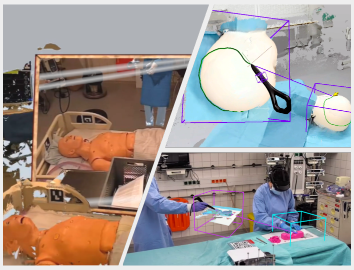
Computer Aided Medical Procedures
Prof. Dr. Nassir Navab



Dissertation

Advanced 3D UI for Immersive AR/VR Medical Teleconsultation

Kevin Yu

TUM School of Computation, Information and Technology
Technische Universität München

Technische Universität München
TUM School of Computation, Information and Technology

Advanced 3D UI for Immersive AR/VR Medical Teleconsultation

Kevin Yu

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Hans Michael Gerndt

Prüfende der Dissertation:

1. Prof. Dr. Nassir Navab
2. Prof. Henry Fuchs
3. Prof. Dr. Daniel Roth

Die Dissertation wurde am 25.04.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 24.01.2024 angenommen.

Kevin Yu
Advanced 3D UI for Immersive AR/VR Medical Teleconsultation
Dissertation, Version 1.0

Technische Universität München
TUM School of Computation, Information and Technology
Lehrstuhl für Informatikanwendungen in der Medizin
Boltzmannstraße 3
85748 and Garching bei München

Abstract

Humans interact with three-dimensional (3D) surroundings at every moment and where ever possible. We learn from the beginning of our lives to touch, grasp, and move anything within reach. Nevertheless, when it comes to digital information later in life, we interact with it through two-dimensional monitors and similar displays. Advanced Virtual Reality (VR) and Augmented Reality (AR) headsets have disrupted this traditional understanding of interfaces by allowing users to perceive digital content in three dimensions. 3D technology allows us to make one of the humans' earliest science fiction dreams come true, to visit distant places instantaneously. This concept is called telepresence and has an immeasurable value in the medical domain as it enables doctors, among other benefits, to aid other surgeons remotely while keeping them safe from transmissive diseases. This dissertation focuses on developing advanced tools for the user interface (UI) in a medical teleconsultation so that users can understand and work more dynamically. In particular, this dissertation presents three interaction techniques to augment the digital representation of the real-world location and the patient to allow users to perceive and interact with it more efficiently and intuitively. The first two concepts are inspired by dioramas miniaturizing the real world into a pocket-sized representation. In the scientific VR community, this is often related to World-in-Miniatures. However, Magnoramas and Duplicated Realities magnify the selected region instead of miniaturizing the world. Therefore, a user study has been conducted for each of the two presented methods. The first study shows that Magnoramas can increase the precision of hand-drawn annotations to communicate ideas between both parties. The second study evaluates the ability of Duplicated Reality to separate real-world space in a co-located scenario while maintaining the spatial context between the real and virtual worlds. The third concept is inspired by real mirrors, creating a secondary view of objects or patients. The Projective Bisector Mirror employs a live camera stream projected onto the bisector plane constructed from the camera's and the user's viewpoint to visualize a mathematically correct virtual mirror view. This method allows users to see a high-resolution view of the real world as a mirror alongside the virtual remote representation. All presented methods are functional and integrated into an immersive medical teleconsultation system called ArtekMed that utilizes AR/VR to connect local users with remote doctors. Furthermore, we discuss that all presented methods have an uncommon property of being capable of visualizing digital content outside the actual operating region while allowing users to maintain their spatial relations inside the real world. With these works, we aim to improve future 3D teleconsultation systems and extend our comprehension of interacting with 3D UIs.

Zusammenfassung

Wir interagieren mit unserer dreidimensionalen Umgebung jeden Moment und jederzeit möglich. Wir lernen von klein an, alles in Reichweite anzufassen, zu ergreifen und zu bewegen. Wenn es jedoch später im Leben um digitale Informationen geht, interagieren wir damit durch zweidimensionale Monitore und Displays. Virtual Reality (VR) und Augmented-Reality (AR) Headsets haben jedoch dieses traditionelle Verständnis geändert, indem sie Benutzern ermöglichen, digitale Inhalte in drei Dimensionen wahrzunehmen. 3D-Technologie ermöglicht es uns, einen der ältesten Science-Fiction-Träume der Menschheit wahr werden zu lassen und entfernte Orte jederzeit und nahezu ohne Aufwand zu besuchen. Dieses Konzept wird als Telepräsenz bezeichnet und hat im medizinischen Bereich einen unschätzbaren Wert. Ärzten wird unter anderem ermöglicht, entfernte Patienten und Ärzte zu besuchen und sich gleichzeitig vor übertragbaren Krankheiten zu schützen. Diese Dissertation konzentriert sich darauf, fortgeschrittene Werkzeuge für die Benutzerschnittstelle einer medizinische Telekonsultation zu entwickeln und dadurch eine dynamischere Arbeitsweise zu ermöglichen. Insbesondere werden drei Methoden präsentiert, um die digitale Darstellung eines Ortes zu erweitern und um Benutzern eine bessere Interaktion mit diesen zu ermöglichen. Die ersten beiden Konzepte sind von Dioramas inspiriert, die die reale Welt in taschengröße verkleinern. In der wissenschaftlichen VR-Community wird dies oft als World-in-Miniatures bezeichnet. Anstatt die Umgebung zu verkleinern, vergrößern Magnoramas und Duplicated Realities jedoch die ausgewählte Region. Eine erste Studie konzentriert sich darauf, die Präzision von handgezeichneten Annotationen in VR zu erhöhen, um Ideen zwischen beiden Parteien zu kommunizieren. Eine zweite Studie bewertet die Fähigkeit der Duplicated Reality Methode, im selben Raum reale und virtuelle Räume voneinander zu trennen, während man den räumlichen Kontext zwischen Realität und Virtualität aufrechterhält. Das dritte Konzept ist von Spiegeln inspiriert, die eine sekundäre Ansicht von Objekten oder Patienten erstellen. Der Projective Bisector Mirror verwendet einen Live-Kamerastream, der auf die von der Benutzerperspektive konstruierte Bisektorebene projiziert wird, um eine mathematisch korrekte Spiegelansicht zu visualisieren. Diese Methode ermöglicht es Benutzern, die reale Welt als Spiegel neben der virtuellen Fernrepräsentation zu sehen. Alle vorgestellten Methoden sind funktional und in ein immersives medizinisches Telekonsultationssystem namens ArtekMed integriert. Darüber hinaus wird diskutiert, wie alle vorgestellten Methoden eine ungewöhnliche Eigenschaft besitzen, digitale Inhalte außerhalb der tatsächlichen Operationsbereiches visualisieren zu können, während Benutzer ihre räumlichen Beziehungen zur realen Welt aufrechterhalten können. Mit diesen Arbeiten zielen wir darauf ab zukünftige 3D-Telekonsultationssysteme zu verbessern und unser Verständnis für die Interaktion mit 3D-Benutzeroberflächen er erweitern.

Acknowledgments

No work is the product of a single person. Neither is this dissertation, as it is a culmination of resilient hard work, endurance, and support from many colleagues, friends, and family that came, went, or stayed along the way.

Firstly, I extend my heartfelt thanks to my mentors, Dr. Ulrich Eck and Prof. Dr. Nassir Navab, whose continuous support from the beginning to the end of this research journey was invaluable. Without their guidance, ideas, funding, and foundation, this dissertation would not have been possible. I am also grateful to Prof. Hubertus Feußner and Prof. Dirk Wilhelm for their valuable consultation and perspectives from the medical side.

I thoroughly enjoyed my time there during my initial three years as a Ph.D. student at MITI. In addition to my aforementioned medical mentors from MITI, I would like to thank Anna Zapaichshykova and Regine Hartwig for brightening up my daily hours and contributing valuable insights to my early ideas. Tereza Baude also deserves thanks for her constant mental support and assistance with office-related matters. Furthermore, I would like to express my appreciation to my colleagues at the CAMP chair, especially those in the Narvis lab, including Prof. Dr. Daniel Roth, Alexander Winkler, Tianyu Song, Michael Sommersperger, and the many students who make Narvis the hub of innovative ideas. Additionally, the INM, especially Dr. Frieder Pankratz and Dr. Marc Lazarovici, deserve equal recognition for supporting me at every step during my studies.

I am deeply grateful to the German Federal Ministry of Education and Research for funding (Funding-No.: 16SV8092) for most of my time as a Ph.D. student. Furthermore, I appreciate medPhoton GmbH's support during my final year. In particular, I thank Phil Steininger and Heinz Deutschmann for showing me the world outside academia and teaching me how to bridge the gap between both.

If you are reading this, please accept my humble gratitude towards you. Your interest in reading my work has made it worth my effort and sweat.

Leaving the most important part for the end, I would like to express my eternal gratitude towards my dear wife, Mengjiao Guo, for her support in difficult and enjoyment in more delightful times. She serves as my anchor, and has even aided in acquiring media material for my second work in this dissertation. So please take a look and see if you can spot her.

Thank you all!

Contents

Part I

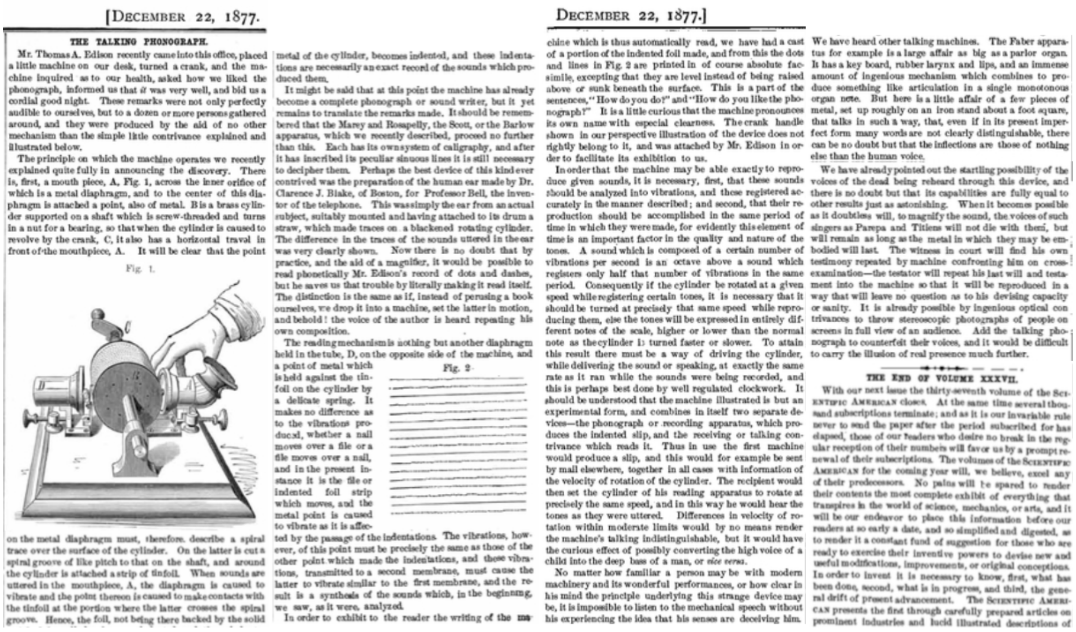
Introduction

1 Introduction

> *You've got to get the fundamentals down, because otherwise the fancy stuff is not going to work.*
>
> — **Prof. Randy Pausch**
> (The Last Lecture [1], p.36)

"Hello!"
"*Hello, can you hear me?*"
"Yes, I can. But I cannot see you."
"*Wait a moment... Now you should be able to see me.*"
"Yes! Thanks."
This is how many online meetings are initiated in the current year in the 21st century. As users, we strive to achieve seamless and uninterrupted online communication, not only in terms of connection stability but also in terms of effective social communicative interaction. Video conferencing tools, such as Zoom, Skype, and many more, unite us with distant people from different countries, cities, or just people who are a few streets apart from our location. However, these tools may not or only partially replicate the level of presence and interactivity experienced during face-to-face communication. Face-to-face communication involves the transmission of subtle verbal and non-verbal information of not just facial expressions but furthermore body language and subtle nuances in behavior and appearance. The enormity of artificially creating a virtual representation that adequately compensates for the missing gaps in human-to-human communication cannot be overstated. Such communicational gaps are evident in video-mediated online meetings where non-verbal cues such as body language, facial expressions, and tone of voice are often lost or inadequately transmitted, sometimes leading to challenges in communication, especially during the discussion of complex topics. *Is the person I am talking to nervous, or do they even look at me if I am talking to them? Should I skip to the next topic?*

In the modern, fast-paced world, meeting in person is often challenging due to logistical constraints, such as geographical distance, tight calendars, or personal reasons. Hence, there is a growing interest in utilizing immersive technology to enhance telecommunication and collaboration. Virtual Reality (VR) and Augmented Reality (AR) are emerging technologies that hold significant potential in this regard, with commercially available devices making them accessible to as many people as ever in human history. The concept of instantaneously relocating to a remote location is called teleportation. The concept of *telepresence* is closely related to teleportation. It is often referred to as the feeling of *being there* at a remote location and is much more feasible regarding ethics and practicability. Telepresence occurs in the earliest science fiction literature. It is unclear who proposed the concept first; However, it likely existed since around the late 19th century. One of the earliest mentions is when Alexander Graham Bell and Elisha Gray independently filed patents for a sonograph as early



[DECEMBER 22, 1877.

THE TALKING PHONOGRAPH.

Mr. Thomas A. Edison recently came into this office, placed a little machine on our desk, turned a crank, and the machine inquired as to our health, asked how we liked the phonograph, informed us that *it* was very well, and bid us a cordial good night. These remarks were not only perfectly audible to ourselves, but to a dozen or more persons gathered around, and they were produced by the aid of no other mechanism than the simple little contrivance explained and illustrated below.

The principle on which the machine operates we recently explained quite fully in announcing the discovery. There is, first, a mouth piece, A, Fig. 1, across the inner orifice of which is a metal diaphragm, and to the center of this diaphragm is attached a point, also of metal. B is a brass cylinder supported on a shaft which is screw-threaded and turns in a nut for a bearing, so that when the cylinder is caused to revolve by the crank, C, it also has a horizontal travel in front of the mouthpiece, A. It will be clear that the point

Fig. 1.

on the metal diaphragm must, therefore, describe a spiral trace over the surface of the cylinder. On the latter is cut a spiral groove of like pitch to that on the shaft, and around the cylinder is attached a strip of tinfoil. When sounds are uttered in the mouthpiece, A, the diaphragm is caused to vibrate and the point thereon is caused to make contacts with the tinfoil at the portion where the latter crosses the spiral groove. Hence, the foil, not being there backed by the solid metal of the cylinder, becomes indented, and these indentations are necessarily an exact record of the sounds which produced them.

It might be said that at this point the machine has already become a complete phonograph or sound writer, but it yet remains to translate the remarks made. It should be remembered that the Marey and Rosapelly, the Scott, or the Barlow apparatus, which we recently described, proceed no further than this. Each has its own system of caligraphy, and after it has inscribed its peculiar sinuous lines it is still necessary to decipher them. Perhaps the best device of this kind ever contrived was the preparation of the human ear made by Dr. Clarence J. Blake, of Boston, for Professor Bell, the inventor of the telephone. This was simply the ear from an actual subject, suitably mounted and having attached to its drum a straw, which made traces on a blackened rotating cylinder. The difference in the traces of the sounds uttered in the ear was very clearly shown. Now there is no doubt that by practice, and the aid of a magnifier, it would be possible to read phonetically Mr. Edison's record of dots and dashes, but he saves us that trouble by literally making it read itself. The distinction is the same as if, instead of perusing a book ourselves, we drop it into a machine, set the latter in motion, and behold! the voice of the author is heard repeating his own composition.

The reading mechanism is nothing but another diaphragm held in the tube, D, on the opposite side of the machine, and a point of metal which is held against the tinfoil on the cylinder by a delicate spring. It makes no difference as to the vibrations produced, whether a nail moves over a file or a file moves over a nail, and in the present instance it is the file or indented foil strip which moves, and the metal point is caused to vibrate as it is affected by the passage of the indentations. The vibrations, however, of this point must be precisely the same as those of the other point which made the indentations, and these vibrations, transmitted to a second membrane, must cause the latter to vibrate similar to the first membrane, and the result is a synthesis of the sounds which, in the beginning, we saw, as it were, analyzed.

Fig. 2

In order to exhibit to the reader the writing of the machine which is thus automatically read, we have had a cast of a portion of the indented foil made, and from this the dots and lines in Fig. 2 are printed in of course absolute facsimile, excepting that they are level instead of being raised above or sunk beneath the surface. This is a part of the sentences, "How do you do?" and "How do you like the phonograph?" It is a little curious that the machine pronounces its own name with especial clearness. The crank handle shown in our perspective illustration of the device does not rightly belong to it, and was attached by Mr. Edison in order to facilitate its exhibition to us.

In order that the machine may be able exactly to reproduce given sounds, it is necessary, first, that these sounds should be analyzed into vibrations, and these registered accurately in the manner described; and second, that their reproduction should be accomplished in the same period of time in which they were made, for evidently this element of time is an important factor in the quality and nature of the tones. A sound which is composed of a certain number of vibrations per second is an octave above a sound which registers only half that number of vibrations in the same period. Consequently if the cylinder be rotated at a given speed while registering certain tones, it is necessary that it should be turned at precisely that same speed while reproducing them, else the tones will be expressed in entirely different notes of the scale, higher or lower than the normal note as the cylinder is turned faster or slower. To attain this result there must be a way of driving the cylinder, while delivering the sound or speaking, at exactly the same rate as it ran while the sounds were being recorded, and this is perhaps best done by well regulated clockwork. It should be understood that the machine illustrated is but an experimental form, and combines in itself two separate devices—the phonograph or recording apparatus, which produces the indented slip, and the receiving or talking contrivance which reads it. Thus in use the first machine would produce a slip, and this would for example be sent by mail elsewhere, together in all cases with information of the velocity of rotation of the cylinder. The recipient would then set the cylinder of his reading apparatus to rotate at precisely the same speed, and in this way he would hear the tones as they were uttered. Differences in velocity of rotation within moderate limits would by no means render the machine's talking indistinguishable, but it would have the curious effect of possibly converting the high voice of a child into the deep bass of a man, or *vice versa*.

No matter how familiar a person may be with modern machinery and its wonderful performances, or how clear in his mind the principle underlying this strange device may be, it is impossible to listen to the mechanical speech without his experiencing the idea that his senses are deceiving him. We have heard other talking machines. The Faber apparatus for example is a large affair as big as a parlor organ. It has a key board, rubber larynx and lips, and an immense amount of ingenious mechanism which combines to produce something like articulation in a single monotonous organ note. But here is a little affair of a few pieces of metal, set up roughly on an iron stand about a foot square, that talks in such a way, that, even if in its present imperfect form many words are not clearly distinguishable, there can be no doubt but that the inflections are those of nothing else than the human voice.

We have already pointed out the startling possibility of the voices of the dead being reheard through this device, and there is no doubt but that its capabilities are fully equal to other results just as astonishing. When it becomes possible as it doubtless will, to magnify the sound, the voices of such singers as Parepa and Titiens will not die with them, but will remain as long as the metal in which they may be embodied will last. The witness in court will find his own testimony repeated by machine confronting him on cross-examination—the testator will repeat his last will and testament into the machine so that it will be reproduced in a way that will leave no question as to his devising capacity or sanity. It is already possible by ingenious optical contrivances to throw stereoscopic photographs of people on screens in full view of an audience. Add the talking phonograph to counterfeit their voices, and it would be difficult to carry the illusion of real presence much further.

THE END OF VOLUME XXXVII.

With our next issue the thirty-seventh volume of the SCIENTIFIC AMERICAN closes. At the same time several thousand subscriptions terminate; and as it is our invariable rule never to send the paper after the period subscribed for has elapsed, those of our readers who desire no break in the regular reception of their numbers will favor us by a prompt renewal of their subscriptions. The volumes of the SCIENTIFIC AMERICAN for the coming year will, we believe, excel any of their predecessors. No pains will be spared to render their contents the most complete exhibit of everything that transpires in the world of science, mechanics, or arts, and it will be our endeavor to place this information before our readers at so early a date, and so simplified and digested, as to render it a constant fund of suggestion for those who are ready to exercise their inventive powers to devise new and useful modifications, improvements, or original conceptions. In order to invent it is necessary to know, first, what has been done, second, what is in progress, and third, the general drift of present advancement. The SCIENTIFIC AMERICAN presents the first through carefully prepared articles on prominent industries and lucid illustrated descriptions of

Fig. 1.1. **The Talking Sonograph** [2]. Screenshot of the original publication in Scientific American in December 1877. This article, among others, likely sparked the inspiration for telepresence in science-fiction authors.

as 1876, the precursor of today's telephone. Then, in 1877, Thomas Edison filed a patent for an improvement, which inspired Scientific American to publish an article that likely helped spark the imagination of science fiction authors and supported in popularizing the concept of telepresence. The last two sentences of that article, "The Talking Sonograph" from 1877 (see Figure 1.1), describe an early concept of immersive telepresence: "*It is already possible by ingenious optical contrivances to throw stereoscopic photographs of people on screen in full view of an audience. Add the talking phonograph to counterfeit their voices, and it would be difficult to carry the illusion of real presence much further.*" [2] Fast-forward to today, AR and VR technologies have progressed to a level that would have seemed unimaginable to people 150 years ago. The knowledge of humans capable of seeing depth through stereoscopic images was known long before; for instance, Englishman Charles Wheatstone invented the stereoscope [3] in 1832 to view drawn or printed images with a stereoscopic 3D effect. One of the earliest works in 1992 helping in shaping the landscape of telepresence overall and in particular for the medical use case, providing realistic solutions, was Ulrich Neumann and Henry Fuchs [4], and shortly following up with the work on a "sea of cameras", also from Fuchs et al. [5], to capture an operating theater in 3D for remote consultation. When reviewing those early works, it is astounding how well the concept of 3D teleconsultation has stayed relevant in the last thirty years: Compared to the first pioneering works, the sheer amount of overall experience of today's technology regarding image resolution, real-time capabilities, and usability significantly surpassed the early concepts. However, the general concepts in capturing and viewing 3D environments remained the same.

VR allows people to experience complete immersion in a virtual environment by using stereoscopic displays to replace vision and earphones for delivering authentic spatial coherent sound. As a result, people now can think of traveling to fantastic worlds in a blink of an eye or just socializing with friends and colleagues in a virtual space. VR resonates with many domains outside of entertainment today. Only alone in medicine, students use VR to learn the human anatomy [6, 7, 8, 9], simulate medical interventions [10], later use it to treat phobias [11, 12], or provide rehabilitation therapy [13, 14] in professional environments. The investment in VR technology by global companies helps immensely to accelerate the development and research over the world to explore the possibilities.

AR headsets often use semi-transparent displays and stereoscopy to superimpose virtual 3D content to the user's real-world view for an enhanced visual experience. AR opens up the fusion of humankind's accumulated knowledge and creativity into the real world as we perceive it and opens up novel ways to interact with the real and virtual worlds simultaneously. The medical use cases of AR may shift in many instances towards aiding surgeons during surgeries rather than education and simulation, e.g., for minimal-invasive [15, 16, 17], orthopedic surgery [18, 19], and many more.

Upon exploring the possibilities of AR and VR, the question arises as to how it is possible to instantly create the impression of visiting remote locations. The answer is that depth cameras are likely the key components in capturing the real world for 3D telepresence. Traditional cameras are equipped with a single lens and image sensor, whereas depth cameras can determine the depth for every captured pixel from the captured scene. Rather than simply capturing the red-green-blue (RGB) wavelengths of visible light, depth cameras can deliver RGB plus depth (therefore, RGB-D cameras) information. The telepresence system can display the acquired 3D information of the RGB-D cameras in VR, allowing users to immerse in the captured environment.

Telepresence will likely change the landscape of virtual meetings and remote collaboration once the hardware becomes more lightweight and less cumbersome to wear for long periods. Moreover, telepresence has excellent potential for medical usage, among other benefits, in creating the availability of medical expertise anywhere and anytime, particularly for ambulatory and trauma interventions [20]. In Germany, for instance, the shortage of paramedics and skilled doctors is a growing concern in many regions. According to the German Federal Health Monitoring, the statistics show that the number of doctors is rising [21]. However, according to the German National Association of Statutory Health Insurance Physicians, most doctors move towards high-density areas, leaving rural areas bare of highly proficient expertise [22]. We can observe the same movement in other countries in Europe and the US. In cases of accidents involving many victims, such as mass casualty incidents, or natural disasters, it is often logistically impossible to have the necessary workforce and expertise on-site. With a telepresence system, doctors can join local paramedics and surgeons in remote hospitals to assist them in triage and necessary first-contact treatments from anywhere. In a lengthy structured interview in 2014, 40 paramedics were interviewed regarding their thoughts on 3D telepresence [20], with the observation that paramedics believe 3D telepresence would increase the trust in them by other medical professionals, and allowing remote consultants to "paint the picture" (the condition) of the patient with few or no words. In a study of 2020, a study showed that from all questions paramedics, 35% to 41% admit that the conventional teleconsultation solely on video and audio communication will improve all phases between pre-clinical diagnosis, decision making at the treatment facility, and patient satisfaction [23]. The remaining surveyed mainly stayed neutral; only a few declined the idea.

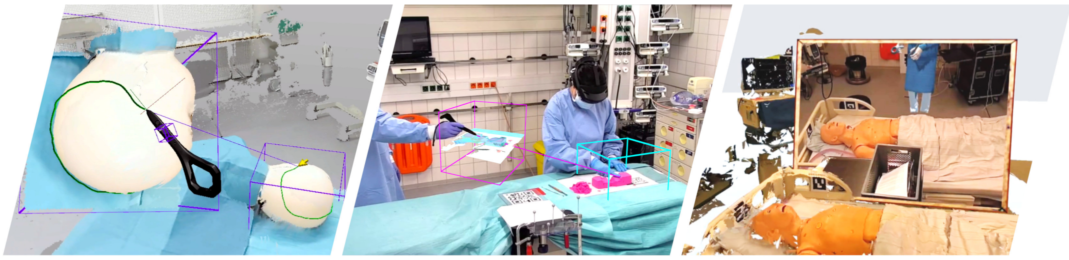


Fig. 1.2. **This Dissertation's Three Interaction Methods to Enhance Medical Consultation.** The Magnorama enhances dynamic interactions with the 3D environment, particularly for creating mid-air annotations. The Duplicated Reality is the continuation of Magnoramas and investigates its ability to utilize 3D space more efficiently and maintain task performance. Finally, the Projective Bisector Mirror provides additional high-fidelity views from the scene reconstructed from RGB-D cameras to enhance the comprehension of the virtual scene.

Making teleconsultation seamless through AR by visually augmenting their natural environment could yield higher acceptance among paramedics and doctors.

What To Expect From This Dissertation

This cumulative dissertation investigates three 3D interaction concepts within a 3D teleconsultation system that utilizes triangulated point clouds from RGB-D cameras to create the feeling of being there. We use AR and VR stereoscopic headsets, allowing local and remote users to communicate naturally and use virtual elements such as annotations and avatars to promote non-verbal communication.

The main contribution of this work focuses on advanced 3D user interaction techniques (Figure 1.2), including the publication of Magnorama [24], Duplicated Reality [25], and Projective Bisector Mirrors [26] to improve workflow and usability of telepresence. In addition, the dissertation presents the *ArtekMed* system (German abbreviation for "**A**ugmented **R**eality **Te**le**k**onsultation für **Med**izin"). ArtekMed is an immersive, asymmetric, real-time, medical 3D teleconsultation system allowing collaboration between multiple remote experts and local users to interact in a face-to-face manner.

The following part of this work discusses and addresses each hardware component's distinct qualities and challenges and their respective roles within the telepresence system. Specifically, the dissertation begins by presenting the camera systems for room reconstruction, followed by a discussion of VR, AR, and telepresence as overarching concepts. Then, as we understand their current state of the art, the content dives deeper into the topic of 3D interaction techniques in virtual and augmented environments and objects as they become increasingly relevant with more complex tasks.

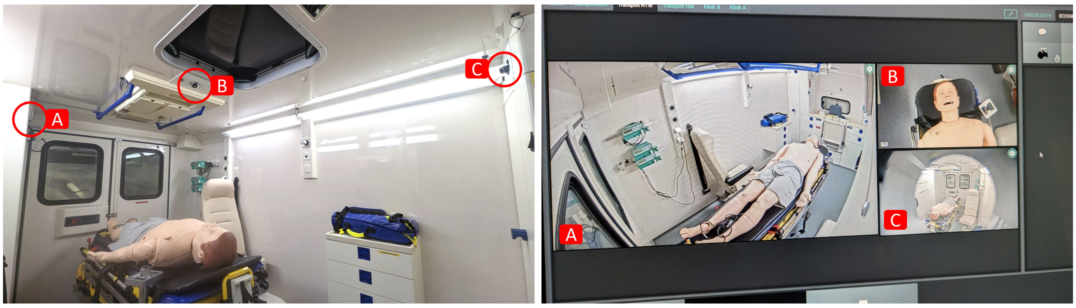


Fig. 1.3. Current Video-Mediated Teleconsultation Inside an Ambulance, captured from the Simulation Center of the Institute of Emergency Medicine in Munich. **Left:** Current ambulances are equipped with cameras on the ceiling. In this setup, two cameras are attached in the corners (A & C), one on top of the patient (B). **Right:** The onboard system streams the live camera streams to the telemedical workstation. The remote expert provides real-time consultation through audio communication.

1.1 Components of 3D Medical Teleconsultation

The following sections present an overview of the components required to develop our immersive 3D teleconsultation system. The development of each component has a rich history of innovation and iteration, leading to the current level of quality. Nonetheless, there remains room for future improvement, particularly for a use case in 3D teleconsultation. Therefore, the following sections detail each technology component's strengths and fundamental issues.

1.1.1 Current State

Many modern ambulance vehicles employ a remote assistance system that combines video and audio to connect remote doctors with local paramedics. The example of Figure 1.3 shows three cameras installed on the ceiling. Two cameras in the corner provide a general overview of the scene, while one above the patient's head delivers a closer view. In addition to the camera streams, remote doctors can access vital patient information, including electrocardiogram (ECG) readings, oxygen levels, and other relevant data. Using this system, remote doctors can consult with local paramedics regarding the necessary procedures. While the system appears complete, several issues become apparent in the field. Firstly, the camera viewpoints are limited and close-up views are partly unavailable. In cases requiring a more detailed view, remote doctors often resort to mobile devices such as smartphones to initiate video calls. In addition, the limited information provided by the camera viewpoints can result in a significant dependence on spoken communication, making it challenging to derive useful information. Furthermore, issues persist with video-mediated teleconsultation. For example, remote doctors often cannot indicate their intention through on-screen drawings, which could provide a persistent form of communication, even if the real-time audio connection or conveyance of information fails. Additionally, in certain situations, such as when using smartphones to communicate, paramedics may want to establish trust with remote doctors by seeing their faces and vice versa. However, the current teleconsultation systems often lack the specific tool to display non-verbal communication, such as gaze or pointing gestures, to guide paramedics.

1.1.2 Virtual Reality

Humans rely heavily on our eyes and ears to perceive the world. Eyesight is one of the primary senses for navigating, anticipating, understanding, and observing space in daily life. A substantial amount of the brain, the visual cortex, is dedicated to processing visual information and building a mental image of our surroundings. Strong visual cues such as stereopsis, where each eye provides a slightly different perspective, allow us to perceive depth immediately. Additionally, weak visual cues such as motion parallax, occlusion, and layers further fine-tune our comprehension of surrounding 3D environments. VR synthesizes these visual cues to create a believable experience that can immerse users into another reality as see in Figure 1.4 (C and D). The concept of immersion is critical to VR and describes the users' belief that they are a part of the virtual world. Regardless, it is easy to break immersion. For instance, current VR systems cannot substitute smell and taste or provide haptic feedback. Especially the inability to provide haptic feedback creates difficulties in 3D teleconsultation, where interactions such as drawing on surfaces or patients become more challenging as users penetrate the virtual surfaces with their hands. There are ways to synthetically create haptic feedback, for instance, through robotic devices [27], though they do not provide a general solution.

The fundamental challenge of haptics in VR is a motivation to develop new 3D interaction techniques. Being not restricted to physical boundaries, we can freely compose virtual content. In anticipation of this dissertation's contribution, Magnoramas can magnify and detach a space within the virtual world, allowing users to interact with the 3D information from dynamic viewpoints. The proposed method significantly improves the perception of the 3D geometry by enhancing weak depth cues through motion parallax and magnification factors as the user gains the ability to quickly change their point of view onto the surface of the reconstruction.

1.1.3 Augmented Reality

AR describes the superimposed visualization of computer-generated virtual objects onto the real world, as seen in Figure 1.4 (A and B). The rendered content may range from single text-based labels attached to physical objects, 3D models, or volumetric medical data. The general consent is that relevant information should be visualized as close as possible to the associated object. The display of virtual data directly atop the relevant physical object is commonly referred to as *in-situ* visualization, a Latin phrase meaning *in place*. In-situ visualization enables users to rapidly and intuitively associate virtual content with the corresponding physical location. Moreover, AR technology can enable users to perceive objects concealed behind physical barriers [28] or even to visualize internal structures of the human body [29, 30] non-invasively. Nonetheless, since AR superimposes virtual content onto the real world, it may obscure the user's view of the real environment, potentially hindering their interaction with the physical world. Therefore, to strike a balance between information delivery and visual occlusion, in-situ visualization must be used adequately, even sparingly, depending on the specific use case. For example, it may be necessary to conceal AR visualization entirely in performance-critical scenarios, such as surgical intervention. Previous research efforts have investigated various visualization techniques to merge virtual content and the physical world into a cohesive, unified view [31, 32, 33, 34]. The antonym of in-situ is ex-situ. Ex-situ

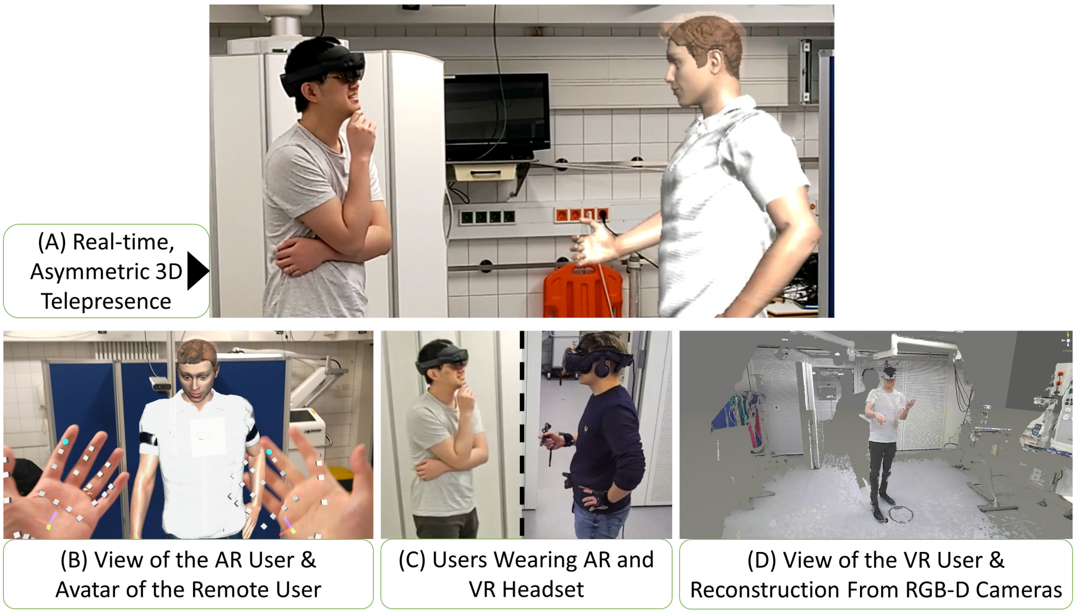


Fig. 1.4. **3D Telepresence using Augmented and Virtual Reality.** (A) Local users wearing an AR headset (here Hololens 2) see the remote-connected VR user. Both users can see the surrounding environment of the local user. (B) The AR user can communicate with the VR user in a face-to-face manner. (C) Both users are physically separated; However, telepresence connects both as if they would share a common space. (D) The feeling of being there is created by capturing the local environment with RGB-D camera and re-project the data into a 3D reconstruction.

visualization often refers to virtual content that is not rendered in a manner to match the position of real-world objects. This type of visualization typically includes virtual content that is presented on a monitor, in VR, or placed in AR without spatial considerations [35, 36]. The optimal visualization strategy may be context-dependent and is often not easily generalized as the focus of attention changes between disciplines and even within disciplines, depending on the focus of the tasks.

Looking forward to the central part of this dissertation, we investigate the visualization of AR content through ex-situ visualization, which refers to the display of information outside of the associated physical space. Our proposed visualization approaches are based on the World-in-Miniature concept and the mirror metaphor to preserve the user's spatial understanding of the real world. As we separate visualization and the physical world, the users' interaction moves further towards the foreground, as the ex-situ visualization may no longer be bound to a single location in space. This work familiarizes the domain of 3D user interfaces (UI) to gain a broad understanding of the context and origin of World-in-Miniature and mirror views.

1.1.4 RGB-D Cameras

The video camera is a crucial invention that has propelled the digital era forward. As the eyes of computer systems, cameras play a vital role in understanding the environment. Computer systems can extract information using computer vision and artificial intelligence to comprehend the semantics of a scene based on a single camera. However, certain information remains hidden even for the most advanced algorithms, such as the 3D geometry of the scene.

Therefore, depth camera systems have been developed to acquire depth and color images simultaneously. There are several approaches to depth camera systems. For example, stereo cameras use two color cameras and compute the disparity from similar features within their captured images. Then, with the distance between the camera (the baseline) and the camera intrinsic known, the depth can be triangulated for each patch based on the disparity. Another type of depth camera system utilizes multiple image sensors, including an infrared sensor. These cameras are usually equipped with an infrared light emitter that illuminates the scene with controlled infrared patterns or light pulses. If the emitter projects a pattern on the scene, the camera can use the projected pattern to estimate the distance of the surface, a method called structured light. Alternatively, if the camera synchronizes the emitter with the infrared sensor, it is possible to measure the distance based on the flight time for the infrared light. This method is called time-of-flight.

Given an RGB-D image pair, depth values can be re-projected into a 3D point cloud and transformed to surfaces by connecting neighboring points with a triangle-based surface. The result of multiple cameras can be seen in Figure 1.4 (D) These surfaces which can then be colorized using the color image. However, point clouds generated from RGB-D cameras are often incomplete, as the camera sensors only capture depth from their point of view. In addition, physical properties such as surface reflectivity may affect the captured data's accuracy. As a result, some surfaces with high light absorption coefficients or high reflectivity may not be visible within the point cloud, leading to missing information and holes in the resulting 3D reconstruction. With hindsight to the contribution of this work, we present the Projective Bisector Mirror method to circumvent the fundamental issue of infrared-based depth cameras. The Projective Bisector Mirror employs the RGB camera stream, in which data is unaffected by the surface properties of the environment, providing a spatially consistent mirror view in cohesion with the point cloud, enabling a more accurate and high-fidelity representation of the local scene.

1.2 Human-Centered Interfaces

Human vision enables depth perception using stereopsis, which relies on slight differences between two viewpoints. This observation suggests that humans naturally interact with their environment in 3D. However, interactions with computer systems are often restricted to two dimensions using monitors and other two-dimensional (2D) displays. While the lack of mature technology to support 3D information visualization is partially to blame, the other side of the truth is that current technology has trained humans to interact with 2D interfaces from an early age. From reading books, writing on paper, interacting with a mouse and keyboard to interact with digital content, and progressing to touch gestures on smartphones and tablets, humans have become accustomed to using such 2D interfaces. These customs are not wrong in any way. However, the computer mouse, in particular, is not an intuitive human-computer interface (HCI), as it disconnects the user's action and the cursor displayed on the monitor. This lack of intuitiveness often requires training to effectively use 2D interfaces, especially for people with low affinity, infants, and older individuals. Even individuals accustomed to using a computer mouse can experience difficulty when the monitor gets rotated. For example, when the mouse movement on a horizontal table no longer matches the accustomed vertical movement on the monitor. In contrast, touch screens are more intuitive as the location of the touch equals the cursor's location. Therefore, designing and evaluating such HCI is crucial to advancing novel 3D display technologies toward general acceptance.

Again, VR and AR refer to computer systems that allow users to interact with information in a 3D space. The human mind can estimate the depth of objects based on the left-right difference from their binocular vision, also known as disparity. One of the first people in history to recognize this effect was the British physicist Sir Charles Wheatstone in 1838 [3], who coined the term stereoscope and described the first stereoscopic viewing device for a set of stereoscopic images. One of the pioneers of computer-based systems is Ivan Sutherland et al.[37] with the system "Sword of Damocles" dating back to 1968. Since then, hardware has continuously improved, leading to off-the-shelve commercially available AR/VR devices today.Unlike monitors, VR and AR adjust their rendering based on the user's movements in 3D, such as their body, head, and hands. Such technology promises that a user can no longer distinguish between what is real and what is a computer generated *(if desired)* but, moreover, can utilize natural gestures to interact with surrounding and virtual information while using the computer-generated views to even further extend beyond what is naturally possible. Independently of whether the user is in VR or AR, the problem arises regarding how users should interact with 3D content, as 2D devices such as computer mice have limited usage in a 3D space. Therefore, the need for intuitive 3DUIs is becoming increasingly critical as hardware for 3D systems continues to improve.

1.3 Intuitive UI

The general behavior and appearance of an intuitive UI are not clearly defined, yet it is normal for the primary goal of any UI is to be *intuitive*. 3DUI naturally should avoid causing user frustration, errors, cybersickness, or other physical discomforts. According to the Oxford Dictionary of Psychology, intuition is the "Immediate understanding, knowledge, or awareness, derived neither from perception nor reasoning" [38]. The word "intuitive" likely originates from the Latin word *intueri*, consisting of the parts *in-* ("into") and *-tueri* ("to look at"). As such, *intuition* describes the link between immediate comprehension and appearance of the concept, e.g., objects or devices, for the user or group. As comprehension does not appear from complete unfamiliarity, comprehension of novel things should be incremental to the existing pre-knowledge and experience of the person [39]. We observe that the smaller the knowledge gap between already familiar concepts and the new concept, the more intuitive the person perceives the new concept [40].

Following up on this observation on the development of intuitive UIs, we further consider the factor of the *user*. As every user may have experienced different challenges in their lifetime, they may also be *familiar* with different concepts. Consequently, developers should consider the target group for which they design the interfaces and incorporate familiar routines or workflows of their users into the systems.

It is a logical belief that individuals belonging to older generations may be more familiar with a diverse range of UI concepts as they have witnessed substantial changes in UI technology throughout their lifetime. It is often assumed that younger generations may be unfamiliar with earlier UIs that utilized manual or physical interfaces, such as dial phones while being

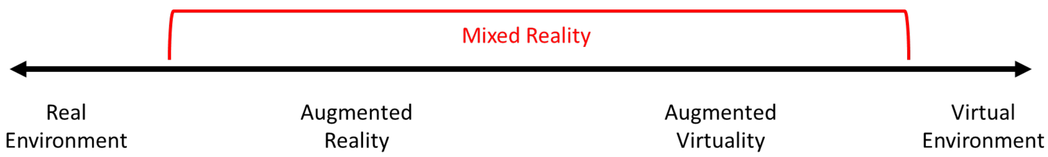


Fig. 1.5. **Virtuality Continuum Based on Milgram et al** [42].

more well-versed with contemporary software-based UIs. However, recent research by Lawry et al. [39] has shown that younger generations can more quickly adapt to any UIs, including those that are thought to be more familiar to older generations. Their study also highlights the existence of a "generational effect", which describes a non-linear correlation between UI types, familiarity, and age, as there is a significant drop in familiarity in the middle-aged (45-49) group. They recommend conducting detailed research on past UIs when designing a device for the middle-aged generation and upward. Sackmann et al. [41] proposed a detailed categorization of generations based on technological advancements, which includes the mechanical (pre-1938), household revolution (1939-1948), technology spread (1949-1963), computer (1964-1980), and internet (>1980) generations.

With the dawn of 3DUIs, it is still being determined whether hardware or software-based concepts provide a smaller knowledge gap for all generations. The significant aspect of 3DUI is the incorporation of the third axis, which was not part of any UIs from previous generations. However, the solutions for 3DUI may eventually be just hidden in plain sight as part of everyday life.

1.4 Domains of 3DUIs

A UI's primary purpose is to communicate the user's intention to the device, and vice-versa, for the device to communicate information back to the user. The characteristic of the user input can significantly vary between systems. In a conventional computer system, user input is, for instance, entered through a keyboard and mouse. On the other hand, specialized use cases such as flight simulators may prefer joysticks, and racing simulators may prefer steering wheels and pedals to acquire more precise inputs and tactile feedback. Specialized input devices increase the immersion of the computer simulation and are an intuitive approach to bridging the gap between reality and virtuality. For example, with VR systems can seamlessly integrate mock-ups of entire plane cock-pits into a 3D simulation environment [43].

3DUI is a general term that includes any interaction a user within a 3D space can perform. A space can consist of a real environment augmented with digital content (AR), a virtual space augmented with real objects (Augmented Virtuality), or be entirely virtually generated (VR). Milgram et al. [42] proposed the virtuality continuum in 1994, which has been widely accepted within the 3DUI community (see Figure 1.5). However, despite the proposed taxonomy, literature often uses the terms MR and AR interchangeably. Moreover, industry and media often prefer to use the term *Extended Reality (XR)* for collectively addressing AR, MR, and VR.

Based on the survey of Jankowski et al. [44], 3DUI methods can be broadly categorized into three categories: *navigation*, *selection/manipulation*, and *control*. In addition, some 3DUI methods, especially for AR, allow users to interact with virtually generated content indirectly by manipulating real-world objects. There is a good chance that the most intuitive interaction techniques are those inspired by nature and have a clear counterpart in everyday life or can be described with a simple metaphor. Healthcare is one of the slowest commercial fields to adapt to novel ideas and technologies. As such, it is possible that optimal visualization and interaction techniques for healthcare-related systems may already exist in other domains, which can be translated and extended into medically feasible solutions. In this article, we explore some of the most prominent types of 3DUI used outside of healthcare and highlight notable work in their respective areas.

1.4.1 Navigation

Prominently relevant for VR is when the virtual space is larger than the user's physical environment or users want to move between locations quickly. Hence, users want to move in the virtual space without moving the same distance or direction in the physical space. The act of traveling describes the locomotion aspect of navigation, also understood as moving the user's point of view to a different location within the virtual world. We can coarsely classify the research in the navigation into walking, steering, and selection, or manipulation-based travel. Traveling for navigation is a very elaborated and highly researched topic. To keep it brief in this overview, we present only the ontology of existing directions. Navigating through walking is further categorized into real [45, 46], redirected [47], and scaled walking [48]. Furthermore, with additional tracked body parts such as the arm and legs of the user, walking-in-space [49, 50] or the human joystick [51] becomes eligible. Finally, we categorize steering-based travel into gaze-directed [52], hand-directed [53], torso-directed [54], and lean-directed steering [55, 56]. While walking and steering focus on moving in a specified direction, selection and manipulation-based traveling focus on reaching a particular target location. Users may use a minimap-representation showing the bird view of their current environment [57, 58] to move their view directly [59, 60] or plan a route towards it [53, 61, 62]. Manipulation-based travel allows users to reposition the camera viewpoint [63] or an avatar (e.g., combined with the World-in-Miniature [64, 65, 66]) within a map or digital representation of their environment. There are more methods for navigation inside a virtual environment. A more detailed overview is provided by LaViola et al. [67], while Boletsis et al. [68] published an extensive literature review on locomotion within a virtual environment.

1.4.2 Selection and Manipulation

Due to the inclusion of depth in 3D environments, the users must bridge the spatial distance to interact with virtual objects. With conventional rendering for 2D monitors, the rendering pipeline projects 3D objects into the 2D screen space. Users can select 3D content on the screen, as often seen in video games or 3D modeling software, using the mouse that implicitly performs a ray-picking algorithm to decide which virtual object the user selected. When users perceive 3D content with a stereoscopic display, they can select objects behind other objects by moving their controllers in three dimensions. Such direct interaction does not allow

Fig. 1.6. **The Concept of World-in-Miniatures** refers to the miniaturization of the virtual world into a smaller 3D representation. This concept was first published by Stoakley et al. [69]. This image is taken from the exectuable demo of Yu et al. [70].

manipulation of distant objects outside the natural reach of the user. The study of selection and manipulation techniques for 3DUI classifies approaches in isomorphic and non-isomorphic views. The isomorphic view correlates perceived actions one-to-one with the user's inputs, such as hand motion. This approach is often more accessible for humans to comprehend as it is similar to how humans naturally interact with the real-world outside virtual environments. Examples are the virtual hand [71], finger interaction [72, 73, 74], bubble cursors [75], and scaled manipulation such as PRISM [76]. However, the limitation of such isomorphic methods is often limited by the input device, e.g., its tracking range and the reach of the human range of motion. In contrast, non-isomorphic views provide the user with supernatural interaction techniques that may only exist in virtual environments. Paradigms to interact with distant objects even without moving physically inside the virtual environment include WiMs [69, 77], GoGo interactions [78], and ray-casting methods [79] including Hand-centered Object Manipulation Extending Ray-casting (short: HOMER) [80, 81] and the fishing reel [82]. Moreover, volume-based pointing techniques allow more efficient selection, such as the flashlight [83, 84] and sphere-casting [85]. While supernatural interaction techniques do not follow natural realism, these interaction techniques maintain usability by imitating metaphorical ideas within the user's imagination, such as laser beams or elastic arms. These presented methods are a subset of all existing interaction techniques and, among others, do not include methods specialized for touch or gesture-based interaction.

Manipulation covers user interaction with virtual objects, e.g., repositioning them, changing their shape, or triggering function calls. A unique manipulation technique uses indirect proxies, particularly WiMs [69] (as depicted in Figure 1.6) and Voodoo Dolls [86]. First, the user interacts with a digital replication. The system then replicates any interaction onto the original location or object. The variations on WiMs are further iterated in section 2.1.3, as it is the fundament of two main contributions of this dissertation.

1.4.3 Control

Often users want to perform more complex interactions that go beyond 3D manipulation. For example, users should be able to initiate the execution of commands controlling the system, including but not restricted to saving and loading a system state, starting a calibration routine, opening files, and similar actions. In conventional applications, *windows, icons, menus, pointer* (WIMP) interfaces allow users to communicate with the underlying system. The pioneers of WIMPs had 2D applications in mind. Towards 3D interfaces that include gesture recognition, users could trigger controlling functions upon performing pre-defined hand gestures [87, 88, 89, 90] or free-hand sketches [91] and avoid conventional WIMP interfaces. Moreover, the concept of diegetic UI [92, 93] - UI elements that are integrated seamlessly into the 3D environment or objects and are part of the virtual world - could replace WIMP-based UIs to facilitate immersion in 3D space.

1.4.4 Interaction Through Real World Manipulation

AR systems that track the nearby environment of the user allow them to place virtual content seamlessly in the real world. When the AR systems acquire sufficient data to understand the real world, they enable users to communicate with the computer system by interacting with the objects inside the real world instead of interacting with virtual WIMP elements. A system can accomplish one of the simplest forms of such AR by positioning a 3D model on top of a tracked marker with optical marker tracking [94]. As it tracks markers in real-time, users can use the physical location of the markers to communicate with the computer system, e.g., by moving and rotating them. For instance, Cai et al. [95] used this method to demonstrate the structure of chemical compounds playfully, Restivo et al. [96] for teaching power circuits by arranging and switching out different AR markers, or enable 3D pop-up for educational books [97, 98].

Physical world interaction allows interaction paradigms that are far more abstract than WIMP interfaces. Related work demonstrates how tracking the 3D geometry of the real world allows users to control virtual content through different sensor modalities. A frequently investigated method includes body-tracking. For example, the estimated position of human joints enables systems to replace or overlay the user's body pose with computer-generated content, such as for anatomical education [6, 99, 100] or virtual fitting rooms [101]. In other cases, a depth sensor captures the geometrical profile of an area, e.g., to playfully visualize how elevations in terrain affect the emergence of lakes in a sandbox, and in-situ visualize the topography with a projector [102].

1.5 3DUI in Medicine

UI is designed with the target group in mind, meaning there are different focuses between patients, nurses, and surgeons. Moreover, there are differences in focus between different fields of surgery, e.g., to name a few but not restricted, orthopedics focuses on bone visualization and targeting, neurosurgery on precision tasks, and visceral surgery for artery and vein detection.

There are distinct differences between a patient interacting with the UI compared to medical experts. First and foremost, UIs for patients are ideally designed to be one-click wonders and reduced towards the minimum required user input with predictable interface elements since the system has to look out for any age group and people of any technical proficiency. Telemedicine is the most frequent scenario in which patients interact extensively with a digital user interface. Modern telemedicine systems provide apps for commercial hand-held devices which allow the upload of documents, monitor health values, and initiating video calls with the responsible medical expert. A recent scientific review by Nguyen et al. [103] indicates that patients are satisfied with existing telemedical apps. More interestingly, they found that providers were only satisfied if they were involved in the development process.

When medical staff interacts with technical devices, they want to avoid bothering with complex UIs due to often limited time and training. For example, intra-operative user interfaces for surgeons may not use any active interaction at all to avoid occupying the hand of the surgeon. Instead, they may rely on feet pedals or gaze interactions to control devices or ask an assistant to interact with surrounding machines. However, precisely this is the crux of the matter. Designers and Developers should first understand the requirement of the medical procedure to decide what might be complex and which designs are easy to use. There exist approaches to circumvent this problem, e.g., an approach with rapid prototyping [104], eventually using a Wizard of Oz approach [105] to manually imitating the seemingly automated functions. The general guideline in designing intuitive UI is to regularly involve the UI's target group in the development process and performing controlled user studies.

Established medical systems frequently use 3D visualization to show doctors volumetric data, e.g., acquired from Computer Tomography (CT) or Magnetic Resonance Imaging (MRI). For example, a typical view for visualizing and exploring CT data consists of four separate views to fully understand the spatial relationship between them. It is a combination of views for each of the three axes and a volume rendering of the CT, as seen in Figure 1.7. Surgeons rarely use 3D visualization during surgery since the slices provide a more detailed view and allow conventional WIMP-based control. However, some navigation systems may use the volume coupled on tracked needles to slice through the volume automatically. Methods exist to visualize acquired 3D data in situ on the patient, especially with the help of head-mounted displays, for instance, for orthopedic surgery [106] or spine surgery [107, 108]. However, without advanced occlusion or perceptual rendering techniques [109], overlaid virtual objects are more like to obstruct the surgeon's view of the patient and increase visual clutter [110, 111, 112]. Immersive telepresence challenges a remote user with similar issues. VR and AR provide users with often unfamiliar controllers and input schemes. A short training would suffice to familiarize a user with the control schemes. However, with 3DUI, they would lose WIMP interfaces due to missing mouse and keyboard. To replace or even create a more efficient working scheme, such VR and AR systems implement 3D interaction techniques that go beyond natural interaction to work with the virtual space faster, more accurately, and more intuitively than they could if they were in the local area in person. The input scheme is a critical element within the design for a 3D interaction technique and must be cautiously chosen.

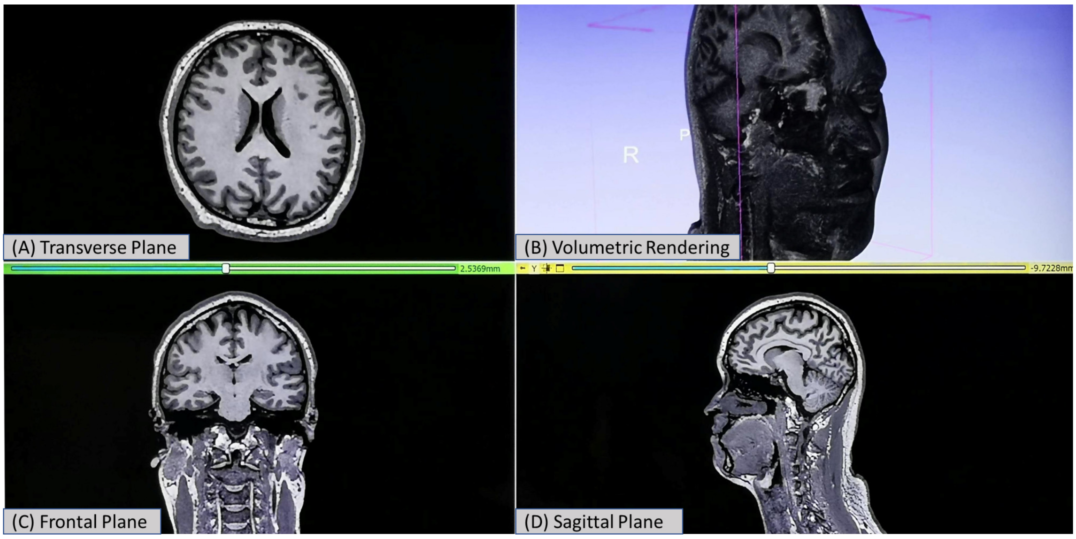


Fig. 1.7. **Exemplary UI for the Three Cardinal Planes and the 3D Rendering** of the author's head MRI data in 3D Slicer. Surgeons use the cardinal planes (A, C, D) to target anatomical structures during the surgery. The volumetric visualization (B) is rarely used during surgery for guidance. Volumetric data is currently used most for pre-operative planning to fully understand the patients' anatomy.

