reality in surgery: World's first "real" holographically navigated spine surgery". In: *Balgrist University Hospital* (Dec. 2020). URL: `https://www.balgrist.ch/fileadmin/user_upload/Aktuelles/Aktuelles/2020/Hololens/AR_Operation/Balgrist_Augmented_reality_in_surgery.pdf` (cit. on p. 119).

[3] B. Avery, C. Sandor, and B. H. Thomas. "Improving spatial perception for augmented reality x-ray vision". In: *2009 IEEE Virtual Reality Conference*. IEEE. 2009, pp. 79–82 (cit. on p. 62).

[4] R. T. Azuma. "A Survey of Augmented Reality". In: *Presence: Teleoperators & Virtual Environments* 6.4 (1997), pp. 355–385 (cit. on p. 19).

[5] M. Bajura, H. Fuchs, and R. Ohbuchi. "Merging virtual objects with the real world: Seeing ultrasound imagery within the patient". In: *ACM SIGGRAPH Computer Graphics* 26.2 (1992), pp. 203–210 (cit. on pp. 61–63).

[6] M.-O. Berger. "Resolving Occlusion in Augmented Reality: A Contour Based Approach without 3D Reconstruction". In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE. 1997, pp. 91–96 (cit. on p. 124).

[7] C. Bichlmeier, S. M. Heining, M. Feuerstein, and N. Navab. "The virtual mirror: a new interaction paradigm for augmented reality environments". In: *IEEE Transactions on Medical Imaging* 28.9 (2009), pp. 1498–1510 (cit. on pp. 89, 90, 107, 111).

[8] C. Bichlmeier, F. Wimmer, S. M. Heining, and N. Navab. "Contextual anatomic mimesis hybrid in-situ visualization method for improving multi-sensory depth perception in medical augmented reality". In: *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*. IEEE. 2007, pp. 129–138 (cit. on pp. 34, 61–63).

[9] M. Björkman and D. Kragic. "Active 3D Scene Segmentation and Detection of Unknown Objects". In: *2010 IEEE international conference on robotics and automation*. IEEE. 2010, pp. 3114–3120 (cit. on p. 125).

[10] M. Blackwell, C. Nikou, A. M. DiGioia, and T. Kanade. "An image overlay system for medical data visualization". In: *Medical image analysis* 4.1 (2000), pp. 67–72 (cit. on p. 61).

[11] K. C. Blits. "Aristotle: form, function, and comparative anatomy". In: *The Anatomical Record: An Official Publication of the American Association of Anatomists* 257.2 (1999), pp. 58–63 (cit. on p. 60).

[12] T. Blum, V. Kleeberger, C. Bichlmeier, and N. Navab. "mirracle: An Augmented Reality Magic Mirror System for Anatomy Education". In: *2012 IEEE Virtual Reality Workshops*. IEEE. 2012, pp. 115–116 (cit. on pp. 107, 110).

[13] C. Botella, R. M. Baños, C. Perpiñá, H. Villa, M. Alcañiz, and A. Rey. "Virtual reality treatment of claustrophobia: a case report". In: *Behaviour research and therapy* 36.2 (1998), pp. 239–246 (cit. on p. 19).

[14] B. P. Brain, P. Brain, et al. *Galen on Bloodletting: A Study of the Origins, Development and Validity of his Opinions, with a Translation of the Three Works*. Cambridge University Press, 1986 (cit. on p. 60).

[15] T. P. Breckon and R. B. Fisher. "Amodal volume completion: 3d visual completion". In: *Computer Vision and Image Understanding* 99.3 (2005), pp. 499–526 (cit. on p. 48).

[16] D. E. Breen, R. T. Whitaker, E. Rose, and M. Tuceryan. "Interactive Occlusion and Automatic Object Placement for Augmented Reality". In: *Computer Graphics Forum* 15.3 (1996), pp. 11–22. DOI: 10.1111/1467-8659.1530011. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-8659.1530011. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-8659.1530011 (cit. on p. 124).

[17] J. Brooke et al. "SUS-A quick and dirty usability scale". In: *Usability evaluation in industry* 189.194 (1996), pp. 4–7 (cit. on pp. 38, 96).

[18] V. Buchmann, T. Nilsen, and M. Billinghurst. "Interaction with partially transparent hands and objects". In: *Proceedings of the Sixth Australasian conference on User interface-Volume 40*. Australian Computer Society, Inc. 2005, pp. 17–20 (cit. on pp. 31, 32, 34, 47).

[19] T. Buhrmann, E. A. Di Paolo, and X. Barandiaran. "A dynamical systems account of sensorimotor contingencies". In: *Frontiers in psychology* 4 (2013), p. 285 (cit. on pp. 104, 109).

[20] E. A. Bustamante and R. D. Spain. "Measurement invariance of the Nasa TLX". In: *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 52. 19. SAGE Publications Sage CA: Los Angeles, CA. 2008, pp. 1522–1526 (cit. on p. 67).

[21] F. A. Casari, N. Navab, L. A. Hruby, P. Kriechling, R. Nakamura, R. Tori, F. d. L. dos Santos Nunes, M. C. Queiroz, P. Fürnstahl, and M. Farshad. "Augmented reality in orthopedic surgery is emerging from proof of concept towards clinical studies: a literature review explaining the technology and current state of the art". In: *Current Reviews in Musculoskeletal Medicine* (2021), pp. 1–12 (cit. on p. 119).

[22] S. Chakraborty, J. K. Stefanucci, S. Creem-Regehr, and B. Bodenheimer. "Distance Estimation with Mobile Augmented Reality in Action Space: Effects of Animated Cues". In: *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE. 2021, pp. 144–147 (cit. on p. 26).

[23] J. E. Cutting. "Reconceiving perceptual space." In: (2003) (cit. on p. 23).

[24] 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 (cit. on pp. 16, 17, 22, 74, 108).

[25] B. Delaunay et al. "Sur la sphere vide". In: *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk* 7.793-800 (1934), pp. 1–2 (cit. on p. 54).

[26] C. Dennler, D. E. Bauer, A.-G. Scheibler, J. Spirig, T. Götschi, P. Fürnstahl, and M. Farshad. "Augmented reality in the operating room: a clinical feasibility study". In: *BMC musculoskeletal disorders* 22.1 (2021), pp. 1–9 (cit. on p. 119).

[27] A. Dey, A. Cunningham, and C. Sandor. "Evaluating depth perception of photorealistic mixed reality visualizations for occluded objects in outdoor environments". In: *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*. 2010, pp. 211–218 (cit. on p. 26).

[28] J. F. Dobson. "Herophilus of alexandria". In: *Proceedings of the Royal Society of Medicine* 18 (1925), pp. 19–32 (cit. on p. 60).

[29] D. Drascic and P. Milgram. “Perceptual issues in augmented reality”. In: *Stereoscopic displays and virtual reality systems III*. Vol. 2653. International Society for Optics and Photonics. 1996, pp. 123–134 (cit. on pp. 14, 15, 31).

[30] W. J. Dunnican, T. P. Singh, A. Ata, E. E. Bendana, T. D. Conlee, C. J. Dolce, and R. Ramakrishnan. “Reverse Alignment "Mirror Image" Visualization as a Laparoscopic Training Tool Improves Task Performance”. In: *Surgical Innovation* 17.2 (2010), pp. 108–113 (cit. on p. 109).

[31] P. J. Edwards, A. P. King, C. R. Maurer, D. A. De Cunha, D. J. Hawkes, D. L. Hill, R. P. Gaston, M. R. Fenlon, A. Jusczyzck, A. J. Strong, et al. “Design and evaluation of a system for microscope-assisted guided interventions (MAGI)”. In: *IEEE Transactions on Medical Imaging* 19.11 (2000), pp. 1082–1093 (cit. on p. 60).

[32] P. Edwards, D. Hill, D. Hawkes, and R. Spink. “Stereo overlays in the operating microscope for image-guided surgery”. In: *Computer Assisted Radiology 1995* (1995), pp. 1197–1202 (cit. on p. 60).

[33] P. Eisert, J. Rurainsky, and P. Fechteler. “Virtual Mirror: Real-Time Tracking of Shoes in Augmented Reality Environments”. In: *2007 IEEE International Conference on Image Processing*. Vol. 2. IEEE. 2007, pp. II–557 (cit. on pp. 107, 110).

[34] S. R. Ellis and B. M. Menges. “Localization of virtual objects in the near visual field”. In: *Human factors* 40.3 (1998), pp. 415–431 (cit. on pp. 23, 31, 59).

[35] T. Engen. “Psychophysics. i. discrimination and detection. ii. scaling”. In: *Woodworth & Schlosberg’s Experimental Psychology* (1972), pp. 11–86 (cit. on p. 64).

[36] J. M. Enoch. “History of mirrors dating back 8000 years”. In: *Optometry and vision science* 83.10 (2006), pp. 775–781 (cit. on p. 108).

[37] S. Feiner, B. Macintyre, and D. Seligmann. “Knowledge-based augmented reality”. In: *Communications of the ACM* 36.7 (1993), pp. 53–62 (cit. on pp. 31, 47).

[38] J. A. Ferwerda. “Psychophysics 101: How to run perception experiments in computer graphics”. In: *ACM SIGGRAPH 2008 classes*. 2008, pp. 1–60 (cit. on p. 66).

[39] M. Fiala. “Magic Mirror System with Hand-Held and Wearable Augmentations”. In: *2007 IEEE Virtual Reality Conference*. IEEE. 2007, pp. 251–254 (cit. on pp. 107, 110).

[40] M. Fischer, C. Leuze, S. Perkins, J. Rosenberg, B. Daniel, and A. Martin-Gomez. “Evaluation of Different Visualization Techniques for Perception-Based Alignment in Medical AR”. In: *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE. 2020, pp. 45–50 (cit. on pp. 45, 135).

[41] J. Fotouhi, C. P. Alexander, M. Unberath, G. Taylor, S. C. Lee, B. Fuerst, A. Johnson, G. Osgood, R. H. Taylor, H. Khanuja, et al. “an augmented reality system for total hip arthroplasty”. In: *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*. Vol. 10576. International Society for Optics and Photonics. 2018, 105760J (cit. on p. 47).

[42] J. Fotouhi, T. Song, A. Mehrfard, G. Taylor, Q. Wang, F. Xian, A. Martin-Gomez, B. Fuerst, M. Armand, M. Unberath, et al. “Reflective-ar display: An interaction methodology for virtual-to-real alignment in medical robotics”. In: *IEEE Robotics and Automation Letters* 5.2 (2020), pp. 2722–2729 (cit. on pp. 90, 107, 111, 120, 135).

[43] H. Fuchs, M. A. Livingston, R. Raskar, K. Keller, J. R. Crawford, P. Rademacher, S. H. Drake, A. A. Meyer, et al. “Augmented reality visualization for laparoscopic surgery”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 1998, pp. 934–943 (cit. on pp. 62, 63).

[44] C. Furmanski, R. Azuma, and M. Daily. “Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information”. In: *Proceedings. International Symposium on Mixed and Augmented Reality*. IEEE. 2002, pp. 215–320 (cit. on pp. 26, 31, 61).

[45] J. L. Gabbard, J. E. Swan, D. Hix, S.-J. Kim, and G. Fitch. "Active text drawing styles for outdoor augmented reality: A user-based study and design implications". In: *2007 IEEE Virtual Reality Conference*. IEEE. 2007, pp. 35–42 (cit. on p. 31).