1.5.1 User Input

A computer can only interpret the information it receives through its connected sensors. Whether a gaming console or a high-end virtual reality simulator for laparoscopic surgery, humans must use some form of Human-Computer Interface (HCI) to communicate with the machine. Different types of HCI have their advantages and disadvantages. In the case of AR/VR technologies, standard HCI options include hand gestures, controllers, voice, and body language. While hand gestures are intuitive for humans as they are a part of our daily lives, the tracking and interpretation of these gestures may not be as precise as using a controller. Similarly, body language is also challenging to track and interpret. Voice input allows for hands-free interactions but can be challenging to use in noisy environments. In addition, it may become tedious if the voice recognition system frequently misunderstands or misinterprets commands, especially after repeated or ongoing usage.

The choice of the HCI can significantly impact the users' perceived usability of the application. There are no formulas or guidelines for choosing the correct HCI, as it generally depends on the use case. In healthcare, however, usability is only one factor in the decision. For example, in use cases with close contact with patients, surgeons must adhere to asepsis guidelines to maintain a sterile environment for patients and surgical tools. This restriction implies that non-sterilizable technical equipment is generally not allowed on the surgeon. However, when a surgeon can think of controlling special input devices with their tongue or feet, they may be allowed to be used. In all other cases, surgeons must rely on using hand gestures and voice commands or ask nurses to interact with devices in the operating theater, including AR systems.

Controllers, particularly for VR systems, have undergone significant advancements in recent years. Some of the most well-known VR controllers, such as the Vive and Oculus controllers, are designed to replicate natural hand posture when held. The developers positioned buttons and triggers such that clenching the hand to a fist would pressure all important buttons for the system to register the pose of the hand. Hence, the system could still estimate a hand posture from the button sensors without tracking each finger individually. In addition, particular VR controllers, such as the Logitech Ink Pen, specialize in precise pointing inputs by imitating a conventional pen. A study [113] in 2022 found that the VR pen was the most accurate HCI to create accurately drawn annotations when compared to hand tracking and VR controllers.

1.5.2 Responsibility of Virtual Content

The presentation of virtual content sets AR and VR apart from the real world, but it also means that AR and VR may encounter similar issues in the real world. For example, imagine an old basement storage room that has not been organized for an extended period. As new objects are carelessly added into the storage room, it becomes increasingly cluttered. At the same time, it becomes increasingly difficult to identify individual objects, and it converges into one object with the name of "The Rage Cave"[1]. Translating this metaphor back to AR and VR, it is crucial to carefully choose the presentation of information to the user to avoid overwhelming their perception. In AR, visual clutter is particularly delicate as the real world is the primary focus of the user's attention. Presenting too much information can make it difficult for users to distinguish different objects and their relative positions in space. Further, it is imperative that the AR content in surgical environments, both open surgery or video-based, does not unnecessarily occlude the actual view of the patient and endanger the interventional procedure.

The UI designer's responsibility is to avoid visual clutter and ensure a clear and concise presentation of virtual content. This guideline goes for both visual and auditory perception. In this context, over-reliance is another significant issue in AR that people often dismiss. Over-reliance refers to the users' heavy dependency on performing their tasks and becomes a problem when they can only perform the task with it. Additionally, they may become unaware of critical factors for their task when the AR system does not indicate them [112]. Therefore for surgeons, the responsible way to aid them with AR is not a complete guidance system dictating how to precisely position their tools and follow a rigid recipe for the medical procedure. Rather, AR should aid them in performing tasks more efficiently and accurately, which they could accomplish even without AR. For example, a teleconsultation system could even integrate Artificial Intelligence to aid remote doctors in consultations. Nevertheless, for the best of the surgeon responsible for the patient's well-being, AI should only partially automatize the process since, otherwise, there will be no room to react to unexpected problems.

Visual clutter and over-reliance can arise from poor design choices made during the conception phase of the UI's design. Like most user-centered applications, virtual content should consider the intended use case and the user's requirements.

[1] Needless to say, this is a reference to the Narvis Lab's deepest of treasure rooms.

1.6 Outline of the Dissertation

Until now, we have discussed the benefits and unresolved issues associated with hardware and software-based concepts in developing a 3D telepresence system. It is crucial that every component of such a system, including VR, AR, RGB-D camera, 3DUI, and HCI, play hand-in-hand effectively to provide an uninterrupted user experience. The present cumulative dissertation unifies three research publications into a cohesive and coherent narrative. The introduction chapter section 1 unfolds the most notable history of 3D telepresence, from its conceptualization over the first prototypes to the current state. Further, it provides a brief overview of existing 3D UI followed by a focused view of the current state of UIs in medicine. Section 2 provides an in-depth view into state of the art, including telepresence, teleconsultation and collaboration, World-In-Miniatures, and mirrors in AR/VR. Before handing over to the core publications, section 2.2.1 describing the state of our 3D teleconsultation system that is utilized for the following works. Following the presentation of all three core publications, section 6 discusses topics building on top of the core publications, including the method's influence on presence, immersion, visual overload, and over-reliance on virtual content.

The first publication, "Magnoramas: Magnifying Dioramas for Precise Annotations in Asymmetric 3D Teleconsultation" [24], presents a novel approach to immersive telepresence that shares similarities with the World-in-Miniatures concept. However, Magnoramas differs by selectively magnifying a local 3D region of interest instead of miniaturizing the virtual world. The Magnorama method indicates the magnified region with a 3D wireframe cube that the user can manipulate via a VR controller. When the user creates annotations within the Magnorama, they are instantly replicated to the region of interest and appear as if the user created them directly. When the Magnorama's size is larger than the region of interest, the Magnorama magnifies the content by the size difference. Due to the magnification and dynamic free choice of viewpoint, users can precisely place annotations within the Magnorama. Hence, Magnoramas distances itself from the well-known World-in-Miniature concept due to magnifying content rather than miniaturizing them and replicating user-created content.

The second paper, entitled "Duplicated Reality for Co-located Augmented Reality Collaboration" [25], builds upon the concept of Magnoramas introduced in the first paper. However, this work shifts the focus from VR to AR and emphasizes on co-located interactions. Duplicated Reality creates a movable and scalable virtual replication of the real world based on a user-selected region of interest. This approach enables co-located teleconsultation, facilitating a teacher-student relationship between users without physical interference. This concept is particularly interesting for complex or rare surgical procedures requiring expertise from different surgical specializations.

The third and final paper presented in this dissertation is entitled "Projective Bisector Mirror (PBM): Concept and Rationale" [26], which primarily focuses on addressing one of the remaining challenges of real-time reconstruction-based telepresence, namely, the low geometrical fidelity and texture resolution of point cloud reconstructions. The PBM tackles this issue by generating a virtual mirror image within the 3D reconstruction space. The PBM projects the image onto the 3D bisector plane between the user's viewpoint and the capturing camera to create a mathematically correct mirror image from an image. This novel approach aims to improve the user's comprehension of the reconstructed environment and synergizes well with any sea of camera setup of the 3D teleconsultation systems.

Through the three papers presented in this dissertation, we address some of the most significant challenges in the field of immersive 3D telepresence. The Magnorama and Duplicated Reality introduce dynamic tools that enhance the precision of user input despite the lack of haptic feedback. Moreover, Duplicated Reality eliminates the risk of visual clutter on the patient by providing a lightweight AR visualization and moving the consultation space to the proximity. Then finally, the PBM resolves the issue of low-fidelity real-time reconstruction by providing a spatially consistent virtual mirror image of the camera feed within the virtual space.

1.6.1 Present Context

As of now, in the year 2023, healthcare is one of the slowest advancing domains in integrating novel technologies. We can likely attribute this observation to the high precision requirement and the continuous validation and certification of devices for clinical use. The proper observance is necessary to ensure patient safety, and the established safety guidelines in modern medicine are a positive development compared to practices in the history of medicine. The healthcare industry, particularly surgery and emergency medicine offers a promising application for 3D telepresence. Although telepresence has gained considerable interest, telemedicine still typically refers to a remote doctor's visit, akin to a video conference with transmission and safe-keeping of confidential patient information [114, 115]. To integrate 3D telepresence into daily routine, further advancements in ergonomics, computer networking, visualization, and 3D input methods are necessary. When these challenges are overcome, 3D telepresence could prove to be a valuable asset in the future of healthcare. Surgeons have expressed interest in incorporating 3D VR and AR tools to aid them during procedures, as visible from mentioned related work for various medical disciplines; however, the challenge lies in ensuring the patient's well-being is not being compromised.

This dissertation has been mainly funded by the Germany Ministry for Education and Research (BMBF Grant No. 16SV8092) as part of the ArtekMed project. The main contribution of ArtekMed is its real-time point cloud capture system coupled with remote 3D telepresence for a virtual face-to-face consultation. In particular, to allow remote teleconsultation for medical emergencies such as mass-casualty incidents. We explain the system in detail in section 2.2.1.

1.6.2 Scientific Context

This document follows the guidelines on publication-based, cumulative dissertations. It includes three primary publications from 2021 and 2022 that were published in the International Symposium on Mixed and Augmented Reality (ISMAR) and its sister conference IEEE Conference on Virtual Reality + User Interfaces (IEEE VR), from which one is a conference paper. Two were published as Journal articles in the IEEE Transactions on Visualization and Computer Graphics (TVCG). The highlighted publications include

- Magnoramas: Magnifying Dioramas for Precise Annotations in Asymmetric 3D Teleconsultation [24] @ IEEE VR 2021
- Duplicated Reality for Co-located Augmented Reality Collaboration [25] @ IEEE VR 2022 & TVCG
- Projective Bisector Mirror (PBM): Concept and Rationale [26] @ ISMAR 2022 & TVCG

The publication on Magnoramas received the annual "Honorable Mention Demo" award.

Additionally, the following articles relevant to ArtekMed were published from 2020 to 2022 with this dissertation's author as a first or co-author; However, due to the thematic context cumulative dissertation, those articles were not selected as core publications:

- Mixed Reality 3D Teleconsultation for Emergency Decompressive Craniotomy: An Evaluation with Medical Residents [116] @ ISMAR 2023
- Real-Time 3D Reconstruction Pipeline for Room-Scale, Immersive, Medical Teleconsultation [117] @ Applied Science 2023
- Medical Augmented Reality: Definition, Principle Components, Domain Modeling, and Design-Development-Validation Process [118] @ Journal of Imaging 2022
- Avatars for Teleconsultation: Effects of Avatar Embodiment Techniques on User Perception in 3D Asymmetric Telepresence [119] @ ISMAR 2021 & TVCG
- Real-Time Mixed Reality Teleconsultation for Intensive Care Units in Pandemic Situations [120] @ 3DUI contest of IEEE VR 2021
- Comparison Between Video-Mediated and Asymmetric 3D Teleconsultation During a Preclinical Scenario [121] @ Mensch und Computer 2021
- Spatial Exploration With a WiM for Capturing 3D Dioramic Snapshots [70] @ 3DUI contest of IEEE VR 2022
- Eyerobot: Enabling Telemedicine Using a Robot Arm and a Head-Mounted Display [122] @ CURAC 2020

Most of these publications are placed and referenced throughout the dissertation.

Part II

3D Teleconsultation System

A Closer Look at Medical 3D Teleconsultation

2

> *It is already possible by ingenious optical contrivances to throw stereoscopic photographs of people on screen in full view of an audience. Add the talking phonograph to counterfeit their voices, and it would be difficult to carry the illusion of real presence much further.*
>
> — **Scientific American 1877**

2.1 State of the Art

The first part of this chapter analyses the related work on telepresence, teleconsultation and collaboration. The second part of this chapter details 3DUIs, particularly emphasizing on World-in-Miniatures, as they are highly relevant for developing Magnoramas and Duplicated Reality. Additionally, a detailed examination of the most relevant work in the mirror paradigm in AR/VR is presented, as this is an essential component of the final work on Projective Bisector Mirrors.

While the main focus of this dissertation is 3D UI in medical 3D teleconsultation, the following overview of the literature will cover the most relevant work independently of the domain. The concept of asynchronous general doctor visits for patients [123] is outside the scope of this work.

2.1.1 Immersive 3D Telepresence

The early conceptualization of telepresence started in the late 19th century, as stated in the introductory chapter, with the advance of the sonograph as the earlier version of the telephone. Since then, the concept of teleportation, i.e., moving to another location in a blink of an eye, has been an integral part of science fiction. While teleportation, to this date, has only been successfully demonstrated on the quantum level [124], immersive 3D telepresence is the closest feasible concept of instantaneous "relocation" known to humanity. The active research in the modern era likely gained traction around the publication of Marvin Minsky [125] in the OMNI magazine in June 1980. Inspired by his friend Patrick Gunkel, Minsky proposed the term telepresence to create the sense of *being there* at a remote location. He indicates earlier work on teleoperators, including Handiman and other robotic systems, but mentions their halt in development around 1960. He says the cost of development was the primary block in progression. IJsselsteijn et al. [126] published a comprehensive article on the early steps of telepresence.

Fast forward to the late 20th century, Henry Fuchs et al. [4, 5] propose to use a "sea of cameras" to capture the 3D geometry of an operating theater and to visualize the acquired 3D data in a remotely connected VR headset or with stereoscopic projections to virtually extend the wall of offices with the acquired 3D information [127, 128, 129]. One of the early bottlenecks was the availability of high-resolution depth cameras. Lidar (derived from "light detection and ranging") sensors were invented in 1961 by Hughes Aircraft Company and experienced a range of iterations afterward; However, due to the nature of laser-based range acquisition, it was not possible to acquire entire 2D images with depth information in a reasonable time and resolution. As real-time acquisition was not possible, multiple works have investigated the use of accumulated data over time from Lidar sensors and less accurate depth from photogrammetry (extraction of depth from 2D images, such as through stereopsis or texture) to build a complete reconstruction of a room and register the 3D data with a color camera [130, 131]. Before 2010, depth cameras increasingly gained attention when the first RGB-D cameras were released [132]. Unfortunately, they did not inhabit the required quality for telepresence.

The release of Microsoft's Kinect v1 camera in 2010 is a significant milestone, as it was an easily accessible depth camera using structured light with near-infrared light to create high-quality 3D point clouds with 30 frames per second [133]. Maimone et al. [134] and Dou et al. [135] demonstrate a bi-directional telepresence system using the point cloud data from the Kinect v1 and a wall of monitors. Additionally, the Kinect v1 was capable of body tracking, allowing users to change the view on the point cloud based on their viewpoint in front of the wall of monitors. A symmetric setup on both sides of a Kinect-Monitor setup allowed users to look at the person on the other side into the eye and further created a sense of depth perception through motion parallax. Pejsa et al. [136] created a similar experience using a projector, while Beck et al. [137] created it with a cave automatic virtual environment (CAVE) system. The latter approach elevated the 3D perception through stereoscopic 3D glasses. Outside the application for telepresence, such 3D reconstruction could be recorded and utilized for education and training, e.g., of surgical procedures [138], allowing for manipulating the temporal aspect through the means of rewinds, pause, and speed of playback. The point cloud provided by the Kinect v1 was far from perfect, as it contained depth inaccuracies, noise, and missing data that resulted in holes within the reconstruction. As a response or moving forward, a series of publications investigate the fusion of point clouds from multiple cameras [139] or a moving RBG-D camera, and methods to clean up the acquired 3D representations [140, 141]. Further, the history of AR headset technology is relevant for 3D telepresence as it dictated the accessibility of its visual medium. AR headsets exist as optical see-through (OST) and video see-through (VST) variants. Early adopters of heads-up displays are the military institution Hughes Aircraft Company for aviation [142], which showed simple shapes such as text and lines in front of a user's eye. However, it was Ivan Sutherland [37] to create the pioneering work of OST stereoscopic displays, similar to today's AR headsets. The first commercially available OST headsets include the Epson Moverio series, to display static AR screens. When coupled with external outside-in tracking, these glasses could be used to create a free-viewpoint AR telepresence system [143].

2016 marked another milestone for immersive technologies through the commercial release of Microsoft's Hololens. The Hololens is an optical see-through head-mounted display with robust SLAM tracking[144], capable of showing virtual objects stabilized in the real world. In addition, Microsoft itself published an impressive demonstration of its capabilities for telepresence, named Holoportation [145], merging a remote user into the local space of a second user. At the same time, two major companies, Facebook and HTC, released powerful tethered VR headsets, the Oculus Rift and HTC Vive, allowing more streamlined research into immersive technologies and 3D interaction techniques. The advances in accessible headsets kicked off many newer telepresence systems, demonstrating the use of telepresence in maintenance [146], remote teaching [147], and medical teleconsultation [148, 149, 150]. A recurring critique against VR headsets is their bulkiness and stationarity. This argument is valid before and even after the increased accessibility of VR headsets. Research has led to a diversification in approaching telepresence using mobile and hand-held devices [151, 152, 153]. In the long term, 3D solutions are likely more desirable than mobile camera-based solutions since hand-held cameras often restrict the remote user to its single, inflexible view.

The challenge of a sea of cameras is that the 3D geometrical information of the reconstructed environment will be prone to occlusion and errors in depth acquisition and estimation. There are several causes for such degradation of the 3D reconstruction. Issues could be related to a surface's reflection and absorption characteristics of light with certain wavelengths or inaccuracies within the depth sensor itself. Virtual telepresence with a 360° camera, e.g., [154, 155], circumvents issues in missing or inaccurate information as they capture stitched 2D video sequences from the local space. Such camera systems use multiple cameras pointing in all directions. Imaging processing software then stitches them to a 360° panorama. When immersed, the VR user can freely choose their view orientation. However, their movement is limited to the rotational aspect. A completely free choice of their point of view inside the local space is impossible. Further, the view of the local space is limited in its 3D representation. As the cameras do not acquire 3D geometrical information, there are no direct methods to interact with the captured environment, and they are often limited to pointing gestures. Therefore, despite the drawbacks of 3D RGB-D cameras for acquiring the reconstruction of the local space, it is yet more desirable to have free choice in perspective and intuitive interaction with the 3D reconstruction in the medical domain due to the complex geometry of human anatomy.

2.1.2 Teleconsultation and Collaboration

While telepresence often covers the idea of being there at another location, expanding it with the *intention* of consulting and collaborating adds a whole new layer on top of exclusive telepresence. During a consultation, we can assume an asymmetric relationship between both users, for instance, in the form of teacher and student or expert and novice. Conversely, collaboration implies a balanced relationship where both users are peers in terms of knowledge but can likewise involve two experts with different areas of expertise. Both types of telepresence with intention in solving a task involve communication beyond visual and auditory cues from *being there*. They may include active interactions, including manipulation of objects and the environment, or creation of new content (often drawn annotations [147]) to convey knowledge. Techniques to elevate consultation and collaboration may find usage in co-located AR scenarios simultaneously with remote telepresence. For instance, the "Studierstube" [156] by Szalavári et al. demonstrate co-located users with AR headsets to inspect and interact with a shared virtual representation of a mathematical structure. This concept could likely be replicated within virtual 3D telepresence while maintaining the benefits of collaborative learning. Shaping the landscape of shared interaction, Billinghurst and Kato [157] present a variety of techniques for displaying virtual elements in collaborative environments. These techniques include floating virtual displays, drawing boards, and dynamic video panels, among others.

The use of drawn annotations in VR/AR consultation systems provides a non-verbal, visual element for communication between users. However, while humans interact with their environment in 3D daily, achieving precise depth perception in virtual environments remains a recurring challenge. This issue is reflected in creating drawn annotations, as users would draw them closer and further than they intended [158]. Multiple studies have replicated the findings with mid-air methods and compared them with drawing with physical feedback [159, 160]. Hence, related work on user interaction techniques has investigated various approaches for creating mid-air annotations more intuitively and precisely in 3D. The typical task in these studies is to trace along given virtual lines. While the precision can be trained and improved through repetition to some degree [161], systems can further aid users with constraining [162] and non-constraining guides [163]. Machine Learning and gesture-based methods may help detect common or known shapes and improve the quality of the annotation [164, 165]; however, they would not recognize any shapes unknown to their algorithm. Previous work used annotation frequently for immersive consultation systems with a collaborative intention [166, 167, 168]. The most significant difference in use cases for their related work to ours is that for medical consultation, drawn shape and precision of unfiltered mid-air annotations are essential for unambiguous knowledge transfer. This missing research is a substantial gap in the literature.

The increased accessibility of technology, including AR/VR headsets and mobile devices, has enabled researchers to shift their focus towards studying 3D interaction techniques in telepresence rather than solely on the hardware itself. As a result, there has been a considerable increase in the number of publications on this topic from around 2015 to the present. Such work may include consultation with a digital twin, also known as virtual replicas [169, 170, 171], avatar variations [172, 173], and 360° cameras [154]. In addition, several works further investigate the communicative aspect of collaboration, investigating the perception of other users [174], sharing awareness cues [175], gaze [176, 177] or simple annotations [151].

2.1.3 World in Miniature

In our work, we employ the concept of WiMs to achieve a magnified view of a specific region of interest rather than the more common approach of miniaturizing the entire environment. This section provides a detailed overview of existing research on WiMs.

The term World-in-Miniatures was coined first by Richard Stoakley, alongside Matthew J. Conway and Randy Pausch [69] in 1995. Initially designed for remote interaction via a digital twin of a virtual environment, WiM's enables users to quickly select and manipulate distant objects within the environment by interacting with their miniaturized counterparts. Such an approach naturally minimizes the effort required to manipulate objects within the virtual environment while allowing users to inspect the virtual environment from different perspectives. They describe the WiM as "*a single unifying metaphor for such application independent interaction techniques as object selection, navigation, path planning, and visualization*". They laid the groundwork for the following works and defined several ideas for improving the workflow with WiMs, e.g., scrolling, clipping, zooming, and multi-user interactions. Shortly after the first publication, Pausch et al. used the same select and move interaction to re-locate a virtual avatar for virtual locomotion [64]. When users lift the miniature avatar from the ground of the WiM, their viewpoint starts to lift similarly, making users think they are flying. Several more works were published since then looking into navigation through virtual space using WiMs [65, 178, 179, 180, 181, 182, 183] and in-door navigation using AR [184].

Notably, Wingrave et al. [185] overcame the limitation of WiMs in environments of varying scale, e.g., room-scale versus city-scale, by investigating the effects of scaling and scrolling of its content. As virtual worlds expand in size and complexity over the years of increasing computation power, the need for adaptive scaling becomes increasingly important. In particular, complex virtual spaces, such as indoor environments with multiple rooms, should enable users to gain a comprehensive overview of the entire space while also providing the ability to zoom in on specific areas of interest: Trueba et al. [186, 187] and Andujar et al [188] demonstrated how WiMs might be extended with rendering techniques such as clipping, slicing planes, and hand-based disocclusion to limit the size of the visualization. They further added culling to prevent walls from occluding more interesting virtual objects.

WiMs decouple the user's position from the interaction. For instance, as described in the original work of Stoakley et al. [69], several users may use their own WiM to collaboratively interact with the same original objects simultaneously without colliding with each other. Equally, WiMs allow asymmetrical collaboration in which user one interacts with the environment in immersive VR while a second user interacts through the WiM. This is, for example, investigated for spatial exploration [189, 137].

Another yet exciting characteristic of WiMs is their recursiveness. Since the user resides within the original-sized virtual world as the WiM depicts it, actions performed on the original world affect the WiM, which may affect the original world. Bluff [77] iterated on this idea and brought forward even more thought-provoking phenomenons of WiMs, such as changing dimensions when users fall out or enter the range of a WiM, the interaction between multiple WiMs, and edge-cases where virtual objects should get duplicated.

When transferring the concept of WiMs for replicating spaces to single objects within the physical or virtual world, we approach the concept of Digital Twins. In the strict meaning of Digital Twins, they are virtual replicas of physical objects in the real world interconnected through (real-time) data streams [190]. As the shape and functionalities of the Digital Twin are optimally converging towards the original real objects, they serve as a remote proxy with which users can interact. An exemplary and notable demonstration of Digital Twins for teleconsultation is presented by Oda et al. [169]. Digitally replicating existing virtual models are known from the work of Pierce et al. on Voodoo-Dolls [86].

The concept of scaling adds an interesting layer worth investigation for collaboration, as they change the perception of virtual objects relative to the user. For example, minifying virtual rooms could imply that the user increases in size, effectively changing them into a giant. In reverse, magnifying space would let users appear to be shrinking in size. This effect is investigated for asymmetric telepresence systems in both directions, i.e., AR seeing multi-scale VR users [191] and VR user seeing multi-scale AR users [192].

2.1.4 Mirrors in AR/VR

Mirrors have been part of AR since its beginning and find early usage in controlling the path of light for AR visualization devices. For instance, Italian architect Filippo Brunelleschi demonstrated linear perspective in 1415 using a mirror and a painting, a principle crucial for modern computer-based rendering techniques [118]. Further, Charles Wheatstone's stereoscope [3] from 1838 utilizes two mirrors to produce stereoscopic images. Half-transparent mirrors can also create AR views comparable to Dr. Pepper's Ghost illusion [193] and are leveraged to superimpose display content into the user's view [194]. The Sonic Flashlight [195] presents a unique application of this effect.

The mirror paradigm in AR/VR refers to the idea of reflecting a dynamic scene in such a way as to closely resemble or even fully imitate the physical properties of a mirror in real time. This principle enables users to acquire a different viewpoint of the scene when looking at the mirror, leading to an improved spatial understanding of the environment, similar to how humans use stereopsis to understand depth more clearly. The mirror paradigm can take on various forms, ranging from entirely virtual mirrors to pseudo mirrors and even real mirrors with augmented images.

A virtual mirror exists as a virtual plane. It is usually only capable of reflecting virtual elements within surroundings since the view of its provided viewpoint is unknown to the system. One implementation of the virtual mirror founds its usage for AR visualizations of virtual anatomical structures [196]. The idea is to allow surgeons to fully understand the overlaid volumetric medical data on the patient without moving the patient or their own view. A virtual mirror delivers the necessary second view to explore the data from all sides. Monitor-based pseudo mirrors usually consist of a large, in some instances, body-sized monitor and a camera. Positioning the camera next to the monitor, facing it in the same direction, and displaying the camera stream on the monitor lets the user believe they are standing in front of a real mirror. Naturally, many characteristics of a real mirror are not present with these pseudo mirrors, including viewpoint-dependent parallax, stereoscopic depth perception, and spatial consistency between actual and mirror images. A widely known example is the mirracle [99], later known as Magic Mirror [100, 6]. Similarly, the same effect is producible with the help of immersive technology. For example, the Reflective-AR display [199] visualizes snapshots of the real world with augmented virtual content based on the camera built into a head-mounted display. Their visualizations appear like a mirror at first glance but are, in fact, picture-in-picture as they do not dynamically change according to the real world and the user's point of view. Moving further away from the pseudo mirror, showing pictures within the primary view of the user are similarly covered under the term of picture-in-picture [200, 201, 202], such that they would no longer perceive as mirrors.

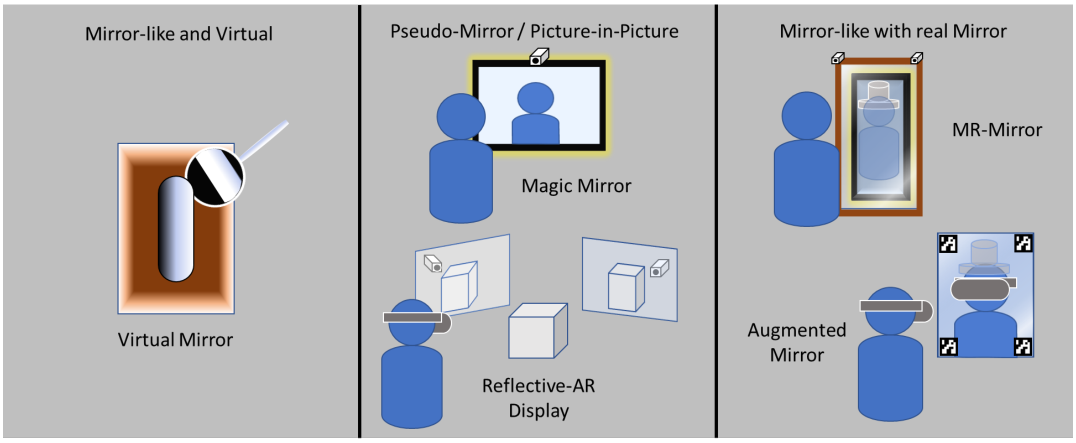


Fig. 2.1. **Selection of Notable Mirror Paradigms** in Virtual and Augmented Reality. The Virtual Mirror [196], MR-Mirror [197], and Augmented Mirror [198] appear as real mirrors in their respective domain. Other solutions, such as the Magic Mirror [100], or Reflective-AR [199] appear like mirrors at first glance; however, a closer look reveals that the mirror image does not dynamically change based on the user's viewpoint.

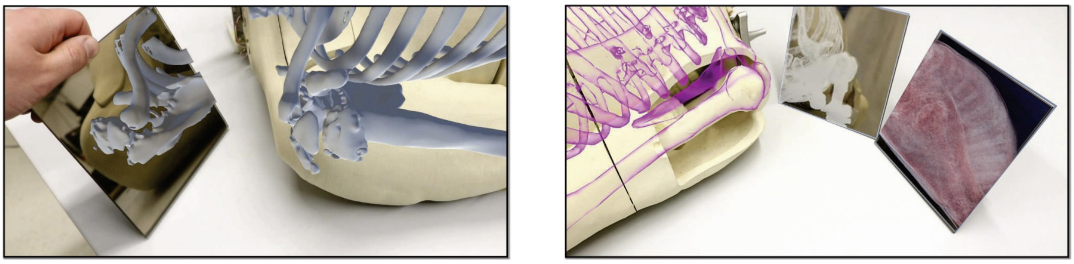


Fig. 2.2. **Augmented Mirrors for Medical Use Cases**. Image courtesy of Martin et al. [198] © 2020 IEEE. Left: Two perspectives of the same region of interest in a single view. Right: Additionally to different perspectives, each mirror shows a different imaging modality.

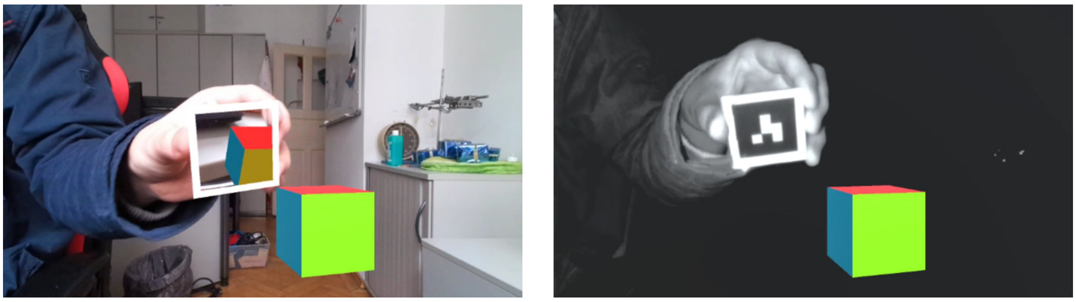


Fig. 2.3. **A Cold Mirror Allows Seamless Mirror Tracking** when combined with an RGB-D camera. The colored cube is a virtual object. The left image shows the view provided to the user, while the right infrared image of the depth camera provides the tracking data.

The final categorization includes work in which the authors combine a physical mirror with displays. The MR-Mirror [197] depicts such a system that uses a half-transparent mirror. To augment virtual objects on the reflected image, they deploy tracking cameras to provide the system on the surrounding environment and allow perspective-correct visualizations based on the acquired information. Finally, Augmented Mirrors [198] use the Hololens, an AR headset, and tracking information of a real mirror to show augmentations in the mirror image (see Figure 2.2). One could further enhance the tracking method with so-called “Cold Mirrors”, which are transparent for infrared light but reflective for the light within visible wavelength to provide seamless tracking of the physical mirror (see Figure 2.3).

2.2 Augmented Reality Teleconsultation for Medicine (ArtekMed)

This chapter presents the combination of previously presented components of AR/VR, RGB-D cameras, and 3D interaction techniques in our 3D teleconsultations system *ArtekMed*. Figure 2.4 illustrates the necessary structure focused on the software components in establishing the fundament of ArtekMed, including avatar animation, annotation, interaction, and audio communication. Figure 2.5 illustrates an example simulated use case of deploying 3D teleconsultation within an ambulance and Figure 2.6 shows a real setup of the ArtekMed system. ArtekMed is an asymmetric telepresence system as it utilizes AR for local users and VR for remote users. We designed it to be synchronous, promoting real-time communication between users. The opposite would be asynchronous communication, such as chats or recorded data for asynchronous playback. Our telepresence system is made possible by preparing the local space with multiple RGB-D cameras installed on the ceiling. The remote user gains the feeling of being in the locally captured space by seeing and hearing the captured content from it within their VR headset.

2.2.1 Implementation

ArtekMed’s implementation requires careful consideration of bandwidth, data management, and hardware selection to handle the amount of data coming from the RGB-D camera. For demonstration, the data load of a typical configuration of the Kinect4Azure RGB-D camera could use 640x576 pixels depth image resolution and 1080p color resolution. Then, the generated data accumulates to roughly 208 Megabytes per second at a 30fps capture rate. Therefore, data from six cameras accumulates quickly and surpasses the bandwidth of networking and internal structures. Further, camera calibration and stream output synchronization are essential to consolidate all 3D data into a unified scene reconstruction. For that, the VR workstation receives the data stream and reprojects the depth and color values into a triangulated, colored point cloud.

The in-depth technical details concerning the network and architecture required for the system are beyond the scope of this dissertation as we study the software components utilized in user interactions. However, the utilized hardware and setup are described in the core publications again. During the course of development, we deployed a preliminary architecture

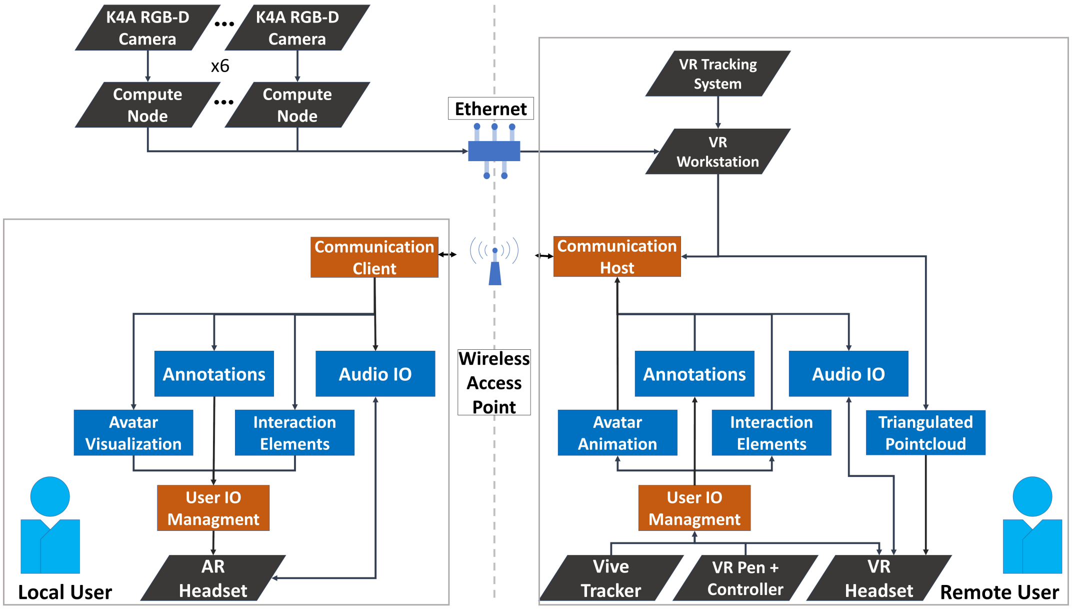


Fig. 2.4. **(Simplified) Data Management of the ArtekMed System** focusing on the synchronization of virtual content. The local space (left) is captured by Kinect4Azure RGB-D cameras, compressed, and streamed to the VR workstation of the remote user (right) to be visualized within their VR headset. In the investigated use cases of ArtekMed, the remote user will be the expert and primary creator of virtual content, such as being the driver for animating their avatar, creating annotations, and interacting with advanced UIs. The local user wearing the AR headset will be the primary consumer of the virtual content, often only passively interacting with them, as they have to focus on physical tasks in front of them. Artekmed supports audio in both directions. For a future deployment, the transmission of the point clouds from the RGB-D camera likely has to be replaced with a wireless connection, e.g., using 5G or 6G network bandwidth to minimize latency.

and a later more unified approach. The preliminary approach matches the illustration of Figure 2.4, using compute nodes for every Kinect camera. In the later unified approach, all Kinect cameras directly connect to a single, powerful reconstruction server. The common property of both versions is the immediate h.264 GPU-based compression of the depth and color image. The main difference is that the reconstruction server allows the choice to forward the compressed RGB-D data to the VR workstation or preprocess it within a voxelizer to reduce overlapping geometry and the size of transmitted data. Both core studies of this dissertation on Magnoramas and Duplicated Reality were conducted with the preliminary version as depicted in Figure 2.4 with compute nodes forwarding the RGB-D data to the VR workstation via real-time streaming protocol (RTSP) [203]. It is noteworthy that all presented interaction methods in this dissertation are functional with any hardware and network architecture as long as a 3D representation of the local scene exists.

We position the RGB-D cameras depending on the respective use case; however, we always mount at least two close above the patient for a high-density reconstruction. There was no consideration of optimal coverage; Nevertheless, we used additional cameras to capture the peripheral room geometry as completely as possible. Future settings may deploy structured planning in selecting the position of the RGB-D cameras, e.g., under consideration of interference between multiple depth cameras and occlusion [204], or the camera's field of view [205]. We calibrate the Azure4Kinect cameras with an L-target and a hand-held target with retro-reflective spheres attached. The L-target defines the origin of the calibrated 3D space, while

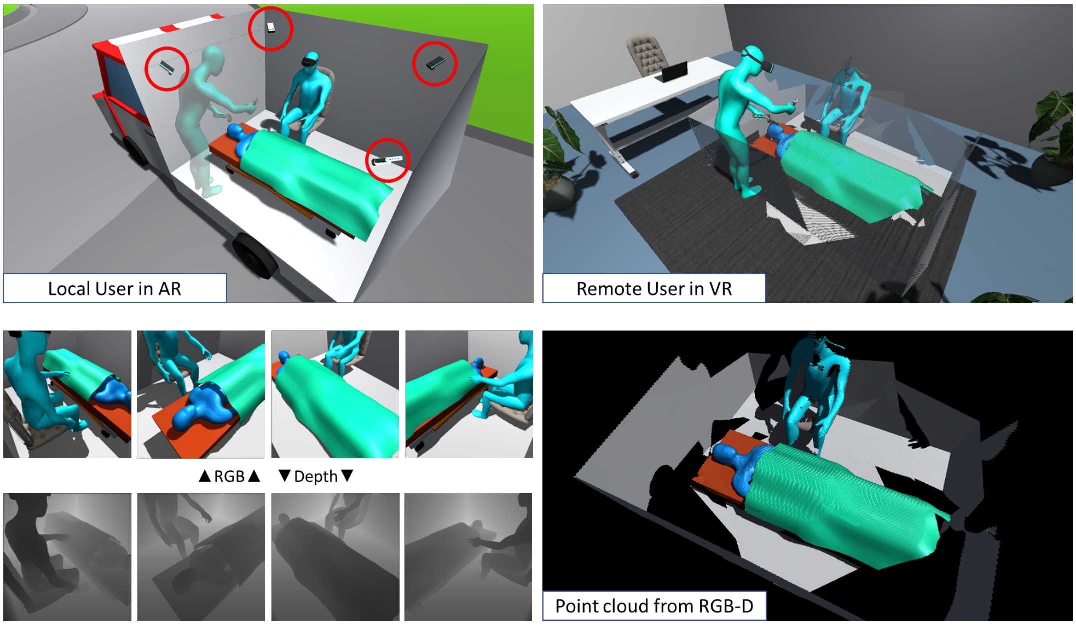


Fig. 2.5. **Illustration of the ArtekMed System** captured from a virtual simulation to demonstrate the concept. The local user is located inside the ambulance, which is fully captured by RGB-D cameras (red encircled). The reconstruction based on the image and depth image pairs provides the remote user with comprehensive 3D information, allowing the remote users to gain the feeling of *being there* alongside the local user.

the hand-held target allows for point-to-point correspondences between cameras for a bundle adjustment calibration [206]. This method follows the principle used for the room calibration of commercial Motion Tracking systems such as OptiTrack, ART, and Vicon. The Hololens 2 of the local user is calibrated differently for each of the core publications. Detailed descriptions of these calibrations are found inside the publication. We set up our prototype within the Simulation Center of the Institute of Emergency Medicine of Munich. A wall separates the local site and the remote site. Therefore, the ethernet connection between Kinect capture units and the VR workstation is realized with a 10GB glass fiber cable, which minimizes transmission losses and maximizes the transmission speed. In a realistic environment with an ambulance, this transmission would require high-energy transmissions, such as 5G or 6G technology, to stem the amount of data at minimal latency.

The VR workstation immediately initiates the update routine upon receiving the RGB-D data for rendering the point cloud in the VR space. As the original depth and color image, as well as the extrinsic and intrisic parameters, are known to the VR station, it can reproject the depth images into a triangulated colored point cloud by creating faces between neighboring depth pixels. We omit faces where the difference in depth exceeds a certain threshold. This threshold is usually between two and five centimeters.

Next, communication and interaction with the reconstruction are only possible after implementing a few basic features. These features include the animation of a digital avatar for the remote user in the local user's space and the ability of the remote user to create shared annotations. ArtekMed uses persistent 3D annotations allowing the VR user to highlight essential anatomical structures, indicate steps or directions relevant to the surgical task, or even write words within the 3D space. The VR user uses a Vive controller and a Logitech VR

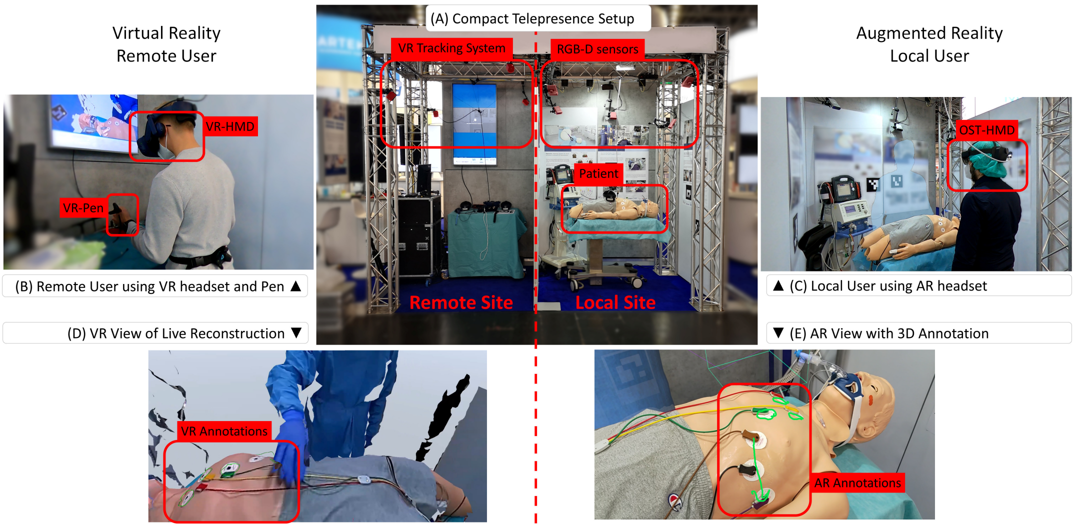


Fig. 2.6. **The ArtekMed System at Medica.** (A) Demonstrator for the Medica fair. (B-C) View on the users in VR and AR. (D-E) First-person view of the users in VR and AR. The demonstrator reveals issues of the telepresence system on a fundamental level. When attempting to work for causes outside low-level communication, users desire higher usability of annotations and fidelity of the scene.

Inkpen to interact with the virtual world. VR pens have proven superior to VR controllers in precise interactions with virtual content [113, 207]. In ArtekMed, we use both controllers to drive the avatar animation but only use the VR pen for creating 3D annotations. Regarding the process of drawing annotations, we decided to allow free mid-air annotations to allow the user greater flexibility in creating content independent from the geometry of the reconstruction. This choice allows users to use the annotation to better convey intentions and draw underneath the surface to indicate cuts and incisions. For the opposite to work, we could compute the collision between the user's VR pen and force annotations to be consistently above the surface. However, we dismissed this approach due to the noise of the RGB-D data and associated latency in the collision computation. Moreover, the real-time requirement of low latency limits the possibilities of traversing the scene's geometry to understand the relative positioning of the pen and a surface and the appropriate corrective action needed. Corrective action may entail adjusting the pen position along the line between the headset and the pen tip or adjusting it on the next adjacent surface. Each line annotation consists of 3D tubes, as they are better suited for stereoscopic displays than 2D billboard lines. Further, we employ Bézier curves to reduce the number of points necessary for multi-user synchronization. Therefore, our approach minimizes redundancy in straight lines by eliminating the need to store the pen's input for every frame. For instance, a straight line segment would require only a few control points, while sharp corners have two additional points to control the curve.

2.2.2 Avatars and Presence

ArtekMed employs a digital representation of the VR user in the form of an animated avatar and is most relevant for the first core publication. We use three Vive trackers, Vive controller, Logitech Inkpen, and the VIVE VR headset coupled, with inverse kinematic to drive the animation. The details are found in the respective core publication.

Fig. 2.7. **Avatars of ArtekMed** create a sense of co-presence and facilitate non-verbal communication through body posture, gaze direction, and pointing gestures. Image courtesy of Roth et al. [120] © 2021 IEEE

An avatar provides users with a sense of body ownership, self-perceived co-presence, and self-agency in virtual environments [208, 209, 210]. Presence, a topic of research in the scientific community, encompasses various types, including self-perceived and perceived others' co-presence, telepresence, and social presence [211]. Co-presence refers to the sense of being together with another person in space and time, which can be further distinguished between one's own perceived presence and the presence of others. Conversely, telepresence refers to the feeling of being there independently of others' presence or existence. Lastly, social presence captures the sense of being co-present with others and having the ability to interact and build interpersonal intimacy. As presence is a subjective experience, it cannot be measured quantitatively. Instead, questionnaires such as those developed by Nowak et al. [211] are used to assess the level of presence at a given time. Closely linked to the concept of presence is the notion of trust, which is of utmost importance when decisions that impact patients' long-term well-being are involved. Trust in the expert providing the consultation can be influenced by various factors such as the expert's identity, focus, and conscious and subconscious attention to the patient. Conversely, a lack of attention or absence of signals indicating attentiveness may lead to a diminished sense of trust and create an environment of intimidation for the paramedic or less experienced surgeon seeking consultation [212, 213].

Furthermore, avatars play a crucial role in a VR system by providing feedback during user interaction. This feedback may include visual indicators of collision with the environment [214] and other users [215], and contextual information [216]. In many VR applications, the user is often represented by two hands, especially in entirely virtual worlds for chatting, meeting other people, or playing video games. Commercial VR applications are typically designed to recognize only the spatial pose of the headset and two controllers as they are most accessible compared to additional motion tracking sensors. A user representation that only shows the hands without the upper body may result in a reduced sense of self-presence, as the user misses information about their body [217]. However, the absence of body pose information is not the only factor that can affect the user's experience. For example, the Proteus effect [218], a widely observed phenomenon in VR systems, suggests that users can change their behavior and actions based on their perceived appearance of themselves. Considering the related work, ArtekMed utilizes full-body avatars controlled by the user's entire body.

In addition to animation fidelity, the visualization of avatars can vary depending on their type. Avatars encompass numerous styles, ranging from photo-realistic to minimalist or cartoonish techniques. Using avatars in telepresence involves several facets, such as manipulating abnormal scales to influence perceptions [172, 173, 174, 219] or utilizing different levels of detail to convey specific messages between users [220, 221]. Depending on the context of telepresence, it may be crucial to emphasize facial expressions to convey intimacy, whereas this may not be necessary for task-oriented telepresence sessions. Large companies often prefer low-poly representations for their users since they are compatible with low-end devices that lack particular face-tracking technology and computational power. Furthermore, non-realistic avatars are more likely to avoid the "uncanny valley" effect. The avatars of ArtekMed are created with the software Reallusion with the Headshot Plugin to compute a meshed avatar based on a portrait image of the head and animated through inverse kinematic on the head pose, both controllers and optional body trackers. Figure 2.7 show the resulting avatars. Our goal was to create a realistic representation of the remote user without touching the uncanny valley.

ArtekMed represents local users with the point cloud automatically generated by the live room reconstruction. In a recent study, we demonstrated that a point cloud is more effective at conveying movements and communication while appearing more human-like compared to avatars [119] even without full reconstruction of the Hololens-obstructed face of the local user.

2.2.3 Video vs 3D Consultation

After the prior sections outlining the numerous details required for each component of telepresence, it appeals to the question of whether we should invest this effort in creating immersive telepresence, not just whether we can. One possible argument of critical voices may be that the benefits of telepresence may not appear as significant as the requirements and considerations necessary to make it work. As always, the answer is tricky and depends on various factors. For example, telepresence is unlikely to replace a quick call to coordinate the evening's dinner with a family member. However, more extended meetings and calls initiated with a physical task are likely to benefit from telepresence as a substitute for face-to-face meetings. In addition, telepresence allows for interaction with the natural world, such as our presented methods for increased comprehension, accurate conveyance of information, and communication through body gestures, creating a sense of togetherness.

In the context of teleconsultation, an additional layer of intention comes into play. Nearly all medical or industrial calls are initiated with a specific purpose. For example, in the medical field, the objective is often to seek consultation from medical experts for a patient. In other disciplines, the intention could be to receive guidance on assembling, moving, or modifying nearby objects. The ability to freely choose a view and perceive 3D environments are two compelling reasons to opt for immersive teleconsultation. For instance, the benefits include that the local users' perspective no longer constrains remote users as they may occlude their view or hold a hand-held camera at an unfavorable angle. Further, the 3D reconstruction enables remote users to communicate with local users without requiring significant training in choosing a point of view as it is easy as moving their head and body. Finally, it allows them to inspect a patient's condition freely next to prior mentioned advantages of virtual systems.

Previous studies have explored the efficacy of medical consultations conducted via 2D video and 3D telepresence systems. Results indicated that telepresence technology increased self-efficacy and decreased harmfulness, favoring its adoption over 2D video systems [222]. To reproduce and understand the difference between type of media better, we evaluated the ArtekMed system with a video-mediated consultation system for the placement of ECG electrodes on a patient dummy, as described in [121]. The evaluation resulted in a tie between the ArtekMed system and a state-of-the-art video-based medical remote consultation system in terms of task outcome. It is noteworthy that, during the evaluation, we only allowed annotations, and no additional 3D interaction and visualization techniques, such as Magnorama or PBM. Hence, despite the disadvantage of being a novel technology for the participants and lack of potential essential features for remote 3D consultation, ArtekMed has shown promising potential for future consultations.

2.3 UI Challenges for Medical Teleconsultation

This section is the binding piece that connects known characteristics of a 3D reconstruction-based telepresence system, observation as a user and developer of the immersive teleconsultation system ArtekMed, and the approaches to solving the fundamental issues of such approach on the conceptual level. We observed the following problems due to repeated testing of our prototype in hindsight to its use case for medical interventions. A prototype of such a focused system setup is seen in Figure 2.6 for placing a 12-channel ECG in a pre-clinical scenario.

Natural Scale Is Not Suited for Annotations, Both Technically and Ergonomically.

Problem: The patient is often not located at a high or ergonomic location that allows for creating precise annotations as the user intends. The non-existence of haptic feedback, which a user would have when working with physical surfaces, further enhances this issue. In VR, a user has to create the annotations in mid-air, using only visual cues from a single view to decide if they are far above or on the surface.

Proposed Solution: We maintain the view of the point cloud in its original scale but add a 3D cutout of a user's selected region of interest. That 3D cutout can be freely moved, rotated, and scaled. In addition, as the view onto the situs can be more dynamically changed than changing the user's viewpoint, they can use motion parallax to gain a depth cue to understand the scene's geometry. The cutout replicates all annotations drawn within it again at the original location within the region of interest. This cutout is called *Magnorama*, a portmanteau of *Mag*nifying Dio*ramas*.

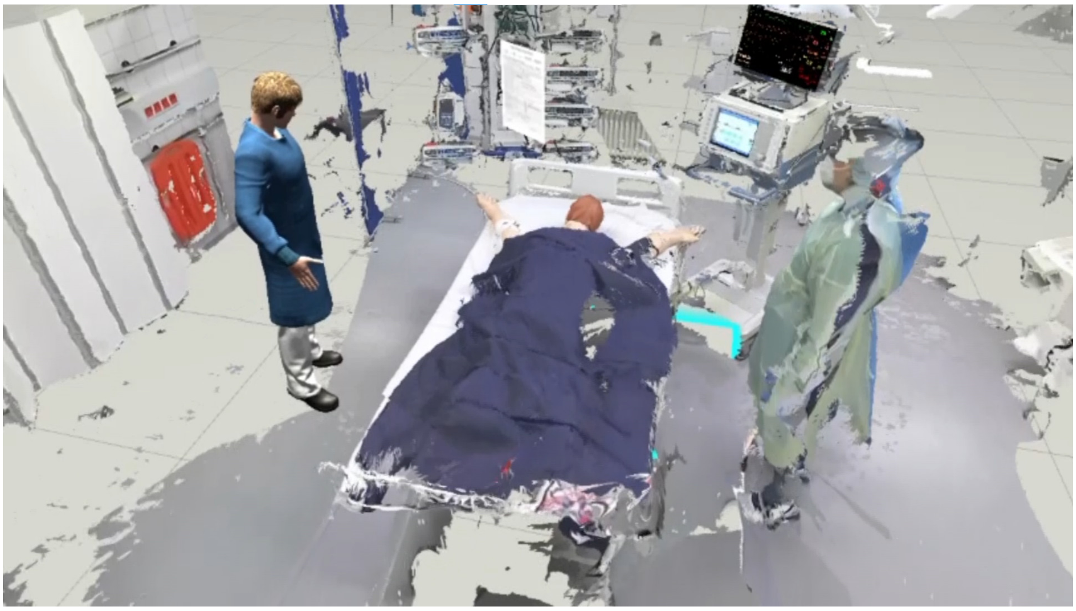


Fig. 2.8. **The Top-Down View of the 3D Reconstruction** provide a good view on the entirety of the operating room. However, a closer look reveal the missing gaps in the reconstruction and noisy surfaces.

Simultaneous Consultation and Performance Interfere With Each Other Even More if Several Users Share a Co-Located Space.

Problem: The ArtekMed system supports multiple users at the local site, allowing even multiple co-located users to initiate a co-located teleconsultation session. Users may start this configuration if time is in high demand. In such cases, the user providing consultation through drawn annotations and the user operating at the region of interest act simultaneously, which results in interference of physical space. Moreover, indicating actions underneath the surface, as often necessary in surgical procedures, are not easily created in situ. Switching operating person and consulting is usually not an option, while the expertise of the current task oscillates between both.

Proposed Solution: We extend the concept of Magnorama into AR and investigate it for multiple users sharing a co-located space. The concept changes, as now we duplicate the real world rather than a point cloud into a virtual object. Instead of working inside a common region of interest, the consulting user cuts out the reality within a region of interest and creates expert annotations within the duplicated reality. The cutout replicates all annotations drawn within it again at the original location within the region of interest. Since the duplicated reality is a virtual object, users can easily indicate action below the physical surface.

The 3D Reconstruction Suffer From Low Resolution and Missing Geometric Fidelity.

Problem: Due to the requirement of maintaining low latency (<500ms end-to-end) for seamless real-time communication, ArtekMed only deploys a minimal amount of point cloud processing techniques. The technical limitations of RGB-D cameras greatly influence the quality of scene reconstruction as seen in Figure 2.8, i.e., the capture of surfaces with different infrared absorption rates, positioning and occlusion of the viewpoint, and depth accuracy. These limitations result in holes within the reconstruction, with little to no real-world information. Medical diagnosis often relies on detecting color or shape abbreviations on the surface of the human body. Therefore, maintaining a clear and authentic view of the patient is essential for a teleconsultation system.