[46] J. L. Gabbard, J. E. Swan, J. Zedlitz, and W. W. Winchester. "More than meets the eye: An engineering study to empirically examine the blending of real and virtual color spaces". In: *2010 IEEE Virtual Reality Conference (VR)*. IEEE. 2010, pp. 79–86 (cit. on p. 31).

[47] H. C. Gagnon, L. Buck, T. N. Smith, G. Narasimham, J. Stefanucci, S. H. Creem-Regehr, and B. Bodenheimer. "Far distance estimation in mixed reality". In: *ACM Symposium on Applied Perception 2020*. 2020, pp. 1–8 (cit. on p. 26).

[48] H. C. Gagnon, C. S. Rosales, R. Mileris, J. K. Stefanucci, S. H. Creem-Regehr, and R. E. Bodenheimer. "Estimating distances in action space in augmented reality". In: *ACM Transactions on Applied Perception (TAP)* 18.2 (2021), pp. 1–16 (cit. on p. 26).

[49] A. Garcia-Palacios, H. Hoffman, A. Carlin, T. A. Furness III, and C. Botella. "Virtual reality in the treatment of spider phobia: a controlled study". In: *Behaviour research and therapy* 40.9 (2002), pp. 983–993 (cit. on p. 19).

[50] J. Gershon, E. Zimand, M. Pickering, B. O. Rothbaum, and L. Hodges. "A pilot and feasibility study of virtual reality as a distraction for children with cancer". In: *Journal of the American Academy of Child & Adolescent Psychiatry* 43.10 (2004), pp. 1243–1249 (cit. on p. 19).

[51] G. A. Gescheider et al. *Psychophysics: Method, theory, and application*. Psychology Press, 1985 (cit. on p. 66).

[52] G. A. Gescheider. *Psychophysics: the fundamentals*. Psychology Press, 2013 (cit. on p. 66).

[53] J. J. Gibson. *The ecological approach to visual perception: classic edition*. Psychology Press, 2014 (cit. on p. 11).

[54] J. J. Gibson. "The perception of the visual world." In: (1950) (cit. on pp. 9, 10).

[55] J. J. Gibson and L. Carmichael. *The senses considered as perceptual systems*. Vol. 2. 1. Houghton Mifflin Boston, 1966 (cit. on p. 12).

[56] J. G. Grandi, H. G. Debarba, I. Bemdt, L. Nedel, and A. Maciel. "Design and assessment of a collaborative 3D interaction technique for handheld augmented reality". In: *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2018, pp. 49–56 (cit. on p. 89).

[57] R. L. Gregory. "Mirrors in mind". In: (1998) (cit. on pp. 108, 109).

[58] R. L. Gregory. "The intelligent eye." In: (1970) (cit. on p. 11).

[59] W. E. L. Grimson, G. Ettinger, T. Kapur, M. E. Leventon, W. M. Wells III, and R. Kikinis. "Utilizing segmented MRI data in image-guided surgery". In: *International Journal of Pattern Recognition and Artificial Intelligence* 11.08 (1997), pp. 1367–1397 (cit. on p. 61).

[60] A. Grundhofer and O. Bimber. "Real-time adaptive radiometric compensation". In: *IEEE transactions on visualization and computer graphics* 14.1 (2007), pp. 97–108 (cit. on p. 57).

[61] J. Hajek, M. Unberath, J. Fotouhi, B. Bier, S. C. Lee, G. Osgood, A. Maier, M. Armand, and N. Navab. "Closing the calibration loop: an inside-out-tracking paradigm for augmented reality in orthopedic surgery". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2018, pp. 299–306 (cit. on p. 47).

[62] C. Hansen, J. Wieferich, F. Ritter, C. Rieder, and H.-O. Peitgen. "Illustrative visualization of 3D planning models for augmented reality in liver surgery". In: *International journal of computer assisted radiology and surgery* 5.2 (2010), pp. 133–141 (cit. on p. 82).

[63] S. G. Hart. "NASA-task load index (NASA-TLX); 20 years later". In: *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 50. 9. Sage publications Sage CA: Los Angeles, CA. 2006, pp. 904–908 (cit. on p. 67).

[64] S. G. Hart and L. E. Staveland. “Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research”. In: *Advances in psychology*. Vol. 52. Elsevier, 1988, pp. 139–183 (cit. on p. 67).

[65] G. Hatfield. “Perception: History of the concept”. In: (2001) (cit. on p. 9).

[66] F. Heinrich, K. Bornemann, K. Lawonn, and C. Hansen. “Depth perception in projective augmented reality: An evaluation of advanced visualization techniques”. In: *25th ACM Symposium on Virtual Reality Software and Technology*. 2019, pp. 1–11 (cit. on p. 82).

[67] H. v. Helmholtz. “Physiological optics (Vol. 3)”. In: *Rochester, NY: Optical Society of America* (1925) (cit. on p. 15).

[68] S. J. Henderson and S. K. Feiner. “Augmented reality in the psychomotor phase of a procedural task”. In: *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*. IEEE. 2011, pp. 191–200 (cit. on p. 32).

[69] J. L. Hintze and R. D. Nelson. “Violin plots: a box plot-density trace synergism”. In: *The American Statistician* 52.2 (1998), pp. 181–184 (cit. on pp. 69, 97).

[70] K. Hirose, T. Ogawa, K. Kiyokawa, and H. Takemura. “Interactive reconfiguration techniques of reference frame hierarchy in the multi-viewport interface”. In: *3D User Interfaces*. IEEE. 2006, pp. 73–80 (cit. on p. 90).

[71] T. N. Hoang and B. H. Thomas. “Augmented Viewport: An action at a distance technique for outdoor AR using distant and zoom lens cameras”. In: *International Symposium on Wearable Computers (ISWC) 2010*. IEEE. 2010, pp. 1–4 (cit. on p. 89).

[72] T. Hoang and B. H. Thomas. “Multiple Camera Augmented Viewport: An Investigation of Camera Position, Visualizations, and the Effects of Sensor Errors and Head Movement”. PhD thesis. Virtual Reality Society of Japan, 2011 (cit. on p. 89).

[73] H. G. Hoffman. “Virtual-reality therapy”. In: *Scientific American* 291.2 (2004), pp. 58–65 (cit. on p. 19).

[74] H. G. Hoffman, A. Garcia-Palacios, V. Kapa, J. Beecher, and S. R. Sharar. “Immersive virtual reality for reducing experimental ischemic pain”. In: *International Journal of Human-Computer Interaction* 15.3 (2003), pp. 469–486 (cit. on p. 19).

[75] H. G. Hoffman, D. R. Patterson, G. J. Carrougher, and S. R. Sharar. “Effectiveness of virtual reality–based pain control with multiple treatments”. In: *The Clinical journal of pain* 17.3 (2001), pp. 229–235 (cit. on p. 19).

[76] A. Holynski and J. Kopf. “Fast Depth Densification for Occlusion-aware Augmented Reality”. In: *ACM Transactions on Graphics (TOG)* 37.6 (2018), pp. 1–11 (cit. on p. 124).

[77] H. Ibayashi, Y. Sugiura, D. Sakamoto, N. Miyata, M. Tada, T. Okuma, T. Kurata, M. Mochimaru, and T. Igarashi. “Dollhouse vr: a multi-view, multi-user collaborative design workspace with vr technology”. In: *SIGGRAPH Asia 2015 Emerging Technologies*. 2015, pp. 1–2 (cit. on p. 89).

[78] S. Ishihara. “Series of plates designed as tests for colour-blindness”. In: (1936) (cit. on p. 36).

[79] S. Ishihara. *Test for colour-blindness*. Kanehara Tokyo, Japan, 1987 (cit. on p. 64).

[80] Y. Itoh, M. Dzitsiuk, T. Amano, and G. Klinker. “Semi-parametric color reproduction method for optical see-through head-mounted displays”. In: *IEEE transactions on visualization and computer graphics* 21.11 (2015), pp. 1269–1278 (cit. on p. 53).

[81] J. S. Jang, G. S. Jung, T. H. Lee, and S. K. Jung. “Two-Phase Calibration for a Mirror Metaphor Augmented Reality System”. In: *Proceedings of the IEEE* 102.2 (2014), pp. 196–203 (cit. on pp. 107, 110).

[82] J. A. Jones, J. E. Swan, G. Singh, and S. R. Ellis. “Peripheral visual information and its effect on distance judgments in virtual and augmented environments”. In: *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization*. 2011, pp. 29–36 (cit. on pp. 25, 31).

[83] 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 (cit. on pp. 25, 31).

[84] D. Kalkofen, E. Mendez, and D. Schmalstieg. “Interactive focus and context visualization for augmented reality”. In: *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE Computer Society. 2007, pp. 1–10 (cit. on pp. 61–63).

[85] B. M. Khuong, K. Kiyokawa, A. Miller, J. J. La Viola, T. Mashita, and H. Takemura. “The effectiveness of an AR-based context-aware assembly support system in object assembly”. In: *2014 IEEE Virtual Reality (VR)*. IEEE. 2014, pp. 57–62 (cit. on pp. 31, 32, 47).

[86] H. K. Kim, J. Park, Y. Choi, and M. Choe. “Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment”. In: *Applied ergonomics* 69 (2018), pp. 66–73 (cit. on p. 67).

[87] K. Kiyokawa and H. Takemura. “A tunnel window and its variations: Seamless teleportation techniques in a virtual environment”. In: *HCI International*. Citeseer. 2005 (cit. on pp. 89, 118).

[88] G. Klein and T. Drummond. “Sensor Fusion and Occlusion Refinement for Tablet-based AR”. In: *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE. 2004, pp. 38–47 (cit. on p. 124).

[89] K. Koffka. *Principles of Gestalt psychology*. Routledge, 2013 (cit. on p. 12).

[90] J. Kruger, J. Schneider, and R. Westermann. “Clearview: An interactive context preserving hotspot visualization technique”. In: *IEEE Transactions on Visualization and Computer Graphics* 12.5 (2006), pp. 941–948 (cit. on p. 61).

[91] A. Kunert, T. Weissker, B. Froehlich, and A. Kulik. “Multi-window 3D interaction for collaborative virtual reality”. In: *IEEE transactions on visualization and computer graphics* (2019) (cit. on p. 89).

[92] M. E. Latoschik, J.-L. Lugrin, and D. Roth. “FakeMi: A Fake Mirror System for Avatar Embodiment Studies”. In: *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. 2016, pp. 73–76 (cit. on p. 107).

[93] K. Lawonn, I. Viola, B. Preim, and T. Isenberg. “A survey of surface-based illustrative rendering for visualization”. In: *Computer Graphics Forum*. Vol. 37. 6. Wiley Online Library. 2018, pp. 205–234 (cit. on pp. 74, 82).

[94] C. Lee, S. Bonebrake, T. Hollerer, and D. A. Bowman. “A replication study testing the validity of ar simulation in vr for controlled experiments”. In: *2009 8th IEEE International Symposium on Mixed and Augmented Reality*. IEEE. 2009, pp. 203–204 (cit. on p. 92).