Proposed Solution: We suggest showing a mirror view in addition to the reconstruction for the VR user. The unique property of this mirror view is that the content uses the real-time color stream of a camera. To visualize a color stream as a real-time mirror, we forward-project it onto the 3D bisector plane, which is precisely in the middle and orthogonal to the viewer and capturing camera. This *Projective Bisector Mirror*, therefore, provides a highly detailed view of the local scene compared to the reconstruction that is still spatially coherent with it.

Part III

Magnoramas

3 Introduction to Magnoramas

The motivation for improving the precision of annotations was initiated by investigating the use of 3D teleconsultation for patient treatment after traumatic accidents. In particular, decompressive craniectomy (Figure 3.1) procedures may be necessary after a traumatic car accident. For example, victims suffering from a traumatic brain injury due to blunt or penetrating trauma and abnormal stark de-or acceleration can experience various forms of neurological illnesses and death in the worst-case scenario. One of the possible causes for this condition is increased pressure within the skull (acute intracranial hypertension [223]) due to fluid build-up within the brain. Decompressive craniectomy involves the removal of a portion of the scalp and skull to access the brain, which is a complex procedure. Surgeons without domain-specific knowledge of neurosurgery may fall into the pitfall of removing too little of the scalp or incorrectly positioning the drill holes in the skull which serve as the guide to disconnect the bone flap from the skull. Additionally, there are areas on the head where surgeons should avoid cutting, e.g., due to sensitive areas with many facial nerve cords anterior and posterior to the ears or for aesthetic reasons. The size of the initial incision on the scalp is particularly critical, as it determines the accessibility of the head for the subsequent

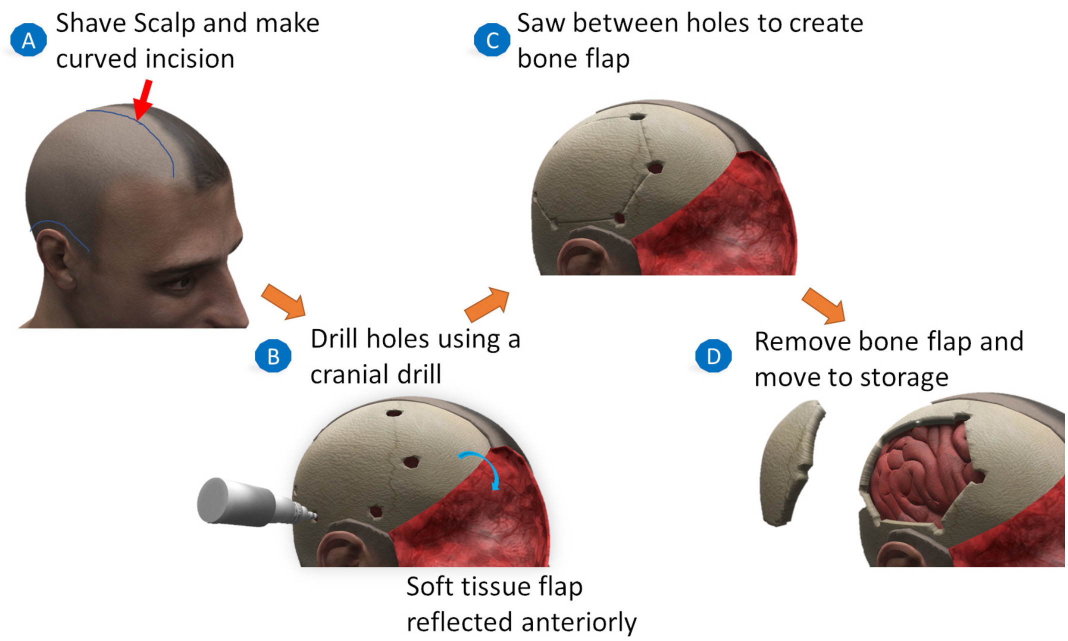


Fig. 3.1. **Relevant Steps of a Craniectomy for the User Study** include the annotations for step (A) for planning the incision on the scalp, step (B) for planning the burr holes, and (C) for the saw path between the holes. (D) In the final step, the surgeon removes the bone flap into storage. This step is not part of the user study.

steps of the procedure. Several studies have highlighted the importance of appropriately sizing the scalp incision and the bone flap removal for optimal patient outcomes [224, 225, 226].

Annotating the patient's head can pose ergonomic challenges for the remote expert, as achieving the necessary angle and height for accurate annotations can be difficult. Additionally, technical limitations such as tracking precision may impede the intended process of creating 3D annotations along anatomical landmarks. Since the remote user creates the virtual annotations without haptic feedback, they may be hovering slightly above or underneath the surface of the reconstruction and, consequently, the corresponding surface in the physical world. As a result, the perception of the annotation's position may be distorted if viewed from an angle other than the one in which it was created, leading to potential misinterpretation of its position. Magnoramas aims to solve these challenges by relocating and magnifying the real-world reconstructions, empowering the remote expert to annotate with greater precision and accuracy, and easily adjustable rotation, while maintaining the link to the original position.

This innovative technique draws inspiration from the WiMs [69] described in "3D User Interfaces: Theory and Practices" [67]. WiMs have traditionally been used for navigation or creating interactive mini-maps within virtual environments, as documented in previous research (section 2.1.3). However, navigation and map creation may be unnecessary in ArtekMed, which embeds a point cloud of an easily overseeable room from the real world. Instead, selecting a region of interest from which the user creates a live copy for interaction is more appealing. The 2D WIMP counterpart most similar to a Magnorama is the regional screen magnifier window built into operating systems as an accessibility tool. However, these essential differences lift Magnoramas to the next level: A user can move the magnified 3D content from the Magnorama anywhere within the virtual environment; however, it replicates any annotations they draw within its initially selected region of interest. Furthermore, the Magnorama can replicate any 3D objects, including virtual objects and real-time point clouds, making it a versatile and powerful tool for immersive teleconsultation.

The Magnorama represents an advanced interaction tool designed for users in VR. The term "advanced" indicates that novice users of an immersive teleconsultation system may not require it initially. For instance, users can create and position annotations without resorting to Magnoramas. Nonetheless, as users gain familiarity with the system, they may seek tools that enhance their efficiency and precision. In traditional teleconsultation settings, which rely on 2D cameras in the local environment, users frequently resort to digital zoom or adjusting the camera's optical zoom and orientation. However, these conventional 2D magnification approaches do not change the point of view. Suppose the user desires enhanced 3D views of the reconstruction after being used to the system. In that case, the Magnoramas prove invaluable by allowing users to magnify a region of interest dynamically and move it around by manipulating the grabbing input, thereby significantly enhancing the user's visual perspective.

According to the results of our user study on Magnoramas, we have observed a significant improvement in annotation precision across all annotation types when using this method. Specifically, the magnification of distances has increased the time spent on tasks for line-based annotations. Furthermore, removing the view of the real-world reconstruction has decreased the perceived social presence of the local user while the levels of co-presence and telepresence remain relatively stable.

Author’s Contribution to the Publication

The author of the dissertation (Kevin Yu) conceived the method, planned and conducted the user study, wrote the paper, and presented the work at the conference of IEEE VR 2021. Alexander Winkler assisted in conducting the dual-participant user study and collected the necessary formalities, such as the consent form and printing questionnaires. Dr. Frieder Pankratz and Dr. Marc Lazarovici are collaborators in the ArtekMed project from the Institute of Emergency Medicine in Munich. The system and user study were located at the institute’s Human simulation center. Prof. Dr. Dirk Wilhelm is the author’s medical consultant for the medical use case and medical supervisor. Dr. Ulrich Eck, Prof. Dr. Daniel Roth, and Prof. Dr. Nassir Navab are the technical supervisors of the author.

The following article is the accepted version acquired through the online library “IEEE Xplore”. The right of reuse is included in the appendix (section VIII).

Magnoramas: Magnifying Dioramas for Precise Annotations in Asymmetric 3D Teleconsultation

Kevin Yu*
Research Group MITI, Hospital Recht der Isar
TUM, Munich, Germany

Alexander Winkler†
Chair of Computer Aided Medical Procedures
TUM, Munich, Germany

Frieder Pankratz‡
Institute for Emergency Medicine
LMU, Munich, Germany

Marc Lazarovici§
Institute for Emergency Medicine
LMU, Munich, Germany

Dirk Wilhelm¶
Research Group MITI, Hospital Recht der Isar
TUM, Munich, Germany

Ulrich Eck‖
Chair of Computer Aided Medical Procedures
TUM, Munich, Germany

Daniel Roth**
Chair of Computer Aided Medical Procedures
TUM, Munich, Germany

Nassir Navab††
Chair of Computer Aided Medical Procedures
TUM, Munich, Germany

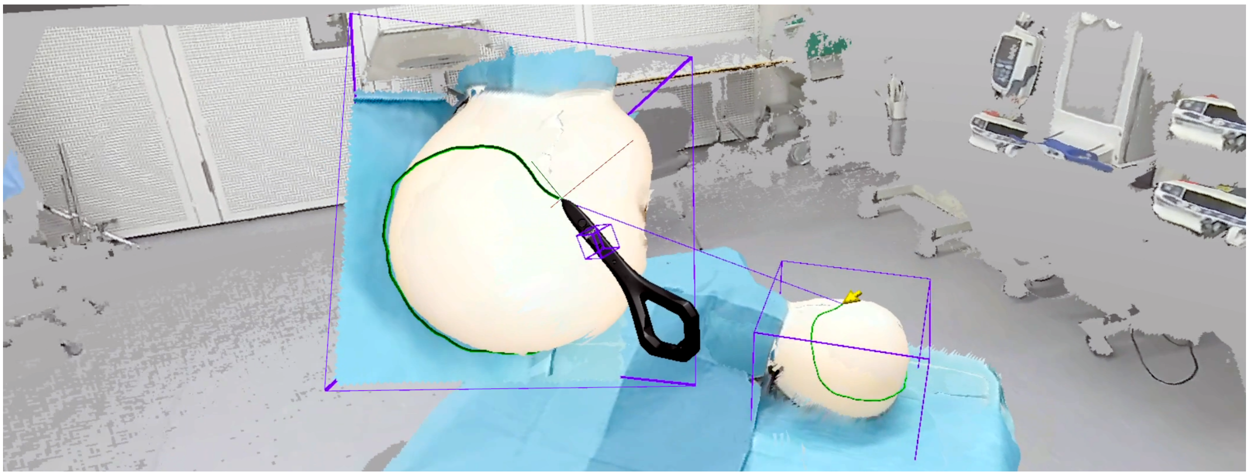


Figure 1: **Our proposed method: Magnorama**. Magnoramas allow the flexible extraction, transformation, and annotation of a region of interest (right) inside the real-time captured point cloud. A Magnorama can be interactively positioned, rotated, and scaled by the user. Increasing the size of the Magnorama provides the user with a magnified view of the region of interest. By that, it supernaturally augments the precision of annotations while remaining in the scene context.

Abstract

When users create hand-drawn annotations in Virtual Reality they often reach their physical limits in terms of precision, especially if the region to be annotated is small. One intuitive solution employs magnification beyond natural scale. However, scaling the whole environment results in wrong assumptions about the coherence between physical and virtual space. In this paper, we introduce *Magnoramas*, a novel interaction method for selecting and extracting a region of interest that the user can subsequently scale and transform inside the virtual space. Our technique enhances the user's capabilities to perform supernaturally precise virtual annotations on virtual objects. We explored our technique in a user study within a simplified clinical scenario of a teleconsultation-supported craniectomy procedure that requires accurate annotations on a human head. Teleconsultation was performed asymmetrically between a remote expert in Virtual Reality that collaborated with a local user through Augmented Reality. The remote expert operates inside a reconstructed environment, captured from RGB-D sensors at the local site, and is embodied by an avatar to establish co-presence. The results show that Magnoramas significantly improve the precision of annotations while preserving usability and perceived presence measures compared to the baseline method. By hiding the 3D reconstruction while keeping the Magnorama, users can intentionally choose to lower their perceived social presence and focus on their tasks.

*e-mail:kevin.yu@tum.de
†e-mail:alexander.winkler@tum.de
‡e-mail:frieder.pankratz@med.uni-muenchen.de
§e-mail:marc.lazarovici@med.lmu.de
¶e-mail:dirk.wilhelm@tum.de
‖e-mail:ulrich.eck@tum.de
**e-mail:daniel.roth@tum.de
††e-mail:nassir.navab@tum.de

Keywords: Interaction techniques, medical information system, virtual reality.

Index Terms: [3D user interaction]: Human factors and ergonomics—Teleoperation and telepresence;

1 Introduction

Immersive virtual environments hold great potential to support collaborative and assistive tasks, such as joint exploration [1] or collaborative medical procedures [2]. They can provide avatar embodiment [3, 4] and augmented forms of interaction in ways that would not be possible in the physical world or traditional media [5, 6, 7, 8].

2021 IEEE Virtual Reality and 3D User Interfaces (VR) | 978-1-6654-1838-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/VR50410.2021.00062

2642-5254/21/$31.00 ©2021 IEEE
DOI 10.1109/VR50410.2021.00062

Such collaborative environments can consist of purely virtual environments (e.g., [9, 10, 11, 12]), augmented environments (e.g., [1, 13]) or asymmetric combinations that merge virtual as well as augmented reality aspects (e.g., [14, 15, 16, 8, 17, 18, 19], see [20, 21] for further systematic reviews). One of the applications of the latter class is telepresence and, more specifically, teleconsultation [2, 18, 22], in which two or more users, physically apart from each other, can interact and guide another through a specific procedure.

To provide a mixed reality or asymmetric teleconsultation, as in the case of medical emergencies, it is necessary to provide bidirectional communication, visualization or replication of the situation, and context [22, 2], for example, by 3D reconstruction [23, 24]. Despite recent progress, remote collaboration in virtual- or mixed reality scenarios still faces several challenges that consider the coherence of shared environments (and the relation to the physical space), such as sharing awareness [25] or avoiding collisions [26].

Moreover, it can rationally be assumed that interactions, especially drawing in 3D reconstructions, are error-prone either due to the technical artifacts of noise and reconstruction errors or, more importantly, the lower accuracy when compared to drawing with a physical pen and paper, which arises from the lack of physical support [27] and the fact that drawing in 3D has higher manual effort or cognitive and sensorimotor demands [28].

In medical teleconsultation, however, precision in interaction and guidance can be critical to ensure a patient's survival. Surgeons, paramedics, and first responders are likely to encounter injuries in which immediate treatment is of paramount importance. Still, they might not be trained to or not possess enough experience to perform certain interventions. Even trauma surgeons may lack specialized skills for specific procedures. One of these emergency surgery procedures is craniectomy, where the patient's skull needs to be opened to promptly release pressure from a swelling of the brain. Teleconsultation may be used for marking the steps to perform the craniectomy but requires exceptionally accurate annotations as guidance, which would directly relate to interventional incisions.

Little work has explored how to support such high-precision interactions while ensuring important communicative functions for successful collaboration. In this regard, common tools for consultation include virtual avatars that can point and gesture at real-world objects or draw mid-air annotations, which both users can see. When users need to draw a line accurately, an intuitive approach is moving closer. However, moving closer may still limit the precision capabilities due to factors such as jitters of the drawing device resulting from tracking errors. These reduced sensorimotor capabilities will become apparent in mid-air drawing tasks. Yet, adapting common methods from purely virtual applications for sketching (e.g., VRSketch) or drawing (e.g., TiltBrush), such as scaling the whole environment depicted as a real-time point cloud, would most likely result in sickness effects, mislead the perception of size and spatial relation between objects and context, and may hinder necessary communicative interactions such as mutual gaze, joint attention [29] and shared awareness [25].

1.1 Contribution

To address this problem, we propose Magnoramas (see Figure 1, which can be described as interactive dioramas for selectively magnifying regions of interest of a real-time 3D reconstruction for remote teleconsultation. We describe our approach and compare our method to i) a baseline and ii) to a variant of our method where users can only see the Magnorama, but the context is masked. Our method outperforms the baseline in terms of precision while having similar usability and task load ratings, thus providing initial evidence for the applicability. Our findings show that removing the scene context (and hence the partner's avatar) reduces social presence. This novel interaction method and its evaluation provide valuable insights, demonstrate high potential, and guide the design of future telepresence systems.

2 Related Work

We divide the related work into three major categories that present the related context and previous work for our approach: (i) Virtual Reality (VR) interaction with a World-In-Miniature (WiM), (ii) drawing precision in VR, and (iii) co-interaction between multiple parties during teleconsultation.

2.1 Interaction with a World-In-Miniature

The well-known work on WiMs by Stoakley et al. [30] follows a related concept and utilizes a tracked physical clipboard. In VR, the entire room is down-scaled and attached to the clipboard inside the virtual environment (VE). Users could move furniture in the miniaturized version and observe the original furniture moving inside the actual room. The authors recognize the potential of enlarging the WiM for more fine-grain control of manipulation in exchange for range. However, to the best of our knowledge, they do not follow up on this idea and neglect the potential of detail selection and improved precision.

In the follow-up works, the metaphor of WiM is primarily researched for interaction [31], spatial locomotion and navigation [32, 33, 34, 35, 36]. Wingrave et al. [37] added scaling and scrolling functionality to the WiM and investigated the use of WiMs for spatial navigation. They, however, used scaling only to shrink the environment. They found that users rarely re-scale the WiM and often leave it at a comfortable size.

Pierce et al. [38] introduce an interaction method alluding to Voodoo-dolls, which applies the idea of indirect object manipulation that is present as well in a WiM. In this method, users can create copies of virtual objects. Any interaction performed on the copy is simultaneously performed on the original object.

In contrast, our method allows such interactions on any virtual subspace and its content, therefore, is not limited to particular virtual objects. Additionally, no magnification was used for their method.

While WiMs and Magnoramas have common characteristics – such as the duplicated view on virtual space and indirect manipulation – the core aspect is that the scaling factor is inverted.

2.2 Precision of Drawing and Annotating

In the following section, we provide an overview of related work regarding (the improvement of) freehand drawing and annotation inside a three dimensional (3D) VR environment. In this regard, we consider drawing to be a context free action and annotating to be an object-centered application of drawing or object placements/attachments. One of the re-occurring difficulties during unconstrained drawing in a 3D environment is the inclusion of the third dimension. One common pitfall is the misjudgment of depth such that drawn line strokes may appear closer or farther than intended by the user [39]. Additionally, drawing a straight line poses a challenge since no haptic feedback nor cognitive, nor sensorimotor aids are provided, unlike drawing on a physical surface [40, 27, 28]. Multiple related works investigated the assistance in freehand 3D manipulations or drawing with a tracked pen, either by including purely visual non-constraining guides [41, 42], constraining guides [43], or haptic modalities [40, 44].

Barrera et al. [45] investigate the relationship between spatial ability and the user's proficiency to redraw a given shape in VR. They found that helping the user identify the correct viewpoint and starting point of the next stroke positively affects line precision. Additionally, they conclude that dynamic viewpoints and feedback on head-movements via a compass or a map can improve the sense of depth.

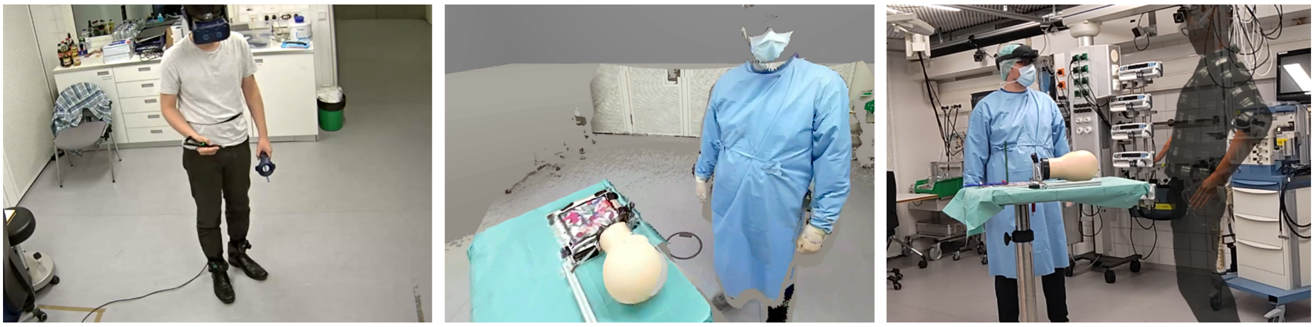


Figure 2: **A scenario using the asymmetric telepresence system**. From left to right: The remote expert wearing the VR tracking setup for animating his avatar and allowing annotations (left). A view on the local user from the first-person perspective of the remote expert in VR (center). The HoloLens 2 is not visible in the point cloud due to the high reflection coefficient of the transparent display. 3rd person AR view on the avatar of the remote expert and the local user in the shared environment (right).

Since drawings in these works and other VR applications are anchored statically in the virtual space, users changed their body position to gain a different perspective. A Magnorama is a cut-out of the drawing region. Users can transform it with their hands and quickly change their point of view to gain a better understanding of the geometry as well as to object details. Simultaneously, the final drawing results will not be changed in position and maintain the spatial correctness.

2.3 Teleconsultation and Collaborative Interaction

Collaborative virtual environment approaches can be distinguished between avatar-mediated systems (e.g., [46, 47, 6, 48]), 3D-reconstruction-based telepresence approaches (e.g., [13, 49, 50, 51, 24, 1]), and mixed/asymmetric approaches (e.g., [17, 8]). These provide the basic context for an application use-case. Research in object manipulation, shared drawings, or annotations for remote guidance is central to teleconsultation systems. The next paragraphs discuss methods in a shared teleconsultation system using annotations.

Oda et al. [16] present a method for VR/Augmented Reality (AR) teleconsultation to guide the consultee in placnig a real physical object onto a second real physical object. They introduce the concept of Virtual Replicas, which is an instantiated virtual version of the physical object to be moved. By defining points on the virtual replica, the consultant can create links connecting the physical object to the replica. Unlike our method, virtual replica require knowledge and 3D model of the annotated object and does not provide methods on increasing the precision while defining the annotations. Oda and colleagues further [52] use a cutout from a real-time captured point cloud in a collaborative AR collaboration system for a more precise pointing gesture of distant objects. Kolkmeier et al. [53] use an RGB-D camera to capture the 3D environment of the consultee in real-time and visualize it for the consultant inside a VR head-mounted display (HMD). Their presented work incorporates a real-time captured point cloud and an abstracted user representation (head and hands) of the consultant drawing annotations. Weibel et al. [2] present ARTEMIS, an asymmetric telepresence system using VR and AR HMDs. Drawing annotations is possible in this system as well but with no additional solution for increased precision.

These works indicate the need for precise annotations in teleconsultation systems. However, none of the the systems helps the users to draw annotations that are more accurate than they could achieve with their natural dexterity, which presents a gap in research.

2.4 Hypotheses

Our review shed light on three major areas of related work. Our research goal was to provide a method that would successfully assist the presented use-case or related requirements. Reviewing the literature on drawing precision and projecting the findings on our proposed method, we assumed that ***H1: The magnification of details in the form of Magnoramas increases freehand drawing precision*** since Magnoramas aim to improve information detail but also act as "lever" for motor precision. Further, since the interaction method is novel and less natural than more coherent interaction, we also assumed that ***H2: Interacting with Magnoramas is inferior in terms of usability compared to context coherent interactions***. Finally, considering the importance of co-location, joint attention, communicative cues, and collaborative verbal and nonverbal interaction (broadly discussed e.g., [5, 54, 55, 8, 6], one could fear that with our method ***H3: The perception of the interaction in terms of co-presence and social presence aspects is inferior when using Magnoramas*** since the remote user would change focus to other parts in the scene when modifying the Magnorama, or completely lose the context (Magnorama-only).

3 Methods

We present a solution for the simultaneous view on the original sized, virtual depiction of the real-world environment and a user-controlled and rigidly transformable duplicate of a region of interest (ROI). As seen in Figure 1, the ROI is visualized as a transparent cube with opaque edges. A duplicate of the same region is created in front of the user, which can be moved, rotated, and scaled. We call this duplicated cut-out, which the user can interact with, a *Magnorama*, as a portmanteau of "magnified" and "diorama". Magnoramas allow the users to focus on their actions but still be aware of their surroundings at different scales and points of view in the remote space. This is especially true in their interaction with other users in the same space.

For further addressing, we refer to the consultant working in the VR environment as the remote expert (RE) and the consultee in the AR environment as the local user (LU). They represent both sides of the teleconsultation system in our study and are subject to measuring their perceived co-presence, telepresence, and social presence of their partner.

3.1 Implementation

Although the implementation can be done using different techniques, we present our solution to this concept, implemented in Unity3D. We used an HTC Vive Pro Eye HMD, together with three HTC Vive trackers, a tracked HTC Vive Controller, and a Logitech VR Ink pen for the RE to realize inverse kinematic tracking [56]

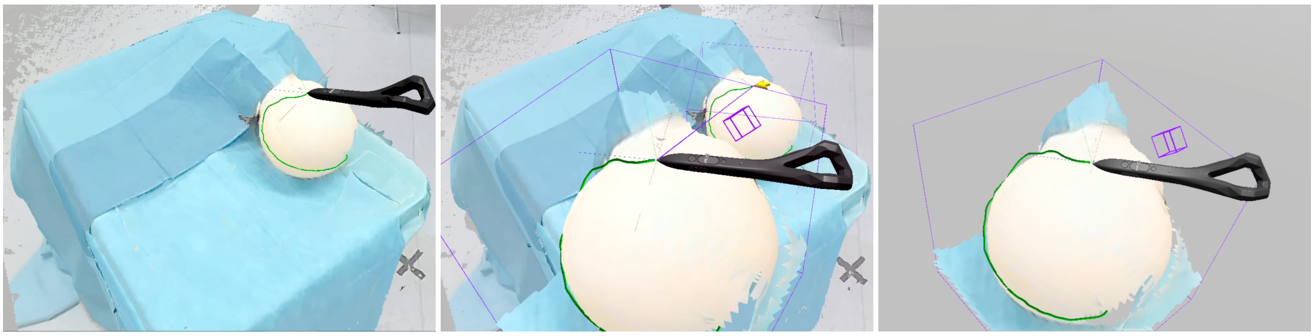


Figure 3: **Three experimental conditions**. From left to right: Baseline (left), Magnorama (center), Magnorama-Only (right). In Magnorama-Only, all context is masked, including the collaboration partner. The purple cube on the controller indicates that the Magnorama is active.

on the VR side. We used a Microsoft HoloLens 2 for the LU. The Magnorama implementation consists of these components:

- **ROI**: A transparent cube with the transformation ${}_{\mathrm{W}}T_{\mathrm{ROI}}$ which encapsulates the ROI inside the virtual environment.
- **Magnorama**: A placeholder object with the transformation ${}_{\mathrm{W}}T_{\mathrm{Magnorama}}$ to provide the user with an interactable object.
- **Magnorama Camera**: A camera positioned with the same relation to the Magnorama as the rendering camera ${}_{\mathrm{World}}\mathrm{T}_{\mathrm{Cam}}$, but in relation to the ROI:

$$
{}_{\mathrm{W}}T_{\mathrm{MagnoramaCam}} = {}_{\mathrm{W}}T_{\mathrm{ROI}} \cdot ({}_{\mathrm{W}}T_{\mathrm{Cam}}{}^{-1} \cdot {}_{\mathrm{W}}T_{\mathrm{Magnorama}})^{-1}
$$

The rendering of the Magnorama camera has to accommodate our method. Objects and reconstructed point clouds need to be scaled around the center and clipped at the ROI's border, which can be done inside their shader used for rendering. The camera itself should be rendered using the same depth buffer as used by the camera of the HMD for the correct occlusion with the scene.

In our implementation, as seen in Figure 1, it appears that annotations drawn in the Magnorama are also directly and in real-time drawn at the original position inside the ROI. However, the opposite is the case. By drawing inside the Magnorama, the pose of the pen is transformed into the coordinate system of the ROI where the line is drawn. Since the Magnorama is a detached camera view of the ROI, the newly drawn line appears simultaneously in the Magnorama. This approach of implementing Magnoramas avoids duplicating objects in the scene since any interactions are directly performed at the original location.

3.2 Digital Representation of the Remote Expert

The RE can directly see the LU in the real-time captured point cloud (see Figure 2); however, the LU cannot see the RE without a virtual representation. For this reason, the RE is represented as a generic male or female avatar, to himself and to the LU, to allow for avatar embodiment [3, 4, 57] and (social)-presence [55]. The avatar's pose is transmitted to the LU and visualized as seen in Figure 4 and calculated in real-time through inverse kinematics. Parallel to the avatar representation, both participants were able to discuss the task using an external audio communication channel.

3.3 Appearance of the Magnorama for the Local User

As soon as the RE creates a Magnorama and proceeds to annotate the ROI, the user simultaneously detaches himself from the region at the on-site location. To communicate the use of the Magnorama for the LU, we added visual indicators. Two boxes depicting the selected ROI and the Magnorama are rendered for the LU while

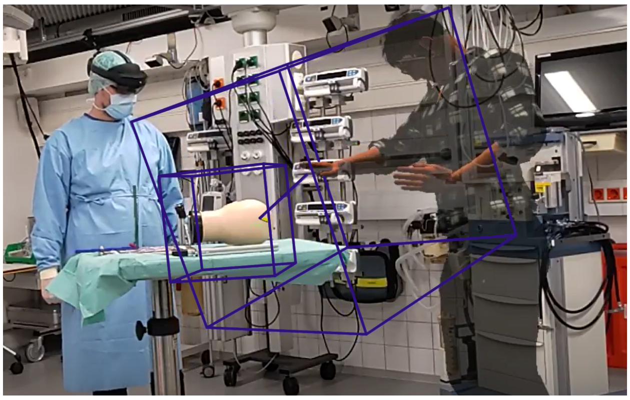


Figure 4: **Magnoramas as seen in shared Augmented Reality**. This view is captured at the local site from an additional HoloLens 2. The local user (left) observes the avatar of the remote expert (right) while drawing annotations using the Magnorama. The image was post-processed for better visibility of the Magnorama (purple).

the VR pen is inside the Magnorama. For the LU, the reconstruction inside of the boxes is not visualized because transmitting the content of the Magnorama as seen in VR would occupy an excessive amount of network bandwidth, memory capacity, and compute capabilities of the HoloLens 2. A link is rendered between both boxes that connect the location of the pen tip within the Magnorama and the corresponding back-transformed position inside the ROI. This link aids the LU to find the RE's avatar representation, even if it moves away from the scene during the annotation process. This link is also visible in Figure 4. We measure potential adverse effects from this solution by including the role of the LU.

3.4 Asymmetric Teleconsultation System

The proposed interaction methods were implemented in an asymmetric telepresence system inspired by Maimone et al. [24]. The system consists of three stationary Azure Kinect RGB-D cameras attached to dedicated camera PCs (MSI Trident 3, 16GB RAM, NVidia Geforce RTX 2060 GPU) and a high-performance rendering workstation (Intel Core I7, 64GB RAM, NVidia Geforce RTX 2080Ti). The computers communicate via a dedicated 1Gbps local area network. Each camera PC captures the color-image (BGRA, 1536p) and depth-image (NFOV Unbinned) with 30 FPS, encodes both image streams to H264 using the hardware encoders on the GPU (color: lossy compression/RGBA, depth: lossless compression/UINT16), and provides these streams with low latency as RTSP endpoints. Furthermore, the sensor calibration (intrinsics and extrinsics) is supplied as Capnproto RPC endpoint from each camera PC. The image streams and calibration data are then received by the rendering workstation using a custom, native Unity3D plugin, decoded using the hardware decoders of the GPU and directly streamed to DirectX textures on the GPU to achieve low latency on the receiver side as well. First, each depth-image is unprojected into a structured point-cloud using the respective sensors' intrinsic parameters.Next, the individual point-clouds are converted to surface meshes [58] in a geometry shader by creating triangles from neighbored values of each depth-image and textured using the respective color images. Edges inside the depth image are handled by only allowing triangles to be generated if all three corners have at most a 2 centimeters difference in depth. The resulting meshes are positioned using their respective camera extrinsic parameters.

The extrinsics of the three RGB-D cameras for 3D reconstruction are estimated using a procedure similar to the room calibration of commercial optical tracking systems. In this process, we use the infrared images from the Azure Kinect sensors since they correlate directly with the generated depth-image for best precision. We use an L-shaped target with four reflective spheres placed on the floor to define the world origin and roughly estimate the camera poses. Next, we collect a video sequence using a calibration wand with two reflective spheres and use bundle-adjustment to refine the estimation of extrinsics. We register the Microsoft HoloLens 2 into the same world coordinate frame using a fiducial marker that is calibrated within the room using an external tracking system.

4 User Study

An extensive user study was performed for the evaluation of our methods on Magnoramas. In the following, we describe the design of our user study and its related components.

4.1 Design

The experiment was designed as a one-factor (*Experimental Condition*) within-subjects experiment. Pairs of two participants performed a semi-collaborative task in an asymmetric VR/AR telepresence setting. The situation reflects a medical scenario with a LU requiring assistance for a surgical task and a RE assisting by annotating procedure steps. Participants experienced both the AR side as a LU and the VR side as a RE in three trials each, differing in the experimental condition. Our primary research goal was to confirm our hypothesized benefits of improved precision of the annotations and investigate potential downsides regarding presence and usability arising from the new methods and communicative inconsistencies that emerge from the two proposed novel interaction concepts. The object of interest for the study is a model of a head that is rigidly fixated in the room.

4.2 Experimental Conditions

We compare three conditions which we refer to as "baseline", "Magnorama", and "Magnorama-Only". We theorize that each condition has advantages and disadvantages regarding the perceived presence and precision of the drawing task.

Baseline When the RE draws annotations using the baseline method for our comparison, it refers to the act of directly drawing on the visualized head in its original pose and size, as seen in Figure 3 (left). This represents the drawing methodology of similar telepresence systems with annotations with no option for magnification. In this condition, the user in VR can only see the 3D reconstruction but no magnification.

Magnorama The RE draws annotations inside the Magnorama but can still see the annotations on the real head. The RE is still able to see the body language of the LU in the point cloud. The RE can use the controller of their non-dominant hand to grab, rotate, and scale the Magnorama. In this condition, the user in VR can see both the 3D reconstruction and the magnification. This method can be seen in Figure 3 (center).

Magnorama-Only Similar to the previous condition, the user draws the annotations inside the Magnorama. However, the user cannot see the original point cloud that is depicting the real-world, as seen in Figure 3 (right). Again, the user can use the controller of their non-dominant hand to grab, rotate, and scale the Magnorama. In this condition, the user in VR cannot see the 3D reconstruction but only the magnification.

4.3 Three Tasks Performed Per Condition

Our user study imposes a simplified scenario of a craniectomy. Craniectomy was identified as one of many potential use-cases for life-supporting remote telepresence systems in exchange with doctors and medical specialists. For this procedure, the surgeon must act both quickly and precisely in order to prevent life-critical damage. In medical terms, a craniectomy describes the removal of a part of the skull for releasing built-up pressure from a swelling of the brain after a traumatic brain injury. Three main tasks are necessary during the procedure: (1) Cut open the scalp of the injured person, (2) use a medical-grade drill to prepare holes in the skull (craniotomy), (3) use a medical-grade saw to disconnect the bone tissue between the holes.

For this study, we reduced the complexity of the tasks into abstracted color-coded tasks. The colors green, blue, and red each indicate one of the craniectomy tasks: a green line for outlining the cut on the scalp, blue pins for marking the drilling spots, and a red line for outlining the saw paths on the skull. The green line task covers a large area from the forehead to the ear. Users only require a single tap on the controller to place a pin during the blue pin placement task, which may provide insight into the precision of single-action tasks. In the red circle task, the guiding line covers a relatively small area, which is also passing through the positions of the pins. The guiding lines appear as blue lines, as seen in Figure 5, and not in the color assigned to the task to avoid confusion during the drawing procedure. All guiding elements are visible inside the Magnorama to the RE. Therefore medical expertise was not required for participation in the study as the participants were only required to redraw predefined guiding elements, as seen in Figure 5. The tasks will be referred to as 'line', 'pin', and 'circle' task further in this work.

4.4 Study Procedure

The user study was conducted in pairs. Each participant experienced both parts of the study paradigm (i.e., RE and LU).We welcomed participants separately and guided them to separate rooms. The study began with the visual tests and an initial demographics questionnaire, followed by the mental rotation questionnaire further described in subsection 4.7. The first participant on VR dons three Vive trackers for controlling their digital representation, which is visible for both RE and LU as described in subsection 3.2. The participants hold the VR pen for drawing annotations in their dominant hand, while they use their non-dominant hand for the controller to move the Magnorama. Each participant had the chance to become acquainted with the system for a maximum of 10 minutes, including creating annotations and interacting with the VR Ink pen and the Magnorama. No participant exhausted the full 10 minutes of familiarization to feel confident with the interactions. The order of the three experimental conditions (Baseline, Magnorama, Magnorama-Only) and the order of the color-coded tasks are randomized. The LU communicates the order of the tasks to the RE over an audio-communication channel. Additionally, the LU decides on a preference for the drawing direction of the annotation. This is done to encourage communication between both parties.

COVID-19 measures: Experimenters wore masks during the experiment and kept a safe distance from the participants. Partici-

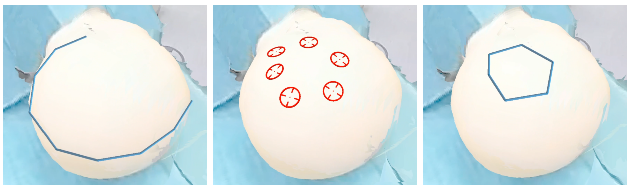


Figure 5: **Guiding elements of the three tasks**. 1. Cut on the scalp (left), 2. drilling locations marked with cross-hairs (center), and 3. saw paths to disconnect bone tissue (right). All guiding elements are also visible in the Magnorama.

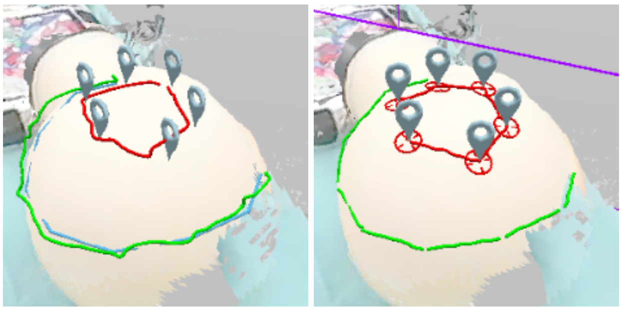


Figure 6: **Exemplary hand-drawn annotations** of the green line, light-blue pins, and red circle as seen in VR. Annotations are drawn by one of the participants based on the guiding elements for a baseline condition (left) and a Magnorama condition (right).

pants wore masks except for the time of the task. All equipment and contact surfaces were carefully disinfected after each trial block, and exchange devices were prepared for the participant switch. Rooms were sufficiently ventilated and participants were located in separate rooms. Strict exclusion criteria for the study were previous visits to risk areas and any symptoms or contact with infected persons. Participants were clarified of these conditions, and all participants consented. The study was conducted in accordance with the local COVID-19 regulations with necessary precautions and in accordance with the declaration of Helsinki.

4.5 Objective Performance Measures

The simulation logged error measurements of the drawings. In the green line and red circle task, the user redraws guiding lines. The error is calculated as the distance between the pen-tip and the closest line segment. In the blue pin task, the error is calculated using the distance between the pin and the closest target cross-hair. Inputs with an error greater than five centimeters are discarded during the evaluation. This excludes the annotations created by accident or for testing. Additionally, we recorded the time to task completion between the first and last valid user input for each task.

4.6 Subjective Measures

Participants are asked to complete a questionnaire consisting of five parts after completing each experimental condition. We assess copresence, telepresence, and social presence based on the factors proposed by Nowak & Biocca [55]. The scales are adjusted to a 7-point Likert scale to ease the interpretation. We assessed the perceived usability by including the system usability scale (SUS) [59]. The SUS was evaluated using a 5-point scale (1 - strongly disagree, 5 - strongly agree). Further, we assessed the perceived task load using the NASA task load index (TLX) [60]. We evaluated the raw TLX total score (see [61]) and the sub-scores. A single question regarding the potential symptoms of cyber-sickness was added (Fast Motion Sickness Scale (FMS) [62]). After each condition, we asked free-text answers for specific advantages and disadvantages of the method. At the end of the study, we asked participants for their method preference, the underlying reason, and comments.

4.7 Participants

In total, $N = 24$ participants ($M_{age} = 23.63$, $SD_{age} = 3.03$) were recruited via mailing lists and campus announcements. Of those, 23 were students of various fields, including medicine (3) and computer science (2). 8 participants were female, 16 male. Participants stated to spend time with digital media (PC, mobile phone, etc.) for about 34.21 hours per week ($SD = 3.85$). 19 participants noted to have used VR systems before, and 8 participants noted to have used AR systems before. The average amount of previous VR usage was $M = 4.46$ times, ranging between 0 and 30. The average amount of AR usage was $M = 2.17$ times, ranging between 0 and 30. 6 participant pairs have known each other before, 6 pairs did not know each other and were matched together on a first-come-first-serve basis.

To avoid any bias from visual impairments, we assessed a Landolt C-Test (EN ISO 8596) for acuity, an Ishihara Color test for color deficiency [63], and a Titmus test for stereo vision. All participants had normal or corrected-to-normal vision regarding acuity. One participant had slightly reduced stereo vision. Two participants had a slight red-green color weakness. Since there were no color mixtures involved in the experiment, we decided to include these in the analysis. We found that all participants were capable of performing the experiment. The average interpupillary distance of the sample was $M = 62.66$ *mm*, measured by the HoloLens 2 device. The mental rotation test [64] confirmed that none of the participants had severe mental rotation deficits.

5 Results

5.1 Objective Performance Measures

The annotation performance was analyzed by calculating the minimum, maximum, and mean error of the deviation from the performed annotations from target shapes/pin positions and their standard deviations. We analyzed the annotation performance by the participants using a one-way repeated measures analysis of variance (ANOVA), with the method of annotation as the factor. Greenhouse-Geisser corrected values are reported in the case the assumption of sphericity was violated. Bonferroni corrected pairwise comparisons are reported for significant main effects. Descriptive results are depicted in Figure 7.

Pin Task Performance The results showed a significant main effect for the mean error of the pin placement measure; $F(1.44, 33.08) = 3.89$, $p = .043$, $\eta_p^2 = .145$. Pairwise comparisons revealed a significant difference between the baseline method ($M = 5.22$ *mm*, $SD = 5.24$ *mm*) and the Magnorama method, which resulted in a statistically significant smaller error ($M = 2.56$ *mm*, $SD = 2.32$ *mm*; $p < .05$). The Magnorama-Only method ($M = 3.02$ *mm*, $SD = 2.73$ *mm*) outperformed the baseline, but not to a significant level.

The analysis revealed a significant main effect for the minimal error of the pin placement measure; $F(2, 46) = 6.57$, $p = .003$, $\eta_p^2 = .222$. Bonferroni corrected pairwise comparisons revealed that the minimal error was significantly lower in the Magnorama-Only condition ($M = 0.98$ *mm*, $SD = 0.75$ *mm*), compared to the baseline condition ($M = 1.18$ *mm*, $SD = 1.38$ *mm*; $p = .012$). The Magnorama condition ($M = 1.18$ *mm*, $SD = 1.38$ *mm*) showed a lower error than the baseline condition, but not to a significant level.

Circle Task Performance Greenhouse Geisser corrected results for the main effect of the mean error of the circle task were statistically significant $F(1.55, 28.28) = 3.93$, $p = .038$, $\eta_p^2 = .146$. Pairwise comparisons showed that both the Magnorama condition ($M = 4.17$ *mm*, $SD = 4.44$ *mm*) as well as the Magnorama-Only

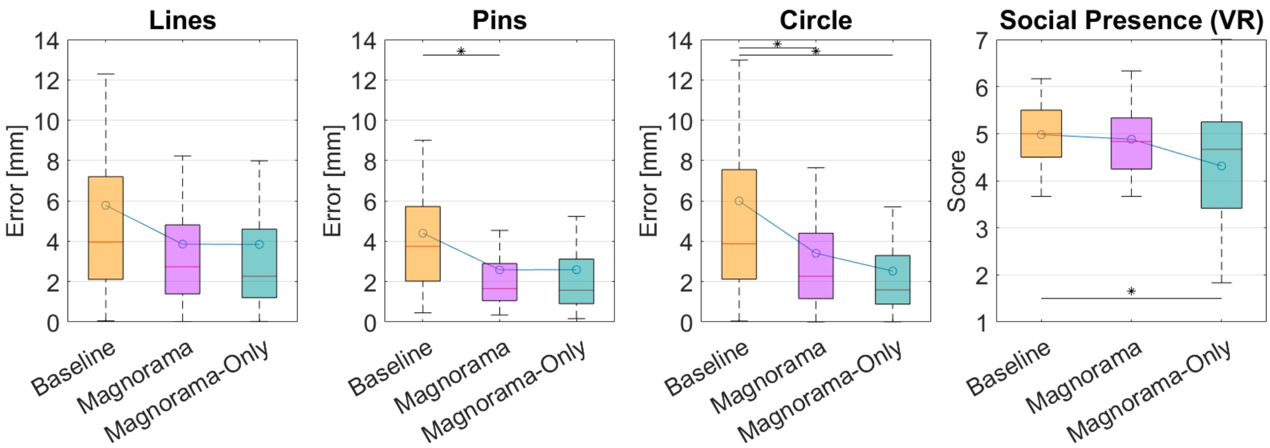


Figure 7: **Subjective and objective results of the study.** From left to right: (1-3) Box plots for the annotation errors from all participants. The red line indicates the median. The lower limit and upper limit of the box represent the 25th and 75th percentile. (4) Social presence as perceived by the participants when being the RE. Significance between conditions ($p < 0.05$) are marked with *.

condition ($M = 3.01$ mm, $SD = 2.33$ mm) significantly outperformed the baseline ($M = 6.43$ mm, $SD = 5.72$ mm ; all $p <= .003$). In this task, the Magnorama-Only condition performed significantly better than the Magnorama condition ($p = .001$).

Line Task Performance Greenhouse Geisser corrected values for the main effect on the mean error of the line error measurement showed no significant difference $F(1.14, 28.28) = 3.49$, $p = .068$, $\eta_p^2 = .132$. The baseline resulted in the highest mean error ($M = 6.79$ mm, $SD = 6.25$ mm), following the Magnorama condition ($M = 4.55$ mm, $SD = 3.42$ mm). The Magnorama-Only condition showed the lowest mean error ($M = 4.09$ mm, $SD = 2.60$ mm). No further significant effects were observed.

In summary, both Magnorama methods outperformed the baseline in all assessments, partly to a significant level. Regarding the mean error for drawing related tasks, the Magnorama-Only condition seems to outperform the Magnorama condition. However, the pin placements were more successful in the Magnorama condition.

Timing Results We recorded the time in which the participants performed each annotation task. We found a significant main effect for the line task; $F(2, 46) = 10.66$, $p < .001$, $\eta_p^2 = .317$. Pairwise comparisons revealed that the baseline method ($M = 18.43$, $SD = 9.14$) outperformed the Magnorama method ($M = 30.59$, $SD = 12.05$) as well as the Magnorama-Only method ($M = 31.35$, $SD = 19.61$, all $p <= .003$). There was no significant difference between Magnorama and Magnorama-Only.

This main effect was similarly present for the circle task with a slightly smaller effect size; $F(2, 46) = 3.78$, $p = .030$, $\eta_p^2 = .141$. Pairwise comparisons showed that the timing for the baseline ($M = 13.49$, $SD = 8.76$) was lower than for the Magnorama method ($M = 18.70$, $SD = 8.70$) as well as lower than the Magnorama-Only method ($M = 20.51$, $SD = 12.93$), but not to a statistically significant level.

Interestingly in the pin placement task, this effect was not present; $F(1.42, 32.66) = 1.30$, $p = .282$, $\eta_p^2 = .054$. Baseline ($M = 11.56$, $SD = 5.88$), Magnorama ($M = 15.06$, $SD = 13.82$), and Magnorama-Only ($M = 13.03$, $SD = 7.09$) were almost at the same level.

5.2 Subjective Results

We performed Friedman tests with consecutive Bonferroni adjusted pairwise comparisons for the subjective measures. We were mainly interested in the VR side (executing the annotation actions through the different methods) of the telepresence system.

Table 1: SUS Score (0-100) and Raw total TLX (0-100), ($M \pm SD$)

	Baseline	Magnorama	Magnorama-Only
SUS RE	74.17 ± 12.74	73.96 ± 11.86	74.17 ± 9.37
SUS LU	71.04 ± 13.91	71.35 ± 13.70	70.62 ± 13.40
TLX RE	26.00 ± 16.25	22.19 ± 16.03	21.90 ± 14.50
TLX LU	16.00 ± 11.22	16.60 ± 10.58	16.56 ± 9.93

For the VR side (RE), we found that the three conditions significantly impacted the level of social presence perceived by the participants; $\chi^2(2) = 6.66$, $p = .036$. Bonferroni corrected pairwise comparisons revealed that the baseline condition ($MDN = 5.00$) showed a significantly higher social presence than the Magnorama-Only condition ($MDN = 4.67$ $p = 0.30$). Differences between the baseline and the Magnorama condition ($MDN = 4.83$) or between the two Magnorama conditions were not significant. No significant differences were observed for the co-presence or telepresence measures. Further, no significant impacts on the presence factors on the AR side resulting from the different methods were found.

A Friedman's test for the SUS showed no significant difference in the usability assessment. All techniques were rated above a score of 70 for both the AR and the VR assessments (see Table 1. Friedman tests for the raw NASA TLX score (see [61] for a discussion), the NASA TLX subscales, and the FMS measure did not show significant differences between the conditions.

5.3 Preference and Comments

In VR, 12 participants preferred the Magnorama condition, 7 are undecided or did not answer, 4 preferred the baseline, and 1 preferred the Magnorama-Only condition. The participants liked the Magnorama condition because they can annotate on the magnified head while still being able to see the LU. Participants preferring the baseline condition perceived it as more natural compared to the other conditions. 18 participants in AR were unsure to pinpoint differences between the three conditions. 6 participants liked the baseline because the avatar directly worked on the head.

6 Discussion

Our study compared the proposed Magnorama technique to a baseline and a Magnorama variant that masks the situative context. Our results support *H1: The magnification of details in the form of Magnoramas increases free-hand drawing precision*, in the sense that the Magnorama conditions clearly outperformed the baseline conditions in many evaluated aspects to a significant level. The results show an improvement of the drawing precision using the magnified view of Magnoramas. This does not contradict one finding of Arora et al. [27], that drawing errors of larger objects in VR are higher, as the user-inputs are scaled back while using Magnoramas.

Magnoramas increased the time required for the line-tracing annotation tasks. We did not specifically draw a hypothesis on this aspect. Still, we suspect the reason for the increased time to be the increased length of the guiding lines inside the magnified region while users draw at similar speed in all conditions. This interpretation is backed by the fact that the time for placing pins did not show any significant difference between conditions. The green line task did not show significant improvement in precision, although having overall lower mean errors. This may be caused by the large region spanning from the forehead to the back of the ear and forces the user to change the point of view multiple times, whereas both, pin and circle tasks covered smaller regions.

The Magnorama-Only condition tends to yield lower error values. We assume by anecdotal observation that by hiding the environment, users are more likely to choose a larger scale of the Magnorama as space occupied by the original point cloud can be used to place the Magnorama. The precision can be further increased by choosing a smaller ROI and a larger Magnorama scale. This may be another starting point for further investigation.

Based on the TLX scores, our findings did not support *H2: Interacting with Magnoramas is inferior in terms of usability compared to context coherent interactions*. This was surprising since we would not have expected the Magnorama condition to be perceived similarly usable. One argument for the result may be that the users also perceived increased performance and, therefore, higher usability. For tasks where continuous lines need to be drawn (such as the line and circle task), users should consider a trade-off between an increased time-on-task and the magnification value.

The evaluation of the questionnaires showed further that among the three types of perceived presence, social presence perception was impacted by the conditions and found to be significantly lower for the RE in the Magnorama-Only condition compared to the baseline. This only partially supports *H3: The perception of the interaction in terms of co-presence and social presence aspects is inferior when using Magnoramas*. In addition, the Magnorama condition was able to maintain its perceived social presence while increasing the precision during the annotation tasks. We interpret this as the cause of the partial remaining coherence.

During the study, we observed that both participants exceedingly focus on the head during the drawing task and rarely look up at the participant. Therefore, we propose the use of Magnoramas when precise annotations or interactions are required. Before and after each task, automatic mechanisms could be incorporated to toggle the visualization of Magnoramas to regain a better perception of the communication partner. Magnoramas have a positive aspect on synchronicity, as the users can gesture, utilize non-verbal communication, etc., compared to the Magnorama-Only condition. On the other hand, in tasks requiring utmost concentration, such as the craniectomy, the Magnorama-Only setting can provide intentional concealment of the periphery as fewer distractions divert attention from the precision task and thus allowing the focus on the region of interest.

6.1 Limitations

There are some limitations. First, our study measured only the drawings from virtual ground truth to virtual space annotations. Therefore we cannot conclude the precision of annotations between virtual and physical relations. However, this was a conscious design choice since we did wanted to exclude additional noise from tracking and calibration errors from the experiment. For the same reason, only the precision for RE annotations was measured but not the precision of the drawings at the LU.Further research should investigate the error of LU annotations and the physical-to-virtual discrepancies. The number of left-handed participants was low (3 out of 24) for concluding its impact on the measurements. However, the randomized drawing direction dictated by the LU should mitigate the effect of handedness when reproduced on a larger sample. We did not explicitly measure the correlation between the degree of magnification and the time-on-task. Future studies using Magnoramas should monitor the drawing speed in the combination with the magnification. We are also aware that the quality of the real-time captured point cloud may introduce artifacts. Therefore, our findings with regard to the presence measures should be subject to further validation. Finally, for simplification and experimental control, we pre-defined the position of both the ROI and Magnorama. Users were neither required to choose the position and the initial sizes of the ROI by themselves. This may partially explain the usability results. In the desired target use-case, the ROI can be automatically selected through object detection based on the point cloud or opened manually by the user.

7 Future Work

Future work could integrate more compatible interactions for Magnoramas besides creating annotations, e.g., selecting and manipulating objects. We imagine Magnoramas hold potential as interactive second viewpoints, similar to Augmented Mirrors [65] to perform specific tasks, such as alignment or multi-modal visualization, more efficiently that are otherwise difficult. In the present work, we focused on the interaction of the VR user. In the future, we would like to compare new approaches in representing the avatar of the expert in AR since 25% of the participants in AR preferred it when the avatar directly annotated on the head. An exciting solution includes the attachment of the avatar at the real head in combination with the scaling of the avatar corresponding to the scale of the Magnorama. A similar approach has been investigated by Piumsomboon et al.[8] under the name of Mini-me. Consequently, rotating the Magnorama could have the avatar fly through the scene with a "jetpack" inside the AR view, presented by Piumsomboon et al. [66]. In a scenario with more than two users, the perceived coherence and social presence may be impacted. Future work could therefore consider augmenting both, social behavior [67, 11] and appearance [8] of the avatars, to potentially compensate for missing coherence.

8 Conclusion

We proposed Magnoramas as a selective magnifier of a region of interest inside a 3D telepresence system using a real-time captured point-cloud. In our study we found that the magnification through a Magnorama allows a user to draw annotations more precisely in trade for a lower perceived social presence of the communication partner. This effect was mostly mitigated when using the Magnorama along-side the original point-cloud. The increased precision from Magnoramas can be incredibly impactful for any teleconsultation system which allows freehand interactions. Moreover, they can be generalized to manifold use-cases but could be specifically beneficial for medical or industrial scenarios. We conclude that the value of Magnoramas is substantial for our scenario of a craniectomy and successful in increasing the precision and quality of the annotations, which opens a path for future endeavors.

Acknowledgements

The authors wish to thank Andreas Keller and Gleb Gorbachev for their support during the software development process. This work was funded by the German Federal Ministry of Education and Research (BMBF), No.: 16SV8092.

References

[1] S. Beck, A. Kunert, A. Kulik, and B. Froehlich, "Immersive Group-to-Group Telepresence," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 4, pp. 616–625, 2013.

[2] N. Weibel, D. Gasques, J. Johnson, T. Sharkey, Z. R. Xu, X. Zhang, E. Zavala, M. Yip, and K. Davis, "ARTEMIS: Mixed-Reality Environment for Immersive Surgical Telementoring," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–4, 2020.

[3] M. Slater, B. Spanlang, M. V. Sanchez-Vives, and O. Blanke, "First Person Experience of Body Transfer in Virtual Reality," *PloS one*, vol. 5, no. 5, p. e10564, 2010.