[95] G. A. Lee, H. S. Park, and M. Billinghurst. “Optical-Reflection Type 3D Augmented Reality Mirrors”. In: *25th ACM Symposium on Virtual Reality Software and Technology*. 2019, pp. 1–2 (cit. on p. 110).

[96] S. Lehar. “Gestalt isomorphism and the quantification of spatial perception”. In: *Gestalt theory* 21 (1999), pp. 122–139 (cit. on p. 48).

[97] M. Lerotic, A. J. Chung, G. Mylonas, and G.-Z. Yang. “Pq-space based non-photorealistic rendering for augmented reality”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2007, pp. 102–109 (cit. on pp. 61–63).

[98] C. Leuze, C. Neves, A. M. Gomez, B. L. Daniel, N. Navab, N. H. Blevins, V. Yona, and J. A. McNab. “Augmented Reality Guided Retrosigmoid Approach”. In: *Journal of Neurological Surgery Part B: Skull Base* 82.S 02 (2021), S025 (cit. on pp. 45, 135).

[99] W. Li, L. Ritter, M. Agrawala, B. Curless, and D. Salesin. “Interactive cutaway illustrations of complex 3d models”. In: *ACM SIGGRAPH 2007 papers on - SIGGRAPH ’07*. ACM Press. 2007, 31–es (cit. on p. 82).

[100] J. M. Liu, G. Narasimham, J. K. Stefanucci, S. Creem-Regehr, and B. Bodenheimer. “Distance perception in modern mobile augmented reality”. In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE. 2020, pp. 196–200 (cit. on p. 26).

[101] M. A. Livingston, A. Dey, C. Sandor, and B. H. Thomas. “Pursuit of "X-ray vision" for augmented reality”. In: *Human Factors in Augmented Reality Environments*. Springer, 2013, pp. 67–107 (cit. on p. 62).

[102] P. Maes, T. Darrell, B. Blumberg, and A. Pentland. “The ALIVE system: Wireless, Full-Body Interaction with Autonomous Agents”. In: *Multimedia Systems* 5.2 (1997), pp. 105–112 (cit. on pp. 107, 110).

[103] N. Maltby, I. Kirsch, M. Mayers, and G. J. Allen. “Virtual reality exposure therapy for the treatment of fear of flying: A controlled investigation.” In: *Journal of consulting and clinical psychology* 70.5 (2002), p. 1112 (cit. on p. 19).

[104] S. Mann. “Mediated reality with implementations for everyday life”. In: *Presence Connect* 1 (2002) (cit. on p. 20).

[105] A. Martin-Gomez, C. Hill, H. Lin, J. Fotouhi, S. Han-Oh, K.-H. Wang, N. Navab, and A. Narang. “Towards Exploring the Benefits of Augmented Reality for Patient Support During Radiation Oncology Interventions”. In: *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* (2020), pp. 1–8 (cit. on p. 135).

[106] A. Martin-Gomez, U. Eck, J. Fotouhi, and N. Navab. “Looking Also From Another Perspective: Exploring the Benefits of Alternative Views for Alignment Tasks”. In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE. 2020, pp. 811–812 (cit. on p. 135).

[107] A. Martin-Gomez, U. Eck, and N. Navab. “Visualization techniques for precise alignment in VR: A comparative study”. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2019, pp. 735–741 (cit. on pp. 47, 54, 135).

[108] A. Martin-Gomez, J. Fotouhi, U. Eck, and N. Navab. “Gain A New Perspective: Towards Exploring Multi-View Alignment in Mixed Reality”. In: *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2020, pp. 207–216 (cit. on p. 135).

[109] A. Martin-Gomez, J. Weiss, A. Keller, U. Eck, D. Roth, and N. Navab. “The Impact of Focus and Context Visualization Techniques on Depth Perception in Optical See-Through Head-Mounted Displays”. In: *IEEE Transactions on Visualization and Computer Graphics* (2021) (cit. on p. 135).

[110] A. Martin-Gomez, A. Winkler, K. Yu, D. Roth, U. Eck, and N. Navab. “Augmented Mirrors”. In: *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2020, pp. 217–226 (cit. on pp. 118, 135).

[111] J. Mercier-Ganady, F. Lotte, E. Loup-Escande, M. Marchal, and A. Lécuyer. “The Mind-Mirror: See your Brain in Action in your Head Using EEG and Augmented Reality”. In: *2014 IEEE Virtual Reality (VR)*. IEEE. 2014, pp. 33–38 (cit. on pp. 107, 110).

[112] P. Milgram and D. Drascic. “Perceptual effects in aligning virtual and real objects in augmented reality displays”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 41. 2. SAGE Publications Sage CA: Los Angeles, CA. 1997, pp. 1239–1243 (cit. on p. 47).

[113] P. Milgram and F. Kishino. “A taxonomy of mixed reality visual displays”. In: *IEICE Transactions on Information and Systems* 77.12 (1994), pp. 1321–1329 (cit. on pp. 3, 19).

[114] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. “Augmented reality: A class of displays on the reality-virtuality continuum”. In: *Telemanipulator and telepresence technologies*. Vol. 2351. International Society for Optics and Photonics. 1995, pp. 282–292 (cit. on pp. 18, 20).

[115] M. R. Mine, F. P. Brooks Jr, and C. H. Sequin. “Moving objects in space: exploiting proprioception in virtual-environment interaction.” In: *SIGGRAPH*. Vol. 97. 1997, pp. 19–26 (cit. on p. 92).

[116] F. Müller, S. Roner, F. Liebmann, J. M. Spirig, P. Fürnstahl, and M. Farshad. “Augmented reality navigation for spinal pedicle screw instrumentation using intraoperative 3D imaging”. In: *The Spine Journal* 20.4 (2020), pp. 621–628 (cit. on p. 119).