[4] K. Kilteni, R. Groten, and M. Slater, "The Sense of Embodiment in Virtual Reality," *Presence: Teleoperators and Virtual Environments*, vol. 21, no. 4, pp. 373–387, 2012.

[5] J. Hollan and S. Stornetta, "Beyond Being There," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 119–125, 1992.

[6] D. Roth, C. Kleinbeck, T. Feigl, C. Mutschler, and M. E. Latoschik, "Beyond Replication: Augmenting Social Behaviors in Multi-user Virtual Realities," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 215–222, IEEE, 2018.

[7] R. Fribourg, N. Ogawa, L. Hoyet, F. Argelaguet, T. Narumi, M. Hirose, and A. Lécuyer, "Virtual Co-Embodiment: Evaluation of the Sense of Agency while Sharing the Control of a Virtual Body among Two Individuals," *IEEE Transactions on Visualization and Computer Graphics*, 2020.

[8] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst, "Mini-me: An Adaptive Avatar for Mixed Reality Remote Collaboration," in *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–13, 2018.

[9] M. Slater and A. Steed, "Meeting People Virtually: Experiments in Shared Virtual Environments," in *The social life of avatars*, pp. 146–171, Springer, 2002.

[10] I. Podkosova, K. Vasylevska, C. Schoenauer, E. Vonach, P. Fikar, E. Bronederk, and H. Kaufmann, "ImmersiveDeck: A large-scale Wireless VR System for Multiple Users," in *2016 IEEE 9th Workshop on Software Engineering and Architectures for Realtime Interactive Systems (SEARIS)*, pp. 1–7, IEEE, 2016.

[11] D. Roth, G. Bente, P. Kullmann, D. Mal, C. F. Purps, K. Vogeley, and M. E. Latoschik, "Technologies for social augmentations in user-embodied virtual reality," in *25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19, (New York, NY, USA), Association for Computing Machinery, 2019.

[12] M. E. Latoschik, F. Kern, J.-P. Stauffert, A. Bartl, M. Botsch, and J.-L. Lugrin, "Not Alone Here?! Scalability and User Experience of Embodied Ambient Crowds in Distributed Social Virtual Reality," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 5, pp. 2134–2144, 2019.

[13] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou, *et al.*, "Holoportation: Virtual 3D Teleportation in Real-time," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pp. 741–754, 2016.

[14] K. Kiyokawa, H. Takemura, and N. Yokoya, "A Collaboration Support Technique by Integrating a Shared Virtual Reality and a Shared Augmented Reality," in *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)*, vol. 6, pp. 48–53, IEEE, 1999.

[15] M. Billinghurst, H. Kato, and I. Poupyrev, "The MagicBook: A Transitional AR Interface," *Computers & Graphics*, vol. 25, no. 5, pp. 745–753, 2001.

[16] 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*, pp. 405–415, 2015.

[17] A. Steed, W. Steptoe, W. Oyekoya, F. Pece, T. Weyrich, J. Kautz, D. Friedman, A. Peer, M. Solazzi, F. Tecchia, *et al.*, "Beaming: An Asymmetric Telepresence System," *IEEE computer graphics and applications*, vol. 32, no. 6, pp. 10–17, 2012.

[18] M. Le Chénéchal, T. Duval, V. Gouranton, J. Royan, and B. Arnaldi, "Vishnu: Virtual Immersive Support for Helping Users. An Interaction Paradigm for Collaborative Remote Guiding in Mixed Reality," in *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE)*, pp. 9–12, IEEE, 2016.

[19] T. Piumsomboon, Y. Lee, G. Lee, and M. Billinghurst, "CoVAR: A Collaborative Virtual and Augmented Reality System for Remote Collaboration," in *SIGGRAPH Asia 2017 Emerging Technologies*, pp. 1–2, 2017.

[20] R. A. J. de Belen, H. Nguyen, D. Filonik, D. Del Favero, and T. Bednarz, "A Systematic Review of the Current State of Collaborative Mixed Reality Technologies: 2013–2018," *AIMS Electronics and Electrical Engineering*, vol. 3, no. 2, p. 181, 2019.

[21] P. Ladwig and C. Geiger, "A Literature Review on Collaboration in Mixed Reality," in *International Conference on Remote Engineering and Virtual Instrumentation*, pp. 591–600, Springer, 2018.

[22] H. Fuchs and U. Neumann, "A Vision of Telepresence for Medical Consultation and other Applications," in *Sixth International Symposium of Robotics Research*, pp. 555–571, 1993.

[23] H. Fuchs, G. Bishop, K. Arthur, L. McMillan, R. Bajcsy, S. Lee, H. Farid, and T. Kanade, "Virtual Space Teleconferencing Using a Sea of Cameras," in *Proc. First International Conference on Medical Robotics and Computer Assisted Surgery*, vol. 26, 1994.

[24] A. Maimone and H. Fuchs, "A First Look at a Telepresence System with Room-sized Real-time 3D Capture and Life-sized Tracked Display Wall," *Proceedings of ICAT 2011, to appear*, pp. 4–9, 2011.

[25] T. Piumsomboon, A. Dey, B. Ens, G. Lee, and M. Billinghurst, "The Effects of Sharing Awareness Cues in Collaborative Mixed Reality," *Frontiers in Robotics and AI*, vol. 6, p. 5, 2019.

[26] I. Podkosova and H. Kaufmann, "Mutual Collision Avoidance During Walking in Real and Collaborative Virtual Environments," in *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, pp. 1–9, 2018.

[27] R. Arora, R. H. Kazi, F. Anderson, T. Grossman, K. Singh, and G. W. Fitzmaurice, "Experimental Evaluation of Sketching on Surfaces in VR.," in *CHI*, vol. 17, pp. 5643–5654, 2017.

[28] E. Wiese, J. H. Israel, A. Meyer, and S. Bongartz, "Investigating the Learnability of Immersive Free-hand Sketching," in *Proceedings of the seventh sketch-based interfaces and modeling symposium*, pp. 135–142, 2010.

[29] C. Moore, P. J. Dunham, and P. Dunham, *Joint Attention: Its Origins and Role in Development*. Psychology Press, 2014.

[30] 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 - CHI '95*, (Denver, Colorado, United States), pp. 265–272, ACM Press, 1995.

[31] A. Bluff and A. Johnston, "Don't Panic: Recursive Interactions in a Miniature Metaworld," in *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, VRCAI '19, (New York, NY, USA), pp. 1–9, Association for Computing Machinery, Nov. 2019.

[32] R. Trueba, C. Andujar, and F. Argelaguet, "Complexity and Occlusion Management for the World-in-Miniature Metaphor," in *International Symposium on Smart Graphics*, pp. 155–166, Springer, 2009.

[33] R. Trueba, C. Andujar, and F. Argelaguet, "Multi-scale Manipulation in Indoor Scenes with the World in Miniature Metaphor," in *Proceedings of the 15th Joint virtual reality Eurographics conference on Virtual Environments*, pp. 93–100, 2009.

[34] A. Mulloni, H. Seichter, and D. Schmalstieg, "Indoor Navigation with Mixed Reality World-in-Miniature Views and Sparse Localization on Mobile Devices," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pp. 212–215, 2012.

[35] F. Bacim, D. Bowman, and M. Pinho, "Wayfinding Techniques for Multiscale Virtual Environments," in *2009 IEEE Symposium on 3D User Interfaces*, pp. 67–74, IEEE, 2009.

[36] D. Coffey, N. Malbraaten, T. Le, I. Borazjani, F. Sotiropoulos, and D. F. Keefe, "Slice WIM: a Multi-surface, Multi-touch Interface for Overview+ Detail Exploration of Volume Datasets in Virtual Reality,"

in *Symposium on Interactive 3D Graphics and Games*, pp. 191–198, 2011.
[37] C. A. Wingrave, Y. Haciahmetoglu, and D. A. Bowman, "Overcoming World in Miniature Limitations by a Scaled and Scrolling WIM," in *3D User Interfaces (3DUI'06)*, pp. 11–16, IEEE, 2006.
[38] J. S. Pierce, B. C. Stearns, and R. Pausch, "Voodoo Dolls: Seamless Interaction at Multiple Scales in Virtual Environments," in *Proceedings of the 1999 symposium on Interactive 3D graphics*, pp. 141–145, 1999.
[39] R. S. Renner, B. M. Velichkovsky, and J. R. Helmert, "The Perception of Egocentric Distances in Virtual Environments - A Review," *ACM Computing Surveys (CSUR)*, vol. 46, no. 2, pp. 1–40, 2013.
[40] H. Elsayed, M. D. Barrera Machuca, C. Schaarschmidt, K. Marky, F. Müller, J. Riemann, A. Matviienko, M. Schmitz, M. Weigel, and M. Mühlhäuser, "VRSketchPen: Unconstrained Haptic Assistance for Sketching in Virtual 3D Environments," in *26th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11, 2020.
[41] M. D. B. Machuca, W. Stuerzlinger, and P. Asente, "Smart3DGuides: Making Unconstrained Immersive 3D Drawing More Accurate," in *25th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–13, 2019.
[42] M. D. Barrera-Machuca, A. Cassinelli, and C. Sandor, "Context-Based 3D Grids for Augmented Reality User Interfaces," in *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pp. 73–76, 2020.
[43] M. D. B. Machuca, P. Asente, W. Stuerzlinger, J. Lu, and B. Kim, "Multiplanes: Assisted Freehand VR Sketching," in *Proceedings of the Symposium on Spatial User Interaction*, pp. 36–47, 2018.
[44] T. Drey, J. Gugenheimer, J. Karlbauer, M. Milo, and E. Rukzio, "VRSketchIn: Exploring the Design Space of Pen and Tablet Interaction for 3D Sketching in Virtual Reality," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2020.
[45] M. D. Barrera Machuca, W. Stuerzlinger, and P. Asente, "The Effect of Spatial Ability on Immersive 3D Drawing," in *Proceedings of the 2019 on Creativity and Cognition*, pp. 173–186, 2019.
[46] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse, "The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 529–536, 2003.
[47] D. Roth, J.-L. Lugrin, D. Galakhov, A. Hofmann, G. Bente, M. E. Latoschik, and A. Fuhrmann, "Avatar Realism and Social Interaction Quality in Virtual Reality," in *2016 IEEE Virtual Reality (VR)*, pp. 277–278, IEEE, 2016.
[48] D. Roth, K. Waldow, M. E. Latoschik, A. Fuhrmann, and G. Bente, "Socially immersive avatar-based communication," in *2017 IEEE Virtual Reality (VR)*, pp. 259–260, 2017.
[49] P. Stotko, S. Krumpen, M. B. Hullin, M. Weinmann, and R. Klein, "Slamcast: Large-scale, Real-time 3D Reconstruction and Streaming for Immersive Multi-client Live Telepresence," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 5, pp. 2102–2112, 2019.
[50] T. Pejsa, J. Kantor, H. Benko, E. Ofek, and A. Wilson, "Room2room: Enabling Life-size Telepresence in a Projected Augmented Reality Environment," in *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*, pp. 1716–1725, 2016.
[51] B. Thoravi Kumaravel, F. Anderson, G. Fitzmaurice, B. Hartmann, and T. Grossman, "Loki: Facilitating Remote Instruction of Physical Tasks using bi-directional Mixed-Reality Telepresence," in *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 161–174, 2019.
[52] O. Oda and S. Feiner, "3D Referencing Techniques for Physical Objects in Shared Augmented Reality," in *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 207–215, IEEE Computer Society, 2012.
[53] J. Kolkmeier, E. Harmsen, S. Giesselink, D. Reidsma, M. Theune, and D. Heylen, "With a Little Help from a Holographic Friend: The OpenIMPRESS Mixed Reality Telepresence Toolkit for Remote Collaboration Systems," in *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11, 2018.
[54] R. L. Daft and R. H. Lengel, "Organizational Information Requirements, Media Richness and Structural Design," *Management science*, vol. 32, no. 5, pp. 554–571, 1986.
[55] K. L. Nowak and F. Biocca, "The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments," *Presence: Teleoperators & Virtual Environments*, vol. 12, no. 5, pp. 481–494, 2003.
[56] D. Roth, J.-L. Lugrin, J. Büser, G. Bente, A. Fuhrmann, and M. E. Latoschik, "A Simplified Inverse Kinematic Approach for Embodied VR Applications," in *2016 IEEE Virtual Reality (VR)*, pp. 275–276, IEEE, 2016.
[57] D. Roth and M. E. Latoschik, "Construction of the virtual embodiment questionnaire (veq)," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 12, pp. 3546–3556, 2020.
[58] G. Turk and M. Levoy, "Zippered Polygon Meshes from Range Images," in *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pp. 311–318, 1994.
[59] J. Brooke, "SUS: a "Quick and Dirty'Usability," *Usability evaluation in industry*, p. 189, 1996.
[60] 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, pp. 139–183, Elsevier, 1988.
[61] 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, pp. 904–908, Sage publications Sage CA: Los Angeles, CA, 2006.
[62] B. Keshavarz and H. Hecht, "Validating an Efficient Method to Quantify Motion Sickness," *Human factors*, vol. 53, no. 4, pp. 415–426, 2011.
[63] S. Ishihara, "Series of Plates Designed as Tests for Colour-blindness," 1936.
[64] S. G. Vandenberg and A. R. Kuse, "Mental Rotations, a Group Test of Three-dimensional Spatial Visualization," *Perceptual and motor skills*, vol. 47, no. 2, pp. 599–604, 1978.
[65] 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)*, pp. 217–226, IEEE, 2020.
[66] T. Piumsomboon, G. A. Lee, B. Ens, B. H. Thomas, and M. Billinghurst, "Superman vs Giant: A Study on Spatial Perception for a Multi-scale Mixed Reality Flying Telepresence Interface," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 11, pp. 2974–2982, 2018.
[67] D. Roth, M. E. Latoschik, K. Vogeley, and G. Bente, "Hybrid avatar-agent technology – a conceptual step towards mediated "social" virtual reality and its respective challenges," *i-com*, vol. 14, no. 2, pp. 107 – 114, 01 Aug. 2015.

Part IV

Duplicated Reality

Introduction to Duplicated Reality

4

Physical objects do not constrain VR users except those in their real-world environment. Although the lack of haptic feedback is a significant challenge in VR, it allows for a free range of motion and positioning within the virtual space. This assumption does not hold in co-located consultations where multiple users share a physical space. For example, co-located teleconsultation can occur in surgical environments when an unexpected issue arises and a surgeon calls in an expert on short notice. Since entering and leaving the sterile environment surrounding the patient is connected with the preparation of sterile surgical gowns and coordination between involved hospital staff, nurses, and the consulting expert, they may desire an alternative way to interact with the patient in 3D. Co-located consultations may also be necessary for planned surgeries that require a diverse range of expertise, necessitating multi-disciplinary teams. Multiple studies have shown that two or more specialized surgeons yield higher patient satisfaction regarding the outcome [227, 228] and faster operations [229], leading to higher cost-efficiency and patient satisfaction.

In our user study, we investigate the collaborative surgical environment of Pedicle Subtraction Osteotomy. This complex surgery requires extensive spinal area knowledge and would benefit from the combined expertise of an orthopedist and a neurosurgeon. Therefore, a multi-disciplinary surgical team can improve patient outcomes and reduce operating times by working in tandem.

The next question is whether Magnoramas, initially designed for teleconsultation, can also be utilized in a co-located environment. This adaptation would aim to replicate the experience of a remote consultant being physically present and moving in and out of spaces occupied by other individuals while maintaining co-locality for direct actions. I present the term "Duplicated Reality" to describe this method, which differs fundamentally from Magnoramas as it involves replicating the real world into augmented virtual content rather than replicating virtual to virtual content.

The user study findings indicate that participants intentionally use more space when utilizing the Duplicated Reality method to avoid physical interference, unlike when in absence, participants would frequently change their position around the patient to coordinate consultation with intervention. While no significant improvements in task outcomes were observed with the use of co-located consultation for shaping a block of kinetic sand, the results suggest a potential enhancement in task performance when repeated with a larger number of participants.

Author's Contribution to the Publication

The author of the dissertation (Kevin Yu) conceived the method, planned and conducted the user study, wrote the paper, and presented the work at the conference of IEEE VR 2022 as part of the journal track and later to be published as a special issue in the Transactions on

Visualization and Computer Graphics (TVCG). Dr. Frieder Pankratz and Dr. Marc Lazarovici are collaborators in the ArtekMed project from the Institute of Emergency Medicine in Munich. The system and user study were located at the institute's Human simulation center. Prof. Dr. Dirk Wilhelm is the author's medical consultant for the medical use case and medical supervisor. Dr. Ulrich Eck and Prof. Dr. Nassir Navab are the technical supervisors of the author.

The following article is the accepted version acquired through the online library "IEEE Xplore". The right of reuse is included in the appendix (section VIII).

Duplicated Reality for Co-located Augmented Reality Collaboration

Kevin Yu, Ulrich Eck, Frieder Pankratz, Marc Lazarovici, Dirk Wilhelm, and Nassir Navab

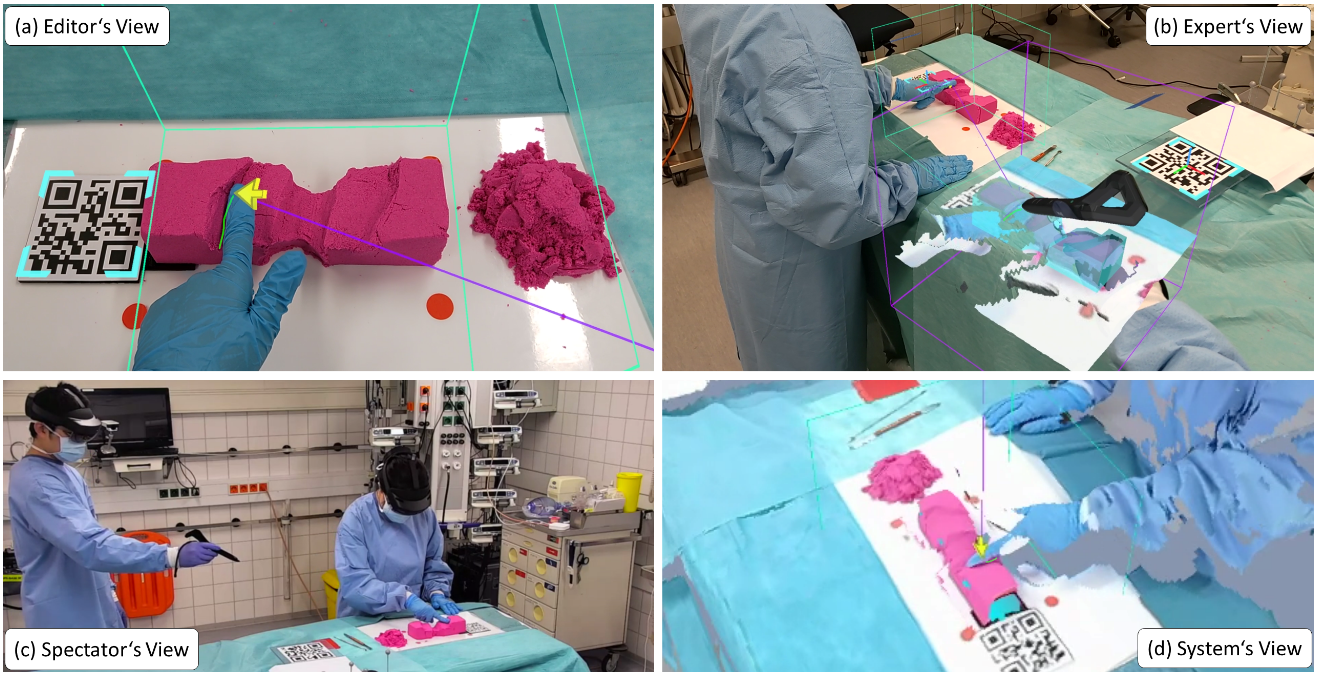


Fig. 1: **The Concept of Duplicated Reality** from multiple views at the same time. Duplicated Reality (DR) transforms co-located face-to-face encounters by replicating the physical world into a digital representation. (a-b) An editor operates in the region of interest under the real-time guidance of the co-located expert through the DR. Annotations made by an expert inside the DR are recreated in front of the editor. (c) A spectator's view (d) System's view on the reconstruction for creating the DR.

Abstract— When two or more users attempt to collaborate in the same space with Augmented Reality, they often encounter conflicting intentions regarding the occupation of the same working area and self-positioning around such without mutual interference. Augmented Reality is a powerful tool for communicating ideas and intentions during a co-assisting task that requires multi-disciplinary expertise. To relax the constraint of physical co-location, we propose the concept of Duplicated Reality, where a digital copy of a 3D region of interest of the users' environment is reconstructed in real-time and visualized in-situ through an Augmented Reality user interface. This enables users to remotely annotate the region of interest while being co-located with others in Augmented Reality. We perform a user study to gain an in-depth understanding of the proposed method compared to an in-situ augmentation, including collaboration, effort, awareness, usability, and the quality of the task. The result indicates almost identical objective and subjective results, except a decrease in the consulting user's awareness of co-located users when using our method. The added benefit from duplicating the working area into a designated consulting area opens up new interaction paradigms to be further investigated for future co-located Augmented Reality collaboration systems.

Index Terms—User Interaction, Mixed Reality, 3D Reconstruction

1 Introduction

Interdisciplinary collaboration is a substantial part of work and life, including various branches of industry and healthcare. Augmented Reality (AR) can provide intuitive visualization and is an interactive media for communicating complex three-dimensional (3D) connections during face-to-face encounters. One of the strengths of AR is its capability to directly overlay computer-generated graphics onto physical objects to provide additional information. Additionally, allowing users to annotate scenes with real-time annotations enhances collaborative experiences immensely as users gain a method for intuitively visualizing their ideas. An example of an interdisciplinary encounter is the surgical operating room. Due to the complexity of human anatomy, surgeons develop highly focused expertise in their discipline. However, patients with rare diseases or operations which require rare procedures demand the expertise of different branches of medicine. For instance, spine surgeries may demand knowledge of a orthopedic- and a neurosurgeon. Recent work has found improvement of the patient outcome when surgeons of multiple disciplines work collaboratively [2]. The usage of conventional in-situ AR interaction encounters several challenges during such space-critical use-cases. The challenges are notable when users need to annotate anatomical structures, such as: (1) AR annotation tools cannot be used in sterile areas unless certified, (2) an-

• *Kevin Yu and Dirk Wilhelm are with the University Hospital Rechts der Isar, Technical University Munich, Germany. E-mail: kevin.yu@tum.de.*
• *Frieder Pankratz and Marc Lazarovici are with the Institute for Emergency Medicine, Ludwig-Maximilian University, Munich, Germany.*
• *Ulrich Eck and Nassir Navab are with the Chair for Computer Aided Medical Procedure, Technical University Munich, Germany.*

Manuscript received 6 Sept. 2021; revised 3 Dec. 2021; accepted 7 Jan. 2022. Date of publication 11 Feb. 2022; date of current version 29 Mar. 2022. Digital Object Identifier no. 10.1109/TVCG.2022.3150520

notations are not easily placed beneath surfaces for indicating incisions, (3) blocking the space for the acting surgeon, and (4) having to walk around the operating table to create desired annotations. For general face-to-face encounters, the collaborative experience may similarly be hindered if several co-located users attempt to simultaneously interact or annotate the same space.

Contribution We propose the concept of a Duplicated Reality (DR) for enhancing co-located collaboration in AR by duplicating the physical world, including augmented information, into a separate and interactive copy. DR maps user interactions performed on the copy, including annotations, into the location from where it was duplicated (see Fig. 1). Our method solves multiple challenges mentioned above: (a) users can remotely annotate the region of interest while remaining in a co-located space, (b) annotations can be placed beneath physical surfaces since the duplication is rendered mid-air through AR, (c) relaxes spatial constraints such that users can interact with the region of interest from a distance, (d) consultants remain eligible for solving local tasks unlike with remote consultation, and (e) allows users to view the region of interest from different angles by simply rotating the duplication in AR.

We put forward a possible implementation to our proposed concept and conducted a within-subject user study to compare DR with in-situ AR visualization for a collaborative task. We show that our proposed concept of DR is interchangeable with no adverse effects except a decreased awareness of the consulting user on the task, while providing solutions to the aforementioned challenges for co-located AR collaboration.

2 Related Work

The novelty of this work builds on top of a large variety of ideas originating from AR, Virtual Reality (VR), and known concepts for Mixed Reality (MR) collaboration. Related work can be grouped into the categories of World-in-Miniature, Digital Twin, the topic of real-time reconstructions, and MR interaction concepts for single and multi-user applications. In the following, we discuss prior work and justify our novelty.

World-In-Miniature An early concept of remote manipulation of objects in VR is the World-in-Miniature (WiM) by Stoakley et al. [60]. WiMs are used most commonly in VR systems to manipulate distant virtual objects [11,30] and for navigation [26,41,66,75]. The concepts can be extended to exploration of data, as shown by Wingrave et al. [17]. As the name suggests, WiMs shrink large virtual environments into a miniature size to allow users to interact with the environment as a proxy. The concept of keeping the original or increasing the scale of the environment, such as with DR, is typically not further explored.

Digital Twin An alternative to WiMs that retain the original size, the concept of digital twins is the closest interaction paradigm filling this gap. In VR systems, a copy of a virtual object can be created to manipulate the original object remotely. This concept is presented as Voodoo dolls by Pierce et al. [48]. Later work combined virtual models, often CAD models, with MR visualization for enabling guided assembly [52] or teleconsultation using a Virtual Replica [44]. We utilize a digital twin as the baseline condition in our work and compare it with our presented method.

Environment Reconstruction Our proposed method utilizes real-time captured depth images from RGB-D sensors to compute a real-time 3D reconstruction. In related work, real-time reconstruction is used for asymmetric VR/AR telepresence system [5, 19, 46, 54, 57, 65, 72], or screen-based telepresence [38,77]. Non-real-time reconstructions can be combined with real-time interaction methods to generate a higher grade of detail of the reconstruction while allowing teleconsultation in mostly static scenes [61,62]. Further work has investigated AR as an interactive method in creating environment reconstructions [47], or as a medium to use reconstruction data to superimpose real-world locations, objects, or patients [27,51,76].

Multi-view Interaction Methods Interaction methods utilizing multiple views on one region have been investigated with the combination of 3D reconstruction in VR and AR environments. Yu et al. [78] proposed the method of Magnoramas, which duplicates a cubic region of a real-time reconstruction as a method to create precise annotations in an immersive virtual environment. In opposition, DR duplicates the physical world rather than the reconstructed world in VR. A further difference between both methods is that physical space is abundantly available in [78] and no physical encounters may occur since users are connected through a digital medium, unlike in co-located spaces. In AR environments, Martin et al. [39] proposed the use of actual physical mirrors to gain multiple views onto an object of interest for superior scene understanding and added benefits of computer-generated visualizations. Additional work includes the usage of 3D scene capture for the selection of distant objects in AR [6,45]. The latter work uses point clouds acquired through RGB-D sensors, allowing users to select and highlight distant real-world objects in AR more accurately. There are few studies that address multiple-view methods in AR/VR and even fewer that additionally address collaborative use-cases.

Collaboration in Mixed Reality We summarize related work on symmetric and asymmetric AR collaborative systems in the following paragraphs. Symmetric systems are frequently deployed for tasks that involve both users in the same location. For example, Billinghurst et al. [9, 10] showed AR allows users to solve a collaborative task faster when compared to face-to-face and using a projection-based visualization. A lower time-on-task has been repeatedly observed for single-user systems, such as assembly [3, 64]. In AR systems, virtual objects are typically placed and sized in the context of the physical world and commonly occupy only small parts of the overall perceived environment. Examples are the joint exploration of virtual objects [35,63] or drawing together in 3D AR space [31].

In contrast, asymmetric collaboration using AR and a VR technology relaxes the constraint of synchronicity in location. Asymmetry for immersive collaboration is created by capturing the scene of a local user in AR through external cameras or transmitting the egocentric view directly to the remote user. Teleconsultation is one of the typical use-cases for which such asymmetry is favourable [21,22,32,49,50,53,78], equally as for time-asynchronous consultation [28]. A disadvantage of asymmetric teleconsultation is the rigid role distribution. The remote user is assumed to be the person with greater knowledge on the collaborative task and a quick role reversal is usually not possible.

Interaction Tools During MR Collaboration Two recurrent methods in communicating intentions, besides voice, in collaborative MR system, are gesturing and drawing shapes, such as 3D lines. A remaining problem of 3D applications is the perception of depth [56], and along the context, the creation of 3D shapes and lines in MR. An approach for improving the spatial ability [4] is the addition of non-constraining [37] or constraining [36] guides. Moreover, post-processing [14] of user-input can further improve the quality of 3D drawings. Besançon et al. [7] published an in-depth review of existing interaction tools.

Novelty of Duplicated Reality DR fills the gap of co-located collaboration allowing both users to work at the same physical place without being confined to the same location. This enables novel types of interaction with reality since users no longer operate solely on the virtual world and neither solely on the augmented world. Instead, DR can be seen as a hybrid of both, utilizing methods from both AR and VR disciplines.

2.1 Hypotheses

We duplicate the region of interest of the physical world into a separate, sensor-captured copy represented by a triangulated point cloud. This attempts to fill the gap between both types of collaboration, allowing both users to work at the same physical place without being confined to the same location. We compare our method with the baseline of in-situ augmentations using AR. We presume users will perceive the baseline method as more natural for solving our task due to the tangible interface metaphor [9]; however, based on previous findings [78], we

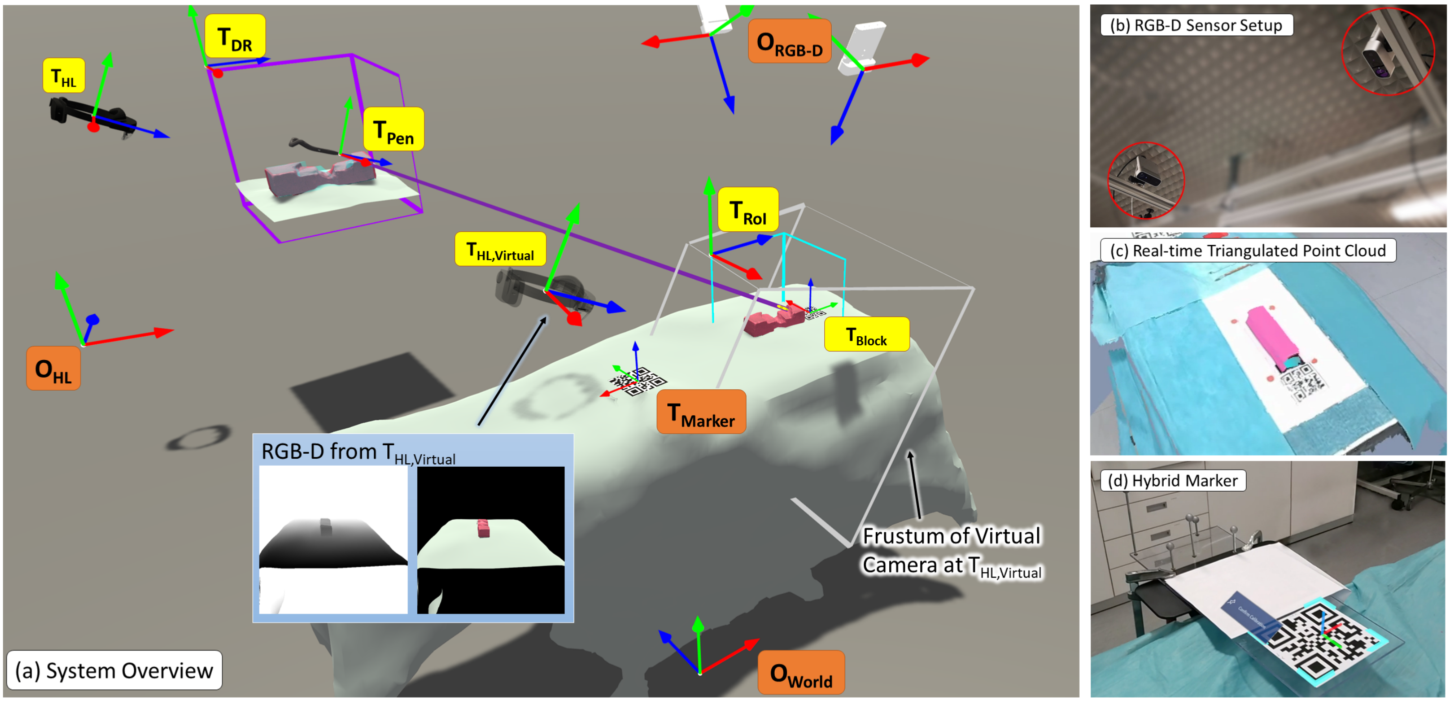


Fig. 2: **Overview of the System Setup** (a) A diagram providing of the most important transformations. Orange elements mark fixed transformations and poses, calibrated offline or before the task. Yellow transformations are continuously tracked and synchronized between all users. We calibrate all devices involved into the common origin of O_{World}, either by bundle adjustment (RGB-D cameras), QR marker detection (HMDs), or absolute orientation calibration (pen). The virtual camera acquires the content of the Duplicated Reality by simulating an RGB-D sensor. (b) Two Azure Kinect RGB-D cameras are rigidly attached to the ceiling, looking down onto the table with the region of interest. (c) Real-time triangulated point cloud from the RGB-D cameras, as the ground data for Duplicated Reality. (d) Multi-modal marker for the calibration of the head-mounted displays and the stylus. An AR button confirms correct tracking of the marker and re-calibrates the HMD origin O_{HL}.

anticipate better results regarding the task using our DR. Therefore, we anticipate lower subjective results in trade for better qualitative results for our proposed method on the collaborative task. In particular, we hypothesize that

- H1: Collaboration, effort, and awareness will be impacted negatively when using the DR since the baseline follows the more intuitive approach of in-situ augmentation with a digital twin.
- H2: the task-load and system usability are better in the baseline condition since users can draw directly on the object to be edited rather than on an intermediate proxy such as the DR.
- H3: the qualitative result of the task will improve when using DR. The precision of the tasks improves since users can magnify the region of interest. The time-on-task will decrease when using DR since both users can work in parallel without mutual interference.

3 Methods

We present the concept of DR for duplication of a physical space into a virtual replication of the real world that acts as a virtual proxy to perform interactions inside a region of interest. A DR requires at least two components: 3D environment capturing sensors and AR-enabled devices, preferably head-mounted displays for perceiving computer-generated content in stereovision. In the following paragraphs, we use the notation O to indicate device internal origins and T for transformations relative to the common origin O_{World} and as seen in Fig. 2. We deploy two Azure Kinect cameras above the operating table for capturing two perspectives of the region of interest with its origin T_{RoI}.

Referring to Fig. 2, a virtual bounding box at T_{DR} containing a subset of the 3D reconstruction framed by the region of interest at T_{RoI} is shown for the users. Users can interact with the DR at T_{DR} by moving, rotating, and scaling. DR allows users in Augmented Reality to participate in the activities of the physical world while simultaneously providing a portable, replicated 3D view capable of relaxing spatial constraints of co-location and supernatural precise interactions (e.g. 3D annotations) through increasing the scale of the 3D reconstruction.

In our user study, we investigate the effects of our system in a co-located collaborative task of a Pedicle Subtraction Osteotomy (PSO) procedure, originated from orthopedic surgery and abstracted to accommodate participants without medical training. In this context, we divide the collaborative task into the role of the editor and the expert for the remaining content. The editor represents the acting surgeon operating on the patient, while the expert represents a co-surgeon, or consultant, with knowledge of a complementary expertise during the operation. We provide a description of the use-case in Sect. 4.1.

3.1 Implementation

We utilize two Microsoft Hololens 2 head-mounted displays (HMDs) for our setup. The Hololens 2 is an optical see-through HMD with six degrees of freedom tracking using Simultaneous Mapping and Localization (SLAM [13]). To reduce the bandwidth requirements, we first stream the sensor data of multiple RGB-D sensors to a powerful workstation (Intel Core I7, 64GB RAM, NVidia RTX 2080Ti) with rendering capabilities for processing and hosting the network connection to the HMDs. We implemented the applications running on the processing unit and the HMDs with Unity3D and C#. We further integrate a VR stylus (LogitechVR Ink pen) into the system, allowing users to create 3D line annotations in AR more precisely than using the built-in hand-tracking of the Hololens 2. The number of co-located users in AR is only restricted by the bandwidth of the local wireless network and computational power of the workstation.

Interaction Users can directly interact with the wireframe box of the DR with the hand-tracking capabilities of the Hololens 2. By entering and pinching the box at T_{DR} with both hands, users can move, rotate, and scale the box. The gestures for each type of rigid transformation resemble touch controls for smartphone devices, that is, moving two hands in parallel moves the box, moving one hand while keeping the other hand fixed rotates it around the fixed point, and moving hands in opposite or closing the distance resizes the box. The box marking the region of interest at T_{RoI} is rendered light-blue to differentiate between the DR box at T_{DR}, rendered in purple (see Fig. 2(a)).

To raise the awareness of the editor on the experts interaction, the position of the stylus at $T_{Pen,DR}$, when inside the box at T_{DR}, is shared with the editor by augmenting a cursor object at the position of $T_{Pen,RoI}=T_{RoI} \bullet (T_{DR}^{-1} \bullet T_{Pen})$. Furthermore, $T_{Pen,DR}$ and $T_{Pen,RoI}$ are connected with a line. The cursor has the added benefit of allowing pointing gestures without the necessity of 3D annotation.

Calibration The extrinsic parameters of the RGB-D cameras are estimated using a wanding procedure similar to commercial IR tracking systems. The initial pose O_{World} is detected through template matching of a known target in the IR image (five IR spheres) and refined through optimizing the tracked target with bundle-adjustment to estimate their pose O_{RGB-D} within O_{World}. A hybrid marker at T_{Marker}, consisting of a rigid construction with a QR marker and five reflective spheres (see Fig. 2(d)), is used to establish the extrinsic calibration between the stylus at T_{Pen}, and the origins of the HMDs O_{HL}. Next, we rigidly estimate the pose of the hybrid marker to the common origin using the reflective spheres. This is done using Hand-Eye Calibration [67].

In the beginning of every collaborative task, we calibrate the HMDs with the QR marker of the hybrid marker, since the internal origin of the Hololens 2 is different every time our application is launched. We use the built-in QR marker detection and estimate the necessary transformation to calibrate the origins O_{HL} for each user to the common origin O_{World}. Finally, we calibrate the pen to retrieve 3D coordinates by selecting the four corners of the QR marker and transform the origin of the pen to match O_{World}. This method is called absolute orientation and was proposed by Tuceryan et al. [68].

Duplication Pipeline for the HMDs Major challenges in the implementation of DR on untethered devices, such as the Hololens 2, are the network transmission of the 3D reconstruction, the optimization for a reasonable update rate, and the quality of visualization. The amount of data acquired by both RGB-D sensors easily exceeds the wireless network bandwidth connecting the HMDs. We exploit the fact that the DRs only require a subset of the reconstruction and tailor a user-specific optimization step.

We use the workstation to receive the video streams of the RGB-D cameras and render them into triangulated point clouds. The host application receives the pose of the HMDs in the common calibrated coordinate system O_{World} and proceeds to compute a subset of the reconstruction based on the updated pose. To retrieve this subset, we simulate a virtual RGB-D camera from the pose of the HMDs that renders a field of view slightly larger than the rendering camera of the Hololens 2. We render the reconstruction into the depth buffer of the GPU, which is then retrieved by the virtual RGB-D cameras from pose $T_{HL,Virtual}=T_{RoI} \bullet (T_{DR}^{-1} \bullet T_{HL})$. The color image is compressed using lossy jpeg. In contrast, the compression of the depth image is aimed for minimal loss in depth accuracy and maximum compression rate, as described in the following: Every depth image pixel encodes a single byte that represents the z-value at the given point in camera coordinates with values between 0 and 255. Thus, every value decodes a distance between the near and far clipping distance of the capturing virtual camera. The accuracy of depth is therefore compromised. Under normal circumstances of transmitting depth, this method would introduce a significant loss in accuracy for precision tasks; however, the DR only requires information of the region of interest. Hence, we dynamically adjust the near and far clipping plane only to fully encapsulate the needed size to capture the region of interest. An estimate of the depth accuracy can be calculated with

$$x = \frac{d_{far} - d_{near}}{256}$$

For example, a 30cm per edge region of interest results would have a depth accuracy of around 1mm precision. Further compression is achieved by run-length encoding (RLE). Our implementation transmits a 640x640 pixel depth image with approximately 7kB to 20kB, depending on the compression rate of RLE. We considered RVL compression [74] as an alternative; however, the compression rate for empty or near-empty depth images was lower compared to our implementation. The Temporal-RVL method [29] improves the compression rate of RVL but requires small deltas in image sequences to be effective.

The workstation sends the compressed buffers of the virtual RGB-D camera back to the respective Hololens 2 via the Unity3D network library Mirror using the kcp network protocol on the reliable sequenced channel. Additional meta information, such as the intrinsic parameter, position, and orientation of the simulated depth camera alongside the received RGB-D stream, allow the Hololens 2 app to reassemble the depth image into a triangulated ordered point cloud using a compute shader. The app further textures the resulting point cloud with the transmitted color image. The resulting point cloud is transformed to T_{DR}, scaled from it's center, and clipped to the box.

Our method is robust against the jitter of head movement and allows low refreshing intervals of the reconstruction to accommodate limited network bandwidth. We measure an end-to-end latency between capture and visualization inside the DR of 900ms. Of these 900ms, 320ms is used for internal processing of the RGB-D camera, 180ms for transmission and rendering inside the processing unit. The remaining 400ms is used for wireless transmission and concurrent preparation for visualization inside the HMD. The resulting application on the HMD runs on average with 45 fps with a one second update interval of the DR content.

4 Study

We conducted a within-subjects, dual-user study to evaluate the benefits of DR compared to in-situ augmentation based on perceived collaboration, effort, awareness, usability [34], task load [25], and qualitative measures. The measure on collaboration, effort, and awareness follow the scheme as proposed by Schafer et al. [59]. We require the pairs of participants to be familiar with each other to reduce shyness and communication barriers during tasks.

The task for a participant is derived from an orthopedic spine surgery procedure, namely Pedicle Subtraction Osteotomy (PSO). This type of surgery is proven to be more effective, both in time and patient outcome, with two operating surgeons of different expertise [2], i.e. a neuro and an orthopedic surgeon.

4.1 Use-Case: Pedicle Subtraction Osteotomy with a Co-Surgeon

Typical PSO patients are aged approximately 60 years. Among them, around 60-70% are women. While a healthy person's spine has the natural shape of an S-shaped curve, patients suffering kyphosis (also known as "roundback") have a more excessive curvature and are unable to assume the healthy posture. In some cases, patients require PSO to correct the spine's shape.

Ames et al. [2], and Cheng et al. [15] have investigated PSO surgeries and compared the outcome of multiple operations with a single surgeon and with two attending surgeons. They found that conducting this type of surgery with two surgeons results in significantly reduced blood loss and time while improving the procedure outcome. Furthermore, the concept of two attending surgeons has been extended to other types of medical disciplines such as breast surgery [24], and patient care [18] with similar observations. The procedure of PSO (also shown in Fig. 3) includes the following steps:

1. Open the back of the patient to access the spine (most commonly around vertebra L3 [23])
2. Insertion of stainless rods and guides to shape the spine later on
3. Removal of the lamina in a V-shape (pointy side of the vertebra)
4. Removal of the pedicles on both sides of the vertebra
5. Adjusting guide and closing

Mummaneni et al. [42] provide a detailed description on PSO. Since engaging multiple surgeons in a user study is challenging, we simplify the task, with the consultation of a senior physician, by abstracting the steps 1, 3, and 4 into a modeling task where users have to shape a block of kinetic sand (5cm x 6cm x 20cm) into the desired form. During the remaining parts, we refer to the user shaping the kinect sand as *editor* and the user that provides guides as the *expert*. The editor represents the

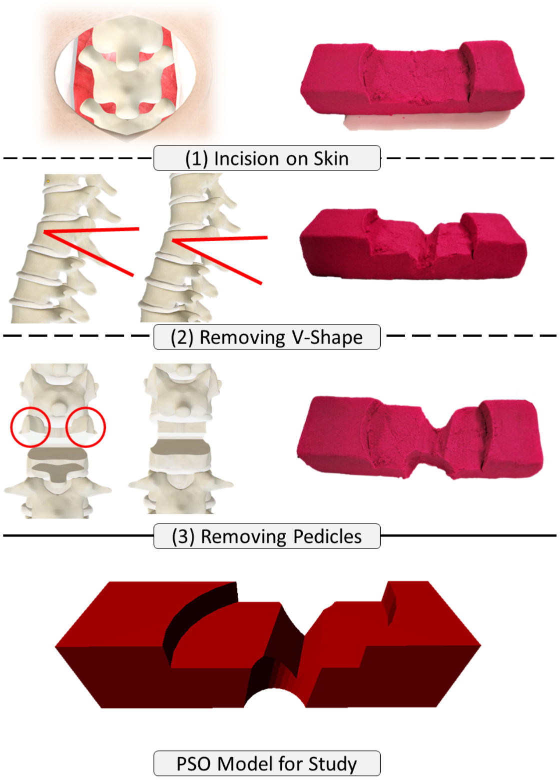


Fig. 3: **Abstraction of a Pedicle Subtraction Osteotomy** into a single PSO model. During the user study, the expert communicates the shape of the model to the editor. (1) Incision of the skin is represented as an oval disk. (2) Partial removal of the vertebra is represented by a V-shaped notch. (3) Removal of both pedicles is represented by cylinders on each side of the block.

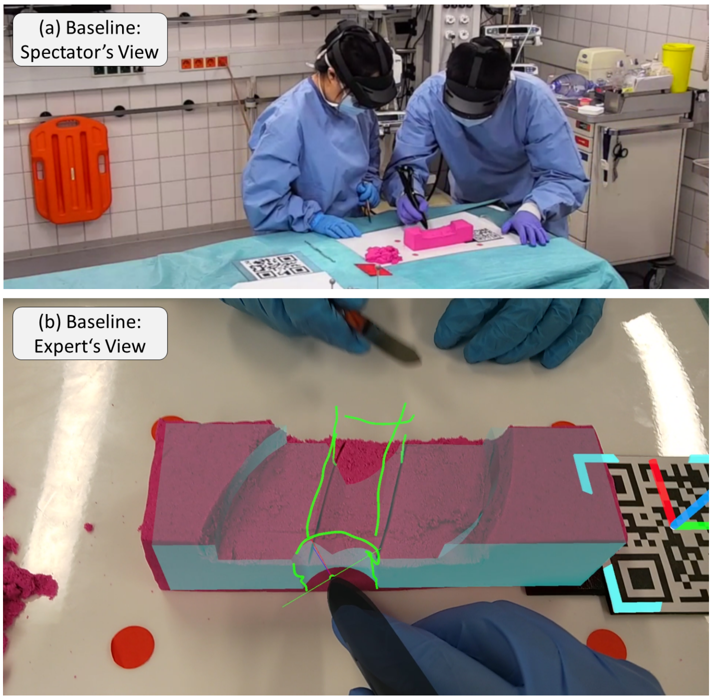


Fig. 4: **The Baseline Condition** (a) The baseline condition forces both users to work in close proximity. (b) The expert during the baseline condition has to annotate directly on the block. Annotations appear misaligned when they are not on the same height as the surface and viewed from a different angle than they were drawn.

currently acting surgeon during the PSO, while the expert represents the acting co-surgeon, consulting the surgeon in their next steps.

We show the overlay of the virtual guide that depicts the desired outcome only to the expert - either in-situ, or on the reconstruction shown in DR. The desired form is shaped by the editor through removing a disk-like segment on top, a stretched triangle shape, and two half-cylinder shapes on the long sides (see Fig. 3). The material of kinetic sand is a compound of fine sand and non-toxic Polydimethylsiloxane, whose consistency hardens under physical pressure. Due to its properties, the block can be shaped by blunt tools without great force while maintaining a rigid shape.

A virtual PSO model represents the desired outcome for the procedure and is shown through 3D in-situ augmentations to the expert. It depicts the medical knowledge of the surgeon's expertise. The expert provides guidance to the editor on the parts to be removed in order to reach the desired PSO model. We prepared different models of similar difficulty of the described shapes, only distinguished in position and proportions of the three shape components. Additionally, a training model consisting of simple shapes is provided to gain confidence in the system during the training phase. The communication can be done verbally, through pointing gestures, and by drawing shared annotations.

4.2 Experimental Conditions

The study compares two experimental conditions and two separate roles within a joint collaborative task. Therefore, four rounds of measurements are acquired for each pair of participants. At any point of the tasks, the PSO model is only visible to the expert through 3D augmentations. In addition to the detection of the QR marker on the hybrid target that is used for extrinsic calibration, we use a second QR marker at T_{Block} in both conditions to localise the position and orientation of the block. The PSO model is directly overlaid on the block during the baseline condition. During the DR condition, the PSO model is superimposed on the 3D reconstruction, as seen in Fig. 1(b). Further, we constrain rotation of the DR box to a single degree of freedom around the up-axis for keeping horizontal surfaces of the physical world equally horizontal inside the DR to prevent disorientation.

Finally, the PSO model is rendered transparent using two different shader techniques to improve understanding of spatial relations. Both shaders are rendered on top of the reconstruction, such that the model is always fully visible. Parts of the model below the reconstruction surface appear matte, while parts above the surface are rendered additionally with a Fresnel shader. Through the transition between both slightly different shaders, users can precisely estimate which parts of the model are located beneath the surface of the reconstruction.

4.3 Procedure

We led the participants to the location of the study and introduced them to the medical use-case of PSO and the conditions surrounding the user study. Then, after acquiring their consent, we performed vision tests (Ishihara color blindness test [16], Titmus test for stereo vision, and acuity) with both participants. We excluded participants with corrected vision acuity worse than 20/25 and a stereo vision score of less than three correct answers out of 9 tests. Participants with red-green weakness would use light-blue annotations instead of green. Both participants proceeded to perform the eye-calibration procedure of the Hololens 2.

Initially, the participants familiarised themselves with the experimental conditions, the AR glasses, and the handling of the sand block during a training phase. Participants were allowed to use the system for up to 30 minutes or until they feel confident in controlling the provided tools. Once the participants felt confident, the trial started. The initial role and the starting condition was determined pseudo-randomly. We manually ensured a balanced ordering of both conditions across the whole user study.

Before each trial, we asked the participants to acquire measurements to estimate the end-to-end error of annotations of the system, as described in Sect. 4.5. After the measurement of end-to-end calibration error, participants proceed to the task. We encouraged participants

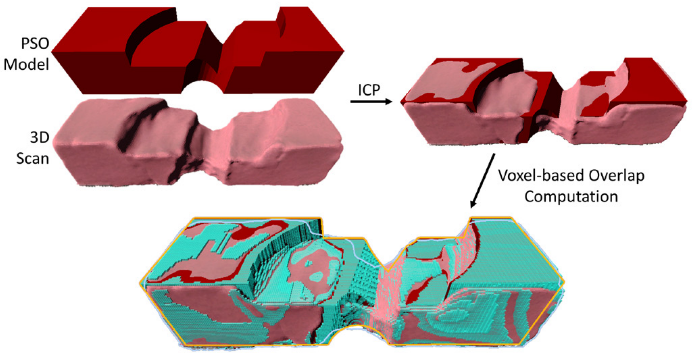


Fig. 5: **Calculating the Overlap** between desired shape and 3D scan of the processed block by the participants using our overlap function. Blue areas encode 1mm voxels that are occupied by only one of the 3D models. We measure an overlap of 82,86% for this example.

to utilize annotations, pointing gestures, and verbal communication during the whole study. After completing the task, the participants are asked to fill out the questionnaire to evaluate subjective measures and to provide free-text feedback. At the same time, we approached the block, cleaned it from loose debris, and used the 3D scanning device to acquire a 3D scan of the final shape. Following this, we replace the block with an unprocessed block, which has been pressed into the exact desired size with 3D printed negatives. The participants change roles upon returning to the subsequent trial. This procedure is repeated once more for the remaining condition. After the final questionnaire of the fourth trial, we asked participants to provide insight into their preferred condition for each role.

Counter-Measures for COVID-19 *Experimenters wore masks and medical gloves during the experiment and kept a safe distance from the participants. Participants wore masks except for the time of the task and wore medical gloves. Only pairs of participants were allowed that are of the same household or were in frequent contact beforehand. All equipment and contact surfaces were carefully disinfected before and after each pair. Rooms were sufficiently ventilated with multiple fans and positioned to avoid circular flows. Strict exclusion criteria for the study were previous visits to risk areas and any symptoms or contact with infected persons. Participants were clarified of these conditions, and all participants consented. The study was conducted in accordance with the local COVID-19 regulations with necessary precautions and in accordance with the declaration of Helsinki.*

4.4 Qualitative Measurements

A questionnaire of 26 entries measures collaboration, effort, awareness, task-load, and system usability. Ten questions are allocated to acquire perception on collaboration, effort, awareness [59]:

C1	I feel the collaboration session overall went great. We had no problems and did not struggle to complete the task.
C2	I could easily understand my partners ideas.
C3	I could easily communicate my ideas.
E1	The task was easy to complete.
E2	The task required little effort.
E3	I did not have to concentrate very hard to do the task.
A1	How often did you know where your partner was located?
A2	During the trial, how often did you know what your partner could see?
A3	During the trial, how often did you know what your partner was directly looking at?
A4	During the trial, how often did you know what your partner was doing? (Standing still, navigating/moving, moving an object, etc.)

Questions on collaboration, effort, and awareness used a 7-Likert scale, where collaboration (C1-3) and effort (E1-3) used polarisation terms of "Strongly disagree" (1) and "Strongly agree" (7), and awareness used the terms "Never" (1) and "Always" (7). Additionally, we include the NASA-TLX questions [25] to calculate the Raw-TLX score, and questions to assess the System Usability Score (SUS) [34].

4.5 Quantitative Measurements

We measure an end-to-end error between both users for annotations. Let $P_A(x)$ be the function mapping the position of a point x to the user A perceived position of x. Then, in the context of Fig. 2(a), this end-to-end error represents $\varepsilon = |P_{Expert}(T_{Pen}) - P_{Editor}(T_{Pen})|$, and therefore, includes perceptual and calibration errors of the virtual pen that participants use for pointing gestures and annotations. To acquire this measurement, we asked each participant to place a virtual disk fixated on the tip of the stylus representation into the middle of four orange circles surrounding the block. They validate the poses with a click on the stylus. In total, we acquire eight points, and therefore, four point correspondences per trial and pair. During the DR condition, the expert aligns the virtual disk with the circles inside the DR reconstruction. In this case, the calculated point corresponds to the back-projected position $T_{RoI} \bullet (T_{DR}^{-1} \bullet T_{Pen})$ that is located at the real block.

We measure the accuracy of the modeling task quantitatively by comparing the shape of the processed block on the operation table with the associated PSO model shown through AR. To compare both, we acquire the 3D scan of the block immediately after the task by using the 3D scanning device, Artec Eva [1]. The similarity of the shape is calculated on a voxel-based overlap function, as described in the following: (1) We align the 3D scan to the PSO model with manually initialized ICP. The PSO model is multiple times sub-divided to increase its vertex counts and roughly matches the number of points inside the 3D scan. (2) We create a boolean voxel grid for each 3D object that fully encapsulates them and mark every voxel inside the object as true, and false otherwise. (3) We calculate the percentage of overlap using the connection

$$A = \{ x \mid x \text{ voxel center inside PSO model} \}$$
$$B = \{ x \mid x \text{ voxel center inside 3D scan} \}$$

$$P_{overlap} = 1 - \frac{|A \triangle B|}{|A \cup B|}, \quad P_{overlap} \in [0,1]$$

where $\triangle$ denotes the symmetric difference, and $\cup$ is the union of two sets. We use a voxel grid with an edge length of 1mm per voxel. Fig. 5 provides a visualization of this principle. Additionally, we measure the average error of the ICP alignment, which simultaneously corresponds to the closest distance from the 3D scan to the PSO model on per point basis, and the Hausdorff distance to acquire the largest distance among all closest distance values per point.

We measure the time-on-task between the point in time when participants started the task and when participants signaled the end of the task. Moreover, we record the users' head movements within the calibrated common coordinate system of O_{World}.

4.6 Participants

We recruited participants from campus announcements and general university social media groups. In total, 26 participants (14♀, 12♂, Age= 25.2 ± 2.8 years) distributed into 13 pairs (3 pairs same-gendered), entered our study. 20 participants are students, of which 6 had a biomedical background, 8 from engineering, and 6 from other disciplines. All participants spend time on digital media to a high amount (48.1 ± 25.6h per week), of which participants spend 4.4h per week on video games. Further, participants disclosed they had used 7.2 ± 12.3 times AR system before, excluding 2 participants who used AR more often than they can recount. AR devices include head-mounted displays and screen-based AR systems such as smartphones.