[117] R. Nakamura, L. L. Lago, A. B. Carneiro, A. J. Cunha, F. J. Ortega, J. L. Bernardes Jr, and R. Tori. “3PI experiment: Immersion in third-person view”. In: *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*. 2010, pp. 43–48 (cit. on p. 89).

[118] N. Navab, M. Feuerstein, and C. Bichlmeier. “Laparoscopic Virtual Mirror New Interaction Paradigm for Monitor Based Augmented Reality”. In: *2007 IEEE Virtual Reality Conference*. IEEE. 2007, pp. 43–50 (cit. on pp. 107, 111).

[119] J. Newman, A. Bornik, D. Pustka, F. Echtler, M. Huber, D. Schmalstieg, and G. Klinker. “Tracking for distributed mixed reality environments”. In: Citeseer (cit. on p. 20).

[120] T.-N. Nguyen, H.-H. Huynh, and J. Meunier. “3D Reconstruction with Time-of-Flight Depth Camera and Multiple Mirrors”. In: *IEEE Access* 6 (2018), pp. 38106–38114 (cit. on p. 124).

[121] J. Norman. “Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches”. In: *Behavioral and brain sciences* 25.1 (2002), p. 73 (cit. on p. 11).

[122] O. Oda, C. Elvezio, M. Sukan, S. Feiner, and B. Tversky. “Virtual replicas for remote assistance in virtual and augmented reality”. In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM. 2015, pp. 405–415 (cit. on pp. 31, 47).

[123] R. Ohbuchi, M. Bajura, and H. Fuchs. “Case study: Observing a volume rendered fetus within a pregnant patient”. In: *Visualization: Proceedings of the IEEE Conference on Visualization*. Vol. 5. Citeseer. 1998 (cit. on p. 61).

[124] M. Otsuki, H. Kuzuoka, and P. Milgram. “Analysis of Depth Perception with Virtual Mask in Stereoscopic AR.” In: *ICAT-EGVE*. 2015, pp. 45–52 (cit. on pp. 62, 64, 73).

[125] M. Otsuki and P. Milgram. “Psychophysical exploration of stereoscopic pseudo-transparency”. In: *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2013, pp. 1–6 (cit. on pp. 64, 84).

[126] F. G. Paas. “Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach.” In: *Journal of educational psychology* 84.4 (1992), p. 429 (cit. on pp. 38, 96).

[127] S. Pardhy, C. Shankwitz, and M. Donath. “A Virtual Mirror for Assisting Drivers”. In: *Proceedings of the IEEE Intelligent Vehicles Symposium 2000 (Cat. No. 00TH8511)*. IEEE. 2000, pp. 255–260 (cit. on pp. 107, 110).

[128] P. Paul, O. Fleig, and P. Jannin. “Augmented virtuality based on stereoscopic reconstruction in multimodal image-guided neurosurgery: Methods and performance evaluation”. In: *IEEE transactions on medical imaging* 24.11 (2005), pp. 1500–1511 (cit. on p. 19).

[129] E. Peillard, Y. Itoh, G. Moreau, J.-M. Normand, A. Lécuyer, and F. Argelaguet. “Can Retinal Projection Displays Improve Spatial Perception in Augmented Reality?” In: *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2020, pp. 80–89 (cit. on pp. 24, 31).

[130] C. Portalés, J. Gimeno, S. Casas, R. Olanda, and F. G. Martínez. “Interacting with Augmented Reality Mirrors”. In: *Handbook of Research on Human-Computer Interfaces, Developments, and Applications*. IGI Global, 2016, pp. 216–244 (cit. on p. 110).

[131] E. Praun, H. Hoppe, M. Webb, and A. Finkelstein. “Real-time hatching”. In: *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. 2001, p. 581 (cit. on p. 75).

[132] D. Reiners, D. Stricker, G. Klinker, and S. Müller. “Augmented reality for construction tasks: Doorlock assembly”. In: *Proc. IEEE and ACM IWAR* 98.1 (1998), pp. 31–46 (cit. on pp. 31–33, 47).

[133] R. S. Renner, B. M. Velichkovsky, and J. R. Helmert. “The perception of egocentric distances in virtual environments-a review”. In: *ACM Computing Surveys (CSUR)* 46.2 (2013), pp. 1–40 (cit. on pp. 17, 23).

[134] G. Riva, S. Raspelli, D. Algeri, F. Pallavicini, A. Gorini, B. K. Wiederhold, and A. Gaggioli. “Interreality in practice: bridging virtual and real worlds in the treatment of posttraumatic stress disorders”. In: *Cyberpsychology, Behavior, and Social Networking* 13.1 (2010), pp. 55–65 (cit. on p. 19).

[135] C. M. Robertson, B. MacIntyre, and B. N. Walker. “An evaluation of graphical context when the graphics are outside of the task area”. In: *Mixed and Augmented Reality, 2008. ISMAR 2008. 7th IEEE/ACM International Symposium on*. IEEE. 2008, pp. 73–76 (cit. on pp. 31, 47).

[136] I. Rock and S. Palmer. “The legacy of Gestalt psychology”. In: *Scientific American* 263.6 (1990), pp. 84–91 (cit. on p. 48).

[137] J. P. Rolland, C. Meyer, K. Arthur, and E. Rinalducci. “Method of adjustments versus method of constant stimuli in the quantification of accuracy and precision of rendered depth in head-mounted displays”. In: *Presence: Teleoperators & Virtual Environments* 11.6 (2002), pp. 610–625 (cit. on p. 64).

[138] T. Ropinski, F. Steinicke, and K. Hinrichs. “Visually supporting depth perception in angiography imaging”. In: *International Symposium on Smart Graphics*. Springer. 2006, pp. 93–104 (cit. on p. 82).