We excluded no participants as a result of the vision tests since none were impeded by their vision regarding color blindness and stereopsis. In addition, all participants had a normal or corrected-to-normal vision regarding acuity.

5 Results

We performed Shapiro-Wilk tests on all measurements, including subjective and qualitative results, showing that the subjective results and time-on-task are not normal distributed. Therefore, we performed Mann-Whitney tests on all measurements. Further, we performed equivalence tests on non-significant findings, where applicable, to show if the observed variables can be considered equivalent. The performed equivalence tests consist of two one-sided t-tests (TOST) with 95% confidence interval and can be applied on non-parametric distributions [43, 73]. A variable is statistically equivalent if both one-sided t-tests are rejected.

5.1 Subjective Results

We evaluated subjective measurements separately between the editor and the expert and probed questions on collaboration, effort, and awareness with a 7-Likert scale, where 7 indicates positive feedback. The Raw-TLX score ranges from 0 to 100, where smaller numbers indicate positive feedback. Lastly, SUS ranges from 0 to 100, where a larger number indicates positive feedback.

Editors' View We found no significance for collaboration, effort, and awareness. The following table summarizes the results from the editors. Fig. 6(left) illustrates the measurements of the questionnaires. For non-significant results of the Mann-Whitney tests, we list the results of the TOST as p^{TOST}.

Table 1: Statistical Evaluation of the Editors

	M_B	Md_B	SD_B	M_{DR}	Md_{DR}	SD_{DR}
Coll.	5.359	6.000	1.414	5.372	5.333	1.318
Effort	4.449	4.833	1.467	4.705	5.000	1.190
Aware.	5.116	5.375	1.273	4.529	5.000	1.668
Raw-TLX	48.6	50.0	16.2	42.6	42.0	16.3
SUS	63.0	63.7	18.1	63.4	62.5	18.6

	z	r	p	p^{TOST}
Coll.	2.014	0.395	0.978	< **0.01**
Effort	0.034	0.007	0.513	**0.025**
Aware.	−0.788	0.154	0.215	0.160
Raw-TLX	−0.749	0.147	0.227	0.864
SUS	1.356	0.266	0.912	0.452

We observe statistical equivalence in collaboration and the effort put by the editor into the task. The editor was slightly more aware of the partner during the baseline condition, in which both participants work in the same space; however, not at a significant level. The different conditions did not significantly impact task-load and usability.

Experts' View We found no significance for collaboration and effort. However, we observed a significant decrease in awareness during the DR condition. The following table summarizes the results from the experts during baseline and DR conditions. Fig. 6(middle) illustrates the measurements of the questionnaires. We show the results of TOST only for non-significant results of the Mann-Whitney tests.

Table 2: Statistical Evaluation of the Experts

	M_B	Md_B	SD_B	M_{DR}	Md_{DR}	SD_{DR}
Coll.	5.282	5.668	1.309	5.295	6.000	1.374
Effort	4.385	4.000	1.206	4.808	4.668	1.208
Aware.	5.375	5.375	1.388	4.337	4.500	1.367
Raw-TLX	45.9	50.0	13.0	46.7	48.0	16.1
SUS	64.1	63.7	18.6	65.1	62.5	15.9

	z	r	p	p^{TOST}
Coll.	1.154	0.226	0.876	< **0.01**
Effort	−0.721	0.141	0.235	**0.045**
Aware.	−2.018	0.396	**0.022**	-
Raw-TLX	1.509	0.296	0.934	0.474
SUS	1.356	0.266	0.912	0.497

We observe statistical equivalence in collaboration and the effort put by the expert into the task. Furthermore, neither task load nor system usability have been impacted by the different conditions.

5.2 Qualitative Results

We report on the average end-to-end error of annotations, the time-on-task, and outcome of the task as indicated by the overlap, average per-point closest distances, and the Hausdorff distance of the 3D scan related to the PSO model. We illustrate heatmaps of the users' locations and gazes in Fig. 7. Participants used a magnification level of M=142%, Md=121%, SD=69%, compared to the original size, to feel confident working with the DR. A chart on the overlap is shown in Fig. 6(right).

We found no significance on either of the qualitative measurements. The error returned by the ICP algorithm provides the first impression on the similarity of the shapes. We observe a slight improvement toward the DR condition when comparing the overlap between both conditions; however, the difference is not significant. We additionally confirm this finding by comparing the Hausdorff distance. The following table summarizes the statistical evaluation. We abbreviate the end-to-end error to EtE Err.

Table 3: Statistical Evaluation on Qualitative Measurements

	M_B	Md_B	SD_B	M_{DR}	Md_{DR}	SD_{DR}
EtE Err.[mm]	11.18	11.11	3.69	12.05	12.31	5.11
ICP Err.[mm]	3.12	3.03	0.62	2.99	2.92	0.63
Hausdorff[mm]	14.59	14.57	2.75	14.49	14.28	3.30
Time[s]	489.25	434.33	230.80	508.91	442.70	222.84
Overlap[%]	73.14	75.06	7.50	75.62	76.96	6.78

	z	r	p	p^{TOST}
EtE Error	1.712	0.365	0.956	<**0.001**
ICP Error	−0.414	0.081	0.340	<**0.001**
Hausdorff	0.918	0.180	0.821	<**0.001**
Time	0.553	0.108	0.710	0.616
Overlap	−0.783	0.153	0.217	0.769

5.3 Free-Text Feedback

Participants ranked their preference of condition once as the editor and once for the expert. 12 out of 26 participants preferred the DR condition as the editor, while exactly half (13) of the participants preferred it as the expert. 7 participants preferred different conditions depending on the role, from which 4 liked the baseline as the editor and DR condition as the expert. Finally, 10 participants consistently liked the baseline condition in both roles, while 9 participants consistently liked the DR condition. In the following table, we summarize both conditions' positive and negative feedback, sorted by the number of occurrences mentioned by the participants. We omit contents that only two or fewer participants stated.

Table 4: Participants' Feedback

Pro: Baseline	#
Easy to explain my ideas directly on the block	9
Easy to understand the progress of the task	4
Easy to learn	3
It had less distraction	3
Contra: Baseline	
The drawing did not align with the block	4
The model covers what the surgeon is doing	3
It felt the annotations were not precise	3
Pro: Duplicated Reality	
We can draw and work on the block at the same time	5
I felt the annotations were precise	5
It felt intuitive, and it was easy to learn	5
It was easy to draw annotations in 3D	4
I can scale the model	4
It is less distracting as the editor	4
Contra: Duplicated Reality	
I felt my annotations were inconsistent with the results	4

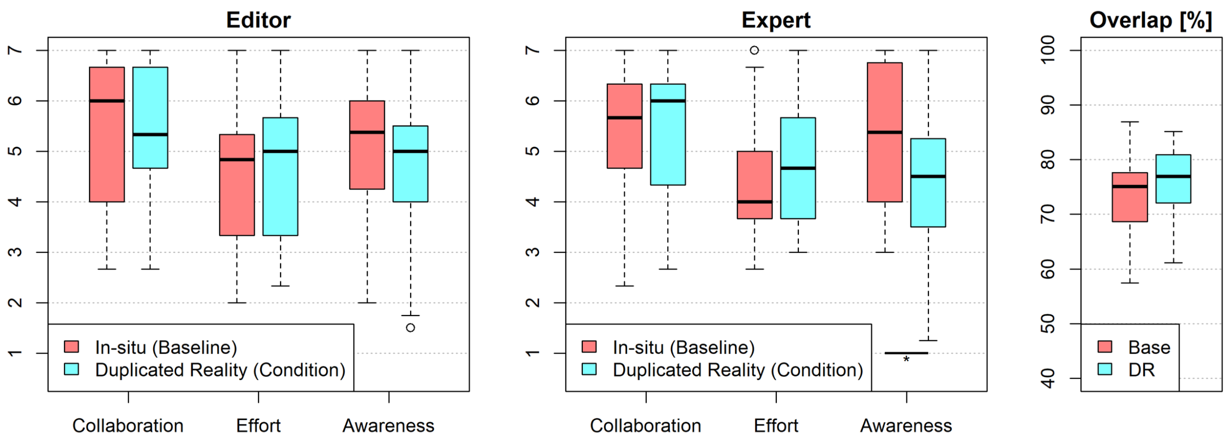


Fig. 6: **Collaboration, Effort, Awareness, and Overlap.** Black bars indicate the median. The lower and upper limit of a box represents the 25th and 75th percentile. Users perceived collaboration and effort similarly across both conditions. The DR condition significantly impacted the expert's (middle) awareness of the editor's actions (left). The difference in overlap (right) shift towards DR; however, was not significant.

6 Discussion

Our study confronted the method of DR to in-situ augmentation using a digital twin. We discuss the hypotheses as stated in Sect. 2.1 based on the acquired results from the user study.

We partially reject ***H1***, as we did not find a significant difference for collaboration and effort except for the awareness between conditions. For the first two measurements, our tests on the equivalence of the mean indicate that collaboration and effort were perceived similarly across both conditions. Thus, we suggest that users had no problem understanding the concept of DR as a proxy for working at the region of interest. The scores on the effort indicate an overall low effort in task completion. While not significant, both mean and median effort of the DR condition from the experts' view seems to shift toward a more favorable score, as seen in Fig. 6(middle).

Moreover, we observe a similar awareness score on both roles; however, only the experts show a significant difference. One reason may be that editors focused on the task rather than paying attention to the working area and actions of the expert. In contrast, we observe a significant loss of awareness from the experts' view, which was anticipated since the DR clips away any representation of the editor besides their hands. The depth sensors may have reinforced this effect as they could not fully capture thin objects such as the cutting tool or single fingers.

Further, we reject ***H2***, as neither results on task-load nor the system usability score provide a precise answer to which condition performed better in this regard. The participants' verbose feedbacks confirm the indecisiveness since their preference for one condition evenly split them into two groups of the same size. Additionally, we did not observe any correlation between prior AR experience to their specified preference. Based on the subjective results, the observations suggest that both conditions are mostly interchangeable for our task.

We reject ***H3***, as we did not find significance for the quality of the task based on computations of overlap with the ground-truth. Therefore, we were not able to confirm findings of similar prior work. The measured end-to-end error of virtual objects may have primarily caused imprecisions during the task. It incorporates a considerable average calibration and perceptual error of 1.1cm and 1.2cm between two users depending on the condition. A possible explanation involves how the Hololens 2 renders virtual objects at a focal plane at a two-meter distance. Stereopsis creates the illusion of depth; however, the accommodation of the human eye remains to focus towards the two meters. When users look at an augmented object closer than two meters, they cannot focus on the physical object and the augmentation simultaneously. This problem leads to a slight misperception of depth due to the vergence-accommodation conflict of optical see-through HMDs.

Further, annotations hovering slightly above or beneath the physical surface may appear shifted along the view direction of the annotating user when seen from an observer's perspective. As a result, users may believe that annotations are displayed inaccurately, or their device has decalibrated. Depending on the awareness of this problem, users may perceive the precision of similar annotations very differently, which would explain the conflicting feedback by the participants regarding the accuracy of annotations. We observe no significant findings in the time-on-task. Initial reasoning for an improvement in the time presupposes participants could provide consultation and edit the block simultaneously. However, we observed the contrary to be the case, independently of the condition. In most cases, the editor will wait for a full step to be communicated before committing to an action.

Next, we discuss findings from recorded user position and gaze direction. We derive from Fig. 7 that users continuously stand closely around the proximity of the task on the table. From the gaze information, we derive they were highly focused on the task and rarely looked up. Editors frequently switch between both sides of the table to take in the best angle for shaping the block. During the baseline condition, experts yield to editors and position themselves either on the small side of the table or opposite the editor. We observed rare cases where the expert will join the editor on the same side of the table whenever they attempt to draw 3D lines to communicate the next step of the task. Further, we find less body movement of the editors during the task by investigating the smaller size of the colored region on the heatmap (Fig. 7(c)) during the DR condition. The illustration coincides with our observation that editors, during DR condition, would still change the side of the operating table; however, they do not need to yield or dodge the experts while changing location. When taking the heatmap of the experts during the DR condition into account, we observe the experts would continue moving in half-circles but relocated to the location of the DR. Moreover, participants seemed to prefer walking around the DR rather than rotating the DR. We had not observed eye contact between pairs during the tasks, even though all of the invited participants knew their partner before the study started. We conclude that DR does not encourage more eye contact during tasks, which does not require eye contact under conventional circumstances. In both conditions, experts primarily use annotations while occasionally verbally emphasizing distinct characteristics of a shape in-between drawing annotations. When editors had questions during the DR condition, both looked at the physical block; however, experts returned to DR for answering and clarifying, e.g., by redrawing annotations.

Finally, we outline further use-cases which may benefit from DR. Telemedicine, for instance, may use AR for displaying remote guidance for teaching medical procedures [8, 70]. In a similar fashion for in-person training [69], DR allows guidance to a trainee without interfering with their workflow while maintaining the ability to intervene physically when necessary. Industrial assembly may benefit from DR by allowing guidance in hard-to-reach areas, on moving tracked objects such as

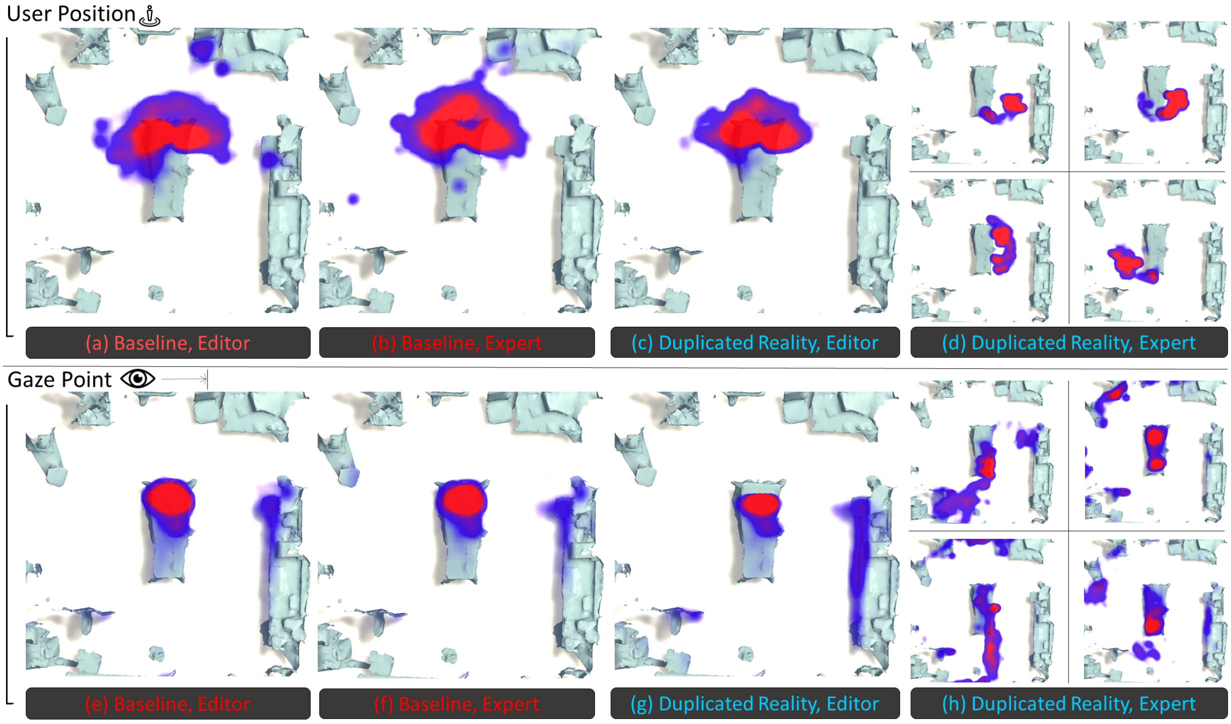


Fig. 7: **Heatmap of Users' Position and Gaze** during both conditions. Red indicates high occupy duration by the users; blue indicates low occupy duration. (a-b, e-f) Users during the baseline condition often switched places and stayed focused on the task. (c, g) Editors during DR condition swung between both sides of the table in a smaller area and stayed focused on the task. (d,h) Position of the expert chosen randomly from four participant pairs: Users swung around the DR box and focused on the DR hovering mid-air.

conveyor belts, or when involved users need to be physically present to assist with the task [58]. Designers and architects may use DR with AR-enabled developer tools [55] for simultaneously altering the model with nearby collaborators.

6.1 Limitations

The realization of DR is non-trivial due to the few technological limitations of currently available system components. For example, our deployed RGB-D cameras were often incapable of capturing the depth of thin objects, and depth values carry a varying range of errors depending on the incident angle and material of the captured surface. Reconstructions of human anatomy additionally require higher resolution in the real use-case. Algorithms for offline or delayed reconstruction collect geometric information from multiple angles and can mitigate sensor errors and increase details. Future work can modify DR to use such methods; however, it would no longer display real-time changes inside the scene. The concept of DR will benefit immediately from improvement in such depth sensor technology or real-time reconstruction methods.

Our implementation refreshed the content of the DR with one frame per second to maintain an operating frequency of around 45 fps on the Hololens 2. Nonetheless, the HMD would frequently overheat during long sessions. During the overheated state, we observed that the HMD would occasionally lose tracking of its current pose and shows increased drifting symptoms, which would induce decalibration of the extrinsics. Therefore, DR would immediately benefit from any improvements in such hardware and display technologies of the HMDs.

Lastly, as our user study focuses on a physical task that users can solve geometrically, verbal communication between participants would rarely add more information than what they can convey better by 3D annotations. Future studies should include verbal-only tasks to measure the impact of DR on verbal interactions between users.

7 Future Work

Open questions emerged from our study for future investigations. In particular, future work should address the decrease of the expert's awareness of the local task. Solutions may include (1) avatars or silhouettes of co-users to indicate their positioning around the region of interest and (2) access to a mini-map-like overview, e.g. using a WiM, to show the location of the region of interest, the DR, and the users. Moreover, 3D annotations created to annotate physical objects but are not touching the actual surface may be misinterpreted or misperceived at the wrong location when seen from different perspectives. Therefore, future implementations should provide additional depth cues or project annotations directly onto surfaces to improve the perception of guiding elements. The latter solution, however, may no longer allow users to indicate below-surface or mid-air drawings. Further research may investigate the effects of role reversal between consultants and consultees on concurrent workflows, particularly regarding 1:N or M:N relations between regions of interest and DRs. For instance, a user-instantiated DR would allow participants to switch between roles by adjusting their focus on either the task or the DR.

8 Conclusion

We proposed Duplicated Reality as an interaction concept for co-located face-to-face encounters in Augmented Reality. Our within-subjects, dual-user study investigated a collaborative task from a surgical procedure on the human spine. We found that deploying DR impacts the expert's perception of co-located users' awareness while being otherwise interchangeable with an in-situ augmentation. In turn, experts gain enhanced flexibility in choosing their position inside the environment as they are no longer bound to the operating area of the editor while maintaining the ability to create in-situ annotations and personally contribute to the task. Thus, we believe Duplicated Reality bears the same exceptional potential for the future as Augmented Reality. DR fills the gap of co-located asymmetric consultation and opens up new ways of understanding Mixed Reality in the future.

Acknowledgments

This work was funded by the German Federal Ministry of Education and Research (BMBF), Grant No.: 16SV8092, 16SV8090, 16SV8088.

References

[1] 3D Object Scanner Artec Eva. https://www.artec3d.com/portable-3d-scanners/artec-eva.

[2] C. P. Ames, J. J. Barry, S. Keshavarzi, O. Dede, M. H. Weber, and V. Deviren. Perioperative Outcomes And Complications Of Pedicle Subtraction Osteotomy In Cases With Single Versus Two Attending Surgeons. *Spine Deformity*, 1(1):51–58, 2013.

[3] K. M. Baird and W. Barfield. Evaluating The Effectiveness Of Augmented Reality Displays For A Manual Assembly Task. *Virtual reality*, 4(4):250–259, 1999.

[4] M. D. Barrera-Machuca, A. Cassinelli, and C. Sandor. Context-Based 3D Grids for Augmented Reality User Interfaces. In *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pp. 73–76, 2020.

[5] S. Beck, A. Kunert, A. Kulik, and B. Froehlich. Immersive Group-to-group Telepresence. *IEEE transactions on visualization and computer graphics*, 19(4):616–625, 2013.

[6] A. Bellarbi, S. Otmane, N. Zenati, and H. Belghit. Design and Evaluation of Zoom-based 3D Interaction Technique for Augmented Reality. In *Proceedings of the Virtual Reality International Conference-Laval Virtual 2017*, pp. 1–4, 2017.

[7] L. Besançon, A. Ynnerman, D. F. Keefe, L. Yu, and T. Isenberg. The State Of The Art Of Spatial Interfaces For 3D Visualization. In *Computer Graphics Forum*, vol. 40, pp. 293–326. Wiley Online Library, 2021.

[8] P. Bifulco, F. Narducci, R. Vertucci, P. Ambruosi, M. Cesarelli, and M. Romano. Telemedicine Supported by Augmented Reality: An Interactive Guide for Untrained People in Performing an ECG Test. *Biomedical engineering online*, 13(1):1–16, 2014.

[9] M. Billinghurst, D. Belcher, A. Gupta, and K. Kiyokawa. Communication Behaviors In Co-located Collaborative AR Interfaces. *International Journal of Human Computer Interaction (IJHCI)*, 2002.

[10] M. Billinghurst, H. Kato, K. Kiyokawa, D. Belcher, and I. Poupyrev. Experiments With Face-to-face Collaborative AR Interfaces. *Virtual Reality*, 6(3):107–121, 2002.

[11] A. Bluff and A. Johnston. Dont Panic: Recursive Interactions in a Miniature Metaworld. In *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, pp. 1–9, 2019.

[12] W. Broll, I. Lindt, J. Ohlenburg, M. Wittkämper, C. Yuan, T. Novotny, C. Mottram, A. Strothmann, et al. Arthur: A Collaborative Augmented Environment for Architectural Design and Urban Planning. *JVRB-Journal of Virtual Reality and Broadcasting*, 1(1), 2004.

[13] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard. Past, Present, And Future Of Simultaneous Localization And Mapping: Toward The Robust-perception Age. *IEEE Transactions on robotics*, 32(6):1309–1332, 2016.

[14] Y. S. Chang, B. Nuernberger, B. Luan, and T. Höllerer. Evaluating Gesture-based Augmented Reality Annotation. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 182–185. IEEE, 2017.

[15] I. Cheng, M. N. Stienen, Z. A. Medress, K. Varshneya, A. L. Ho, J. K. Ratliff, and A. Veeravagu. Single-versus Dual-attending Strategy For Spinal Deformity Surgery: 2-year Experience And Systematic Review Of The Literature. *Journal of Neurosurgery: Spine*, 33(5):560–571, 2020.

[16] J. Clark. The Ishihara Test For Color Blindness. *American Journal of Physiological Optics*, 1924.

[17] D. Coffey, N. Malbraaten, T. Le, I. Borazjani, F. Sotiropoulos, and D. F. Keefe. Slice WIM: A Multi-surface, Multi-touch Interface For Overview+ Detail Exploration Of Volume Datasets In Virtual Reality. In *Symposium on Interactive 3D Graphics and Games*, pp. 191–198, 2011.

[18] M. J. Davis, B. C. Luu, S. Raj, A. Abu-Ghname, and E. P. Buchanan. Multidisciplinary Care In Surgery: Are Team-based Interventions Cost-effective? *The Surgeon*, 19(1):49–60, 2021.

[19] H. Fuchs, G. Bishop, K. Arthur, L. McMillan, R. Bajcsy, S. Lee, H. Farid, and T. Kanade. Virtual Space Teleconferencing Using A Sea Of Cameras. In *Proc. First International Conference on Medical Robotics and Computer Assisted Surgery*, vol. 26, 1994.

[20] G. Gamelin, A. Chellali, S. Cheikh, A. Ricca, C. Dumas, and S. Otmane. Point-cloud Avatars To Improve Spatial Communication In Immersive Collaborative Virtual Environments. *Personal and Ubiquitous Computing*, pp. 1–18, 2020.

[21] L. Gao, H. Bai, G. Lee, and M. Billinghurst. An Oriented Point-cloud View For MR Remote Collaboration. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications*, pp. 1–4. 2016.

[22] R. Grasset, P. Lamb, and M. Billinghurst. Evaluation Of Mixed-space Collaboration. In *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)*, pp. 90–99. IEEE, 2005.

[23] M. C. Gupta, E. Ferrero, G. Mundis, J. S. Smith, C. I. Shaffrey, F. Schwab, H. J. Kim, O. Boachie-Adjei, V. Lafage, S. Bess, et al. Pedicle Subtraction Osteotomy In The Revision Versus primary Adult Spinal Deformity Patient: Is There A Difference In Correction And Complications? *Spine*, 40(22):E1169–E1175, 2015.

[24] N. T. Haddock, S. Kayfan, R. A. Pezeshk, and S. S. Teotia. Co-surgeons In Breast Reconstructive Microsurgery: What Do They Bring To The Table? *Microsurgery*, 38(1):14–20, 2018.

[25] 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, pp. 904–908. Sage publications Sage CA: Los Angeles, CA, 2006.

[26] M. Hirose, D. Schmalstieg, C. Wingrave, and K. Nishimura. Multi-scale Manipulation in Indoor Scenes with the World in Miniature Metaphor.

[27] Y. Huang, Y. Liu, and Y. Wang. AR-view: An Augmented Reality Device For Digital Reconstruction Of Yuangmingyuan. In *2009 IEEE International Symposium on Mixed and Augmented Reality-Arts, Media and Humanities*, pp. 3–7. IEEE, 2009.

[28] A. Irlitti, R. T. Smith, S. Von Itzstein, M. Billinghurst, and B. H. Thomas. Challenges For Asynchronous Collaboration In Augmented Reality. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pp. 31–35. IEEE, 2016.

[29] H. Jun and J. Bailenson. Temporal RVL: A Depth Stream Compression Method. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 664–665. IEEE, 2020.

[30] H. J. Kang, J.-h. Shin, and K. Ponto. A Comparative Analysis Of 3D User Interaction: How To Move Virtual Objects In Mixed Reality. In *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*, pp. 275–284. IEEE, 2020.

[31] S. Kasahara, V. Heun, A. S. Lee, and H. Ishii. Second Surface: Multi-user Spatial Collaboration System Based On Augmented Reality. In *SIGGRAPH Asia 2012 Emerging Technologies*, pp. 1–4. 2012.

[32] J. Kolkmeier, E. Harmsen, S. Giesselink, D. Reidsma, M. Theune, and D. Heylen. With A Little Help From A Holographic Friend: The Openimpress Mixed Reality Telepresence Toolkit For Remote Collaboration Systems. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11, 2018.

[33] P. Ladwig and C. Geiger. A Literature Review On Collaboration In Mixed Reality. In *International Conference on Remote Engineering and Virtual Instrumentation*, pp. 591–600. Springer, 2018.

[34] J. R. Lewis and J. Sauro. The Factor Structure Of The System Usability Scale. In *International conference on human centered design*, pp. 94–103. Springer, 2009.

[35] S. Lukosch, M. Billinghurst, L. Alem, and K. Kiyokawa. Collaboration In Augmented Reality. *Computer Supported Cooperative Work (CSCW)*, 24(6):515–525, 2015.

[36] M. D. B. Machuca, P. Asente, W. Stuerzlinger, J. Lu, and B. Kim. Multiplanes: Assisted Freehand VR Sketching. In *Proceedings of the Symposium on Spatial User Interaction*, pp. 36–47, 2018.

[37] M. D. B. Machuca, W. Stuerzlinger, and P. Asente. Smart3Dguides: Making Unconstrained Immersive 3D Drawing More Accurate. In *25th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–13, 2019.

[38] A. Maimone and H. Fuchs. A First Look At A Telepresence System With Room-sized Real-time 3d Capture And Life-sized Tracked Display Wall. *Proceedings of ICAT 2011, to appear*, pp. 4–9, 2011.

[39] 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)*, pp. 217–226. IEEE, 2020.

[40] P. Milgram and F. Kishino. A Taxonomy Of Mixed Reality Visual Displays. *IEICE Transactions on Information and Systems*, 77(12):1321–1329, 1994.

[41] A. Mulloni, H. Seichter, and D. Schmalstieg. Indoor Navigation With Mixed Reality World-in-miniature Views And Sparse Localization On Mobile Devices. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pp. 212–215, 2012.

[42] P. V. Mummaneni, S. S. Dhall, S. L. Ondra, V. P. Mummaneni, and S. Berven. Pedicle Subtraction Osteotomy. *Neurosurgery*, 63(3 Suppl):171–176, 2008.

[43] G. Norman. Likert Scales, Levels Of Measurement And The Laws Of Statistics. *Advances in health sciences education*, 15(5):625–632, 2010.

[44] 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*, pp. 405–415, 2015.

[45] O. Oda and S. Feiner. 3D Referencing Techniques For Physical Objects In Shared Augmented Reality. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 207–215. IEEE, 2012.

[46] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou, et al. Holoportation: Virtual 3D Teleportation In Real-time. In *Proceedings of the 29th annual symposium on user interface software and technology*, pp. 741–754, 2016.

[47] W. Piekarski and B. H. Thomas. Interactive Augmented Reality Techniques For Construction At A Distance Of 3D Geometry. In *Proceedings of the workshop on Virtual environments 2003*, pp. 19–28, 2003.

[48] J. S. Pierce, B. C. Stearns, and R. Pausch. Voodoo Dolls: Seamless Interaction At Multiple Scales In Virtual Environments. In *Proceedings of the 1999 symposium on Interactive 3D graphics*, pp. 141–145, 1999.

[49] T. Piumsomboon, A. Dey, B. Ens, G. Lee, and M. Billinghurst. The Effects Of Sharing Awareness Cues In Collaborative Mixed Reality. *Frontiers in Robotics and AI*, 6:5, 2019.

[50] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst. Mini-me: An Adaptive Avatar For Mixed Reality Remote Collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–13, 2018.

[51] P. Pratt, M. Ives, G. Lawton, J. Simmons, N. Radev, L. Spyropoulou, and D. Amiras. Through The Hololens Looking Glass: Augmented Reality For Extremity Reconstruction Surgery Using 3D Vascular Models With Perforating Vessels. *European radiology experimental*, 2(1):1–7, 2018.

[52] C. Qiu, S. Zhou, Z. Liu, Q. Gao, and J. Tan. Digital Assembly Technology Based On Augmented Reality And Digital Twins: A Review. *Virtual Reality & Intelligent Hardware*, 1(6):597–610, 2019.

[53] D. Ranatunga, D. Feng, M. Adcock, and B. Thomas. Towards Object Based Manipulation In Remote Guidance. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1–6. IEEE, 2013.

[54] H. Regenbrecht, K. Meng, A. Reepen, S. Beck, and T. Langlotz. Mixed Voxel Reality: Presence And Embodiment In Low Fidelity, Visually Coherent, Mixed Reality Environments. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 90–99. IEEE, 2017.

[55] P. Reipschläger and R. Dachselt. Designar: Immersive 3D-modeling Combining Augmented Reality with Interactive Displays. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*, pp. 29–41, 2019.

[56] R. Renner, B. Velichkovsky, and J. Helmert. The Perception Of Egocentric Distances In Virtual Environments. *ACM Comput. Surv.(to appear, 2014)*, 2014.

[57] D. Roth, K. Yu, F. Pankratz, G. Gorbachev, A. Keller, M. Lazarovici, D. Wilhelm, S. Weidert, N. Navab, and U. Eck. Real-time Mixed Reality Teleconsultation for Intensive Care Units in Pandemic Situations. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 693–694. IEEE, 2021.

[58] H. Salzmann, J. Jacobs, and B. Froehlich. Collaborative Interaction in Co-located Two-user Scenarios. In *Proceedings of the 15th Joint virtual reality Eurographics conference on Virtual Environments*, pp. 85–92, 2009.

[59] W. A. Schafer and D. A. Bowman. Evaluating The Effects Of Frame Of Reference On Spatial Collaboration Using Desktop Collaborative Virtual Environments. *Virtual Reality*, 7(3):164–174, 2004.

[60] 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*, pp. 265–272, 1995.

[61] P. Stotko, S. Krumpen, M. B. Hullin, M. Weinmann, and R. Klein. SlamCast: Large-scale, Real-time 3D Reconstruction And Streaming For Immersive Multi-client Live Telepresence. *IEEE transactions on visualization and computer graphics*, 25(5):2102–2112, 2019.

[62] P. Stotko, S. Krumpen, R. Klein, and M. Weinmann. Towards Scalable Sharing Of Immersive Live Telepresence Experiences Beyond Room-scale Based On Efficient Real-time 3D Reconstruction And Streaming. In *CVPR Workshop on Computer Vision for Augmented and Virtual Reality (Long Beach, CA)*, vol. 3, 2019.

[63] Z. Szalavári, D. Schmalstieg, A. Fuhrmann, and M. Gervautz. Studierstube: An Environment For Collaboration In Augmented Reality. *Virtual Reality*, 3(1):37–48, 1998.

[64] 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*, pp. 73–80, 2003.

[65] B. Thoravi Kumaravel, F. Anderson, G. Fitzmaurice, B. Hartmann, and T. Grossman. Loki: Facilitating Remote Instruction Of Physical Tasks Using Bi-directional Mixed-reality Telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 161–174, 2019.

[66] S. Truman and S. von Mammen. An Integrated Design of World-in-Miniature Navigation in Virtual Reality. In *International Conference on the Foundations of Digital Games*, pp. 1–9, 2020.

[67] R. Y. Tsai and R. K. Lenz. Real Time Versatile Robotics Hand/Eye Calibration Using 3D Machine Vision. In *Proceedings. 1988 IEEE International Conference on Robotics and Automation*, pp. 554–561. IEEE, 1988.

[68] M. Tuceryan, D. S. Greer, R. T. Whitaker, D. E. Breen, C. Crampton, E. Rose, and K. H. Ahlers. Calibration Requirements And Procedures For A Monitor-based Augmented Reality System. *IEEE Transactions on Visualization and Computer Graphics*, 1(3):255–273, 1995.

[69] A. Vidal-Balea, O. Blanco-Novoa, P. Fraga-Lamas, M. Vilar-Montesinos, and T. M. Fernández-Caramés. Creating Collaborative Augmented Reality Experiences for Industry 4.0 Training and Assistance Applications: Performance Evaluation in the Shipyard of the Future. *Applied Sciences*, 10(24):9073, 2020.

[70] S. Wang, M. Parsons, J. Stone-McLean, P. Rogers, S. Boyd, K. Hoover, O. Meruvia-Pastor, M. Gong, and A. Smith. Augmented Reality as a Telemedicine Platform for Remote Procedural Training. *Sensors*, 17(10):2294, 2017.

[71] T.-Y. Wang, Y. Sato, M. Otsuki, H. Kuzuoka, and Y. Suzuki. Effect of Full Body Avatar in Augmented Reality Remote Collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1221–1222. IEEE, 2019.

[72] N. Weibel, D. Gasques, J. Johnson, T. Sharkey, Z. R. Xu, X. Zhang, E. Zavala, M. Yip, and K. Davis. Artemis: Mixed-reality Environment For Immersive Surgical Telementoring. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–4, 2020.

[73] S. Wellek. *Testing Statistical Hypotheses Of Equivalence*. Chapman and Hall/CRC, 2002.

[74] A. D. Wilson. Fast Lossless Depth Image Compression. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, pp. 100–105, 2017.

[75] C. A. Wingrave, Y. Haciahmetoglu, and D. A. Bowman. Overcoming World In Miniature Limitations By A Scaled And Scrolling WIM. In *3D User Interfaces (3DUI'06)*, pp. 11–16. IEEE, 2006.

[76] M.-D. Yang, C.-F. Chao, K.-S. Huang, L.-Y. Lu, and Y.-P. Chen. Image-based 3D Scene Reconstruction And Exploration In Augmented Reality. *Automation in Construction*, 33:48–60, 2013.

[77] J. Young, T. Langlotz, S. Mills, and H. Regenbrecht. Mobileportation: Nomadic Telepresence For Mobile Devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(2):1–16, 2020.

[78] K. Yu, A. Winkler, F. Pankratz, M. Lazarovici, D. Wilhelm, U. Eck, D. Roth, and N. Navab. Magnoramas: Magnifying Dioramas For Precise Annotations In Asymmetric 3D Teleconsultation. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 392–401. IEEE, 2021.

Part V

Projective Bisector Mirror

5 Introduction to PBMs

An immersive teleconsultation system requires multiple cameras to maximize coverage of the environment. Fuchs et al. [5] pioneered the approach of what they called a "sea of cameras" to describe this approach of deploying multiple cameras for coverage. The depth image generated by RGB-D cameras is often less complete than the color image, as it can contain holes in pixels where the camera could not obtain a depth value for various reasons. In previous studies, one recurring criticism is the geometric fidelity of the real-time point cloud data, as it can be noisy and incomplete, hindering a clear understanding of the environment. Several factors contribute to this problem, including the camera's resolution and depth estimation accuracy, which manufacturers may improve through modernized computation power, bandwidth, and imaging sensors. However, depth imaging sensors will presumably inevitably encounter reflection and absorption characteristics of surfaces or shadows from occlusion, which cannot be entirely resolved with current technology. Adding more cameras may fill missing gaps; however, it would increase the amount of data that needs to be processed, cleaned up, and transmitted. The acceptable latency for real-time communication is somewhere between 100ms and 500ms [230, 231], where participants felt that a latency of 100ms was an instantaneous response. This hard requirement filters out the number of possible point cloud processing methods. For example, algorithms for cleaning point cloud data exist [232, 233, 234], but they are still far from performing below the given requirement of a real-time system of a few milliseconds per frame. Therefore, considering being surrounded by a sea of cameras and potentially other users, how can we effectively utilize the high-quality color images from any of those cameras or other users' viewpoints to enhance our perception of the environment? The following work presents the Projective Bisector Mirror, enabling users to see any camera image as a virtual mirror view. We achieve the mathematically accurate representation of a mirror using a 2D image by projecting the image onto the bisector plane between the capturing camera and the viewer's viewpoint. A 3D point and a normal vector define a plane. For the PBM, these correspond to the midpoint and normalized difference vector between the capturer and viewer. We introduce the method of PBMs and discuss its conceptual and theoretical properties while presenting use cases beyond medical teleconsultation.

Author's Contribution to the Publication

The author of the dissertation (Kevin Yu) implemented all use-cases, wrote the paper, and presented the work at the International Symposium for Mixed and Augmented Reality (ISMAR) 2022 as part of the journal track and later to be published as a special issue in the Transactions on Visualization and Computer Graphics (TVCG). Konstantinos Zacharis, M.Sc. aided in correcting and notating mathematical formulas. Dr. Ulrich Eck and Prof. Dr. Nassir Navab are the technical supervisors of the author. The following article is the accepted version acquired through the online library "IEEE Xplore". The right of reuse is included in the appendix (section VIII).

Projective Bisector Mirror (PBM): Concept and Rationale

Kevin Yu, Kostantinos Zacharis, Ulrich Eck, and Nassir Navab

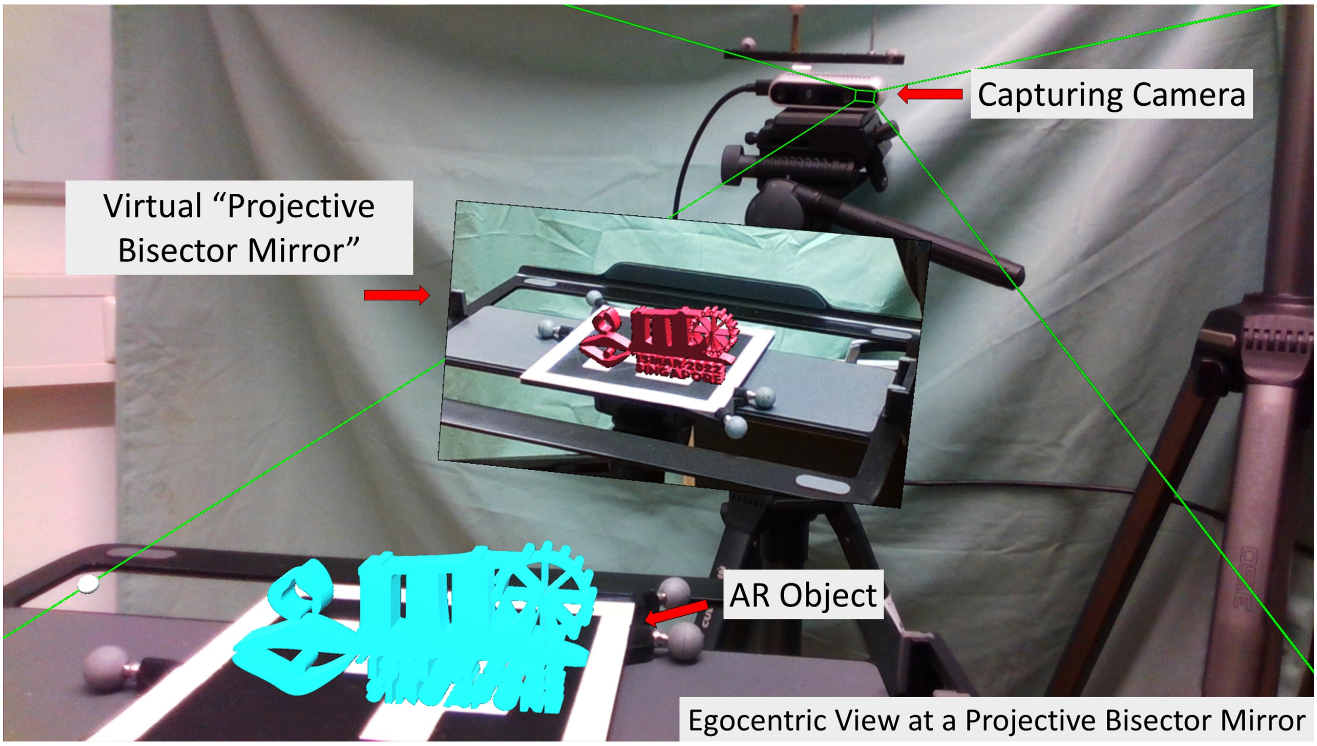


Fig. 1. **The novel concept of a Projective Bisector Mirror** allows users to present views of any secondary calibrated camera as natural mirrors, which reflect both real and virtual environments in real-time without the knowledge of the 3D structure of the scene. Correct placement and projective transformation ensure compliance with the laws of reflection such that users can see the physical and perspective-correct mirror plane from their egocentric view. Furthermore, the Projective Bisector Mirror view does not occupy physical space, while a cropping metaphor and transparency settings can dynamically adjust its visualization.

Abstract— Our world is full of cameras, whether they are installed in the environment or integrated into mobile devices such as mobile phones or head-mounted displays. Displaying external camera views in our egocentric view with a picture-in-picture approach allows us to understand their view; however, it would not allow us to correlate their viewpoint with our perceived reality. We introduce Projective Bisector Mirrors for visualizing a camera view comprehensibly in the egocentric view of an observer with the metaphor of a virtual mirror. Our concept projects the image of a capturing camera onto the bisecting plane between the capture and the observer camera. We present extensive mathematical descriptions of this novel paradigm for multi-view visualization, discuss the effects of tracking errors and provide concrete implementation for multiple exemplary use-cases.

Index Terms—Mirror Geometry, Multi-camera system, Augmented Reality

1 Introduction

How can we understand what another person or another camera is observing? Assuming they can share their view with us by providing a camera stream from their perspective, displaying it on a separate display or using Augmented Reality (AR) to present it as a floating view attached to the second user/camera allows us to try to understand what they observe; however, such picture-in-picture approaches possess no direct geometric spatial relationship, in terms of position but also orientation of the secondary observer, to our own egocentric view. Mirrors are essential objects in our daily life, and we intuitively understand the physical laws of reflection. We can interpret images perceived through the mirror even unconsciously and extrapolate the pose of an object inside the reflected image within a fraction of a second. Therefore, we introduce a novel method to display external camera views in our egocentric view with the metaphor of mirrors without any 3D knowledge of the observed scene.

Mirrors are associated with reflective surfaces, where the angle of incident light to the normal of the mirror surface is equal to the angle of outgoing light. In computer graphics, it is customary to model a mirror with the concept of virtual cameras and virtual images rather than using ray-casting techniques to simulate physical laws. For instance, one can acquire the content of the mirror image by deploying a virtual camera at the reflected position of the viewer to produce the image as seen

- *Kevin Yu is with Computer Aided Medical Procedures, Technical University Munich. E-Mail: kevin.yu@tum.de*
- *Kostantinos Zacharis is with Research Group MITI, Hospital Rechts der Isar, Technical University Munich.*
- *Ulrich Eck and Nassir Navab are with Computer Aided Medical Procedures, Technical University Munich.*

Manuscript received 11 March 2022; revised 11 June 2022; accepted 2 July 2022. Date of publication 01 September 2022; date of current version 03 October 2022. Digital Object Identifier no. 10.1109/TVCG.2022.3203108

on the mirror plane. We observe that cameras could be considered as behaving like a mirror as long as their views are *projected*, *transformed* and *presented on a particular virtual surface* such that the mirror geometry and the laws of reflection could be respected. In this paper, we demonstrate that for creating the metaphor of mirrors with cameras, we only require to know the relative pose between the capturing camera and the pose of the observer's egocentric view. We also show that based on the position, orientation and field of view of the egocentric view, this is only possible when the secondary camera position, orientation and field of view satisfy given geometric conditions.

We observe that these requirements are increasingly easier to meet as new digital systems and solutions integrate self-locating camera systems to gain awareness on environment and events in the real world. For example, Automotives integrate arrays of cameras, allowing them to perceive the surrounding traffic for assisted or self-driving [2]. Seas of cameras that capture environments in real-time enable 3D reconstruction of 3D scenes for example for immersive telepresence [43]. Consumer-friendly hand-held devices with integrated cameras are available at any given time and location, and self-locating camera systems utilizing Simultaneous Localization and Mapping [14] (SLAM) provide dynamic camera pose in addition to the acquired imaging data.

We define a physically plausible dynamic mirror plane to project external camera views through simple planar transformations allowing for their correct perception as mirror images from the egocentric view of the user/camera. This is only possible if the camera image is projected onto the bisecting plane defined by the optical centers of the external and egocentric views. This unique configuration allows us to create Projective Bisector Mirrors (PBMs) that reflect both the real and digital content observed by external cameras while respecting the exact geometric constraints of mirror reflection. In this paper, we discuss this novel concept and the rationale of such PBM visualizations for exemplary use-cases and put forward underlying mathematical constructs and technical implementations.

2 Related Work

Secondary views have been traditionally visualized in form of picture-in-pictures. User interfaces often offer users to click on 2D or 3D camera icons to move to and visualize their views. When the world is built in virtual or a 3D reconstruction of the environment is available virtual mirrors have also been introduced to visualize alternative views more intuitively.

Miyamoto et al. [36][1] developed a floating virtual mirror to allow drivers to intermediate same-direction traffic in close proximity using a surveillance camera installed at a traffic intersection. However, they restrict the method to the use case. Their simplified homography transformation using a single pair of image correspondence requires the driver to be within the camera's field of view and both views to be not rotated along their optical axis. To our knowledge, there has been no other prior work offering the possibility of viewing secondary 2D camera views of unknown 3D scenes as mirror views scene from an egocentric view.

Here, we discuss prior work on mirror viewing in AR and VR, which are not directly but remotely related to the presented novel concept of Projective Bisector Mirrors (PBM). Prior work depicts that the concept of mirrors in Mixed Reality has had a large variety of design and utilization. Here, we summarize some of existing mirror related literature and discuss all characteristics distinguishing them from the proposed concept of PBMs. Virtual Mirrors in related work are a concept for reflecting virtual environments and objects [5, 30, 32, 39]. A virtual camera behind the mirror plane captures the virtual scene from its perspective and renders the result onto the mirror object. Computer graphics generate the mirror plane, reflecting the virtual content with additional digitally available information. The proposed use-case of the related work resides in the medical domain to gain a superior spatial understanding of human anatomy in surgery. However, this method is limited to only virtual environments since this method does not allow for the acquisition of the real-world scene from the viewpoint of the virtual camera for generating the mirror content. A common approach to constructing an Augmented Reality Mirror combines a physical half-silvered mirror with a monitor [6, 37, 45]. While the half-silvered mirror reflects the real-world scene, the monitor additionally displays digital augmentations. Systems with fully opaque mirrors allow similar effects; however, they require a more complex setup [31, 38] and careful control of ambient brightness for the optimal result. Additionally to constructing the mirror system, when requiring augmentation of, e.g., the person standing or interacting with the mirror, the system may require additional sensors to capture the user's pose for adjusting visualizations perspective-correct for their egocentric views. A different approach in creating Augmented Reality Mirrors is to pair a monitor with a camera [15, 26, 29, 41] with no involvement of a physical mirror. Instead, those systems display the camera image on the monitor to create the illusion of a mirror. While these systems are comparably straightforward in their construction, many systems suffer from an incorrect perspective as they do not consider the user's viewpoint. Related work explored such systems for medical anatomy education [3, 8–10]. Furthermore, related work has attempted to solve the problem of perspective correction with robots by dynamically adjusting the viewpoint of the capturing camera [56] or use camera arrays to create a real-time 3D reconstructed to invert and display within the mirror [20]. Finally, augmented physical mirrors [33, 58] allow fully opaque and physical mirrors to reflect AR content, as long users wear an AR-enabled head-mounted display (HMD) or hand-held device and track the mirror in its position and rotation. The deployment of cameras to gain additional viewpoints enhances the perception and spatial understanding of AR users. The method of Augmented Viewports [22] demonstrates this by overlaying a real-time and enlarged view of a region of interest in the user's view through AR. Further work integrates real-time captured camera stream into a user's view to see beyond buildings [25, 27] or to create the effect of Diminished Reality [59]. Reflective-AR displays [16] utilize single-shot RGB captures, visualize them as floating picture-in-picture displays and augment them with a 2D projection of a virtual robot positioned at its target destination for aligning the real robot to such virtual counterparts. However, in this approach the reflectors *do not* satisfy the physical constrains of a mirror, but they are simple projections of the virtual 3D onto floating representation of 2D camera views.

PBM provides a perspectively correct mirror visualization of an external camera within the user egocentric view, seen previously only within systems that incorporate real physical mirrors. The concept of PBM does not restrict to specific display types for the observer and can be presented within most available displays, including but not restricted to hand-held, head-mounted displays, and half-transparent displays.

3 Method

A PBM view displays external camera views with the metaphor of a real mirror and satisfies the required physical constrains. The naming of this method originates from the underlying principle of planar projection of the external camera image onto the bisecting plane (or bisector). PBM computes the bisecting plane between the optical centers of the capturing camera and the egocentric view of the observer. Please note, that for every point on this plane the 3D Euclidean distance to both centers are equal, satisfying therefore the physical constrains of actual mirrors. The mirror plane is a sub-plane of the bisector defined by four corners, the 3D intersection of the camera frustum with the bisecting plane (see Fig. 2). The plane only exists virtually. A PBM fills the gap of displaying external camera views in the egocentric view and further can replace a conventional mirror. The main difference is that for replacing conventional mirrors, users need to carefully position the capturing camera such that the PBM computes the resulting bisecting plane at the desired location.

Systems that integrate PBMs require at least three components: (1) A capturing camera for acquiring the mirror's content, (2) a tracking system or algorithm to estimate the relative position of all involved cameras, and (3) an egocentric display system, e.g. an optical see-through display or an optical camera view. The PBMs can be generated

[1]We thank the anonymous reviewer for bringing this related work to our attention.

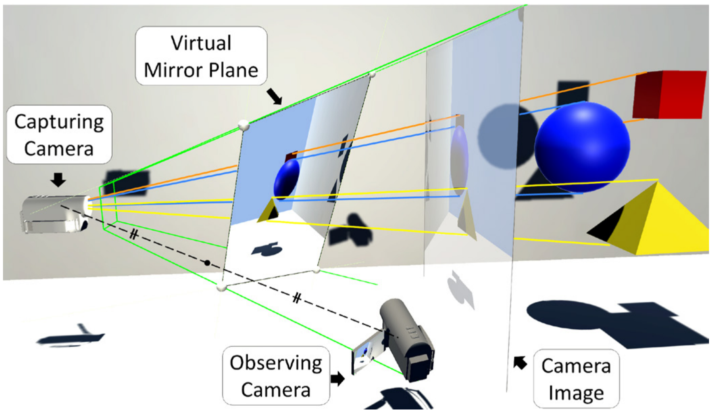


Fig. 2. **3D Illustration of a PBM Plane** in the camera projection frustum. A PBM view generates the mirror image that corresponds to the metaphor of cutting the project frustum of a pinhole camera viewing the 3D scene. We project the camera image onto the bisector plane, maintaining perspective correctness from the view of an observer.

as far as all cameras or optical see-through displays are tracked. A system incorporating a camera and a display can fill the roles of the capturer and the observer. Moreover, users may dynamically reverse their roles between the capturer and the observer depending on the situation during a multi-user scenario. In the following sections, we discuss PBMs in the hindsight of the physical properties of physical mirrors and provide a detailed insight into its concept.

3.1 Projective Bisector Mirror Geometry

A method for rendering mirrors in computer graphics includes the deployment of a virtual second camera additionally to the primary camera for rendering the egocentric view of the observer, and flip the pose along the mirror plane (see Fig. 3 (left)). This secondary camera re-renders the scene from the new perspective and projects the rendered image onto the mirror surface. The concept of the PBM utilizes this approach to position the view of an actual camera such that it appears to be mirror. In particular, the mirror plane is co-planar to the bisecting plane that is constructed from a normal vector and a point on the plane. The normal vector corresponds to the difference vector between both optical camera centers, while the point on plane is equal to the 3D center between both centers (see Fig. 3 (right)). If the observer is unaware of a mirror, they may perceive reflected objects behind the mirror plane rather than as a reflection (e.g., Pepper's Ghost effect in theaters [19]) in both real and virtual environments.

Positioning and warping the image stream of external camera views is crucial for the concept of PBMs to fit physical descriptions of a mirror. We identify two methods for creating the necessary visualization, that is, 2D homography in screen space and 3D planar projection. Both methods require the computation of the 3D corners of the mirror plane and divert in the final steps. In particular, PBM requires the four corners of the transformed image to correspond to the four 3D points from the 3D plane-ray intersection between the bisecting plane and the four rays defined by the camera frustum. The convex shape from four corners corresponds to a convex quadrilateral with no systematic parallel lines or regularities. The approach with homography displays the mirror by overlaying the transformed image as a 2D overlay on the egocentric view of the observer. Therefore, we recommended generating a PBM view with homography for embedded systems with no 3D rendering engine or systems with no access to a GPU. An approach with planar projection generate a textured 3D quad mesh. Linear texture mapping allows affine transformation of a rectangle without artifacts; however, it will show a distorted seam on quadrilateral shapes [21]. Therefore, implementing a 3D quad requires quadrilateral texture interpolation during rendering [23]. Both approaches require PBM to update the 3D quad or the homography matrix when the positions of the capturer or observer change.

3.2 Algorithm

For the remainder of this paper, we define a rigid transformation in homogeneous matrix form T consisting of the rotation matrix R and t, while T_B^A is the transformation from coordinate system A to coordinate system B. Any variable given with p denotes a 2D or a 3D point and $\vec{v}$ a 3D vector. We define the pose of the capturing camera as T_W^{cap} and the pose of the observer as T_W^{obs} (see Fig. 3 (right)) in the world coordinate system W.

We disclose the algorithm to render the dynamic PBM plane from the observer's egocentric view. We compute 3D rays describing the camera frustum of the capturing camera under the requirement of an undistorted camera image and knowledge on the 3D pose of its optical center.