[139] B. O. Rothbaum, L. F. Hodges, R. Kooper, D. Opdyke, J. S. Williford, and M. North. “Virtual reality graded exposure in the treatment of acrophobia: A case report”. In: *Behavior therapy* 26.3 (1995), pp. 547–554 (cit. on p. 19).

[140] N. Sala. “Fractal geometry and architecture: some interesting connections”. In: *WIT Transactions on the Built Environment* 86 (2006), pp. 163–173 (cit. on p. 54).

[141] C. Sandor, A. Cunningham, A. Dey, and V.-V. Mattila. “An augmented reality x-ray system based on visual saliency”. In: *2010 IEEE International Symposium on Mixed and Augmented Reality*. IEEE. 2010, pp. 27–36 (cit. on p. 62).

[142] S. Sangwine and N. Le Bihan. “Quaternion Toolbox for Matlab, Version 2 with support for Octonions”. In: *Software Library* (2013) (cit. on p. 54).

[143] F. R. Sarlegna and R. L. Sainburg. “The Roles of Vision and Proprioception in the Planning of Reaching Movements”. In: *Progress in Motor Control*. Springer, 2009, pp. 317–335 (cit. on pp. 108, 109).

[144] H. Sato, I. Kitahara, and Y. Ohta. “MR-mirror: a Complex of Real and Virtual Mirrors”. In: *International Conference on Virtual and Mixed Reality*. Springer. 2009, pp. 482–491 (cit. on pp. 107, 110).

[145] M. Sauer, F. Leutert, and K. Schilling. “Occlusion Handling in Augmented Reality User Interfaces for Robotic Systems”. In: *ISR 2010 (41st International Symposium on Robotics) and ROBOTIK 2010 (6th German Conference on Robotics)*. VDE. 2010, pp. 1–7 (cit. on p. 124).

[146] G. Schall, E. Mendez, E. Kruijff, E. Veas, S. Junghanns, B. Reitinger, and D. Schmalstieg. “Handheld augmented reality for underground infrastructure visualization”. In: *Personal and ubiquitous computing* 13.4 (2009), pp. 281–291 (cit. on pp. 61, 62).

[147] B. Schauerte. “Spectral Visual Saliency Toolbox”. In: *Software Library* (2021) (cit. on p. 54).

[148] B. Schauerte and R. Stiefelhagen. “Quaternion-based spectral saliency detection for eye fixation prediction”. In: *European Conference on Computer Vision*. Springer. 2012, pp. 116–129 (cit. on p. 54).

[149] C. M. Schneck. “Visual perception”. In: *Occupational Therapy for Children. sixth ed. Mosby Inc* (2010), pp. 373–403 (cit. on pp. 10, 12, 13).

[150] S. M. Schneider and M. Workman. “Virtual reality as a distraction intervention for older children receiving chemotherapy”. In: *Pediatric Nursing* 26.6 (2000), p. 593 (cit. on p. 19).

[151] R. A. Serway and J. W. Jewett. *Physics for Scientists and Engineers with Modern Physics, Ninth Edition*. Cengage Learning, 2013 (cit. on pp. 109, 113).

[152] S. R. Sharar, W. Miller, A. Teeley, M. Soltani, H. G. Hoffman, M. P. Jensen, and D. R. Patterson. “Applications of virtual reality for pain management in burn-injured patients”. In: *Expert review of neurotherapeutics* 8.11 (2008), pp. 1667–1674 (cit. on p. 19).

[153] T. Sielhorst, C. Bichlmeier, S. M. Heining, and N. Navab. “Depth perception–a major issue in medical AR: evaluation study by twenty surgeons”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2006, pp. 364–372 (cit. on p. 62).

[154] K. T. Simsarian and K.-P. Akesson. “Windows on the world: An example of augmented virtuality”. In: (1997) (cit. on p. 19).

[155] G. Singh, S. R. Ellis, and J. E. Swan. “The effect of focal distance, age, and brightness on near-field augmented reality depth matching”. In: *IEEE transactions on visualization and computer graphics* 26.2 (2018), pp. 1385–1398 (cit. on pp. 24, 31, 45, 83).

[156] G. Singh, J. E. Swan, J. A. Jones, and S. R. Ellis. “Depth judgment measures and occluding surfaces in near-field augmented reality”. In: *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*. 2010, pp. 149–156 (cit. on pp. 23, 24, 31, 59, 65, 72, 84).

[157] V. Šoltészová, D. Patel, and I. Viola. “Chromatic shadows for improved perception”. In: *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering*. 2011, pp. 105–116 (cit. on pp. 74, 75).

[158] C. Stapleton and J. Davies. “Imagination: The third reality to the virtuality continuum”. In: *2011 IEEE International Symposium on Mixed and Augmented Reality-Arts, Media, and Humanities*. IEEE. 2011, pp. 53–60 (cit. on p. 20).

[159] R. Stoakley, M. J. Conway, and R. Pausch. “Virtual reality on a WIM: interactive worlds in miniature”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1995, pp. 265–272 (cit. on p. 89).

[160] M. Sukan, S. Feiner, B. Tversky, and S. Energin. “Quick viewpoint switching for manipulating virtual objects in hand-held augmented reality using stored snapshots”. In: *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2012, pp. 217–226 (cit. on pp. 89, 90, 107).

[161] J. E. Swan, A. Jones, E. Kolstad, M. A. Livingston, and H. S. Smallman. “Egocentric depth judgments in optical, see-through augmented reality”. In: *IEEE transactions on visualization and computer graphics* 13.3 (2007), pp. 429–442 (cit. on pp. 22, 25, 31).

[162] J. E. Swan, L. Kuparinen, S. Rapson, and C. Sandor. “Visually perceived distance judgments: Tablet-based augmented reality versus the real world”. In: *International Journal of Human–Computer Interaction* 33.7 (2017), pp. 576–591 (cit. on pp. 25, 26).