1. Calculate bisecting point p_{bisect} and mirror normal $\overrightarrow{v_N}$.
2. Compute 3D plane-ray intersection points for all rays defined by the camera frustum at T_W^{cap} with the bisecting plane.
3. Project 3D points into 2D points in normalized viewport coordinates of the camera at T_W^{obs}.
4. Compute the homography matrix or the 3D quad with planar projection.
5. Render the transformed capture image within the egocentric view of the observer

3.2.1 Computation of the Mirror Projection Points

We put forward the computation of the four 3D corner points of the PBM plane as a result of 3D plane-ray intersection. The capturing camera is defined by its pose in world coordinate $T_W^{cap} = [R_W^{cap} | t_W^{cap}]$, pixel resolution (w, h), and focal-length f. We set the origin of the ray equal to the optical center of the capturing camera at p_{cap} and the rays' directional vector with $\overrightarrow{v_{corner}}$. In case of the top left corner of the camera frustum, we compute $\overrightarrow{v_{corner}}$ with the pixel resolution and focal length

$$\overrightarrow{v_{left,top}} = R_W^{cap} \cdot \left[-\frac{w}{2f}, \frac{h}{2f}, 1\right]^T \tag{1}$$

and the remaining rays similarly by adjusting the sign of the first two elements inside the right vector. Let p_{bisect} be the 3D bisecting point between p_{cap} and p_{obs}, with $p_{bisect} = (p_{cap} + p_{obs})/2$ and $\overrightarrow{v_N} = p_{obs} - p_{cap}$ the normal vector, then the distance at which each frustum ray hits the bisecting plane is given with

$$d_{hit,corner} = \frac{\overrightarrow{v_N} \bullet (p_{bisect} - p_{cap})}{\overrightarrow{v_N} \bullet \overrightarrow{v_{corner}}} \tag{2}$$

, where $\bullet$ denotes the dot-product. Finally, we acquire the 3D plane-ray intersection points with

$$p_{hit,corner} = p_{cap} + d_{hit,corner}\overrightarrow{v_{corner}} \tag{3}$$

For an implementation approach with homography, we additionally project the acquired intersection points into 2D viewport coordinates of the observer camera at T_W^{obs} with the view matrix $view_{obs}$ and projection matrix $proj_{obs}$:

$$p_{obs,corner} = view_{obs} proj_{obs} \cdot \begin{bmatrix} p_{hit,corner} \\ 1 \end{bmatrix} \tag{4}$$

$$p_{obs,viewport} = \begin{bmatrix} \frac{p_{obs,corner,x}}{-2p_{obs,corner,w}} + 0.5 \\ \frac{p_{obs,corner,y}}{-2p_{obs,corner,w}} + 0.5 \end{bmatrix} \tag{5}$$

Values of the 2D point $p_{obs,viewport}$ with $x < 0$ and $x > 1$ refer to pixels outside the image. When using homography, corners outside the image causes the reflection image to be partially outside the egocentric view of the observer. Next, our method uses the four resulting 2D points for each corner to compute the projection of the camera stream with the homography matrix.

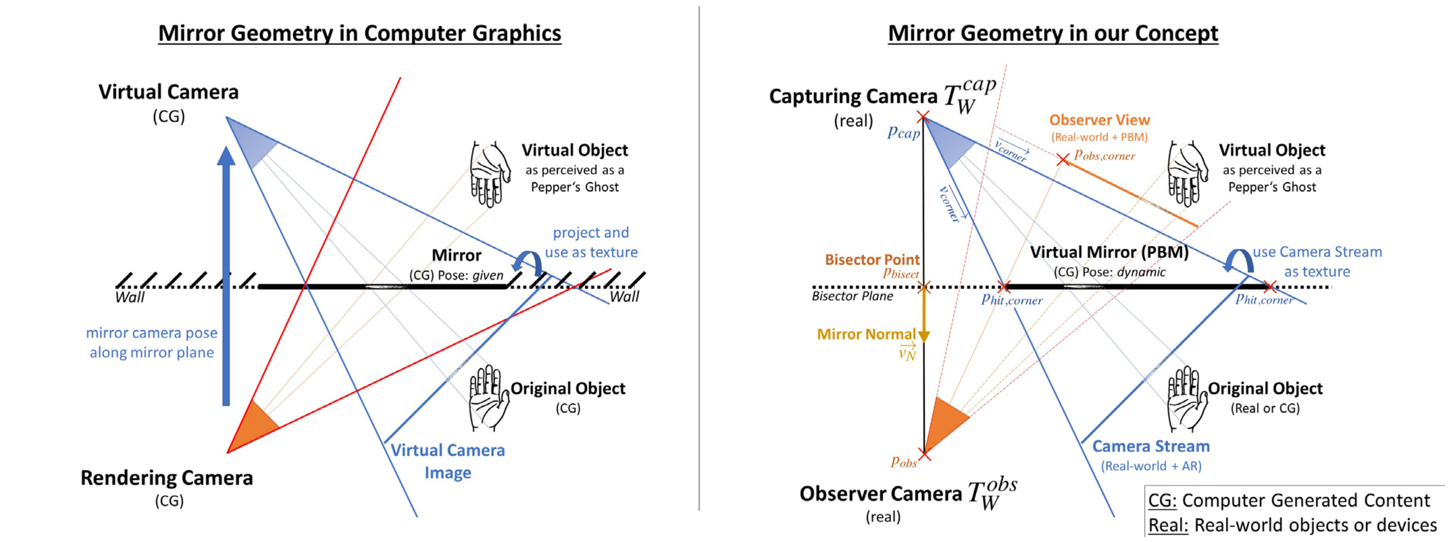


Fig. 3. **Mirror Geometries in Comparison.** Left: Possible mirror implementation for a fully computer generated environment. Right: Our concept of a PBM. Geometrically seen, our concept is equal to conventional mirror implementation; however, our concept show the major differences in the components: We substitute the virtual mirror camera with an actual camera and dynamically adjust the mirror plane. Our concept allows creation of the mirror image combined from real-world capture and the augmented world.

3.2.2 Computation of the Homography Matrix

We transform the original camera stream into the mirror plane and project the corner to create a perception-correct mirror from the view of the user camera. A homography matrix H is conventionally constructed by a 3-by-3 matrix with eight Degree of Freedom (DoF) for a 2D-2D mapping and requires a Singular-Value-Decomposition (SVD) to solve for the unknowns of the form

$$s \cdot \begin{bmatrix} h_{1,1} & h_{1,2} & h_{1,3} \\ h_{2,1} & h_{2,2} & h_{2,3} \\ h_{3,1} & h_{3,2} & h_{3,3} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \tag{6}$$

where s depicts an unknown scale. A point $p = (u,v)$ is projected into point $p'_{(u,v)} = (x'_{(u,v)}, y'_{(u,v)})$ once multiplied with the homography matrix. We assume that the corners of the color image are in camera viewport coordinates $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$, and we calculate the matrix H without SVD. We assume a normalized homography matrix where $h_{3,3} = 1$ and solve the equation for every variable $h_{m,n}$ in Equation 6 with

$$s \cdot (h_{1,1}x + h_{1,2}y + h_{1,3}) = x' \tag{7}$$
$$s \cdot (h_{2,1}x + h_{2,2}y + h_{2,3}) = y' \tag{8}$$
$$s \cdot (h_{3,1}x + h_{3,2}y + 1) = 1 \tag{9}$$

Then solve Equation 9 for s, substitute into Equation 7 and 8 and receive

$$x' = h_{1,1}x + h_{1,2}y + h_{1,3} - h_{3,1}xx' - h_{3,2}yx' \tag{10}$$
$$y' = h_{2,1}x + h_{2,2}y + h_{2,3} - h_{3,1}xy' - h_{3,2}yy' \tag{11}$$

We further simplify the equations by substituting values for x and y with the corners points in viewport coordinates into Equation 10 and 11.

$$x'_{(0,0)} = h_{1,3}, \quad y'_{(0,0)} = h_{2,3} \tag{12}$$
$$x'_{(0,1)} = h_{1,2} + h_{1,3} - h_{3,2}x'_{(0,1)} \tag{13}$$
$$y'_{(0,1)} = h_{2,2} + h_{2,3} - h_{3,2}y'_{(0,1)} \tag{14}$$
$$x'_{(1,0)} = h_{1,1} + h_{1,3} - h_{3,1}x'_{(1,0)} \tag{15}$$
$$y'_{(1,0)} = h_{2,1} + h_{2,3} - h_{3,1}y'_{(1,0)} \tag{16}$$
$$x'_{(1,1)} = h_{1,1} + h_{1,2} + h_{1,3} - h_{3,1}x'_{(1,1)} - h_{3,2}x'_{(1,1)} \tag{17}$$
$$y'_{(1,1)} = h_{2,1} + h_{2,2} + h_{2,3} - h_{3,1}y'_{(1,1)} - h_{3,2}y'_{(1,1)} \tag{18}$$

Finally, we substitute the equations from 12 downward to receive

$$h_{1,1} = x'_{(1,0)} - x'_{(0,0)} + h_{3,1}x'_{(1,0)} \tag{19}$$
$$h_{1,2} = x'_{(0,1)} - x'_{(0,0)} + h_{3,2}x'_{(0,1)} \tag{20}$$
$$h_{1,3} = x'_{(0,0)} \tag{21}$$
$$h_{2,1} = y'_{(1,0)} - y'_{(0,0)} + h_{3,1}y'_{(1,0)} \tag{22}$$
$$h_{2,2} = y'_{(0,1)} - y'_{(0,0)} + h_{3,2}y'_{(0,1)} \tag{23}$$
$$h_{2,3} = y'_{(0,0)} \tag{24}$$
$$h_{3,1} = \frac{a\Delta y'_{(0,1)} - b\Delta x'_{(0,1)}}{\Delta x'_{(1,0)}\Delta y'_{(0,1)} - \Delta x'_{(0,1)}\Delta y'_{(1,0)}} \tag{25}$$
$$h_{3,2} = \frac{a\Delta y'_{(1,0)} - b\Delta x'_{(1,0)}}{\Delta x'_{(0,1)}\Delta y'_{(1,0)} - \Delta x'_{(1,0)}\Delta y'_{(0,1)}} \tag{26}$$
$$h_{3,3} = 1 \tag{27}$$

where $a = x'_{(0,0)} - x'_{(0,1)} - x'_{(1,0)} + x'_{(1,1)}$, and $b = y'_{(0,0)} - y'_{(0,1)} - y'_{(1,0)} + y'_{(1,1)}$, and $\Delta x'_{(u,v)} = x'_{(u,v)} - x'_{(1,1)}$, and $\Delta y'_{(u,v)} = y'_{(u,v)} - y'_{(1,1)}$.

3.3 Quadrilateral Texture Mapping

3D planar projection for the generation of the PBM view allows for a quasi-equal rendering approach compared to the homography-based approach. The 3D quad is the result of triangulation of the four points at $p_{hit,corner}$ (Equation 3). The resulting mesh resides inside the common coordinate system W, same as all involved cameras.

Linear texture mapping allows correct visualization of an affine transformed rectangles; however, it does not apply for quadrilateral shape as in the case of PBMs. The problem is that both co-planar triangles within the quadrilateral remain congruent with the UV space, which results in an unnatural seam at the shared edge of both triangles. Therefore, the system must correct texture mapping with quadrilateral interpolation by associating the depth with the calculation of the UV coordinates [7]. With PBMs, the system must multiply texture UV coordinates with the distance of each corner to the geometrical 3D center. Then, during the rasterization step of the rendering pipeline, the system divides the interpolated distance again from the UV coordinate to receive the corrected UV coordinate of the color pixel.

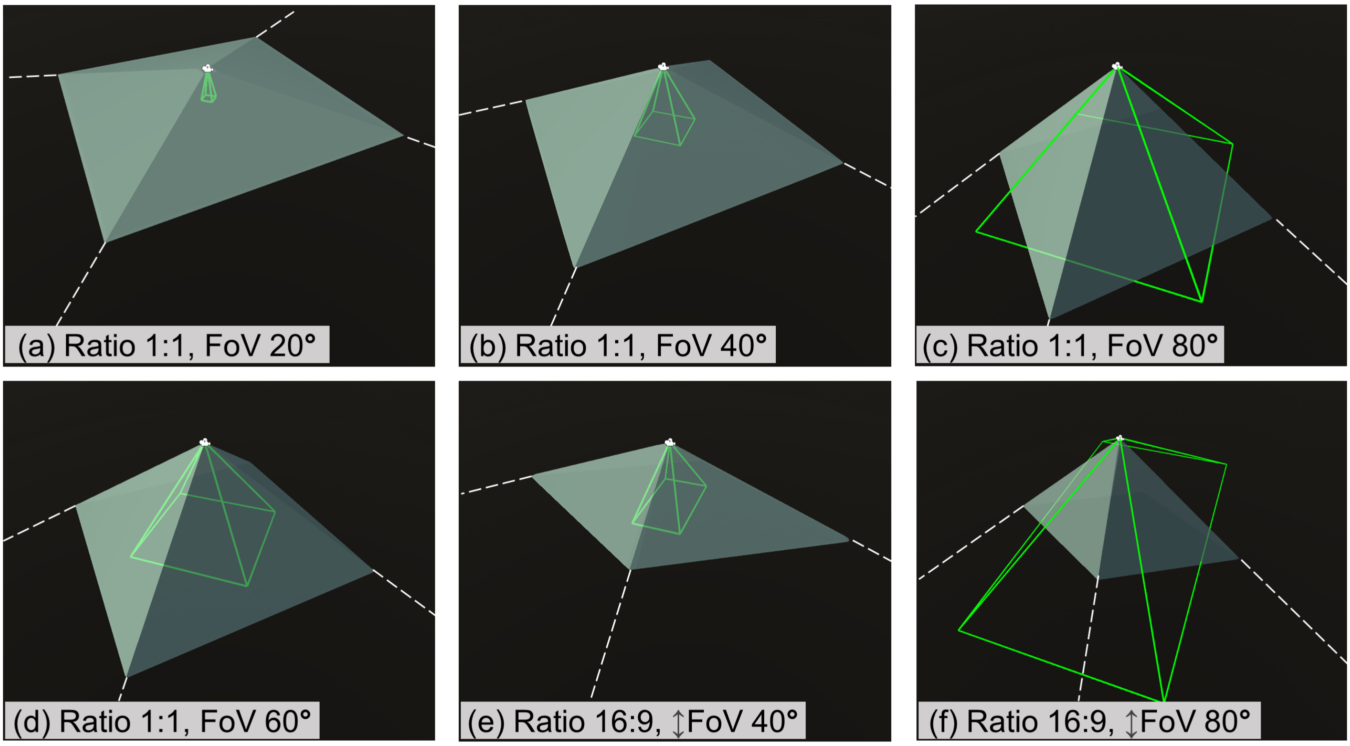


Fig. 4. **Visualized 3D Area for Valid Positions of the Observing Camera** Green lines depict the camera frustum of the capturing camera. Blue volumes depict the area of valid mirror geometry. The area extends linearly into infinity (white dashes). The frustums and valid area are cut-off co-planar for size comparison. We can create a Projective Bisector Mirror as long as the optical center of the observer camera is located inside the valid area. (a-d): The valid area shrinks with an increasing field of view of the capturing camera. (e-f): Camera input with non-uniform pixel ratio result in non-uniform valid areas in the shape of a rhombic pyramid.

3.4 Cropping

PBM do not minify the captured image similar to physical planar mirrors. Therefore, capturing cameras with wide field of view (FoV) generate similarly large PBM views. We introduce cropping as a tool to reduce the size of the PBM view without breaking the physical constrains of a mirror. PBMs allow two types of cropping. The first type refers to cropping the input image of the capturing camera and simultaneously reducing the FoV accordingly. Users can reduce the resolution by cropping uniformly from all side and reduce the FoV value for the PBM computation and increase the area of valid positions for the observer. The second type of cropping refers to reducing the size of the PBM view in the means of masking pixels with transparency. Users can see more of the content from their own camera perspective. The advantage here is that the cropping mask can take in arbitrary 2D shapes. We propose the second type of cropping as a processing step prior to the main algorithm, as described in Sect. 3.2. The cropping center can be selected by the user or automatically depending on the use case. To perform cropping in the screen space of the capturing camera, the system may use a dynamically cropping center selected from the observer or automatically by the system:

1. Select a cropping center in the viewport coordinates of the observer camera at T_W^{obs}
2. Re-project the 2D center into a 3D point where its depth intersects the bisecting plane
3. Project the 3D point into the 2D viewport coordinates of the capturing camera at T_W^{cap}
4. Mark all pixels outside the cropping area surrounding the 2D viewport point as transparent
5. Use the resulting image instead of the originally captured image for the remainder of the algorithm.

The center of cropping can be defined differently for the specific use case. For example, a hand-held device may utilize touch-input to define the center and to adjust size of the cropping area, while HMDs may use eye-tracking or a pre-defined location in space.

3.5 Occlusion Handling

Human perception utilizes several depth cues to estimate the distance of objects. When combining virtual and real-world scenes, occlusion is one of the major depth cues users immediately notice when not correctly handled. Breen et al. [11] present two methods of occlusion handling: A model-based and a reconstruction-based occlusion handling. Both types of occlusion handling work well for PBMs views. With model-based occlusion handling, the system needs to be aware of the poses of relevant real-world objects and render an invisible yet occluding object of the same geometrical shape at the given location. The reconstruction-based solution utilizes geometrical information from depth sensors from the viewpoint of the observer camera to estimate which parts of digital content the system needs to occlude inside the egocentric view.

Further, we consider occlusion handling for homography and 3D quad-based rendering of the PBM. Since homography is a 2D projection, no geometrical data exist to disclose occlusions automatically. One option is to use a 2D occlusion mask that excludes the PBM for particular areas or known objects within the 2D egocentric view of the observing camera. A system that integrate PBMs may include computer vision algorithms [18,42] for estimating 3D poses of objects or depth from 2D images [46] in dynamic scenes with unknown foreign objects.

A textured 3D quad allows automatic occlusion handling during the 3D rendering pipeline. Virtual objects intended for AR visualization naturally occlude with the mirror plane. A system may use the methods presented for homography to allow occlusion with the physical environment. Systems may use 3D reconstruction algorithms such as Kinect Fusion [24] or triangulation of depth images to occlude the PBM in place of the natural environment. A model-based approach with digital twins [48] allows natural occlusion handling through the 3D rendering engine. In-situ rendering of the models allow occlusion free of noise and artifacts which are otherwise common reconstruction-based approaches. The drawback is that the system must know the real-world object's position, rotation, and scale. In order to acquire those parame-

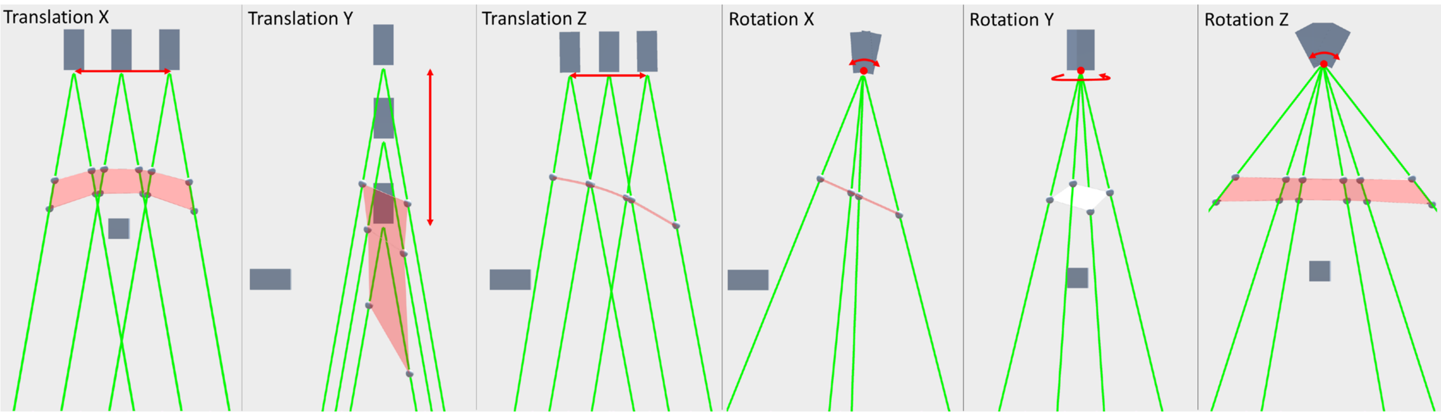


Fig. 5. **Effect of Positional and Rotational Tracking Errors of the Capturing Camera** on the Projective Bisector Mirror. In particular, we investigate positional and rotation errors per axis and visualize them. From these superimposed poses, we derive that positional errors (1-3) directly affects the estimated mirror object's shape, position, and rotation. On the other hand, rotational errors (4-6) causes the mirror to maintain co-planarity with the actual mirror plane.

ters, the system may need to integrate sophisticated tracking systems for six DoF pose estimations of real-world objects.

3.6 Condition of Valid Mirror Geometry

The computation of the PBM requires the 3D plane-ray intersection to compute the corner position with the bisector. Therefore, we can not create the PBM view if the bisecting plane does not intersect with the rays defined from the capturing camera's frustum. This effect occurs when users place the observer camera such that the bisecting plane is parallel to at least one of four frustum rays or place it behind the capturing camera. Consequently, a larger FoV of the capturing camera results in a smaller region where the bisecting plane intersects all corner rays. Let $a = tan(\frac{fov_x}{2} \cdot \frac{\pi}{180^\circ})$ and $b = tan(\frac{fov_y}{2} \cdot \frac{\pi}{180^\circ})$ be variables that involve the FoV angles in degree defined for the horizontal and vertical image axes. Further, let ϕ $(0 \leq \phi \leq \frac{\pi}{2})$ be the angle in radian constructed between the camera's horizontal axis with the camera's forward direction as the rotation axis. Then

$$\gamma = 2 \cdot tan^{-1}(cos(\phi) \cdot a + sin(\phi) \cdot b) \quad (28)$$

is the angle in radian between the vector on the frustum and the optical axis of the camera. Finally, we describe the critical angle $\theta_{critical}$ between the optical axis of the capturing camera and the mirror normal $\vec{v_N}$, where the bisecting plane is parallel to at least one ray constructed by the camera frustum with

$$\theta_{critical} = 180^\circ - \gamma \frac{180^\circ}{\pi} \quad (29)$$

As the FoV angle increases, when considering the diagonal lines inside the frustum, the width of the corresponding area decreases linearly. As seen in Fig. 4, the smallest FoV angles within the frustum, namely for the horizontal and vertical image axes, define the edges of the area of valid mirror geometries. In-between the axes, diagonal FoV angles are larger , which causes the area to form planar triangles in-between the edges. Furthermore, an image ratio unequal to one will result in the area resembling the shape of a rhombic pyramid. According to these descriptions, the method will allow for a larger area of valid positions by artificially cropping the camera input and reducing the values for the FoV accordingly.

3.7 Effect of Tracking Errors

Tracking errors affect the positioning of the PBM plane. An unaccounted tracking error of the camera position may lead to misinterpretation of the actual location of objects seen on the PBM view since the system calculated the mirror plane with an erroneous pose. As depicted in Fig. 5, tracking errors of the capturing camera result in six different types of behavior for each of the six existing DoF. The positional error affects the calculation of the bisecting plane and, therefore, changes the mirror normal. However, since a PBM view computes the bisecting plane from just both positions of the involved cameras, rotational tracking errors only affect the mirror's position, but not its facing direction.

Tracking errors of the observer camera translate to direct offset visible inside its combined view of the original camera image and PBM visualization since the pose of the acquired image from the physical camera in 3D space no longer matches the pose of the corresponding virtual camera of the observer for rendering virtual content. Nonetheless, the observing camera's tracking errors are more lenient than the capturing camera since they will not distort the reflected image.

4 Exploratory Use-cases

We present two use-cases that utilize the concept of PBMs. For each, we provide a brief description of the implementation for estimating the effort and complexity required to create a working system from our proposed concept. Finally, we summarize our observations while interacting with the functioning prototypes and point towards limitations and future work for both use-cases.

4.1 A Mirror into Reality

Virtual Reality (VR)-enabled 3D Telepresence is a promising concept for immersive communication and allows full-body non-verbal communication compared to webcam-based video conferences. Specialized types of telepresence combine VR and AR technologies for asymmetric teleconsultation. Related work investigates such systems' usage for daily communication [40], medical consultation [35, 51, 53], industrial maintenance [50], and collaborative tasks [52]. The systems presented in these works capture a real-world area with depth sensors and reconstruct a 3D environment displayed in a VR application. VR users can enter the reconstruction and perceive the capture 3D environment as if they are physically located at the captured environment. While proposed reconstructions have reached high fidelity, technical limitations remain regarding noise and holes within the 3D reconstruction and the capture range of the depth and color (RGB-D) sensors. Such artifacts reduce the information on the captured scene. Therefore, we propose the usage of PBMs in combination with the 3D reconstruction to provide a real-world view of the original location. In particular, since poses of the RGB-D sensors are often known for 3D reconstruction, we utilize these poses as the capturing camera for our PBMs. While immersed in the 3D reconstructed room, VR users perceive the PBM views with the real-world reflected within, resulting in a metaphorical understanding of a mirror into reality. Furthermore, since the depth-sensor has a limited capture range while the color image does not, this concept allows the perception of the local environment beyond the depth capturing range.

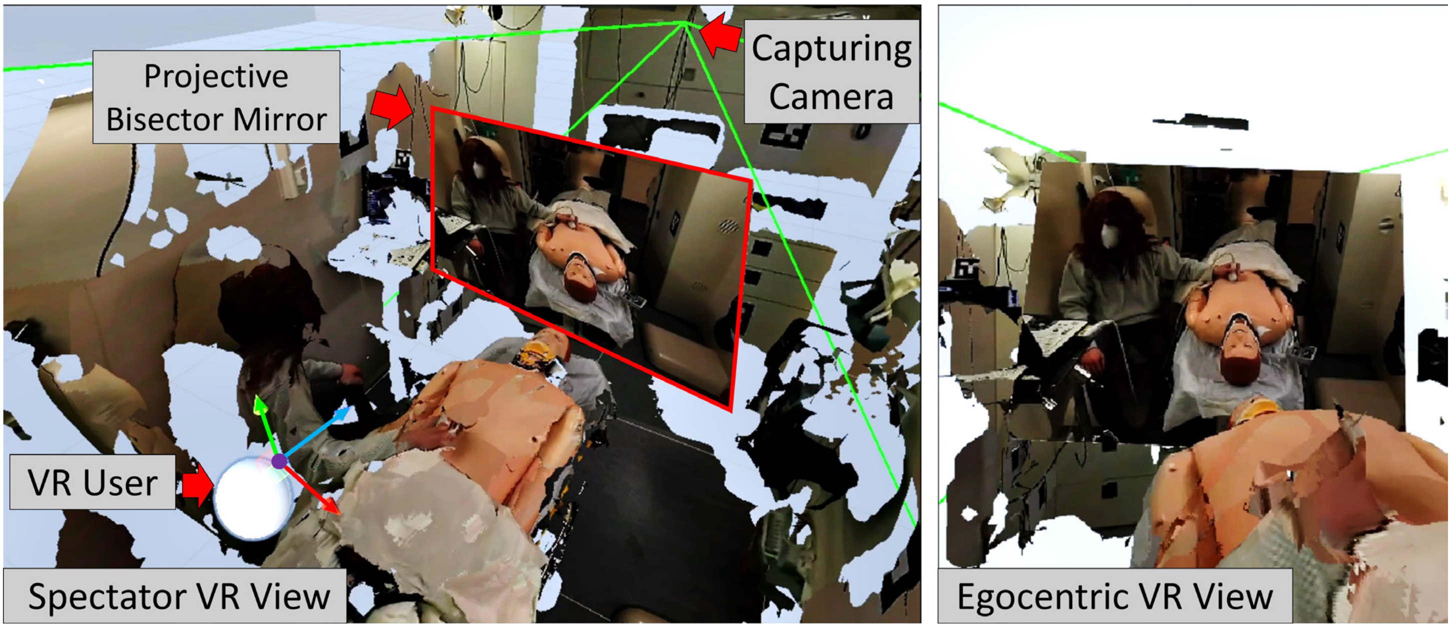


Fig. 6. **Use-case: A Mirror into Reality**. Screenshot taken from our presented prototype in Unity3D which combines a real-time captured 3D reconstruction and a VR headset. The spectator view (left) shows the position of the VR user and one of the capturing RGB-D cameras for reconstructing the inside of an ambulance. The PBM is constructed perspective-correct for the VR user's egocentric view (right). The PBM only uses the 2D color image of the capturing camera for generating the mirror view.

4.1.1 Implementation

We implement a prototype based on our presented algorithms for the concept of a mirror into reality. In particular, we investigate the VR side of a telepresence system. Inspired by the 3D reconstruction setup of [54] and [51], we deploy five Azure Kinect cameras for acquiring the reconstruction of a room by rigidly mounting them onto the ceiling. We calibrate the RGB-D cameras with a bundle adjustment [1] using an L-shaped ground target and moving features acquired through a wanding object, similar to the calibration of commercial multi-camera motion tracking systems. We deploy PBMs for each Kinect camera and use their camera poses and color image streams as input. A workstation receives all Kinect cameras' RGB-D streams and triangulates acquired colored point-cloud data into a triangulated 3D environment representation. A user with a VR headset, in our case a tethered Vive Pro Eye, dives into the 3D reconstructed scene. We use the user's VR camera as the observer camera for the PBMs. Henceforth, we have an N:1 relationship between N capturing cameras against one observer camera. We create the PBMs with the approach of the quadrilateral plane as it naturally regulates occlusion with the 3D reconstructed data.

4.1.2 Observations

Our point cloud reconstruction shows large holes and low geometric fidelity compared to the actual environment; however, the PBM allows the VR user to see the camera-captured scene, as seen in Fig. 6. We utilize a fixed ratio of cropping to decrease the size of the mirror in the final view of the VR user. We observe that the mirror plane increases in size the further the VR user moves away from the capturing camera. An alternative design choice for the integration of PBMs may integrate a dynamic ratio of cropping such that the size of the perceived mirror plane stays invariant towards the user's distance. Since we deploy multiple RGB-D cameras for the 3D reconstruction, we investigated the simultaneous display of PBMs from multiple viewpoints. We observe that the mirror planes may intersect with the PBM views of neighboring cameras. To optimize the interaction of multiple PBMs simultaneously, developers may want to provide users with the ability to precisely select and toggle PBM views on the fly or add a logic to allow the system to handle their visibility automatically. In a ceiling-mounted multi-camera setup surrounding a common area of interest, PBMs never intersect with the central area as long the observer is within the same space, as seen in Fig. 6 (right). Instead, the PBMs of each camera appears to create a dome of PBMs surrounding the observer. Finally, we recommend that users who wish a closer view onto a region of interest install a camera closer to the desired region or use a hand-held camera along with the PBM view.

4.2 Augmented Reality Automotive - Side Mirror

Camera Monitoring Systems (CMSs) installed on an automotive allow drivers to gain additional information on the surrounding area. Related work proposed using such systems to gain additional views in blind zones and investigated replacing conventional rear mirrors entirely [4, 34, 47]. An observed disadvantage of a CMS is the minification of objects due to displaying the acquired camera stream onto a small monitor within the driver's cockpit. We transfer the principle of PBMs into the context of CMS. An advantage of a PBM is its visualization through AR and is, therefore, not limited to a small monitor restricted by the interior design of the driver's cockpit. Instead, the system visualizes the acquired camera view in-situ as an extension to the conventional side rear mirror in mitigating blind zones.

This concept assumes drivers to see the PBM view from their egocentric view. We put forward a prototype of our AR side rear mirror following the principles of PBMs (see Fig. 7), and we further demonstrate the use of an occlusion mask to increase coherence with the real world (see Fig. 9). This prototype solely aims to demonstrate the feasibility of PBM for this use case.

4.2.1 Implementation

We rigidly attach an RGB camera onto an automotive. Since the pose of the PBMs rear mirror should result in approximately the same position as conventional rear mirrors, we attach our camera on the left car fender and in front of the driver. When placed correctly, the camera faces backward while the bisection plane between camera and driver is coplanar with the mirror plane of the conventional rear mirror. For our prototype, we use a Vicon system as an outside-in tracking system that tracks the rigidly attached capturing camera (see Fig. 8) and the hand-held observer camera used to simulate the egocentric view of the driver. Both cameras are Realsense D435 sensors and integrate color and depth sensors. We only use the color stream (1280x720 pixel @ 30Hz, RGB8) for this prototype. A laptop (Intel Core i7 @ 2.20Ghz, NVIDIA Geforce RTX 2060, 16GB RAM) processes both camera inputs simultaneously and handles the necessary computations

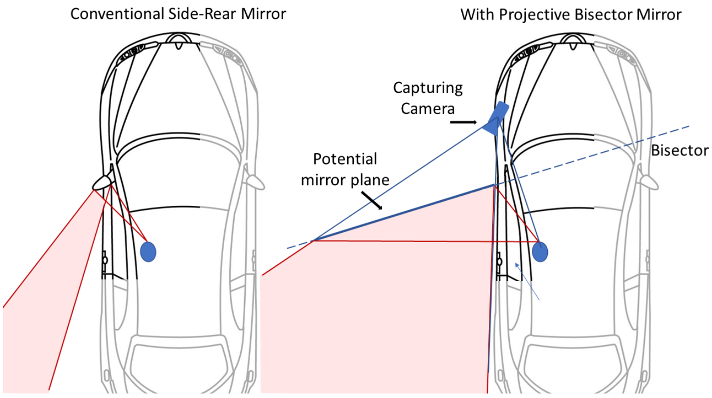


Fig. 7. **Use-case: Augmented Reality-based Side Rear Mirror** to minimize blind spots and increase the field of view. Left: Conventional real mirrors inhabit blind spots and a limited field of view. Right: Projective Bisector Mirrors covers a larger field of view. In contrast to Camera Monitor Systems, our method maintains the mirror metaphor and potentially allows the driver understand their surroundings more quickly.

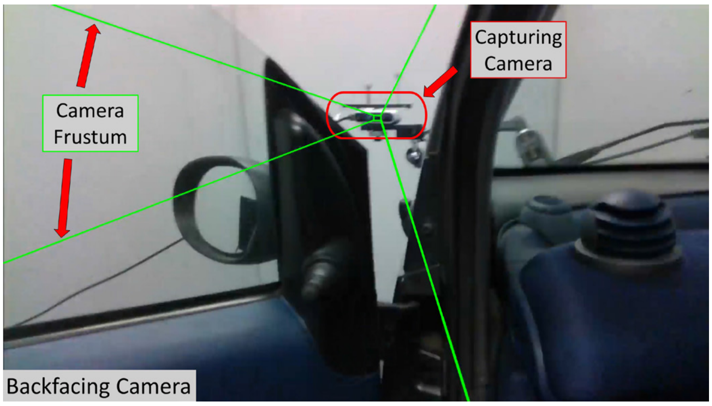


Fig. 8. **A Car-Mounted Camera Facing Backward** captures a wide field of view of the area behind the driver. The frustum (green lines) of the camera visualizes the 3D region inside its field of view.

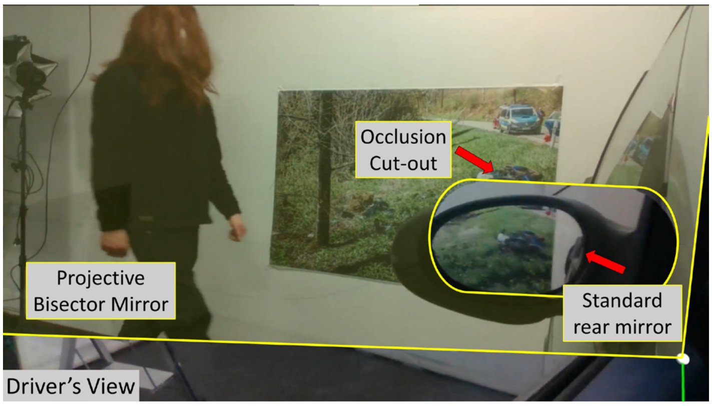


Fig. 9. **The Projective Bisector Mirror as Extended Rear Mirror** (tinted yellow) shows the mirror view towards the back, including areas within the blind zones of the conventional rear mirror. A cut-out within the virtual mirror allows the driver to see the original rear mirror. This image is a snapshot from the egocentric view of the driver.

for the PBM. To allow a view on the conventional side mirror and the PBM simultaneously, we include a 2D cut-out area in the shape of a stadium. We measure the latency of the RGB color stream between capture and PBM display at around 150ms.

4.2.2 Observations

The FoV of the PBM far exceeds the range covered by the conventional side mirror, as seen in Fig. 9, and reveals common blind zones of such mirrors. Depending on the positioning and FoV of the capturing camera, the generated PBM view covers anything from the door to the two neighboring lanes. We applied transparency of 90% on the PBM, which allows the driver to see traffic behind the PBM plane. In this regard, future work of this particular use case includes investigations of optimal visualization techniques to display the PBM without obstructing the natural view. We identified cropping and adjusting transparency as two tools for opposing the occlusion issue. In addition, we consider using an X-Ray visualization [44] or image-based ghosting [60] as advanced solutions for improving the visualization further.

On the opposite side of the driver's cockpit, the rear-facing mirror needs to be installed further on the front part of the automotive such that the bisecting plane is placed at the level of the conventional rear mirror.

Aside from camera quality aspects, including resistance against brightness changes, electromagnetic noise, dirt, and weather, we introduce image capture latency and tracking accuracy to the system. In the actual use case, the system coupled to the automotive should transmit captured camera image to the egocentric view of the driver. A portable HMD in the future may rely on wireless transmission to communicate with a network of surrounding devices, which potentially results in latencies during transmission. Possible solutions include upgrading wireless communication interfaces to support high energy transmission with higher frequency or projecting the view of the PBM with the principle of a Pepper's Ghost onto the glass plane of the side window. A future iteration of this use case may rigidly attach the capturing camera to the chassis and, therefore, relieves the system in tracking the camera's real-time position.

5 Discussion

In the following sections, we discuss the characteristics and remaining challenges of PBMs. Additionally, we focus on presenting features of the PBMs on a meta-level, including collaboration, visualization, and utility.

5.1 Characteristics of the Projective Bisector Mirrors

The PBM inhabits characteristics of both conventional mirrors and virtual mirrors.

Mirror Plane is independent of camera orientations The orientation plane is calculated based on the bisecting plane and only dependent on the camera positions. This point benefits outside-in tracking systems with high positional and lower rotational tracking precisions.

Objects appear laterally inverted Lateral inversion describes the front-back inversion of mirrors, commonly interpreted as a left-right inversion by humans. We maintain this property of physical mirrors with PBMs since we construct them such that the capturing camera is located at the theoretical position of the virtual mirror camera and looking at the scene from the opposing direction.

Reflected objects are the same size as the original objects The distance of the capturing camera to the mirror plane is equal to the distance of the observer. Therefore, all perceived objects within the PBM view appear at the same position and scale as the original objects.

The mirror size depends on the FoV of the capturing camera The size of the mirror plane depends on the FoV of the capturing camera. A large FoV causes a broader spread of the rays describing the camera frustum and, hence, increases the size of the PBM. Furthermore, under the assumption that both cameras have the same FoV, it is possible to rotate the view of the observer camera such that the PBM view perfectly fills the egocentric view.

The PBM plane is calculated for every user individually The system calculates the position of the mirror plane based on the camera positions. Therefore, no two cameras can see the result in the same bisecting plane at any time; however, all cameras see a physically correct mirror representation for their perspective. Therefore, if a second user were to see the PBM view calculated for the first user, the view would not be a mirror.

The content of the reflection is independent of the observer The captured content of the capturing camera is independent of the bisection plane and, therefore, the observer. This implies, that observers always perceive the same image on the PBM as long as the capturing camera remains static and the captured scene does not change.

Knowledge on the 3D geometry of the real-world are optional PBMs only require color streams for creating mirror images. Therefore, 3D geometry of the environment, e.g., acquired through depth sensors, is optional for occlusion handling. This aspect enhances flexibility in choosing the camera type and system design for integrating PBMs.

5.2 Stereopsis

In previous sections, we described the concept of PBMs with monocular cameras. The extension to stereovision, for instance, for AR-enabled HMDs, requires the system to render the PBM once for each eye. Stereovision is a strong depth cue of human perception for medium viewing distances. An approach with the texture 3D quad as PBM allows conventional 3D rendering techniques to handle stereo rendering, and are available for most AR and VR rendering engines. Implementing PBMs with homography for stereoscopic displays require the system to compute the bisecting plane for each eye and compute the corresponding homography matrix.

5.3 Meta Features

The following sections highlight features on a higher level than concrete use-cases. The primary purpose is to communicate the versatility of our concept in everyday life and domain-specific challenges.

5.3.1 Augmented Reality Collaboration

Correct depth perception of AR objects that tightly co-exist with real-world objects is a recurring challenge. In particular, AR HMDs often use focal distances different to the distance of real-world objects resulting in wrong depth estimation caused by the vergence-accommodation conflict [12,57] during near-field tasks. Further, related work shows that accurately aligning real and virtual objects is not a trivial task [13,49] where depth perception is one of the leading issues. A user benefits from PBMs by seeing the egocentric view of another user, represented as a mirror image. Additional views onto an area of interest allow users to compare different perspectives and gain a more comprehensive spatial understanding [33].

Co-located collaborative AR benefit from PBMs as they help two or more users to communicate their ideas. When two or more users are nearby, PBMs provide the colleagues' views and displays them into the egocentric view of the observer. At the same time, the observer may offer their view as an PBM view for others. Aside from the potential benefits for aligning tasks, the additional viewpoints help understand other users' perspectives and prevent miscommunications. In contrast to a picture-in-picture approach, the user can directly correlate objects inside the reflection to the real environment.

5.3.2 Alternative Visualization in the Mirror Image

The view of the PBM is a composite of the captured camera image and the virtual world as perceived from the same pose. Unlike real mirrors, PBMs can augment 3D objects into their view or apply computer vision and machine learning algorithms. Moreover, the capture medium of the cameras is not limited to visible light. It can equally capture alternative optical information such as an infrared camera, an x-ray detector, or a thermal camera. Multi-modal visualization opens up new methods for allowing users to perceive the real-world and PBM view differently beyond of just a one-to-one mirror view. For instance, AR objects can be exclusive rendered for one camera or visualized differently in each PBM view (see Fig. 1).

In use cases where capturer and observer are two separate systems, virtual 3D data may be exclusively available to one system. For instance, such data include complex structures such as 3D volumes, 3D point clouds, and procedurally generated geometries. PBM help users to share complex or large data without transmitting the raw data itself. Similarly, an AR application can intentionally hide AR content in the observer's egocentric view while rendering them only within the PBM view. This method would leave the direct view of the observer free of visual clutter while providing the benefit of AR visualizations for the specific task still available within their view.

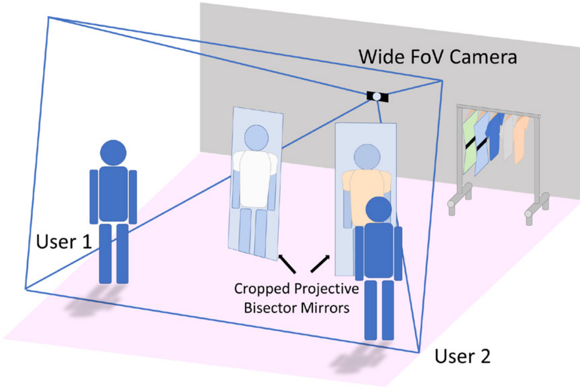


Fig. 10. **Use-case: Full-body Mirror.** Projective Bisector Mirrors replaces conventional mirrors with perspective-correct visualization. A single wide-angle camera can support multiple users simultaneously by cropping out only the relevant areas for each person. Furthermore, the Projective Bisector Mirror can display virtual content such as virtual clothes for a virtual try-on or anatomical structures for medical education on top of the user if the application can acquire the user's body pose.

5.3.3 Full-Body Mirrors

AR mirrors are a popular approach for creating a full-body AR experience, including use-cases in medical education and virtual try-on of clothes [17,28,55]. Besides realistic animation of augmentations, one of the challenge of AR mirrors is the visualization and correction perspective to match the egocentric view of the user. Setups involving monitors, either as the primary display or with a combination with a half-silvered mirror, are often unsuitable for displaying a perspective-correct mirror, even more for displaying correct mirror views for multiple observers simultaneously. Here, PBM provides remedy as its view are calculated for each observer individually. As depicted by Fig. 10, two or more users concurrently interact with a single capturing camera while observing only the PBM view designated for each of them while augmenting the view with additional AR content. The technical requirements are one of the remaining challenges as the system connecting the capturing camera needs to provide the camera stream to the user's wearable or hand-held device. Moreover, the receiving device must be aware of its pose in regard to the capturing camera. This could be done by SLAM with initial registration or outside-in tracking installed in the vicinity.

6 Conclusion

This work introduces the concept of Projective Bisector Mirrors. PBMs allow displaying external camera views in the egocentric views of observers while following the metaphor and mathematical constraints of a mirror. We lay out extensive mathematical descriptions of the underlying constructs and conditions of operation and delivered an insight into the effects of tracking errors of the capturing camera. Systems can use PBMs to provide the view of any locatable cameras in the egocentric view of an observer. Alternatively, the user can position a camera to present a PBM mirror explicitly. To demonstrate both approaches of using a PBM, we present two concrete use cases and their implementation. First, with the Mirror into Reality, users entering the room can immediately use PBMs to perceive the view of the cameras. The second concrete use case demonstrates that the specific placement of the capturing camera allows the construction of a superior mirror compared to conventional mirrors. We believe the concept of PBMs ventures into new possibilities of AR visualization of camera views and allows novel methods for interaction with digital content.

References

[1] S. Agarwal, K. Mierle, and T. C. S. Team. Ceres Solver, 3 2022.

[2] C. Badue, R. Guidolini, R. V. Carneiro, P. Azevedo, V. B. Cardoso, A. Forechi, L. Jesus, R. Berriel, T. M. Paixao, F. Mutz, et al. Self-driving Cars: A Survey. *Expert Systems with Applications*, 165:113816, 2021.

[3] R. Barmaki, K. Yu, R. Pearlman, R. Shingles, F. Bork, G. M. Osgood, and N. Navab. Enhancement Of Anatomical Education Using Augmented Reality: An Empirical Study Of Body Painting. *Anatomical sciences education*, 12(6):599–609, 2019.

[4] C. Bernhard and H. Hecht. Mirror Or Camera? Acceptance And Valuation Of Camera-monitor Systems. *Transportation Research Interdisciplinary Perspectives*, p. 100512, 2021.

[5] C. Bichlmeier, S. M. Heining, M. Feuerstein, and N. Navab. The Virtual Mirror: A New Interaction Paradigm for Augmented Reality Environments. *IEEE Transactions on Medical Imaging*, 28(9):1498–1510, 2009.

[6] O. Bimber, L. M. Encarnaçáo, and D. Schmalstieg. Augmented Reality with Back-Projection Systems using Transflective Surfaces. In *Computer Graphics Forum*, vol. 19, pp. 161–168. Wiley Online Library, 2000.

[7] J. F. Blinn. Hyperbolic Interpolation. *IEEE Computer Graphics and Applications*, 12(4):89–94, July 1992. Conference Name: IEEE Computer Graphics and Applications. doi: 10.1109/MCG.1992.10028

[8] 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 (VRW)*, pp. 115–116. IEEE, 2012.

[9] F. Bork, R. Barmaki, U. Eck, K. Yu, C. Sandor, and N. Navab. Empirical Study Of Non-reversing Magic Mirrors For Augmented Reality Anatomy Learning. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 169–176. IEEE, 2017.

[10] F. Bork, L. Stratmann, S. Enssle, U. Eck, N. Navab, J. Waschke, and D. Kugelmann. The Benefits Of An Augmented Reality Magic Mirror System For Integrated Radiology Teaching In Gross Anatomy. *Anatomical sciences education*, 12(6):585–598, 2019.

[11] D. E. Breen, E. Rose, and R. T. Whitaker. Interactive Occlusion And Collision Of Real And Virtual Objects In Augmented Reality. *Munich, Germany, European Computer Industry Research Center*, 1995.

[12] S. Condino, M. Carbone, R. Piazza, M. Ferrari, and V. Ferrari. Perceptual Limits of Optical See-Through Visors for Augmented Reality Guidance of Manual Tasks. *IEEE Transactions on Biomedical Engineering*, 67(2):411–419, 2019.

[13] A. Dey, G. Jarvis, C. Sandor, and G. Reitmayr. Tablet Versus Phone: Depth Perception In Handheld Augmented Reality. In *2012 IEEE international symposium on mixed and augmented reality (ISMAR)*, pp. 187–196. IEEE, 2012.

[14] H. Durrant-Whyte and T. Bailey. Simultaneous Localization and Mapping: Part I. *IEEE robotics & automation magazine*, 13(2):99–110, 2006.

[15] 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, pp. II–557. IEEE, 2007.

[16] 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. *IEEE Robotics and Automation Letters*, 5(2):2722–2729, 2020.

[17] S. Giovanni, Y. C. Choi, J. Huang, E. T. Khoo, and K. Yin. Virtual Try-on Using Kinect And HD Camera. In *International Conference on Motion in Games*, pp. 55–65. Springer, 2012.

[18] C. Godard, O. Mac Aodha, and G. J. Brostow. Unsupervised Monocular Depth Estimation With Left-right Consistency. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 270–279, 2017.

[19] T. B. Greenslade Jr. Pepper's Ghost. *The Physics Teacher*, 49(6):338–339, 2011.

[20] S. Hauswiesner, M. Straka, and G. Reitmayr. Free Viewpoint Virtual Try-on With Commodity Depth Cameras. In *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry*, pp. 23–30, 2011.

[21] P. S. Heckbert. Fundamentals Of Texture Mapping And Image Warping. 1989.

[22] 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*, pp. 1–4. IEEE, 2010.

[23] K. Hormann and M. Tarini. A Quadrilateral Rendering Primitive. In *Proceedings of the ACM SIGGRAPH/EUROGRAPHICS conference on Graphics hardware*, pp. 7–14, 2004.

[24] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, et al. Kinectfusion: Real-time 3D Reconstruction And Interaction Using A Moving Depth Camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pp. 559–568, 2011.

[25] Y. Kameda, T. Takemasa, and Y. Ohta. Outdoor See-through Vision Utilizing Surveillance Cameras. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 151–160. IEEE, 2004.

[26] M. Kim and K. Cheeyong. Augmented Reality Fashion Apparel Simulation Using A Magic Mirror. *International journal of smart home*, 9(2):169–178, 2015.

[27] K. Kiyokawa and H. Takemura. A Tunnel Window And Its Variations: Seamless Teleportation Techniques In A Virtual Environment. In *HCI International*. Citeseer, 2005.

[28] K. Kjærside, K. J. Kortbek, H. Hedegaard, and K. Grønbæk. ARDressCode: Augmented Dressing Room With Tag-based Motion Tracking And Real-time Clothes Simulation. In *Proceedings of the central european multimedia and virtual reality conference*, pp. 511–515, 2005.

[29] 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*, pp. 73–76, 2016.

[30] N. Li, Z. Zhang, C. Liu, Z. Yang, Y. Fu, F. Tian, T. Han, and M. Fan. vMirror: Enhancing The Interaction With Occluded Or Distant Objects In VR With Virtual Mirrors. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–11, 2021.

[31] F. Marreiros and A. Marcos. Calculating the Stereo Pairs Of A Mirror-based Augmented Reality System. *Atas do 13º Encontro Português de Computação Gráfica, 12-14 outubro 2005, UTAD, Vila Real, Portugal.*, pp. 41–48, 2005.

[32] 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)*, pp. 207–216. IEEE, 2020.

[33] 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)*, pp. 217–226. IEEE, 2020.

[34] E. N. Mazzae, G. Baldwin, A. T. Andrella, et al. Examination of a Prototype Camera Monitor System for Light Vehicle Outside Mirror Replacement. Technical report, United States. Department of Transportation. National Highway Traffic Safety ..., 2018.

[35] L. Mcmillan, K. Arthur, L. College, and R. Bajcsy. Virtual-Space-Teleconferencing-Using-a-Sea-of-Cameras. p. 8.

[36] T. Miyamoto, I. Kitahara, Y. Kameda, and Y. Ohta. Floating Virtual Mirrors: Visualization of the Scene Behind a Vehicle. In *International Conference on Artificial Reality and Telexistence*, pp. 302–313. Springer, 2006.

[37] K. Mukhopadhyay, C. Sinha, H. N. Saha, S. Rakshit, and S. Auddy. Smart Mirror-a Secured Application of Artificial Intelligence Recognizing Human Face and Voice. In *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pp. 1279–1289. IEEE, 2018.

[38] J. D. Mulder. Realistic Occlusion Effects In Mirror-based Co-located Augmented Reality Systems. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, pp. 203–208. IEEE, 2005.

[39] N. Navab, M. Feuerstein, and C. Bichlmeier. Laparoscopic Virtual Mirror New Interaction Paradigm For Monitor Based Augmented Reality. In *2007 IEEE Virtual Reality Conference*, pp. 43–50. IEEE, 2007.

[40] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou, V. Tankovich, C. Loop, Q. Cai, P. A. Chou, S. Mennicken, J. Valentin, V. Pradeep, S. Wang, S. B. Kang, P. Kohli, Y. Lutchyn, C. Keskin, and S. Izadi. Holoportation: Virtual 3D teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, pp. 741–754. Association for Computing Machinery, New York, NY, USA, Oct. 2016.

[41] C. Portalés, J. Gimeno, S. Casas, R. Olanda, and F. G. Martínez. Interacting With Augmented Reality Mirrors. In *Virtual and Augmented Reality: Concepts, Methodologies, Tools, and Applications*, pp. 18–46. IGI Global, 2018.

[42] N. D. Reddy, M. Vo, and S. G. Narasimhan. Occlusion-Net: 2D/3D Occluded Keypoint Localization Using Graph Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*,

pp. 7326–7335, 2019.

[43] D. Roth, K. Yu, F. Pankratz, G. Gorbachev, A. Keller, M. Lazarovici, D. Wilhelm, S. Weidert, N. Navab, and U. Eck. Real-time Mixed Reality Teleconsultation For Intensive Care Units In Pandemic Situations. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 693–694. IEEE, 2021.

[44] 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*, pp. 27–36. IEEE, 2010.

[45] H. Sato, I. Kitahara, and Y. Ohta. MR-Mirror: A Complex Of Real And Virtual Mirrors. In *International Conference on Virtual and Mixed Reality*, pp. 482–491. Springer, 2009.

[46] A. Saxena, J. Schulte, A. Y. Ng, et al. Depth Estimation Using Monocular and Stereo Cues. In *IJCAI*, vol. 7, pp. 2197–2203, 2007.

[47] E. A. Schmidt, H. Hoffmann, R. Krautscheid, M. Bierbach, A. Frey, J. Gail, and C. Lotz-Keens. Camera-monitor systems as a replacement for exterior mirrors in cars and trucks. In *Handbook of camera monitor systems*, pp. 369–435. Springer, 2016.

[48] M. M. Shah, H. Arshad, and R. Sulaiman. Occlusion in Augmented Reality. In *2012 8th International Conference on Information Science and Digital Content Technology (ICIDT2012)*, vol. 2, pp. 372–378. IEEE, 2012.

[49] 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*, pp. 364–372. Springer, 2006.

[50] P. Stotko, S. Krumpen, M. B. Hullin, M. Weinmann, and R. Klein. SLAMCast: Large-Scale, Real-Time 3D reconstruction and streaming for immersive Multi-Client live telepresence. *IEEE Trans. Vis. Comput. Graph.*, 25(5):2102–2112, May 2019.

[51] R. Strak, K. Yu, F. Pankratz, M. Lazarovici, B. Sandmeyer, J. Reichling, S. Weidert, C. Kraetsch, B. Roegele, N. Navab, U. Eck, and D. Roth. Comparison Between Video-mediated and Asymmetric 3D Teleconsultation During a Preclinical Scenario. In *Mensch Und Computer 2021*, pp. 227–235. ACM, Ingolstadt Germany, Sept. 2021. doi: 10.1145/3473856.3473883

[52] B. Thoravi Kumaravel, F. Anderson, G. Fitzmaurice, B. Hartmann, and T. Grossman. Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, pp. 161–174. Association for Computing Machinery, New York, NY, USA, Oct. 2019. doi: 10.1145/3332165.3347872

[53] N. Weibel, D. Gasques, J. Johnson, T. Sharkey, Z. R. Xu, X. Zhang, E. Zavala, M. Yip, and K. Davis. Artemis: Mixed-reality environment for immersive surgical telementoring. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–4, 2020.

[54] K. Yu, A. Winkler, F. Pankratz, M. Lazarovici, D. Wilhelm, U. Eck, D. Roth, and N. Navab. *Magnoramas: Magnifying Dioramas for Precise Annotations in Asymmetric 3D Teleconsultation*. Mar. 2021. doi: 10.1109/VR50410.2021.00062

[55] M. Yuan, I. R. Khan, F. Farbiz, S. Yao, A. Niswar, and M.-H. Foo. A Mixed Reality Virtual Clothes Try-on System. *IEEE Transactions on Multimedia*, 15(8):1958–1968, 2013.

[56] Y. Zhang, N. Uzuegbunam, W. Xu, and S. C. Sen-ching. RoboMirror: Simulating A Mirror With A Robotic Camera. In *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 1734–1738. IEEE, 2016.

[57] Y. Zhou, J. Zhang, and F. Fang. Vergence-Accommodation Conflict in Optical See-Through Display: Review and Prospect. *Results in Optics*, 5:100160, 2021.

[58] 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*, pp. 1–7. 2017.

[59] S. Zokai, J. Esteve, Y. Genc, and N. Navab. Multiview Paraperspective Projection Model For Diminished Reality. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.*, pp. 217–226. IEEE, 2003.

[60] 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*, pp. 19–26. IEEE, 2010.

Part VI

Discussions

6 Discussions

The field of immersive medical teleconsultation is shaped by prior research (section 2.1.1), including notable contributions of teleconsultation systems with end-to-end solutions and implementation proposed by Weibel et al. [148] and Rebol et al. [149]. These systems aim to facilitate communication during ongoing medical procedures through the use of simple tools that assist in conveying the current state of the procedure. While previous research on collaboration in AR/VR has explored collaborative interactions in telepresence and co-located scenarios (see section 2.1.2), to our knowledge, no prior work has focused on investigating interaction techniques to address the specific challenges in enhancing the precision of annotation and interaction with the 3D point cloud inside a medical 3D teleconsultation systems. One significant difference between medical and non-medical consultations is the need for efficient communication and high visual fidelity. In our studies, the inclusion of drawn annotations has proved substantial as they allow users to communicate visually without relying on temporally fleeting signals such as speech and gestures. In the following, I build up and extend the discussion in the work of Magnoramas, Duplicated Reality, and the Projective Bisector Mirror.

6.1 Magnoramas and Duplicated Reality

Increasing the size of virtual objects have a similar effect to approaching the 3D geometry (represented as the baseline of the study); However, being close to virtual objects may induce discomfort and cause the user to experience cross-eye. Moreover, the accuracy of annotations is more likely to be compromised due to the occlusion of outside-in tracking sensors responsible for tracking the user's physical body and controllers. Therefore, magnification is the superior alternative to the simplistic approach of bringing the user closer to the virtual objects.

Our investigation has revealed that Magnoramas can naturally enhance the precision of mid-air annotations by increasing the time on task for line segments. As elaborated in our research, this is due to the magnified distance for users to trace the line. The percentage of improvement and time-on-task appears to increase linearly with the degree of magnification. While the possible correlation between magnification level and precision warrants further exploration, it is a secondary concern for the method's efficacy. Our studies have found that users prefer a magnification level of 1.5x to 2x the original scale to work comfortably. Going beyond this magnification factor could result in a loss of context, as the user may lose sight of the patient's overall anatomy and relevant anatomical landmarks for the task. Implementing virtual RGB-D cameras allows for capturing 3D snapshots [70] from any virtual scene, including un-prepared virtual environments. This contrasts virtual replicas [169] or Voodoo Dolls [86], where virtual objects need to be segmented from their environment or acquire the references to virtual objects through selection techniques. We believe that Magnoramas are, therefore, more generalizable than virtual replicas and voodoo dolls.

Vergence-Accomodation Conflict with Duplicated Reality

The discussion of Duplicated Reality revealed that the effect of the vergence-accommodation conflict (VAC) of optical see-through headsets was causing issues in translating annotations to the shape on the sand block. The Hololens 2, for example, renders all virtual content on a focal plane located two meters before the user, and the human eyes naturally adjust their pupil lens to two meters. The perception of depth is created in the Hololens by utilizing stereo-disparity, resulting in the vergence between both eyes. At a typical interaction range of human hands, which is around one meter, this creates the conflict described above, as our eyes are accommodating for two meters and converge for one meter. Furthermore, since it is impossible for the human eye to accommodate two and one meters, it is equally impossible in this scenario to see the natural world and annotations sharply at the same time. Duplicated Reality circumvents the issue, as annotations and the duplicated real world are both virtual. VAC is an issue in VR but not as stressful as in AR since VR does not blend virtual content with the real world. To a certain degree, we may treat the Duplicated Reality still as a piece of VR embedded into AR space, yet, interconnecting AR and VR through a visual link and replicated interactions in a shared space.

Scalability

In our studies, our system has continuously maintained a single instance of Magnorama or Duplicated Reality. However, this approach is not limited to a single instance and is only restricted by the computation power of the system and the specific use case. Users can simultaneously observe and annotate several regions of interest within their environment by allowing for multiple instances. Furthermore, selecting a region of interest inside an already replicated space and replicating it again is conceivable, providing an even more magnified view. Therefore, the value of Magnorama and Duplicated Reality in VR and AR systems may extend beyond medical teleconsultation, as it presents novel ways of interacting with the real and virtual environment, especially for exploring 3D geometry and interaction with distant objects.

6.2 Projective Bisector Mirrors

Conventional approaches for reflecting the real world alongside augmented objects would either acquire a digital twin of the environment, barring it from capturing real-time changes, or directly use a physical mirror, as mentioned in the related work (section 2.1.4). The mathematics behind the PBM allows the creation of virtual mirror views with any cameras whose 3D position is known to the observer. Therefore, the PBM does not require cumbersome heavy mirror panels or scanning the scene for the digital twin. Instead, it only requires the six degrees of freedom difference in the position of the physical camera and the observer. This requirement allows PBM creation with co-registered stand-alone cameras, smartphones, head-mounted displays, and similar devices with integrated cameras. Furthermore, since the mirror stream uses the camera's live stream, it is well-suited for applications set in a dynamic environment with real-time changes.

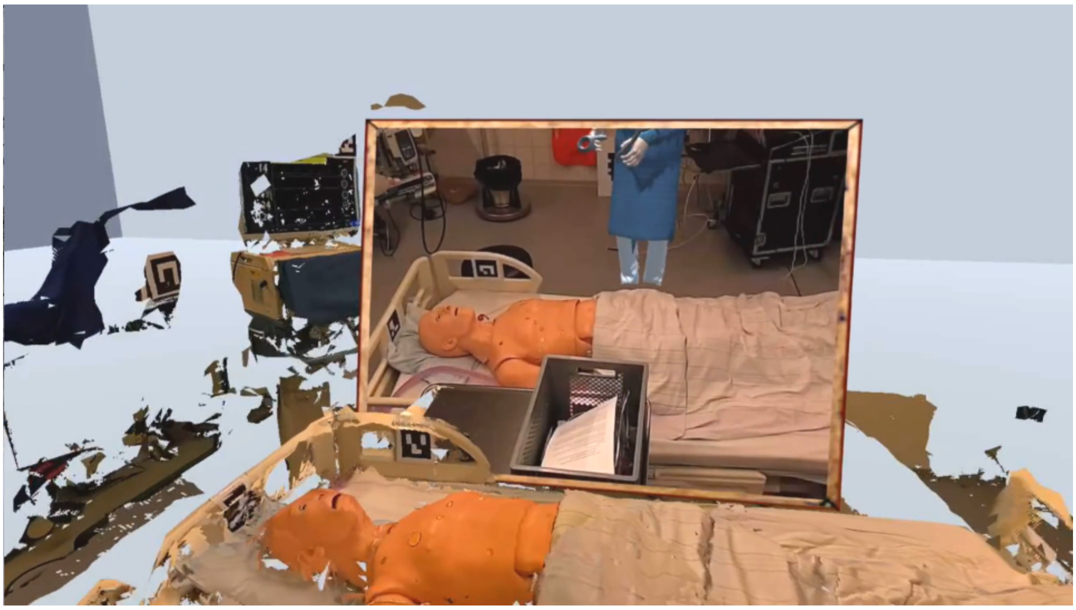


Fig. 6.1. The Projective Bisector Mirror Shows the Real World View, allowing users to perceive a live camera stream added to the virtual reconstruction of the local scene. In addition, the mirror image shows selected virtual objects such as annotations and the virtual 3D avatar of the remote user augmented into its view.

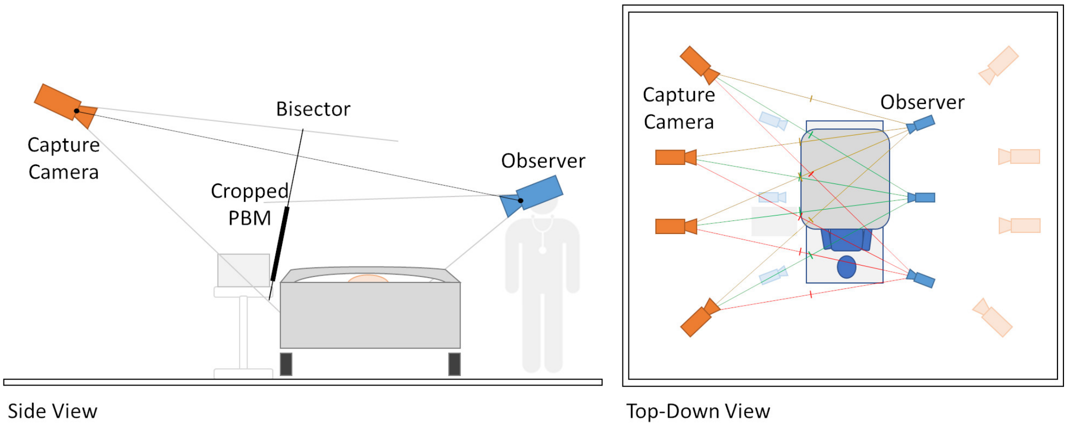


Fig. 6.2. The Scene as Illustration. Left: We attach the capture camera to the ceiling. In turn, the camera delivers the RGB-D point cloud for the reconstruction but further the 2D color image for the PBM at the center. The PBM projects the camera image onto the bisector based on its camera's intrinsic parameters. Cropping the image further allows the user to digest it visually and avoids major occlusions of the background. Right: A sea of cameras enables users to create a PBM from any cameras within the location around the patient.

Exceptions of the Mirror Visualization

One of the unique properties of PBMs is their clear definition on the visualization plane, namely on the bisector. For example, a setup within an operating theatre needs a camera on the opposite side of the patient table where the user stands. Due to the geometrical definition, the PBM will correctly reflect objects between the bisector and the observer; however, objects between the bisector and the capturing camera will not have a correct visualization. Nonetheless, it provides information on the location of those objects since they are located within the field of view of the capturing camera. Figure 6.1 shows an example

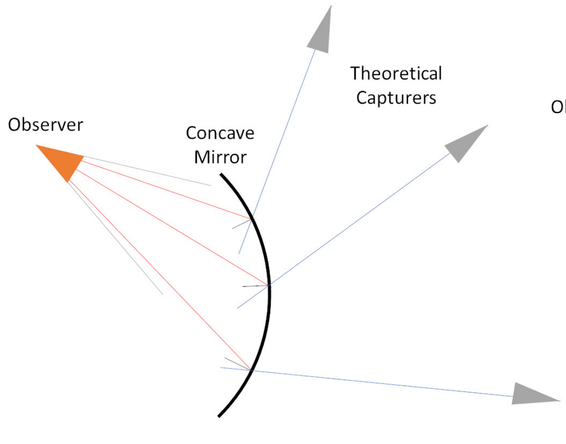


Fig. 6.3. **Theory of a Curved Mirror Based on PBM.** The curved mirror surface reflects every ray from the observer camera differently. By illustrating the necessary capture positions of real cameras of exemplary rays, we quickly observe that we would require an indefinite amount of cameras, or at least the same number of cameras for every pixel of the mirror texture.

demonstrating this issue: The PBM reflects the patient and the bed within its mirror view, where the mirror view shows a high-resolution view of the patient. The PBM shows the gray box we positioned behind the patient's bed. As it is behind the mirror plane, we can no longer see the box inside the 3D reconstruction but remain able to see it in the mirror view. To acquire the view as shown in Figure 6.1, we position the capture camera on the ceiling as seen in Figure 6.2 (left). As the user moves left and right, the angle at which they observe the PBM changes accordingly. For the user to maintain a similar mirror view in different positions, require a moving capture camera or a *sea of cameras* alternating in providing the live stream and pose to create the PBM view, as seen in Figure 6.2 (right). In this scenario, when the observer wishes to see with the occluded box, they need to move the bisector plane backward such that it is located behind the box. There are two approaches: First, the capturing camera must move backward, or second, the observer must move forward. The current mathematical definition of the PBM would not allow any other options. Suppose both cases are unattainable due to physical constraints, then the PBM would require an extension in the assumption of the PBM. In theory, we can freely choose the position of the PBM if we can choose the position of the capture camera. However, as the view of the capture camera is real, it would imply that we need to simulate a real camera, including the visual fidelity of the original camera.

Curved PBMs

The PBM models the mirror view assuming a planar mirror plane. Concave and convex mirrors in the real world reflect the light such that the reflected image appears magnified, minified, or even turned around its head due to the curvature of the reflective surface. What would it take to create a virtual mirror image that resembles a curved mirror? One characteristic of the PBM is that for every point on the planar surface, we correctly model the reflection where the incident angle equals the outgoing angle when tracing viewing rays from the respective cameras. Both angles are dependent on the orientation of the plane's normal vector. On a curved mirror, every point on the surface has a slightly different normal vector. Therefore, by constructing the geometry, as seen in Figure 6.3, we observe that a curved PBM would require a different position of the capture camera to acquire the correct view for every point on the surface. In practice, it is not necessary to deploy an infinite number of cameras, but only one camera for each pixel is necessary to provide a fully rendered view for the observer. Realistically, a construction to provide the required information could resemble a high-density sea of cameras. However, even then, it would be only feasible for a single curved mirror with a particular observer view and distance. We conclude that a hardware-based solution appears impossible. What if we can synthesize virtual camera images between individual cameras within the sea of cameras?

NeRF Extension of PBMs

To accommodate the challenges of both manually adjustable position of the virtual mirror image and imitation of curved surfaces, we require synthetic camera views with the visual fidelity of a physical camera. In this context, Neural Radiance Fields [235] (NeRF) may be the method we require to take PBMs to the next level. NeRFs can synthesize high-fidelity, never-before-seen RGB images based on a camera transform. The precondition is a trained network based on a fully-connected deep network with the environment captured from multiple viewpoints. The primary bottleneck for this idea is that NeRF can not capture real-time scene changes since the training takes excessive time beyond the update rates required for real-time. Furthermore, model training on images generated from a sea of images with parallel image synthetization would require immense optimization and computation power. Such a venture is likely unrealistic to achieve with the current state of the art.

Relaxing Requirement for Valid PBM Geometries

The original paper mentions a condition in the geometrical construction of the intersection points from the capturer camera with the bisector plane: The observer has to stay within the angle spanned by $\theta_{critical}$ which is a hard requirement for the bisector to intersect with each of the four frustum rays of the capturing camera. The following section presents an extension to the original work on PBMs by discussing the necessary computation to compensate for cases where the observer is outside the valid area. This requirement mainly exists because the bisector becomes parallel to at least one of the frustum rays. Intuitively, the bisector not being able to intersect a ray implies that there are pixels that cannot be projected onto the bisector as the distance for the projection goes into infinity. We observe that $\theta_{critical}$ depends on the camera's field of view. To lift the requirement and maintain the maximum availability of the camera image, we need to reduce the field of view from the camera dynamically such that the observer is just within $\theta_{critical}$. A smaller field of view of the capture camera will result in a larger 3D area where the observer can reside for a valid construction of the PBM. Therefore, without dynamically changing the aspect ratio of the capturer's image stream, we compute the ratio by how much we must artificially reduce the camera's field of view. By the same ratio, we cut away the outer edges of the image stream whose pixel would project into infinity otherwise. The approach to compute the ratio re-utilizes the formulas as stated within the original paper; however, replace the field of view fov with twice the angle $\angle A_{optical} P_{observer}$ spanned open by the observer with the optical axis of the capture camera. We compute the required field of view with

$$\theta_{required} = 2 \cdot \angle A_{optical} P_{observer} \cdot \frac{\pi}{180}$$

in radians, and

$$fov_{xy,required,radian} = \pi - \theta_{required}$$

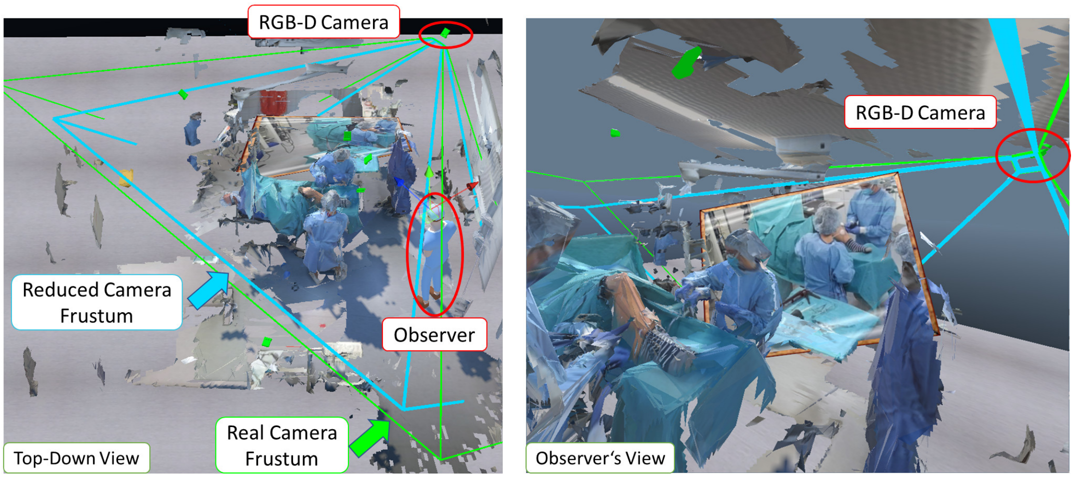


Fig. 6.4. Dynamically Reducing the Field of View Increases the Availability of the PBM due to its dependency on the capturer's field of view. Left: Top-down view of the new geometry with the blue frustum being the reduced FoV and green the actual camera's FoV. Right: The mirrored content within the PBM image does not change appearance. The reduction in the FoV leads the algorithm to crop away the outer edges; However, it becomes irrelevant when the user chooses a manually cropped visualization to avoid visual clutter.

in radians, which allows for the PBM's construction. With ϕ, $(0 \leq \phi \leq \frac{\pi}{2})$ being the angle in radian between the point of the observer projected onto the camera plane and its horizontal axis, we compute

$$f_{y,required} = H / \left(2 \cdot tan\left(\frac{fov_{xy,required,radian}}{2}\right)\right)$$

$$f_{x,required} = W / \left(2 \cdot tan\left(\frac{fov_{xy,required,radian}}{2}\right)\right)$$

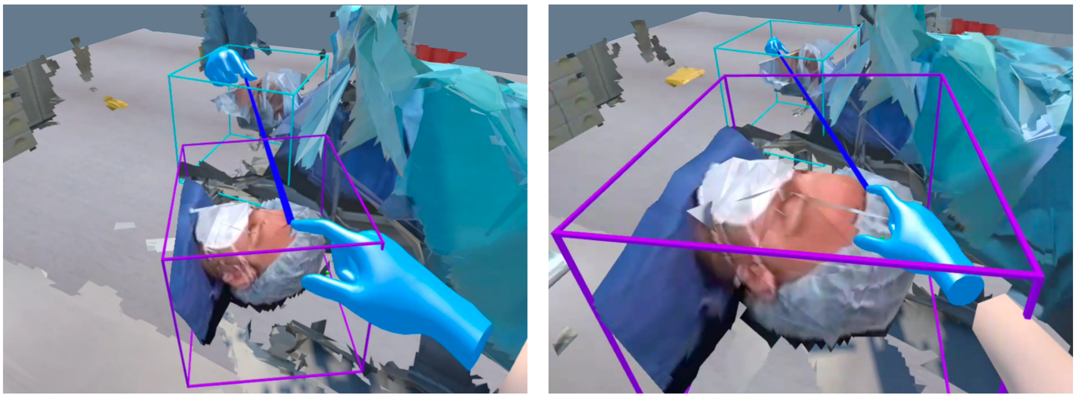
Finally, we compute the ratio r between the focal length of the original and compensated camera setting with

$$r = \frac{sin(\phi) \cdot f_{y,required} + cos(\phi) \cdot f_{x,required}}{f_{capturer}}$$

Figure 6.4 visualizes the result. With this extension, a system can construct a PBM with any observer positioned in front of the capture camera.

6.3 Regarding Presence and Immersion

In a well-cited article by Mel Slater [236], he discusses the relationship between presence and immersion in virtual environments. In his view, presence is a perceived response to a level of given immersion by the system. This is similar to how the perception of color depends on the individual human sensors reacting to light of specific wavelengths. Meanwhile, a system is more immersive, as sensory input and tracking technology increase fidelity to resemble the real world. According to Bulu et al. [237], social presence affects overall satisfaction, while telepresence and co-presence affect satisfaction in the virtual environment. Following the relation between presence and immersion, social presence indicates the perception of a



(A) Magnorama and Region of Interest are roughly the same size

(B) Magnorama is larger compared to the Region of Interest

Fig. 6.5. **A Hand Representation for the VR User** is reflected within the region of interest instead of a cursor object. This image is captured within a re-implementation of the ArtekMed interactions within Oculus Quest with hand-tracking capability and the 4D-OR dataset [238]. Additional to annotations, users can see the real-time gesturing hand within the region of interest. The hand representation is scaled down or up at the original location to allow gestures to be translated correctly. Left: Original and duplication are roughly the same scale. Therefore, the size of the hand does not noticeably change. Right: The Magnorama is substantially larger than the region of interest. For the user to perceive and maintain proportions, the hand at the original location is scaled down.

given immersion to communicate social cues and salience to a person. We can apply the same principle to co-presence and telepresence. Latter is often understood interchangeably with *just* presence or the sense of *being there*. Researchers can only indirectly measure immersion by subjectively measuring exposed users' perceptions. Virtually generated objects, such as Magnoramas and Duplicated Reality, show a non-natural view of the environment that disrupts the natural interaction with space as they partly disconnect locality. One arising question is whether unnatural visualization of space affects immersion. As shown within the first work, it does not impact telepresence and co-presence. Hence, replicating space while keeping the original environment representation does not affect the immersion. Surprisingly, for task-based teleconsultation, removing large parts of the virtual environment does not affect telepresence and co-presence negatively. This finding hints that voice communication may be sufficient to maintain telepresence and co-presence for a consultation system in VR. However, removing the virtual representation of the original environment will negatively impact social presence as it removes the visual cues of the communication partner. Daly-Jones et al. [239] described in 1998 what can be experienced today with online meetings on a near daily basis. In particular, for meetings with more than two participants, having visuals of at least the face and upper body from other people will immensely improve fluidity in communication and awareness. The user study of Magnorama and Duplicated Reality indicates that working in a separate cutout of reality slightly negatively impacts the social aspects and awareness, as they delocalize direct non-verbal communication. This observation co-aligns with the results of Daly-Jones, as participants can no longer directly perceive major communication cues such as gaze, hand gestures, and body rotation. As mentioned within the discussion of the Magnorama publication, one possible way to improve the presence is to add a virtual representation of the remote user to the region of interest regarding their position towards the Magnorama. Based on the related research with avatars [119], it would be preferred to represent the remote user with a real-time captured point cloud rather than a pre-made avatar.

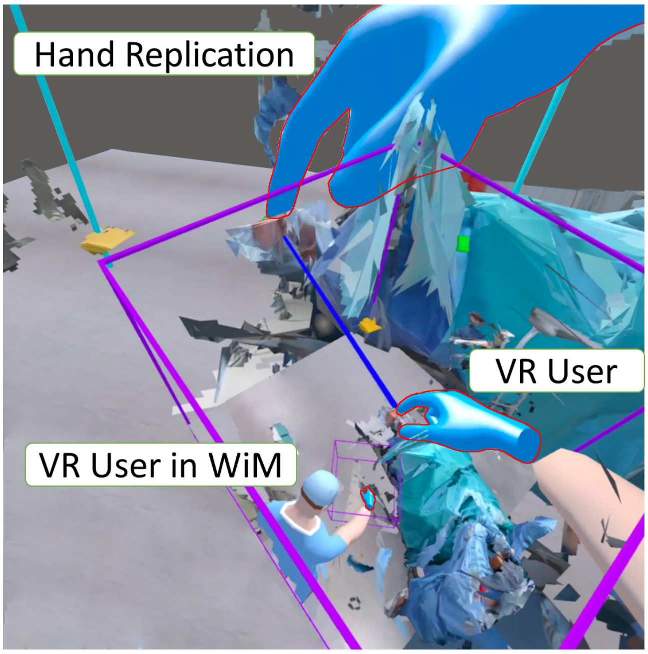


Fig. 6.6. **Recursion with WiMs** highlights the unnatural side of spatial replication. This image is captured within a re-implementation of the ArtekMed interactions within Oculus Quest with hand-tracking capability and the 4D-OR dataset [238]. In a particular case, where the region of interest encapsulates the replication (WiM or Magnorama) with user representations, VR users see their hand three times or more, given on the recursion depth, as it is additionally seen within the miniature and as the replication at the original location.

Recursion and Scale

There are approaches with potentially significant impacts that we have yet to investigate. For example, the section on future work in the publication of Magnoramas stated that a user representation could change the perception of social presence. There are several considerations to keep in mind. First, changing the scale involves user representation. To maintain overall spatial relationships, gestures, and avatar poses must be coherent in all nine degrees of freedom (translation, rotation, scale). A strong magnification of the Magnorama will mean the system will visualize the user in a downscaled form at the region of interest. Figure 6.5 illustrates this matter. Previous work [174] indicates a change in an avatar's size, especially miniaturization, hurts the remote user's attention and authority in the local user's eye. Therefore, an investigation is still needed to understand the requirements of dynamic changes in the user representation's scale in medical teleconsultation while maintaining or enhancing social presence. With the addition of a hand representation instead of universal cursor objects as proposed within the work of Magnoramas, recursions may occur, as seen in Figure 6.6 when the region of interest is scaled up and encapsulates the Magnorama and the VR user. The difference in scale makes this scenario a WiM. When users point within the Magnorama, their hands will be replicated on a giant scale at the original space and miniaturized within the replication. Recursion with WiMs [77] is known to the scientific community. However, whether recursion could provide value for multi-user applications is yet to be investigated.

To conclude the topic of presence for Magnoramas and Duplicated, I propose a pragmatic solution for minimizing the decreased presence when using them in a multi-user setup. We

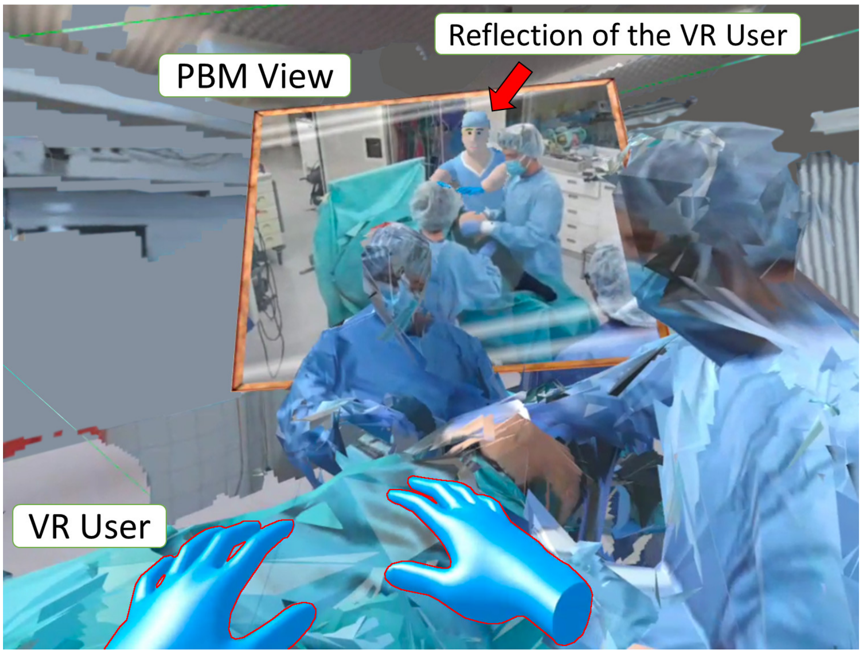


Fig. 6.7. **A VR User See Themselves in the PBM.** This image is captured within a re-implementation of the ArtekMed interactions within Oculus Quest with hand-tracking capability and the 4D-OR dataset [238], and shows the VR user seamlessly integrated into the virtual mirror image. As a result, the user can now easily understand their location within the real-world environment. The PBM view is visualized with a virtual wooden frame and static specular marks to suggest its glass surface properties. As the system has the 3D information on the scene, it uses them to introduce occlusion, which provides users with strong depth cues to understand the PBM's position. Furthermore, knowledge of the scene allows the system to fuse the virtual camera image of the user representation with the original video stream from the PBM's capture camera.

prioritize task performance over inter-personally perceived presence in medical emergencies and surgeries. To gain the best result from Magnorama and Duplicated Reality, users should treat these methods as tools to reach a specific goal rather than permanently offering them as some omnipresent entity within the teleconsultation. A compromise between availability and deactivation is likely the most pragmatic approach of Magnoramas and Duplicated Reality in everyday usage. Toggling them off allows users in VR and AR to focus on the general situation and team movements while having them within the user's view allows them to concentrate on the patient and the task.

Presence with PBMs

The following passage discusses the possible effects of PBM on perceiving presence. The original research on PBM aims at generating spatially coherent views of camera images. The primary use case for PBM inside the ArtekMed system is to obtain a high-resolution view in addition to the noisy reconstruction. The fusion of the virtual avatar into the PBM view was hinted at in Figure 6.1. The benefits should become clear with Figure 6.7. A few properties make PBM a powerful tool for 3D telepresence since mirrors in VR are regarded as important for creating a mental image of virtual self-representation, as detailed in the following. First, the system leverages the 3D geometry of the reconstruction to handle the occlusion of the virtual user representation within the camera image and the mirror plane. Occlusion provides a strong cue for depth estimation, enabling the user to obtain clear information on their position within the space. Second, the VR user can see oneself inside the PBM. Previous studies show that seeing the own avatar with matching movements will significantly enhance self-agency and the ownership of the virtual body [240, 241, 242], further improving the feeling of being part of the environment. For pro-active movements, the view of the virtual self inside the mirror is reported to improve motor actions [243]. Finally, the Proteus effect [218] is more likely to affect the VR user due to increased awareness of the appearance of the self-representation. Guegan et al. [244] showed that the Proteus effect could enhance creativity when users are represented as an avatar that they perceive as creative. It would not be a deliberative hypothesis to suggest that the representation of a doctor or surgeon within VR could induce a working mode in the remote expert, similar to the act of donning a uniform, irrespective of their attire, before entering the telepresence session. Hence, a potential future work is to investigate further the use of virtual uniforms in improving the quality of consultation.

6.4 Regarding Visual Overload and Over-Reliance

The human brain allows automatic visual processing that is separated from the attention [245]; However, according to a study by Charron et al. [246] in 2010, the human mind can only fully attend to a single task at any given time. Visual overload, or visual clutter, refers to the visualization of distracting and misleading visual content besides the focused content of the user, increasing the difficulty for humans to consciously and subconsciously process visual stimuli. Both real and virtual worlds are susceptible to the effect.

As outlined in section 1.5.2, the utilization of AR content in the immediate treatment of patients can result in adverse consequences, as in misunderstood or missed details that could, in severe instances, lead to life-threatening mistakes in the medical procedure. All digital mediums, whether screen-based or immersive applications are liable to generate visual overload. However, AR presents a unique challenge, as the user simultaneously processes both real-world and virtual content. Consequently, there likely exists a heightened risk of inadvertently obstructing contextually vital information to the user. In contrast to in-situ, ex-situ methods disconnect visualization and interaction to the referenced space and avoid occluding relevant real-world objects. Previous work often investigates ex-situ visualization as the conventional way to display information, e.g., comparing novel in-situ AR visualization with an ex-situ screen-based view [36, 247].

Discussing these types of visualization is thought-provoking because Magnoramas, Duplicated Reality, and PBM can match in none of those categories with minor modifications. In particular, these three methods provide additional canvases by duplicating and creating secondary views of the virtual environment. Those canvases can be augmented with virtual content, as seen in the experimental conditions of the user studies. PBMs could include AR content only visible within the mirror image. To path the idea for future work, I propose the term *inter-situ* to describe visualization that lies between in-situ and ex-situ. Inter-situ visualization refers to techniques that reference the real-world location visually and spatially but enable users to perceive and interact with AR content outside that referenced 3D space, such as through the interaction techniques commonly used in VR applications. Figure 6.8 demonstrates the differences between the visualization techniques. It is *essential* to note that the user studies performed in the relevant works were not designed to evaluate and support this idea. Therefore, additional future work focusing on the idea of inter-situ AR visualization would be required to establish this proposal. The following points are potential advantages of inter-situ visualization, based from the findings in this dissertation:

1. *Inter-situ visualization allows multi-user collaboration without interference in spatial occupations and interactions.*
   The property is based on observations from the user study of Duplicated Reality for co-located users. Collaborative experiences in VR are less affected by a shared environment, as there are no physical collisions between the users. Instead, virtual multi-user systems permit users to clip into other users or handle it by pushing themselves or other users away. Inter-situ visualization duplicates the space of the situs and is interlinked with it. Users can fully augment and interact with the space without visually overloading the view at the situs, physically colliding, or coordinating with co-located users as the duplicated space is separated. Virtual interactions, such as the creation of virtual annotations, can be synchronized back to the situs to enable simultaneous multi-user collaboration even in tight spaces.

2. *When inter-situ visualizations maintain the context of the region of interest, they can use the best of both real and virtual worlds.*
   In the method of Duplicated Reality, we utilized the VR pen to interact with the duplicated environment. The duplication is separated from the associated original location within the co-located space. Yet, users can quickly switch their view between the situs and the duplication. Furthermore, since the duplication is an entirely virtual object, it is possible to apply any virtual modifications to its visualization, such as clipping to get an intersection view, changing up the rendering shader, or altering the scaling, which would not be possible with the real world.

3. *Inter-situ visualization provides additional views.*
   Ex-situ visualization creates a duplicated space that represents the real world but in a different shape or with a different focus on data representation. This redundancy of information provides the system with additional canvases to present data in various ways to the user. For instance, AR mirrors like the Projective Bisector Mirror and the Augmented Mirror can display imaging information of a patient in the OR without showing any AR content on the patient. Furthermore, as the mirror image is a function of the position of the mirror plane in the environment, its inter-situ visualization creates a visual-spatial association that enables users to understand the relationship between visualization and the environment immediately.

The location of visualizing virtual content is not limited to impacting the visual processing of the human mind but also the interpretation. When visual AR guides, there is a good chance that users will begin to rely on them to perform tasks. AR guidance has proven to improve medical interventions, e.g., needle incision [19]; However, if both patient and needle are tracked, it would be more feasible to let a robot arm assist or perform the incision than using AR to show surgeons how to align the needle for maximizing insertion precision. If robot-assisted surgery is not an option, visual guides will be susceptible to tracking and perceptual error, which are sometimes not easily recognized. Over-reliance on computer-generated guidance is a concern in surgical procedures, as it may take over the surgeons' decision-making and experience.

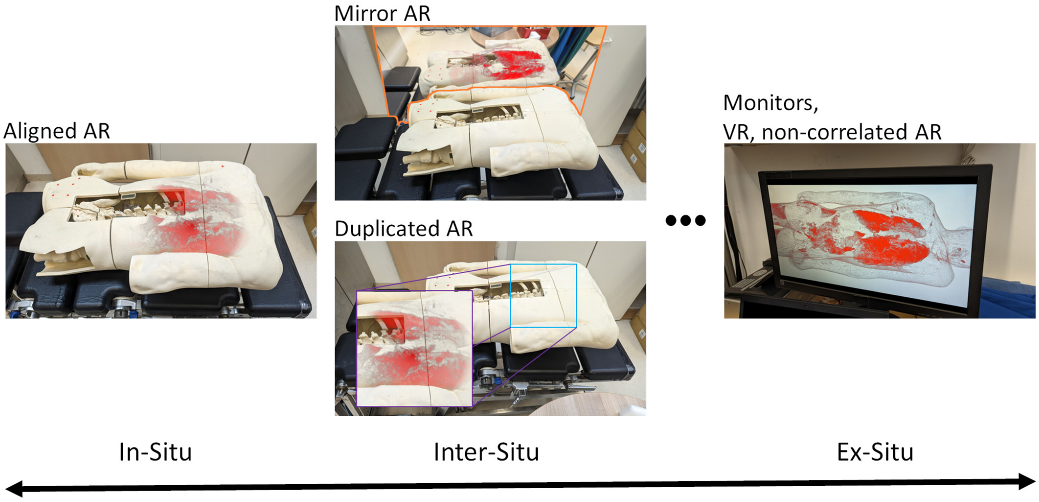


Fig. 6.8. The Proposed Categorization of 3D Visualization Techniques, such as for Duplicated Reality and the Projective Bisector Mirror. In-situ visualizations overlay the real-world content. Frequent issues are related to the occlusion of the working area and over-reliance on virtual content. On the other side, ex-situ visualization does not provide spatial information of the virtual content regarding the real world. In between both, I propose inter-situ visualization to label visualization types such as mirror-based AR and duplicated views, as a canvas to show virtual content. The advantage is an unobstructed view of the situs/patient while still benefiting from the augmentations of virtual content.

Moreover, dependence on such guidance can result in insecurity in surgical performance if the system malfunctions or becomes unavailable for any reason.

Literature on medical AR requires increased documentation on the issue of over-reliance. Although some research has explored this issue in specific contexts such as AR for cars [248], assembly [249, 250], and medical interventions [251], there is a lack of literature on the topic overall. Translating AR research into the operating room, it might become important to establish a dialogue between human and machine interfaces, to ensure user safety and improve patient outcomes rather than relying solely on digital authority to dictate actions. Human-assisted teleconsultation can facilitate this communication between users since involved users could still attempt to mitigate such shortcomings when they are mutually acknowledged. Looking at ArtekMed, one way to validate the correct positioning of provided annotations in local space is a comparison with the PBM view. Inconsistencies may arise due to de-calibrated tracking of the AR headset caused by sensor drift. If the annotations appear differently in the PBM mirror image compared to the in-situ AR visualization, the local user can notice the de-calibration issues and react accordingly.

6.5 Future Work

In addition to the future work discussed in the core publications, it would be valuable to extend the presented 3D UI techniques to clinical trials and it would be important to understand whether any potential decrease in presence and awareness may affect patient treatment. The duplication of content is a common function of digital systems. It would be valuable to explore whether individuals unfamiliar with digital technologies can similarly comprehend the concept of 3D space duplication in a short amount of time.

Alternatively, what if time is shifted within the confined local space of the region of interest rather than the space being replicated and shifted space? If time travel were possible, and people were familiar with the concept, they may find it easier to imagine time travel within a 3D box such as Magnorama.

Part VII

Conclusion

7 Conclusion

Immersive teleconsultation will likely turn into a valuable technology in the digital era, bridging the distance between individuals to almost instantaneous teleportation. The ArtekMed system provides an immersive 3D teleconsultation experience for medical applications, offering a glimpse into the future of communication while simultaneously revealing challenges at both the conceptual and technical levels. An essential tool for communication and conveyance of knowledge is the ability to create virtual 3D annotations shared and displayed in the virtual and augmented environment. However, technical and physiological limitations impede users' ability to create these annotations with the necessary precision and location. Furthermore, digital reconstructions must balance real-time capability and visual fidelity. In this context, the presented works examine and address fundamental issues in our 3D teleconsultation system through the Magnorama, Duplicated Reality, and Projective Bisector Mirror methods. Inspired by the World-in-Miniatures concept, Magnorama, and Duplicated Reality magnify a selected 3D region of interest instead of miniaturizing it. Two user studies demonstrate that these methods enhance hand-made annotation precision and overcome physical constraints in co-located scenarios. Furthermore, the Projective Bisector Mirror method turns tracked cameras into a mirror view, allowing the ArtekMed system to present the remote user with a clear view onto the patient additional to the 3D reconstruction by using the high-resolution real-world camera stream or an external user's viewpoint. The dissertation concludes by discussing the impact of the methods on the multi-user presence and their behavior regarding visual overload and over-reliance on virtual content. The outcome of the discussion leads to the proposal of the term "inter-situ visualization", describing a type of AR visualization that are not overlaid directly in-situ on to the region of interest yet still maintain spatial relevance to the situs such that they could not be labeled as ex-situ visualization.

This dissertation directed its research aspirations toward developing solutions to address fundamental challenges related to 3D user interfaces for medical teleconsultation. Drawn annotations and gestures, coupled with advanced 3D UI such as Magnoramas, Duplicated Reality, and Projective Bisector Mirrors, enable new perspectives in understanding and handling spatial information. The novel techniques developed and evaluated in this dissertation show promising results in advancing the field of medical teleconsultation and 3D user interfaces, providing a foundation for a variety of further research and development in the future.

Part VIII

Appendix

A List of Authored and Co-authored Publications

2023

[116] **K. Yu**, D. Roth, R. Strak, F. Pankratz, J. Reichling, C. Kraetsch, S. Weidert, M. Lazarovici, N. Navab, U. Eck.
"Mixed Reality 3D Teleconsultation for Emergency Decompressive Craniotomy: An Evaluation with Medical Residents", *IEEE International Symposium on Mixed and Augmented Reality (ISMAR) 2023*,

[117] U. Eck, M. Wechner, F. Pankratz, **K. Yu**, M. Lazarovici, N. Navab.
"Real-Time 3D Reconstruction Pipeline for Room-Scale, Immersive, Medical Teleconsultation", *Applied Science 2023 Vol. 13, No. 18*,

[118] N. Navab, A. Martin-Gomez, M. Seibold, M. Sommersperger, T. Song, A. Winkler, **K. Yu**, U. Eck.
"Medical Augmented Reality: Definition, Principle Components, Domain Modeling, and Design-Development-Validation Process", *Journal of Imaging, 9(1), p.4.*,

2022

[26] **K. Yu**, K. Zacharis, U. Eck, N. Navab.
"Projective Bisector Mirror (PBM): Concept and Rationale.", *IEEE Transactions on Visualization and Computer Graphics 28.11 (2022): 3694-3704, 2022 Sep*,

[70] **K. Yu**, U. Eck, N. Navab.
"Spatial Exploration with a WiM for Capturing 3D Dioramic Snapshots", *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE 2022, 2022 Mar*,

[25] **K. Yu**, U. Eck, F. Pankratz, M. Lazarovici, D. Wilhelm, N. Navab.
"Duplicated Reality for Co-located Augmented Reality Collaboration", *IEEE Transactions on Visualization and Computer Graphics 2022, 2022 Mar*,

2021

[252] D. Ostler, J. Steger, L. Bernhard, **K. Yu**, R. Hartwig, H. Feussner, D. Wilhelm. **"'Hybrid' Scientific Conference: Lessons Learned From the Digital Annual Meeting of the CARS International Conference During the COVID-19 pandemic."**, *Innovative Surgical Sciences. 2021 Nov 9.*,

[121] R. Strak, **K. Yu**, F. Pankratz, M. Lazarovici, B. Sandmeyer, J. Reichling, S. Weidert, C. Kraetsch, B. Roegele, N. Navab, U. Eck, D. Roth. **"Comparison Between Video-Mediated and Asymmetric 3D Teleconsultation During a Preclinical Scenario"**, *Mensch und Maschinen, 2021 Sep*,

[253] **K. Yu**, D. Ostler, J. Fuchtmann, A. Zapaishchykova, M. Berlet, N. Navab, H. Feussner, D. Wilhelm. **"Clean-AR: Using Augmented Reality for Reducing the Risk of Contamination from Airborne Disease Agents on Surfaces"**, *Current Directions in Biomedical Engineering, 2021 Aug*,

[119] **K. Yu**, G. Gorbachev, U. Eck, F. Pankratz, N. Navab, D. Roth. **"Avatars for Teleconsultation: Effects of Avatar Embodiment Techniques on User Perception in 3D Asymmetric Telepresence"**, *IEEE Transactions on Visualization and Computer Graphics 2021, 2021 Aug*,

[120] D. Roth, **K. Yu**, F. Pankratz, G. Gorbachev, A. Keller, M. Lazarovici, D. Wilhelm, S. Weidert, N. Navab U. Eck. **"Real-time Mixed Reality Teleconsultation for Intensive Care Units in Pademic Situations"**, *3DUI Contest, IEEE VR 2021, 2021 Mar*, Poster

[24] **K. Yu**, A. Winkler, F. Pankratz, M. Lazarovici, D. Wilhelm, U. Eck, D. Roth, N. Navab. **"Magnoramas: Magnifying Dioramas for Precise Annotations in Asymmetric 3D Teleconsultation"**, *IEEE VR 2021, 2021 Mar*, **Honorable Mention Demo Award**

2020

[198] **K. Yu***, A. Gomez*, A. Winkler*, D. Roth, U. Eck, N. Navab. **"Augmented Mirrors"**, *IEEE International Symposium on Mixed and Augmented Reality (ISMAR) 2020 Nov*, (*Shared Authorship)

[122] **K. Yu**, T. Wegele, D. Ostler, D. Wilhelm, H. Feußner. **"EyeRobot: Enabling Telemedicine Using a Robot Arm and a Head-Mounted Display"**, *Current Directions in Biomedical Engineering. 2020 Sep 17;6(1)*,

[254] **K. Yu***, JPO Held*, C. Pyles, JM Veerbeek, F. Bork, SM Heining, N. Navab, AR Luft. **"Augmented Reality–Based Rehabilitation of Gait Impairments: Case Report"**, *JMIR Mhealth Uhealth 2020;8(5):e17804, URL: https://mhealth.jmir.org/2020/5/e17804, DOI: 10.2196/17804, PMID: 32452815*, (*Shared Authorship)

[255] R. Hartwig, D. Ostler, H. Feußner, M. Berlet, **K. Yu**, J. Rosenthal, D. Wilhelm. **"COMPASS: Localization in Laparoscopic Visceral Surgery"**, *Current Directions in Biomedical Engineering 6.1 (2020)*,

2019

[6] R. Barmaki, K. Yu, R. Pearlman, R. Shingles, F. Bork, G.M. Osgood, N. Navab. **"Enhancement of Anatomical Education Using Augmented Reality: An Empirical Study of Body Painting"**, *Journal of Anatomical Sciences Education 2017*,

2018

[256] G. Deib, A. Johnson, M. Unberath, **K. Yu**, S. Andress, L. Qian, G. Osgood, N.Navab, F. Hui, P. Gailloud. **"Image Guidance for Percutaneous Interventional Spine Procedures using an Optical See-through Head-mounted Display: Proof of Concept and Rationale"**, *Journal of NeuroInterventional Surgery, to appear 2018*,

[257] S. Andress*, A. Johnson*, M. Unberath*, A. Winkler*, **K. Yu**, J. Fotouhi, S.Weidert, G. Osgood, N. Navab. **"On-the-fly Augmented Reality for Orthopaedic Surgery Using a Multi-Modal Fiducial"**, *SPIE Journal of Medical Imaging, vol.5, no. 2, 2018*, (*These authors contributed equally and are listed in alphabetical order.)

[258] **K. Yu***, R. Barmaki*, M. Unberath*, A. Mears M.D., J. Brey M.D., T. Hwan Chung M.D., and N. Navab. **"On the Accuracy of Low-cost Motion Capture Systems for Range of Motion Measurements"**, *SPIE Medical Imaging, to appear 2018.*, *These authors contributed equally.

2017

[100] F. Bork, R. Barmaki, U. Eck, **K. Yu**, C. Sandor, N. Navab. **"Empirical Study of Non-Reversing Magic Mirrors for Augmented Reality Anatomy Learning"**, *International Mixed and Augmented Reality (ISMAR) 2017 IEEE International Symposium on (pp. 169-176). IEEE.*,

[259] L. Qian, M. Unberath, **K. Yu**, B. Fuerst, A. Johnson, N. Navab, G. Osgood. **"Technical Note: Towards Virtual Monitors for Image Guided Interventions - Real-time Streaming to Optical See-Through Head-Mounted Displays"**, *arXiv preprint, 2017, arXiv:1710.00808.*,

RightsLink

Home | Help | Live Chat | Sign in | Create Account

Magnoramas: Magnifying Dioramas for Precise Annotations in Asymmetric 3D Teleconsultation

Conference Proceedings: 2021 IEEE Virtual Reality and 3D User Interfaces (VR)

Author: Kevin Yu

Publisher: IEEE

Date: March 2021

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK | CLOSE WINDOW

Duplicated Reality for Co-located Augmented Reality Collaboration

Author: Kevin Yu
Publication: IEEE Transactions on Visualization and Computer Graphics
Publisher: IEEE
Date: May 2022

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK | CLOSE WINDOW

Projective Bisector Mirror (PBM): Concept and Rationale

Author: Kevin Yu

Publication: IEEE Transactions on Visualization and Computer Graphics

Publisher: IEEE

Date: November 2022

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK | CLOSE WINDOW

Bibliography

[1] R. Pausch. *The Last Lecture*. Hachette UK, 2008 (cit. on p. 2).

[2] "THE TALKING PHONOGRAPH". In: *Scientific American* 37.25 (1877), pp. 384–385 (cit. on p. 3).

[3] C. Wheatstone. "XVIII. Contributions to the Physiology of Vision.—Part the First. On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision". In: *Philosophical transactions of the Royal Society of London* 128 (1838), pp. 371–394 (cit. on pp. 3, 10, 27).

[4] H. Fuchs and U. Neumann. "A Vision of Telepresence for Medical Consultation and Other Applications". In: *Sixth International Symposium of Robotics Research*. 1993, pp. 555–571 (cit. on pp. 3, 23).

[5] H. Fuchs, G. Bishop, K. Arthur, et al. "Virtual Space Teleconferencing Using a Sea of Cameras". In: *Proc. First International Conference on Medical Robotics and Computer Assisted Surgery*. Vol. 26. Citeseer. 1994 (cit. on pp. 3, 23, 67).

[6] R. Barmaki, K. Yu, R. Pearlman, et al. "Enhancement Of Anatomical Education Using Augmented Reality: An Empirical Study Of Body Painting". In: *Anatomical sciences education* 12.6 (2019), pp. 599–609 (cit. on pp. 4, 14, 27, 98).

[7] J. Zhao, X. Xu, H. Jiang, and Y. Ding. "The Effectiveness of Virtual Reality-Based Technology on Anatomy Teaching: A Meta-Analysis of Randomized Controlled Studies". In: *BMC medical education* 20.1 (2020), pp. 1–10 (cit. on p. 4).

[8] J. W. V. de Faria, M. J. Teixeira, L. d. M. S. Júnior, J. P. Otoch, and E. G. Figueiredo. "Virtual and Stereoscopic Anatomy: When Virtual Reality Meets Medical Education". In: *Journal of neurosurgery* 125.5 (2016), pp. 1105–1111 (cit. on p. 4).

[9] R. Kurul, M. N. Ögün, A. Neriman Narin, Ş. Avci, and B. Yazgan. "An Alternative Method for Anatomy Training: Immersive Virtual Reality". In: *Anatomical Sciences Education* 13.5 (2020), pp. 648–656 (cit. on p. 4).

[10] S. Dargar, R. Kennedy, W. Lai, V. Arikatla, and S. De. "Towards Immersive Virtual Reality (iVR): A Route to Surgical Expertise". In: *Journal of computational surgery* 2.1 (2015), pp. 1–26 (cit. on p. 4).

[11] L. Hodges, R Kooper, T. Meyer, et al. "Virtual Environments for Treating the Fear of Heights". In: *Computer* 28.7 (1995), pp. 27–34 (cit. on p. 4).

[12] P. Lindner, A. Miloff, C. Bergman, G. Andersson, W. Hamilton, and P. Carlbring. "Gamified, Automated Virtual Reality Exposure Therapy for Fear of Spiders: A Single-Subject Trial Under Simulated Real-World Conditions". In: *Frontiers in psychiatry* 11 (2020), p. 116 (cit. on p. 4).

[13] K. E. Laver, B. Lange, S. George, J. E. Deutsch, G. Saposnik, and M. Crotty. "Virtual Reality for Stroke Rehabilitation". In: *Cochrane database of systematic reviews* 11 (2017) (cit. on p. 4).

[14] A. Baldominos, Y. Saez, and C. G. Del Pozo. “An Approach to Physical Rehabilitation Using State-of-the-Art Virtual Reality and Motion Tracking Technologies”. In: *Procedia Computer Science* 64 (2015), pp. 10–16 (cit. on p. 4).

[15] H. Fuchs, M. A. Livingston, R. Raskar, et al. “Augmented Reality Visualization for Laparoscopic Surgery”. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI’98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings 1*. Springer. 1998, pp. 934–943 (cit. on p. 4).

[16] L. T. De Paolis and V. De Luca. “Augmented Visualization With Depth Perception Cues To Improve the Surgeon’s Performance in Minimally Invasive Surgery”. In: *Medical & biological engineering & computing* 57 (2019), pp. 995–1013 (cit. on p. 4).

[17] L. Qian, X. Zhang, A. Deguet, and P. Kazanzides. “Aramis: Augmented Reality Assistance for Minimally Invasive Surgery Using a Head-Mounted Display”. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part V 22*. Springer. 2019, pp. 74–82 (cit. on p. 4).

[18] J. Fotouhi, C. P. Alexander, M. Unberath, et al. “Plan in 2-D, Execute in 3-D: An Augmented Reality Solution for Cup Placement in Total Hip Arthroplasty”. In: *Journal of Medical Imaging* 5.2 (2018), pp. 021205–021205 (cit. on p. 4).

[19] Y. Abe, S. Sato, K. Kato, et al. “A Novel 3D Guidance System Using Augmented Reality for Percutaneous Vertebroplasty”. In: *Journal of Neurosurgery: Spine* 19.4 (2013), pp. 492–501 (cit. on pp. 4, 90).

[20] D. H. Sonnenwald, H. M. Söderholm, G. F. Welch, B. A. Cairns, J. E. Manning, and H. Fuchs. “Illuminating Collaboration in Emergency Helath Care Situations: Paramedic-Physician Collaboration and 3D Telepresence Technology”. In: *Information Research* 19.2 (2014) (cit. on p. 4).

[21] *Number of Medical Practitioners With Specialization Registered With the Medical Associations, Absolute, per 100,000 Inhabitants and Inhabitants per Medical Practitioner. Classification: Years, Region, Age, Sex, Specialist Title, Field of Activity*. `https://www.gbe-bund.de/gbe/pkg_isgbe5.prc_menu_olap?p_uid=gast&p_aid=25719483&p_sprache=E&p_help=2&p_indnr=656&p_version=41&p_ansnr=80862442`. Accessed: 2023-01-24 (cit. on p. 4).

[22] *Versorgungsgrade in den Planungsbereichen*. `https://gesundheitsdaten.kbv.de/cms/html/17016.php`. Accessed: 2023-01-24 (cit. on p. 4).

[23] L. E. Simon, J. Shan, A Rauchwerger, et al. “Paramedics’ Perspectives on Telemedicine in the Ambulance: A Survey Study”. In: *JEMS* 4.30 (2020), p. 20 (cit. on p. 4).

[24] K. Yu, A. Winkler, F. Pankratz, et al. “Magnoramas: Magnifying Dioramas for Precise Annotations in Asymmetric 3D teleconsultation”. In: *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. ©2021 IEEE. Reprinted, with permission, from Kevin Yu. IEEE. 2021, pp. 392–401 (cit. on pp. 5, 18, 20, 97).

[25] K. Yu, U. Eck, F. Pankratz, M. Lazarovici, D. Wilhelm, and N. Navab. “Duplicated Reality for Co-located Augmented Reality Collaboration”. In: *IEEE Transactions on Visualization and Computer Graphics* 28.5 (2022). ©2022 IEEE. Reprinted, with permission, from Kevin Yu, pp. 2190–2200 (cit. on pp. 5, 18, 20, 96).

[26] K. Yu, K. Zacharis, U. Eck, and N. Navab. “Projective Bisector Mirror (PBM): Concept and Rationale”. In: *IEEE Transactions on Visualization and Computer Graphics* (2022). ©2022 IEEE. Reprinted, with permission, from Kevin Yu (cit. on pp. 5, 18, 20, 96).

[27] U. Eck, P. Stefan, H. Laga, C. Sandor, P. Fallavollita, and N. Navab. “Exploring Visuo-Haptic Augmented Reality User Interfaces for Stereo-Tactic Neurosurgery Planning”. In: *International Conference on Medical Imaging and Augmented Reality*. Springer. 2016, pp. 208–220 (cit. on p. 7).

[28] S. Zollmann, C. Hoppe, S. Kluckner, C. Poglitsch, H. Bischof, and G. Reitmayr. "Augmented Reality for Construction Site Monitoring and Documentation". In: *Proceedings of the IEEE* 102.2 (2014), pp. 137–154 (cit. on p. 7).

[29] J.-R. Wu, M.-L. Wang, K.-C. Liu, M.-H. Hu, and P.-Y. Lee. "Real-time Advanced Spinal Surgery via Visible Patient Model and Augmented Reality System". In: *Computer methods and programs in biomedicine* 113.3 (2014), pp. 869–881 (cit. on p. 7).