[163] J. E. Swan, G. Singh, and S. R. Ellis. “Matching and reaching depth judgments with real and augmented reality targets”. In: *IEEE transactions on visualization and computer graphics* 21.11 (2015), pp. 1289–1298 (cit. on pp. 23, 24, 31, 72).

[164] A. Tang, C. Owen, F. Biocca, and W. Mou. “Comparative effectiveness of augmented reality in object assembly”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. 2003, pp. 73–80 (cit. on pp. 31, 32, 47).

[165] P. Thomas and W. David. “Augmented reality: An application of heads-up display technology to manual manufacturing processes”. In: *Hawaii international conference on system sciences*. 1992, pp. 659–669 (cit. on pp. 31, 47).

[166] Y. Tian, Y. Long, D. Xia, H. Yao, and J. Zhang. “Handling Occlusions in Augmented Reality Based on 3D Reconstruction Method”. In: *Neurocomputing* 156 (2015), pp. 96–104 (cit. on p. 124).

[167] M. Unberath, J. Fotouhi, J. Hajek, A. Maier, G. Osgood, R. Taylor, M. Armand, and N. Navab. “Augmented reality-based feedback for technician-in-the-loop C-arm repositioning”. In: *Healthcare technology letters* 5.5 (2018), pp. 143–147 (cit. on p. 31).

[168] L. Vera, J. Gimeno, I. Coma, and M. Fernández. “Augmented Mirror: Interactive Augmented Reality System Based on Kinect”. In: *IFIP Conference on Human-Computer Interaction*. Springer. 2011, pp. 483–486 (cit. on pp. 107, 110).

[169] F. Vincelli, L. Anolli, S. Bouchard, B. K. Wiederhold, V. Zurloni, and G. Riva. “Experiential cognitive therapy in the treatment of panic disorders with agoraphobia: a controlled study”. In: *CyberPsychology & Behavior* 6.3 (2003), pp. 321–328 (cit. on p. 19).

[170] S. Vogt, A. Khamene, F. Sauer, A. Keil, and H. Niemann. “A high performance AR system for medical applications”. In: *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.* IEEE. 2003, pp. 270–271 (cit. on p. 61).

[171] H. Von Helmholtz. *Handbuch der physiologischen Optik: mit 213 in den Text eingedruckten Holzschnitten und 11 Tafeln*. Vol. 9. Voss, 1867 (cit. on p. 11).

[172] N. Wang, Y. Zhang, Z. Li, Y. Fu, W. Liu, and Y.-G. Jiang. “Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images”. In: *Proceedings of the European Conference on Computer Vision*. 2018, pp. 52–67 (cit. on p. 124).

[173] R. Wang, Z. Geng, Z. Zhang, and R. Pei. “Visualization techniques for augmented reality in endoscopic surgery”. In: *International Conference on Medical Imaging and Augmented Reality*. Springer. 2016, pp. 129–138 (cit. on pp. 62, 64).

[174] M. Weiser. “The computer for the 21st century”. In: *ACM SIGMOBILE mobile computing and communications review* 3.3 (1999), pp. 3–11 (cit. on p. 20).

[175] G. Westerfield, A. Mitrovic, and M. Billinghurst. “Intelligent augmented reality training for motherboard assembly”. In: *International Journal of Artificial Intelligence in Education* 25.1 (2015), pp. 157–172 (cit. on pp. 31, 47).

[176] W. Wetzlinger, A. Auinger, and M. Dörflinger. “Comparing effectiveness, efficiency, ease of use, usability and user experience when using tablets and laptops”. In: *International Conference of Design, User Experience, and Usability*. Springer. 2014, pp. 402–412 (cit. on p. 67).

[177] C. D. Wickens, J. G. Hollands, S. Banbury, and R. Parasuraman. *Engineering psychology and human performance*. Psychology Press, 2015, pp. 69–118 (cit. on p. 105).

[178] P. Willemsen, M. B. Colton, S. H. Creem-Regehr, and W. B. Thompson. “The effects of head-mounted display mechanics on distance judgments in virtual environments”. In: *Proceedings of the 1st Symposium on Applied perception in graphics and visualization*. 2004, pp. 35–38 (cit. on p. 25).

[179] S. S. Wint, D. Eshelman, J. Steele, and C. E. Guzzetta. “Effects of distraction using virtual reality glasses during lumbar punctures in adolescents with cancer.” In: *Oncology nursing forum*. Vol. 29. 1. 2002 (cit. on p. 19).

[180] E. N. Zalta, U. Nodelman, C. Allen, and J. Perry. *Stanford encyclopedia of philosophy*. 1995 (cit. on p. 9).

[181] J. Zauner, M. Haller, A. Brandl, and W. Hartmann. “Authoring of a mixed reality assembly instructor for hierarchical structures”. In: *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE Computer Society. 2003, p. 237 (cit. on pp. 31, 47).

[182] C. Zimmer, M. Bertram, F. Büntig, D. Drochtert, and C. Geiger. “Mobile Augmented Reality Illustrations That Entertain and Inform: Design and Implementation Issues with the Hololens”. In: *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*. SA ’17. Bangkok, Thailand: Association for Computing Machinery, 2017. ISBN: 9781450354103. URL: `https://doi.org/10.1145/3132787.3132804` (cit. on p. 111).

[183] S. Zollmann, D. Kalkofen, E. Mendez, and G. Reitmayr. “Image-based ghostings for single layer occlusions in augmented reality”. In: *2010 IEEE International Symposium on Mixed and Augmented Reality*. IEEE. 2010, pp. 19–26 (cit. on p. 64).

List of Figures

List of Tables