[30] C. Bichlmeier, B. Ockert, S. M. Heining, A. Ahmadi, and N. Navab. "Stepping Into the Operating Theater: ARAV—Augmented Reality Aided Vertebroplasty". In: *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE. 2008, pp. 165–166 (cit. on p. 7).

[31] 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 p. 7).

[32] 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: *2007 6th IEEE and ACM international symposium on mixed and augmented reality*. IEEE. 2007, pp. 129–138 (cit. on p. 7).

[33] S. Zollmann, T. Langlotz, R. Grasset, W. H. Lo, S. Mori, and H. Regenbrecht. "Visualization Techniques in Augmented Reality: A Taxonomy, Methods and Patterns". In: *IEEE transactions on visualization and computer graphics* 27.9 (2020), pp. 3808–3825 (cit. on p. 7).

[34] 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 (2010), pp. 133–141 (cit. on p. 7).

[35] A. Devaux, C. Hoarau, M. Brédif, and S. Christophe. "Underground Visualization: Web-App, Virtual Reality, Ex Situ and In Situ Augmented Reality." In: *ISPRS Technical Commission IV Symposium 2018*. 2018 (cit. on p. 8).

[36] S. Hubenschmid, J. Wieland, D. I. Fink, et al. "Relive: Bridging In-Situ and Ex-Situ Visual Analytics for Analyzing Mixed Reality User Studies". In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022, pp. 1–20 (cit. on pp. 8, 89).

[37] I. E. Sutherland. "A Head-Mounted Three Dimensional Display". In: *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. 1968, pp. 757–764 (cit. on pp. 10, 23).

[38] A. M. Colman. *A Dictionary of Psychology*. Oxford quick reference, 2015 (cit. on p. 10).

[39] S. Lawry, V. Popovic, A. Blackler, and H. Thompson. "Age, Familiarity, and Intuitive Use: An Empirical Investigation". In: *Applied ergonomics* 74 (2019), pp. 74–84 (cit. on pp. 10, 11).

[40] J. M. Spool. "What Makes a Design Seem 'Intuitive'". In: *User Interface Engineering* 10.01 (2005) (cit. on p. 10).

[41] R. Sackmann and O. Winkler. "Technology Generations Revisited: The Internet Generation". In: *Gerontechnology* 11.4 (2013), pp. 493–503 (cit. on p. 11).

[42] 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 p. 11).

[43] M. Oberhauser, D. Dreyer, R. Braunstingl, and I. Koglbauer. "What's Real About Virtual Reality Flight Simulation? Comparing The Fidelity Of A Virtual Reality With A Conventional Flight Simulation Environment." In: *Aviation Psychology and Applied Human Factors* 8.1 (2018), p. 22 (cit. on p. 11).

[44] J. Jankowski and M. Hachet. "A Survey of Interaction Techniques for Interactive 3D Environments". In: *Eurographics 2013-STAR*. 2013 (cit. on p. 12).

[45] M. Usoh, K. Arthur, M. C. Whitton, et al. "Walking> Walking-in-place> Flying, in Virtual Environments". In: *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. 1999, pp. 359–364 (cit. on p. 12).

[46] S. S. Chance, F. Gaunet, A. C. Beall, and J. M. Loomis. "Locomotion Mode Affects the Updating of Objects Encountered During Travel: The Contribution of Vestibular and Proprioceptive Inputs to Path Integration". In: *Presence* 7.2 (1998), pp. 168–178 (cit. on p. 12).

[47] S. Razzaque. *Redirected Walking*. The University of North Carolina at Chapel Hill, 2005 (cit. on p. 12).

[48] V. Interrante, B. Ries, and L. Anderson. "Seven League Boots: A New Metaphor for Augmented Locomotion Through Moderately Large Scale Immersive Virtual Environments". In: *2007 IEEE Symposium on 3D User interfaces*. IEEE. 2007 (cit. on p. 12).

[49] M. Slater, M. Usoh, and A. Steed. "Taking Steps: The Influence of a Walking Technique on Presence in Virtual Reality". In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 2.3 (1995), pp. 201–219 (cit. on p. 12).

[50] N. C. Nilsson, S. Serafin, M. H. Laursen, K. S. Pedersen, E. Sikström, and R. Nordahl. "Tapping-in-Place: Increasing the Naturalness of Immersive Walking-in-Place Locomotion Through Novel Gestural Input". In: *2013 IEEE symposium on 3D user interfaces (3DUI)*. IEEE. 2013, pp. 31–38 (cit. on p. 12).

[51] R. P. McMahan, D. A. Bowman, D. J. Zielinski, and R. B. Brady. "Evaluating Display Fidelity and Interaction Fidelity in a Virtual Reality Game". In: *IEEE transactions on visualization and computer graphics* 18.4 (2012), pp. 626–633 (cit. on p. 12).

[52] M. R. Mine. "Virtual Environment Interaction Techniques". In: *UNC Chapel Hill CS Dept* (1995) (cit. on p. 12).

[53] D. A. Bowman, E. T. Davis, L. F. Hodges, and A. N. Badre. "Maintaining Spatial Orientation During Travel in an Immersive Virtual Environment". In: *Presence* 8.6 (1999), pp. 618–631 (cit. on p. 12).

[54] D. Zielasko, Y. C. Law, and B. Weyers. "Take a Look Around – The Impact of Decoupling Gaze and Travel-Direction in Seated and Ground-Based Virtual Reality Utilizing Torso-Directed Steering". In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2020, pp. 398–406 (cit. on p. 12).

[55] A. Von Kapri, T. Rick, and S. Feiner. "Comparing Steering-Based Travel Techniques for Search Tasks in a Cave". In: *2011 IEEE Virtual Reality Conference*. IEEE. 2011, pp. 91–94 (cit. on p. 12).

[56] S. Beckhaus, K. J. Blom, and M. Haringer. "ChairIO – The Chair-Based Interface". In: *Concepts and technologies for pervasive games: a reader for pervasive gaming research* 1 (2007), pp. 231–264 (cit. on p. 12).

[57] R. P. Darken and H. Cevik. "Map Usage In Virtual Environments: Orientation Issues". In: *Proceedings IEEE virtual reality (cat. no. 99CB36316)*. IEEE. 1999, pp. 133–140 (cit. on p. 12).

[58] D. A. Bowman, D. B. Johnson, and L. F. Hodges. "Testbed Evaluation of Virtual Environment Interaction Techniques". In: *Proceedings of the ACM symposium on Virtual reality software and technology*. 1999, pp. 26–33 (cit. on p. 12).

[59] E. Bozgeyikli, A. Raij, S. Katkoori, and R. Dubey. "Point & Teleport Locomotion Technique for Virtual Reality". In: *Proceedings of the 2016 annual symposium on computer-human interaction in play*. 2016, pp. 205–216 (cit. on p. 12).

[60] E. Langbehn, P. Lubos, and F. Steinicke. "Evaluation of Locomotion Techniques for Room-Scale VR: Joystick, Teleportation, and Redirected Walking". In: *Proceedings of the Virtual Reality International Conference-Laval Virtual*. 2018, pp. 1–9 (cit. on p. 12).

[61] T. Igarashi, R. Kadobayashi, K. Mase, and H. Tanaka. "Path Drawing for 3D Walkthrough". In: *Proceedings of the 11th annual ACM symposium on User interface software and technology*. 1998, pp. 173–174 (cit. on p. 12).

[62] B. Hagedorn and J. Döllner. "Sketch-Based Navigation in 3D Virtual Environments". In: *International Symposium on Smart Graphics*. Springer. 2008, pp. 239–246 (cit. on p. 12).

[63] C. Ware and S. Osborne. "Exploration and Virtual Camera Control in Virtual Three Dimensional Environments". In: *Proceedings of the 1990 symposium on Interactive 3D graphics*. 1990, pp. 175–183 (cit. on p. 12).

[64] R. Pausch, T. Burnette, D. Brockway, and M. E. Weiblen. "Navigation and Locomotion in Virtual Worlds via Flight Into Hand-Held Miniatures". In: *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. 1995, pp. 399–400 (cit. on pp. 12, 26).

[65] J. J. LaViola Jr, D. A. Feliz, D. F. Keefe, and R. C. Zeleznik. "Hands-Free Multi-Scale Navigation in Virtual Environments". In: *Proceedings of the 2001 symposium on Interactive 3D graphics*. 2001, pp. 9–15 (cit. on pp. 12, 26).

[66] D. Coffey, N. Malbraaten, T. B. Le, et al. "Interactive Slice WIM: Navigating and Interrogating Volume Data Sets Using a Multisurface, Multitouch VR Interface". In: *IEEE Transactions on Visualization and Computer Graphics* 18.10 (2011), pp. 1614–1626 (cit. on p. 12).

[67] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev. *3D User Interfaces: Theory and Practice*. Addison-Wesley Professional, 2017 (cit. on pp. 12, 40).

[68] C. Boletsis. "The New Era of Virtual Reality Locomotion: A Systematic Literature Review of Techniques and a Proposed Typology". In: *Multimodal Technologies and Interaction* 1.4 (2017), p. 24 (cit. on p. 12).

[69] 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 pp. 13, 25, 26, 40).

[70] K. Yu, U. Eck, and N. Navab. "Spatial Exploration with a WiM for Capturing 3D Dioramic Snapshots". In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE. 2022, pp. 900–901 (cit. on pp. 13, 20, 80, 96).

[71] F. Argelaguet, L. Hoyet, M. Trico, and A. Lécuyer. "The Role of Interaction in Virtual Embodiment: Effects of the Virtual Hand Representation". In: *2016 IEEE virtual reality (VR)*. IEEE. 2016, pp. 3–10 (cit. on p. 13).

[72] C. W. Borst and A. P. Indugula. "Realistic Virtual Grasping". In: *IEEE Proceedings. VR 2005. Virtual Reality, 2005.* IEEE. 2005, pp. 91–98 (cit. on p. 13).

[73] J. Jacobs and B. Froehlich. "A Soft Hand Model for Physically-Based Manipulation of Virtual Objects". In: *2011 IEEE Virtual Reality Conference*. IEEE. 2011, pp. 11–18 (cit. on p. 13).

[74] A. Talvas, M. Marchal, and A. Lécuyer. "The God-Finger Method for Improving 3D Interaction With Virtual Objects Through Simulation of Contact Area". In: *2013 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE. 2013, pp. 111–114 (cit. on p. 13).

[75] L. Vanacken, T. Grossman, and K. Coninx. "Exploring the Effects of Environment Density and Target Visibility on Object Selection in 3D Virtual Environments". In: *2007 IEEE symposium on 3D user interfaces*. IEEE. 2007 (cit. on p. 13).

[76] S. Frees and G. D. Kessler. "Precise and Rapid Interaction Through Scaled Manipulation in Immersive Virtual Environments". In: *IEEE Proceedings. VR 2005. Virtual Reality, 2005.* IEEE. 2005, pp. 99–106 (cit. on p. 13).

[77] A. Bluff and A. Johnston. "Don't Panic: Recursive Interactions in a Miniature Metaworld". In: *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*. 2019, pp. 1–9 (cit. on pp. 13, 26, 87).

[78] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa. “The Go-Go Interaction Technique: Non-linear Mapping for Direct Manipulation in VR”. In: *Proceedings of the 9th annual ACM symposium on User interface software and technology*. 1996, pp. 79–80 (cit. on p. 13).

[79] A. Bellarbi, S. Otmane, N. Zenati, and H. Belghit. “Design and Evaluation of Zoom-based 3D Interaction Technique for Augmented Reality”. In: *Proceedings of the Virtual Reality International Conference-Laval Virtual 2017*. 2017, pp. 1–4 (cit. on p. 13).

[80] A. Mossel, B. Venditti, and H. Kaufmann. “3DTouch and HOMER-S: Intuitive Manipulation Techniques for One-handed Handheld Augmented Reality”. In: *Proceedings of the virtual reality international conference: laval virtual*. 2013, pp. 1–10 (cit. on p. 13).

[81] C. Wilkes and D. A. Bowman. “Advantages of Velocity-Based Scaling for Distant 3D Manipulation”. In: *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*. 2008, pp. 23–29 (cit. on p. 13).

[82] D. A. Bowman and L. F. Hodges. “An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments”. In: *Proceedings of the 1997 symposium on Interactive 3D graphics*. 1997, 35–ff (cit. on p. 13).

[83] J. Liang and M. Green. “JDCAD: A Highly Interactive 3D Modeling System”. In: *Computers & graphics* 18.4 (1994), pp. 499–506 (cit. on p. 13).

[84] A. Forsberg, K. Herndon, and R. Zeleznik. “Aperture Based Selection For Immersive Virtual Environments”. In: *Proceedings of the 9th annual ACM symposium on User interface software and technology*. 1996, pp. 95–96 (cit. on p. 13).

[85] R. Kopper, F. Bacim, and D. A. Bowman. “Rapid and Accurate 3D Selection by Progressive Refinement”. In: *2011 IEEE symposium on 3D user interfaces (3DUI)*. IEEE. 2011, pp. 67–74 (cit. on p. 13).

[86] J. S. Pierce, B. C. Stearns, and R. Pausch. “Voodoo-dolls: Seamless Interaction at Multiple Scales in Virtual Environments”. In: *Proceedings of the 1999 symposium on Interactive 3D graphics*. 1999, pp. 141–145 (cit. on pp. 13, 26, 80).

[87] R. Wen, W.-L. Tay, B. P. Nguyen, C.-B. Chng, and C.-K. Chui. “Hand Gesture Guided Robot-Assisted Surgery Based on a Direct Augmented Reality Interface”. In: *Computer methods and programs in biomedicine* 116.2 (2014), pp. 68–80 (cit. on p. 14).

[88] R. Radkowski and C. Stritzke. “Interactive Hand Gesture-Based Assembly for Augmented Reality Applications”. In: *Proceedings of the 2012 International Conference on Advances in Computer-Human Interactions*. Citeseer. 2012, pp. 303–308 (cit. on p. 14).

[89] S. Song, D. Yan, and Y. Xie. “Design of Control System Based on Hand Gesture Recognition”. In: *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*. IEEE. 2018, pp. 1–4 (cit. on p. 14).

[90] I.-A. Zaiţi, Ş.-G. Pentiuc, and R.-D. Vatavu. “On Free-Hand TV Control: Experimental Results on User-Elicited Gestures With Leap Motion”. In: *Personal and Ubiquitous Computing* 19.5 (2015), pp. 821–838 (cit. on p. 14).

[91] R. Hu, S. James, and J. Collomosse. “Annotated Free-Hand Sketches for Video Retrieval Using Object Semantics and Motion”. In: *International Conference on Multimedia Modeling*. Springer. 2012, pp. 473–484 (cit. on p. 14).

[92] P. Salomoni, C. Prandi, M. Roccetti, L. Casanova, L. Marchetti, and G. Marfia. “Diegetic User Interfaces for Virtual Environments with HMDs: A User Experience Study With Oculus Rift”. In: *Journal on Multimodal User Interfaces* 11.2 (2017), pp. 173–184 (cit. on p. 14).

[93] I. Iacovides, A. Cox, R. Kennedy, P. Cairns, and C. Jennett. “Removing the HUD: The Impact of Non-diegetic Game Elements and Expertise on Player Involvement”. In: *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*. 2015, pp. 13–22 (cit. on p. 14).

[94] H. Kato and M. Billinghurst. "Marker Tracking and HMD Calibration for a Video-Based Augmented Reality Conferencing System". In: *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*. IEEE. 1999, pp. 85–94 (cit. on p. 14).

[95] S. Cai, X. Wang, and F.-K. Chiang. "A Case Study of Augmented Reality Simulation System Application in a Chemistry Course". In: *Computers in human behavior* 37 (2014), pp. 31–40 (cit. on p. 14).

[96] M. T. Restivo, M. de Fátima Chouzal, J. Rodrigues, P. Menezes, B. Patrão, and J. B. Lopes. "Augmented Reality in Electrical Fundamentals." In: *Int. J. Online Eng.* 10.6 (2014), pp. 68–72 (cit. on p. 14).

[97] N. Taketa, K. Hayashi, H. Kato, and S. Noshida. "Virtual Pop-up Book Based on Augmented Reality". In: *Symposium on human interface and the management of information*. Springer. 2007, pp. 475–484 (cit. on p. 14).

[98] J. Gausemeier, J. Fründ, and C. Matysczok. "AR-Planning Tool: Designing Flexible Manufacturing Systems With Augmented Reality". In: *ACM International Conference Proceeding Series*. Vol. 23. 2002, pp. 19–25 (cit. on p. 14).

[99] 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 (VRW)*. IEEE. 2012, pp. 115–116 (cit. on pp. 14, 27).

[100] F. Bork, R. Barmaki, U. Eck, P. Fallavolita, B. Fuerst, and N. Navab. "Exploring Non-reversing Magic Mirrors for Screen-based Augmented Reality Systems". In: *2017 IEEE Virtual Reality (VR)*. IEEE. 2017, pp. 373–374 (cit. on pp. 14, 27, 28, 98).

[101] F. Pereira, C. Silva, and M. Alves. "Virtual Fitting Room Augmented Reality Techniques For E-Commerce". In: *International conference on ENTERprise information systems*. Springer. 2011, pp. 62–71 (cit. on p. 14).

[102] T. L. Woods, S. Reed, S. Hsi, J. A. Woods, and M. R. Woods. "Pilot Study Using The Augmented Reality Sandbox To Teach Topographic Maps And Surficial Processes In Introductory Geology Labs". In: *Journal of Geoscience Education* 64.3 (2016), pp. 199–214 (cit. on p. 14).

[103] M. Nguyen, M. Waller, A. Pandya, and J. Portnoy. "A Review of Patient and Provider Satisfaction With Telemedicine". In: *Current allergy and asthma reports* 20.11 (2020), pp. 1–7 (cit. on p. 15).

[104] H. R. Hartson and E. C. Smith. "Rapid Prototyping in Human-Computer Interface Development". In: *Interacting with computers* 3.1 (1991), pp. 51–91 (cit. on p. 15).

[105] N. Larsson and L. Molin. "Rapid Prototyping of User Interfaces in Robot Surgery—Wizard of Oz in Participatory Design". In: *Advances in Information Systems Development*. Springer, 2006, pp. 361–371 (cit. on p. 15).

[106] S. Pokhrel, A. Alsadoon, P. Prasad, and M. Paul. "A Novel Augmented Reality (AR) Scheme for Knee Replacement Surgery by Considering Cutting Error Accuracy". In: *The international journal of medical robotics and computer assisted surgery* 15.1 (2019), e1958 (cit. on p. 15).

[107] J. A. Luzon, B. V. Stimec, A. O. Bakka, B. Edwin, and D. Ignjatovic. "Value of the Surgeon's Sightline on Hologram Registration and Targeting in Mixed Reality". In: *International journal of computer assisted radiology and surgery* 15.12 (2020), pp. 2027–2039 (cit. on p. 15).

[108] B. Carl, M. Bopp, B. Saß, B. Voellger, and C. Nimsky. "Implementation of Augmented Reality Support in Spine Surgery". In: *European Spine Journal* 28.7 (2019), pp. 1697–1711 (cit. on p. 15).

[109] 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 p. 15).

[110] A. W. Stedmon, R. Kalawsky, K Hill, and C. Cook. “Old Theories, New Technologies: Cumulative Clutter Effects Using Augmented Reality”. In: *1999 IEEE International Conference on Information Visualization (Cat. No. PR00210)*. IEEE. 1999, pp. 132–137 (cit. on p. 15).

[111] J. Flittner and J. L. Gabbard. “Adapting Image Analysis Measures of Visual Clutter to Multiple Plane Augmented Reality User Interfaces”. In: *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE. 2021, pp. 92–95 (cit. on p. 15).

[112] B. J. Dixon, M. J. Daly, H. Chan, A. D. Vescan, I. J. Witterick, and J. C. Irish. “Surgeons Blinded by Enhanced Navigation: The Effect of Augmented Reality on Attention”. In: *Surgical endoscopy* 27 (2013), pp. 454–461 (cit. on pp. 15, 17).

[113] C. Chen, M. Yarmand, Z. Xu, V. Singh, Y. Zhang, and N. Weibel. “Investigating Input Modality and Task Geometry on Precision-First 3D Drawing in Virtual Reality”. In: *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2022, pp. 384–393 (cit. on pp. 17, 32).

[114] A. Haleem, M. Javaid, R. P. Singh, and R. Suman. “Telemedicine for Healthcare: Capabilities, Features, Barriers, and Applications”. In: *Sensors international* 2 (2021), p. 100117 (cit. on p. 19).

[115] T. Raza, M. Joshi, R. M. Schapira, and Z. Agha. “Pulmonary Telemedicine—A Model To Access the Subspecialist Services in Underserved Rural Areas”. In: *International journal of medical informatics* 78.1 (2009), pp. 53–59 (cit. on p. 19).

[116] K. Yu, D. Roth, R. Strak, et al. “Mixed Reality 3D Teleconsultation for Emergency Decompressive Craniotomy: An Evaluation with Medical Residents”. In: *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2023, pp. 662–671 (cit. on pp. 20, 96).

[117] U. Eck, M. Wechner, F. Pankratz, K. Yu, M. Lazarovici, and N. Navab. “Real-Time 3D Reconstruction Pipeline for Room-Scale, Immersive, Medical Teleconsultation”. In: *Applied Sciences* 13.18 (2023), p. 10199 (cit. on pp. 20, 96).

[118] N. Navab, A. Martin-Gomez, M. Seibold, et al. “Medical Augmented Reality: Definition, Principle Components, Domain Modeling, and Design-Development-Validation Process”. In: *Journal of Imaging* 9.1 (2023), p. 4 (cit. on pp. 20, 27, 96).

[119] K. Yu, G. Gorbachev, U. Eck, F. Pankratz, N. Navab, and D. Roth. “Avatars For Teleconsultation: Effects Of Avatar Embodiment Techniques On User Perception In 3D Asymmetric Telepresence”. In: *IEEE Transactions on Visualization and Computer Graphics* 27.11 (2021), pp. 4129–4139 (cit. on pp. 20, 34, 86, 97).

[120] D. Roth, K. Yu, F. Pankratz, et al. “Real-Time Mixed Reality Teleconsultation for Intensive Care Units in Pandemic Situations”. In: *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE. 2021, pp. 693–694 (cit. on pp. 20, 33, 97).

[121] R. Strak, K. Yu, F. Pankratz, et al. “Comparison Between Video-Mediated and Asymmetric 3D Teleconsultation During a Preclinical Scenario”. In: *Mensch und Computer 2021*. 2021, pp. 227–235 (cit. on pp. 20, 35, 97).

[122] K. Yu, T. Wegele, D. Ostler, D. Wilhelm, and H. Feußner. “Eyerobot: Enabling Telemedicine Using a Robot Arm and a Head-Mounted Display”. In: *Current Directions in Biomedical Engineering* 6.1 (2020) (cit. on pp. 20, 97).

[123] A. Deshpande, S. Khoja, J. Lorca, et al. “Asynchronous Telehealth: A Scoping Review of Analytic Studies”. In: *Open Medicine* 3.2 (2009), e69 (cit. on p. 22).

[124] H. Qiao, Y. P. Kandel, S. K. Manikandan, et al. “Conditional Teleportation of Quantum-Dot Spin States”. In: *Nature communications* 11.1 (2020), p. 3022 (cit. on p. 22).

[125] M. Minsky. “Telepresence”. In: (1980) (cit. on p. 22).

[126] W. A. IJsselsteijn. “History of Telepresence”. In: *3D Videocommunication: Algorithms, Concepts and Real-Time Systems in Human Centred Communication* (2005), pp. 5–21 (cit. on p. 23).

[127] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs. "The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays". In: *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*. 1998, pp. 179–188 (cit. on p. 23).

[128] H. Fuchs, A. State, and J.-C. Bazin. "Immersive 3D Telepresence". In: *Computer* 47.7 (2014), pp. 46–52 (cit. on p. 23).

[129] W.-C. Wen, H. Towles, L. Nyland, G. Welch, and H. Fuchs. "Toward a Compelling Sensation of Telepresence: Demonstrating a Portal to a Distant (Static) Office". In: *Proceedings Visualization 2000. VIS 2000 (Cat. No. 00CH37145)*. IEEE. 2000, pp. 327–333 (cit. on p. 23).

[130] V. Sequeira, E. Wolfart, E. Bovisio, E. Biotti, and J. G. Goncalves. "Hybrid 3D Reconstruction and Image-Based Rendering Techniques for Reality Modeling". In: *Videometrics and Optical Methods for 3D Shape Measurement*. Vol. 4309. SPIE. 2000, pp. 126–136 (cit. on p. 23).

[131] L. S. Nyland, A. A. Lastra, D. K. McAllister, V. Popescu, C. Mccue, and H. Fuchs. "Capturing, Processing, and Rendering Real-World Scenes". In: *Videometrics and Optical Methods for 3D Shape Measurement*. Vol. 4309. SPIE. 2000, pp. 107–116 (cit. on p. 23).

[132] A. Bogomjakov, C. Gotsman, and M. Magnor. "Free-Viewpoint Video From Depth Cameras". In: *Proc. Vision, Modeling and Visualization (VMV)*. Citeseer. 2006, pp. 89–96 (cit. on p. 23).

[133] A. Maimone and H. Fuchs. "Encumbrance-Free Telepresence System With Real-Time 3D Capture and Display Using Commodity Depth Cameras". In: *2011 10th IEEE International Symposium on Mixed and Augmented Reality*. IEEE. 2011, pp. 137–146 (cit. on p. 23).

[134] A. Maimone and H. Fuchs. "A First Look at a Telepresence System With Room-Sized Real-Time 3D Capture and Life-Sized Tracked Display Wall". In: *Proceedings of ICAT 2011, to appear* (2011), pp. 4–9 (cit. on p. 23).

[135] M. Dou, Y. Shi, J.-M. Frahm, H. Fuchs, B. Mauchly, and M. Marathe. "Room-Sized Informal Telepresence System". In: *2012 IEEE Virtual Reality Workshops (VRW)*. IEEE. 2012, pp. 15–18 (cit. on p. 23).

[136] T. Pejsa, J. Kantor, H. Benko, E. Ofek, and A. Wilson. "Room2room: Enabling Life-Size Telepresence in a Projected Augmented Reality Environment". In: *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. 2016, pp. 1716–1725 (cit. on p. 23).

[137] S. Beck, A. Kunert, A. Kulik, and B. Froehlich. "Immersive Group-to-Group Telepresence". In: *IEEE transactions on visualization and computer graphics* 19.4 (2013), pp. 616–625 (cit. on pp. 23, 26).

[138] Y.-W. Cha, M. Dou, R. Chabra, F. Menozzi, E. Wallen, H. Fuchs, et al. "Immersive Learning Experiences for Surgical Procedures". In: *Medicine Meets Virtual Reality 22*. IOS Press, 2016, pp. 55–62 (cit. on p. 23).

[139] S. Izadi, D. Kim, O. Hilliges, et al. "Kinectfusion: Real-Time 3D Reconstruction and Interaction Using a Moving Depth Camera". In: *Proceedings of the 24th annual ACM symposium on User interface software and technology*. 2011, pp. 559–568 (cit. on p. 23).

[140] M. Dou, L. Guan, J.-M. Frahm, and H. Fuchs. "Exploring High-Level Plane Primitives for Indoor 3D Reconstruction With a Hand-Held RGB-D Camera". In: *Computer Vision-ACCV 2012 Workshops: ACCV 2012 International Workshops, Daejeon, Korea, November 5-6, 2012, Revised Selected Papers, Part II 11*. Springer. 2013, pp. 94–108 (cit. on p. 23).

[141] A. Maimone and H. Fuchs. "Real-Time Volumetric 3D Capture of Room-Sized Scenes for Telepresence". In: *2012 3DTV-conference: the true vision-capture, transmission and display of 3D video (3DTV-CON)*. IEEE. 2012, pp. 1–4 (cit. on p. 23).

[142] H. RJ. "Head Mounted Electrocular Display: A New Display Concept for Specialized Environments." In: *Aerospace Medicine* 34 (1963), pp. 316–318 (cit. on p. 23).

[143] A. Maimone, X. Yang, N. Dierk, A. State, M. Dou, and H. Fuchs. “General-Purpose Telepresence With Head-Worn Optical See-Through Displays and Projector-Based Lighting”. In: *2013 IEEE Virtual Reality (VR)*. IEEE. 2013, pp. 23–26 (cit. on p. 23).

[144] J. J. Leonard and H. F. Durrant-Whyte. “Simultaneous Map Building and Localization for an Autonomous Mobile Robot.” In: *IROS*. Vol. 3. 1991, pp. 1442–1447 (cit. on p. 24).

[145] S. Orts-Escolano, C. Rhemann, S. Fanello, et al. “Holoportation: Virtual 3D Teleportation in Real-Time”. In: *Proceedings of the 29th annual symposium on user interface software and technology*. 2016, pp. 741–754 (cit. on p. 24).

[146] P. Stotko, S. Krumpen, M. B. Hullin, M. Weinmann, and R. Klein. “Slamcast: Large-Scale, Real-Time 3D Reconstruction and Streaming for Immersive Multi-Client Live Telepresence”. In: *IEEE transactions on visualization and computer graphics* 25.5 (2019), pp. 2102–2112 (cit. on p. 24).

[147] B. Thoravi Kumaravel, F. Anderson, G. Fitzmaurice, B. Hartmann, and T. Grossman. “Loki: Facilitating Remote Instruction Of Physical Tasks Using Bi-directional Mixed-reality Telepresence”. In: *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 2019, pp. 161–174 (cit. on p. 24).

[148] N. Weibel, D. Gasques, J. Johnson, et al. “Artemis: Mixed-Reality Environment for Immersive Surgical Telementoring”. In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–4 (cit. on pp. 24, 80).

[149] M. Rebol, K. Pietroszek, C. Ranniger, et al. “Mixed Reality Communication for Medical Procedures: Teaching the Placement of a Central Venous Catheter”. In: *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2022, pp. 346–354 (cit. on pp. 24, 80).

[150] A. Hale, M. Fischer, L. Schütz, H. Fuchs, and C. Leuze. “Remote Training for Medical Staff in Low-Resource Environments Using Augmented Reality”. In: *Journal of Imaging* 8.12 (2022), p. 319 (cit. on p. 24).

[151] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer. “In Touch With The Remote World: Remote Collaboration With Augmented Reality Drawings And Virtual Navigation”. In: *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*. 2014, pp. 197–205 (cit. on pp. 24, 25).

[152] J. Young, T. Langlotz, S. Mills, and H. Regenbrecht. “Mobileportation: Nomadic Telepresence for Mobile Devices”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4.2 (2020), pp. 1–16 (cit. on p. 24).

[153] J. Young, T. Langlotz, M. Cook, S. Mills, and H. Regenbrecht. “Immersive Telepresence and Remote Collaboration Using Mobile and Wearable Devices”. In: *IEEE transactions on visualization and computer graphics* 25.5 (2019), pp. 1908–1918 (cit. on p. 24).

[154] T. Rhee, S. Thompson, D. Medeiros, R. Dos Anjos, and A. Chalmers. “Augmented Virtual Teleportation for High-Fidelity Telecollaboration”. In: *IEEE transactions on visualization and computer graphics* 26.5 (2020), pp. 1923–1933 (cit. on pp. 24, 25).

[155] T. Teo, G. A. Lee, M. Billinghurst, and M. Adcock. “360Drops: Mixed Reality Remote Collaboration Using 360 Panoramas Within the 3D Scene”. In: *SIGGRAPH Asia 2019 Emerging Technologies*. 2019, pp. 1–2 (cit. on p. 24).

[156] Z. Szalavári, D. Schmalstieg, A. Fuhrmann, and M. Gervautz. ““Studierstube”: An Environment for Collaboration in Augmented Reality”. In: *Virtual Reality* 3.1 (1998), pp. 37–48 (cit. on p. 25).

[157] M. Billinghurst and H. Kato. “Collaborative Mixed Reality”. In: *Proceedings of the first international symposium on mixed reality*. 1999, pp. 261–284 (cit. on p. 25).

[158] 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 p. 25).

[159] H. Elsayed, M. D. Barrera Machuca, C. Schaarschmidt, et al. "VRSketchPen: Unconstrained Haptic Assistance for Sketching in Virtual 3D Environments". In: *Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology*. 2020, pp. 1–11 (cit. on p. 25).

[160] R. Arora, R. H. Kazi, F. Anderson, T. Grossman, K. Singh, and G. W. Fitzmaurice. "Experimental Evaluation of Sketching on Surfaces in VR." In: *CHI*. Vol. 17. 2017, pp. 5643–5654 (cit. on p. 25).

[161] E. Wiese, J. H. Israel, A. Meyer, and S. Bongartz. "Investigating the Learnability of Immersive Free-Hand Sketching". In: *Proceedings of the seventh sketch-based interfaces and modeling symposium*. 2010, pp. 135–142 (cit. on p. 25).

[162] M. D. B. Machuca, P. Asente, W. Stuerzlinger, J. Lu, and B. Kim. "Multiplanes: Assisted Freehand VR Sketching". In: *Proceedings of the 2018 ACM Symposium on Spatial User Interaction*. 2018, pp. 36–47 (cit. on p. 25).

[163] M. D. B. Machuca, W. Stuerzlinger, and P. Asente. "Smart3DGuides: Making Unconstrained Immersive 3D Drawing More Accurate". In: *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*. 2019, pp. 1–13 (cit. on p. 25).

[164] Q. Yu, Y. Yang, F. Liu, Y.-Z. Song, T. Xiang, and T. M. Hospedales. "Sketch-a-Net: A Deep Neural Network That Beats Humans". In: *International journal of computer vision* 122 (2017), pp. 411–425 (cit. on p. 25).

[165] Y. S. Chang, B. Nuernberger, B. Luan, and T. Höllerer. "Evaluating Gesture-Based Augmented Reality Annotation". In: *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE. 2017, pp. 182–185 (cit. on p. 25).

[166] J. Zillner, E. Mendez, and D. Wagner. "Augmented Reality Remote Collaboration With Dense Reconstruction". In: *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE. 2018, pp. 38–39 (cit. on p. 25).

[167] M. Adcock, S. Anderson, and B. Thomas. "RemoteFusion: Real Time Depth Camera Fusion for Remote Collaboration on Physical Tasks". In: *Proceedings of the 12th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry*. 2013, pp. 235–242 (cit. on p. 25).

[168] S. Lukosch, M. Billinghurst, L. Alem, and K. Kiyokawa. "Collaboration in Augmented Reality". In: *Computer Supported Cooperative Work (CSCW)* 24 (2015), pp. 515–525 (cit. on p. 25).

[169] 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*. 2015, pp. 405–415 (cit. on pp. 25, 26, 80).

[170] S. Utzig, R. Kaps, S. M. Azeem, and A. Gerndt. "Augmented Reality for Remote Collaboration in Aircraft Maintenance Tasks". In: *2019 IEEE aerospace conference*. IEEE. 2019, pp. 1–10 (cit. on p. 25).

[171] H. Tian, G. A. Lee, H. Bai, and M. Billinghurst. "Using Virtual Replicas to Improve Mixed Reality Remote Collaboration". In: *IEEE Transactions on Visualization and Computer Graphics* (2023) (cit. on p. 25).

[172] T. Piumsomboon, G. A. Lee, J. D. Hart, et al. "Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration". In: *Proceedings of the 2018 CHI conference on human factors in computing systems*. 2018, pp. 1–13 (cit. on pp. 25, 34).

[173] T. Piumsomboon, G. A. Lee, A. Irlitti, B. Ens, B. H. Thomas, and M. Billinghurst. "On the Shoulder of the Giant: A Multi-Scale Mixed Reality Collaboration With 360 Video Sharing and Tangible Interaction". In: *Proceedings of the 2019 CHI conference on human factors in computing systems*. 2019, pp. 1–17 (cit. on pp. 25, 34).

[174] M. E. Walker, D. Szafir, and I. Rae. "The Influence of Size in Augmented Reality Telepresence Avatars". In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2019, pp. 538–546 (cit. on pp. 25, 34, 87).

[175] T. Piumsomboon, A. Dey, B. Ens, G. Lee, and M. Billinghurst. “The Effects of Sharing Awareness Cues in Collaborative Mixed Reality”. In: *Frontiers in Robotics and AI* 6 (2019), p. 5 (cit. on p. 25).

[176] H. Bai, P. Sasikumar, J. Yang, and M. Billinghurst. “A User Study on Mixed Reality Remote Collaboration With Eye Gaze and Hand Gesture Sharing”. In: *Proceedings of the 2020 CHI conference on human factors in computing systems*. 2020, pp. 1–13 (cit. on p. 25).

[177] A. Erickson, N. Norouzi, K. Kim, et al. “Sharing Gaze Rays for Visual Target Identification Tasks in Collaborative Augmented Reality”. In: *Journal on Multimodal User Interfaces* 14.4 (2020), pp. 353–371 (cit. on p. 25).

[178] L. Berger and K. Wolf. “WIM: Fast Locomotion in Virtual Reality With Spatial Orientation Gain & Without Motion Sickness”. In: *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*. 2018, pp. 19–24 (cit. on p. 26).

[179] D. Englmeier, W. Sajko, and A. Butz. “Spherical World in Miniature: Exploring the Tiny Planets Metaphor for Discrete Locomotion in Virtual Reality”. In: *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2021, pp. 345–352 (cit. on p. 26).

[180] D. Valkov, F. Steinicke, G. Bruder, and K. H. Hinrichs. “Traveling in 3D Virtual Environments With Foot Gestures and a Multi-Touch Enabled WiM”. In: *Proceedings of virtual reality international conference (VRIC 2010)*. 2010, pp. 171–180 (cit. on p. 26).

[181] F. Bacim, D. Bowman, and M. Pinho. “Wayfinding Techniques for Multiscale Virtual Environments”. In: *2009 IEEE Symposium on 3D User Interfaces*. IEEE. 2009, pp. 67–74 (cit. on p. 26).

[182] S. Truman and S. von Mammen. “An Integrated Design of World-in-Miniature Navigation in Virtual Reality”. In: *International Conference on the Foundations of Digital Games*. 2020, pp. 1–9 (cit. on p. 26).

[183] V. Chheang, F. Heinrich, F. Joeres, et al. “WiM-Based Group Navigation for Collaborative Virtual Reality”. In: *2022 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE. 2022, pp. 82–92 (cit. on p. 26).

[184] A. Mulloni, H. Seichter, and D. Schmalstieg. “Indoor Navigation With Mixed Reality World-in-Miniature Views and Sparse Localization on Mobile Devices”. In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*. 2012, pp. 212–215 (cit. on p. 26).

[185] C. A. Wingrave, Y. Haciahmetoglu, and D. A. Bowman. “Overcoming World in Miniature Limitations by a Scaled and Scrolling WIM”. In: *3D User Interfaces (3DUI’06)*. IEEE. 2006, pp. 11–16 (cit. on p. 26).

[186] R. Trueba, C. Andujar, and F. Argelaguet. “Multi-Scale Manipulation in Indoor Scenes With the World in Miniature Metaphor”. In: *Proceedings of the 15th Joint virtual reality Eurographics conference on Virtual Environments*. 2009, pp. 93–100 (cit. on p. 26).

[187] R. Trueba, C. Andujar, and F. Argelaguet. “Complexity and Occlusion Management for the World-in-Miniature Metaphor”. In: *International Symposium on Smart Graphics*. Springer. 2009, pp. 155–166 (cit. on p. 26).

[188] C. Andujar, F. Argelaguet, and R. Trueba. “Hand-Based Disocclusion for the World-in-Miniature Metaphor”. In: *Presence: Teleoperators and Virtual Environments* 19.6 (2010), pp. 499–512 (cit. on p. 26).

[189] T. T. H. Nguyen, C. Fleury, and T. Duval. “Collaborative Exploration in a Multi-Scale Shared Virtual Environment”. In: *2012 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE. 2012, pp. 181–182 (cit. on p. 26).

[190] H. Van der Valk, H. Haße, F. Möller, M. Arbter, J.-L. Henning, and B. Otto. “A Taxonomy of Digital Twins.” In: *AMCIS*. 2020 (cit. on p. 26).

[191] T. Piumsomboon, G. A. Lee, and M. Billinghurst. “Snow Dome: A Multi-Scale Interaction in Mixed Reality Remote Collaboration”. In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–4 (cit. on p. 27).

[192] H. Ibayashi, Y. Sugiura, D. Sakamoto, et al. “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. 27).

[193] G. Speaight. “Professor Pepper’s Ghost”. In: *Theatre Notebook* 43.1 (1989), pp. 16–24 (cit. on p. 27).

[194] J. Wind, K. Riege, and M. Bogen. “Spinnstube®: A Seated Augmented Reality Display System”. In: *Proceedings of the 13th Eurographics conference on Virtual Environments*. 2007, pp. 17–23 (cit. on p. 27).

[195] G. Stetten, D. Shelton, W. Chang, et al. “Towards a Clinically Useful Sonic Flashlight”. In: *Proceedings IEEE International Symposium on Biomedical Imaging*. IEEE. 2002, pp. 417–420 (cit. on p. 27).

[196] 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. 27, 28).

[197] H. Sato, I. Kitahara, and Y. Ohta. “MR-Mirror: A Complex of Real and Virtual Mirrors”. In: *Virtual and Mixed Reality: Third International Conference, VMR 2009, Held as Part of HCI International 2009, San Diego, CA, USA, July 19-24, 2009. Proceedings 3*. Springer. 2009, pp. 482–491 (cit. on pp. 28, 29).

[198] 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. 28, 29, 97).

[199] J. Fotouhi, T. Song, A. Mehrfard, 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. 27, 28).

[200] E. Veas, R. Grasset, E. Kruijff, and D. Schmalstieg. “Extended Overview Techniques for Outdoor Augmented Reality”. In: *IEEE transactions on visualization and computer graphics* 18.4 (2012), pp. 565–572 (cit. on p. 27).

[201] K. Lilija, H. Pohl, S. Boring, and K. Hornbæk. “Augmented Reality Views for Occluded Interaction”. In: *Proceedings of the 2019 CHI conference on human factors in computing systems*. 2019, pp. 1–12 (cit. on p. 27).

[202] J. E. Grønbæk, B. Saatçi, C. F. Griggio, and C. N. Klokmose. “MirrorBlender: Supporting Hybrid Meetings With a Malleable Video-Conferencing System”. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, pp. 1–13 (cit. on p. 27).

[203] H. Schulzrinne, A. Rao, and R. Lanphier. *Real Time Streaming Protocol (RTSP)*. Tech. rep. 1998 (cit. on p. 30).

[204] R. Chabra, A. Ilie, N. Rewkowski, Y.-W. Cha, and H. Fuchs. “Optimizing Placement of Commodity Depth Cameras for Known 3D Dynamic Scene Capture”. In: *2017 IEEE Virtual Reality (VR)*. IEEE. 2017, pp. 157–166 (cit. on p. 30).

[205] E. Hörster and R. Lienhart. “On the Optimal Placement of Multiple Visual Sensors”. In: *Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks*. 2006, pp. 111–120 (cit. on p. 30).

[206] S. Agarwal, K. Mierle, and The Ceres Solver Team. *Ceres Solver*. Version 2.1. Mar. 2022 (cit. on p. 31).

[207] D.-M. Pham and W. Stuerzlinger. “Is the Pen Mightier Than the Controller? A Comparison of Input Devices for Selection in Virtual and Augmented Reality”. In: *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*. 2019, pp. 1–11 (cit. on p. 32).

[208] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. E. Latoschik. “The Impact of Avatar Personalization and Immersion on Virtual Body Ownership, Presence, and Emotional Response”. In: *IEEE transactions on visualization and computer graphics* 24.4 (2018), pp. 1643–1652 (cit. on p. 33).

[209] J.-L. Lugrin, J. Latt, and M. E. Latoschik. “Avatar Anthropomorphism and Illusion of Body Ownership in VR”. In: *2015 IEEE virtual reality (VR)*. IEEE. 2015, pp. 229–230 (cit. on p. 33).

[210] P. Heidicker, E. Langbehn, and F. Steinicke. “Influence of Avatar Appearance on Presence in Social VR”. In: *2017 IEEE symposium on 3D user interfaces (3DUI)*. IEEE. 2017, pp. 233–234 (cit. on p. 33).

[211] K. L. Nowak and F. Biocca. “The Effect of the Agency and Anthropomorphism on Users’ Sense of Telepresence, Copresence, and Social Presence in Virtual Environments”. In: *Presence: Teleoperators & Virtual Environments* 12.5 (2003), pp. 481–494 (cit. on p. 33).

[212] P. Henry Fuchs. “3D Telepresence for Medical Consultation: Extending Medical Expertise Throughout, Between and Beyond Hospitals”. In: (2007) (cit. on p. 33).

[213] D. H. Sonnenwald, H. M. Söderholm, J. E. Manning, B. Cairns, G. Welch, and H. Fuchs. “Exploring the potential of video technologies for collaboration in emergency medical care: Part I. Information sharing”. In: *Journal of the American Society for Information Science and Technology* 59.14 (2008), pp. 2320–2334 (cit. on p. 33).

[214] N. Ogawa, T. Narumi, H. Kuzuoka, and M. Hirose. “Do You Feel Like Passing Through Walls?: Effect of Self-Avatar Appearance on Facilitating Realistic Behavior in Virtual Environments”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–14 (cit. on p. 33).

[215] E. Langbehn, E. Harting, and F. Steinicke. “Shadow-Avatars: A Visualization Method To Avoid Collisions of Physically Co-Located Users in Room-Scale VR”. In: *IEEE Workshop on Everyday Virtual Reality*. 2018 (cit. on p. 33).

[216] M. Wirth, S. Gradl, G. Prosinger, F. Kluge, D. Roth, and B. M. Eskofier. “The Impact of Avatar Appearance, Perspective and Context on Gait Variability and User Experience in Virtual Reality”. In: *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2021, pp. 326–335 (cit. on p. 33).

[217] E. Ebrahimi, L. S. Hartman, A. Robb, C. C. Pagano, and S. V. Babu. “Investigating the Effects of Anthropomorphic Fidelity of Self-Avatars on Near Field Depth Perception in Immersive Virtual Environments”. In: *2018 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE. 2018, pp. 1–8 (cit. on p. 33).

[218] N. Yee and J. Bailenson. “The Proteus Effect: The Effect of Transformed Self-Representation on Behavior”. In: *Human communication research* 33.3 (2007), pp. 271–290 (cit. on pp. 34, 89).

[219] G. Brown and M. Prilla. “The Effects of Consultant Avatar Size and Dynamics on Customer Trust in Online Consultations”. In: *Proceedings of Mensch und Computer 2020*. 2020, pp. 239–249 (cit. on p. 34).

[220] S. Kang, B. Yoon, B. Kim, and W. Woo. “Effects of Avatar Face Level of Detail Control on Social Presence in Augmented Reality Remote Collaboration”. In: *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE. 2022, pp. 763–767 (cit. on p. 34).

[221] Y. Choi, J. Lee, and S.-H. Lee. “Effects of Locomotion Style and Body Visibility of a Telepresence Avatar”. In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2020, pp. 1–9 (cit. on p. 34).

[222] H. M. Söderholm, D. H. Sonnenwald, B. Cairns, J. E. Manning, G. F. Welch, and H. Fuchs. “The Potential Impact of 3D Telepresence Technology on Task Performance in Emergency Trauma Care”. In: *Proceedings of the 2007 ACM International Conference on Supporting Group Work*. 2007, pp. 79–88 (cit. on p. 35).

[223] H. Alvis-Miranda, S. M. Castellar-Leones, and L. R. Moscote-Salazar. “Decompressive Craniectomy and Traumatic Brain Injury: A Review”. In: *Bulletin of Emergency & Trauma* 1.2 (2013), p. 60 (cit. on p. 39).

[224] R. Jabbarli, S.-Q. He, M. Darkwah Oppong, et al. “Size Does Matter: The Role of Decompressive Craniectomy Extent for Outcome After Aneurysmal Subarachnoid Hemorrhage”. In: *European Journal of Neurology* 28.7 (2021), pp. 2200–2207 (cit. on p. 40).

[225] S. Schur, P. Martel, and J. Marcoux. “Optimal Bone Flap Size for Decompressive Craniectomy for Refractory Increased Intracranial Pressure in Traumatic Brain Injury: Taking the Patient’s Head Size Into Account”. In: *World Neurosurgery* 137 (2020), e430–e436 (cit. on p. 40).

[226] H. Çavuşoğlu, R. A. Kaya, O. N. Türkmenoğlu, and Y. Aydin. “Value of Early Unilateral Decompressive Craniectomy in Patients With Severe Traumatic Brain Injury”. In: *Ulus Travma Acil Cerrahi Derg* 16.2 (2010), pp. 119–124 (cit. on p. 40).

[227] M. R. Stephens, W. G. Lewis, A. Brewster, et al. “Multidisciplinary Team Management Is Associated With Improved Outcomes After Surgery for Esophageal Cancer”. In: *Diseases of the Esophagus* 19.3 (2006), pp. 164–171 (cit. on p. 53).

[228] C. P. Ames, J. J. Barry, S. Keshavarzi, O. Dede, M. H. Weber, and V. Deviren. “Perioperative Outcomes and Complications of Pedicle Subtraction Osteotomy in Cases With Single Versus Two Attending Surgeons”. In: *Spine Deformity* 1.1 (2013), pp. 51–58 (cit. on p. 53).

[229] M. J. Davis, B. C. Luu, S. Raj, A. Abu-Ghname, and E. P. Buchanan. “Multidisciplinary Care in Surgery: Are Team-Based Interventions Cost-Effective?” In: *The Surgeon* 19.1 (2021), pp. 49–60 (cit. on p. 53).

[230] R. B. Miller. “Response Time in Man-Computer Conversational Transactions”. In: *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. 1968, pp. 267–277 (cit. on p. 67).

[231] M Suznjevic and J Saldana. “Delay Limits for Real-Time Services”. In: *IETF draft* 6 (2016) (cit. on p. 67).

[232] P. Wu, Y. Liu, M. Ye, J. Li, and S. Du. “Fast and Adaptive 3D Reconstruction With Extensively High Completeness”. In: *IEEE Transactions on Multimedia* 19.2 (2016), pp. 266–278 (cit. on p. 67).

[233] X. Long, C. Lin, P. Wang, T. Komura, and W. Wang. “Sparseneus: Fast Generalizable Neural Surface Reconstruction From Sparse Views”. In: *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXII*. Springer. 2022, pp. 210–227 (cit. on p. 67).

[234] S. Nikoohemat, A. A. Diakité, S. Zlatanova, and G. Vosselman. “Indoor 3D Reconstruction From Point Clouds for Optimal Routing in Complex Buildings To Support Disaster Management”. In: *Automation in construction* 113 (2020), p. 103109 (cit. on p. 67).

[235] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. “Nerf: Representing Scenes As Neural Radiance Fields for View Synthesis”. In: *Communications of the ACM* 65.1 (2021), pp. 99–106 (cit. on p. 84).

[236] M. Slater. “A Note on Presence Terminology”. In: *Presence connect* 3.3 (2003), pp. 1–5 (cit. on p. 85).

[237] S. T. Bulu. “Place Presence, Social Presence, Co-Presence, and Satisfaction in Virtual Worlds”. In: *Computers & Education* 58.1 (2012), pp. 154–161 (cit. on p. 85).

[238] E. Özsoy, E. P. Örnek, U. Eck, T. Czempiel, F. Tombari, and N. Navab. “4D-OR: Semantic Scene Graphs for or Domain Modeling”. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VII*. Springer. 2022, pp. 475–485 (cit. on pp. 86–88).

[239] O. Daly-Jones, A. Monk, and L. Watts. “Some Advantages of Video Conferencing Over High-Quality Audio Conferencing: Fluency and Awareness of Attentional Focus”. In: *International Journal of Human-Computer Studies* 49.1 (1998), pp. 21–58 (cit. on p. 86).

[240] E. Wolf, M. L. Fiedler, N. Döllinger, C. Wienrich, and M. E. Latoschik. “Exploring Presence, Avatar Embodiment, and Body Perception With a Holographic Augmented Reality Mirror”. In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2022, pp. 350–359 (cit. on p. 89).

[241] C. Nimcharoen, S. Zollmann, J. Collins, and H. Regenbrecht. “Is That Me?—Embodiment and Body Perception With an Augmented Reality Mirror”. In: *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE. 2018, pp. 158–163 (cit. on p. 89).

[242] B. Gao, J. Lee, H. Tu, W. Seong, and H. Kim. “The Effects of Avatar Visibility on Behavioral Response With or Without Mirror-Visual Feedback in Virtual Environments”. In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE. 2020, pp. 780–781 (cit. on p. 89).

[243] M. Gonzalez-Franco, D. Perez-Marcos, B. Spanlang, and M. Slater. “The Contribution of Real-Time Mirror Reflections of Motor Actions on Virtual Body Ownership in an Immersive Virtual Environment”. In: *2010 IEEE virtual reality conference (VR)*. IEEE. 2010, pp. 111–114 (cit. on p. 89).

[244] J. Guegan, S. Buisine, F. Mantelet, N. Maranzana, and F. Segonds. “Avatar-Mediated Creativity: When Embodying Inventors Makes Engineers More Creative”. In: *Computers in Human Behavior* 61 (2016), pp. 165–175 (cit. on p. 89).

[245] D. LaBerge and S. J. Samuels. “Toward a Theory of Automatic Information Processing in Reading”. In: *Cognitive psychology* 6.2 (1974), pp. 293–323 (cit. on p. 89).

[246] S. Charron and E. Koechlin. “Divided Representation of Concurrent Goals in the Human Frontal Lobes”. In: *Science* 328.5976 (2010), pp. 360–363 (cit. on p. 89).

[247] B. Wu, R. Klatzky, and G. Stetten. “Exploring Here, Seeing Where? Visualization With In-Situ vs. Ex-Situ Viewing”. In: *Journal of Vision* 8.6 (2008), pp. 1066–1066 (cit. on p. 89).

[248] V. Ng-Thow-Hing, K. Bark, L. Beckwith, C. Tran, R. Bhandari, and S. Sridhar. “User-Centered Perspectives for Automotive Augmented Reality”. In: *IEEE international symposium on mixed and augmented reality*. 2013, pp. 13–22 (cit. on p. 91).

[249] J. Arbeláez, R. Viganò, and G. Osorio-Gómez. “Haptic Augmented Reality (HapticAR) for Assembly Guidance”. In: *International Journal on Interactive Design and Manufacturing (IJIDeM)* 13 (2019), pp. 673–687 (cit. on p. 91).

[250] 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*. 2003, pp. 73–80 (cit. on p. 91).

[251] R. M. Dickey, N. Srikishen, L. I. Lipshultz, P. E. Spiess, R. E. Carrion, and T. S. Hakky. “Augmented Reality Assisted Surgery: A Urologic Training Tool”. In: *Asian journal of andrology* 18.5 (2016), p. 732 (cit. on p. 91).

[252] D. Ostler, J. Steger, L. Bernhard, et al. ““Hybrid” Scientific Conference: Lessons Learned From the Digital Annual Meeting of the CARS International Conference During the COVID-19 Pandemic”. In: *Innovative Surgical Sciences* 6.3 (2021), pp. 115–123 (cit. on p. 97).

[253] K. Yu, D. Ostler, J. Fuchtmann, et al. “Clean-AR: Using Augmented Reality for Reducing the Risk of Contamination from Airborne Disease Agents on Surfaces”. In: *Current Directions in Biomedical Engineering* 7.1 (2021), pp. 6–10 (cit. on p. 97).

[254] J. P. O. Held, K. Yu, C. Pyles, et al. “Augmented Reality-Based Rehabilitation of Gait Impairments: Case Report”. In: *JMIR mHealth and uHealth* 8.5 (2020), e17804 (cit. on p. 97).

[255] R. Hartwig, D. Ostler, H. Feußner, et al. “COMPASS: Localization in Laparoscopic Visceral Surgery”. In: *Current Directions in Biomedical Engineering* 6.1 (2020) (cit. on p. 97).

[256] G. Deib, A. Johnson, M. Unberath, et al. “Image Guided Percutaneous Spine Procedures Using an Optical See-Through Head Mounted Display: Proof of Concept and Rationale”. In: *Journal of neurointerventional surgery* 10.12 (2018), pp. 1187–1191 (cit. on p. 98).

[257] S. Andress, A. Johnson, M. Unberath, et al. “On-the-Fly Augmented Reality for Orthopedic Surgery Using a Multimodal Fiducial”. In: *Journal of Medical Imaging* 5.2 (2018), pp. 021209–021209 (cit. on p. 98).

[258] K. Yu, R. Barmaki, M. Unberath, et al. “On the Accuracy of Low-Cost Motion Capture Systems for Range of Motion Measurements”. In: *Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications*. Vol. 10579. SPIE. 2018, pp. 90–95 (cit. on p. 98).

[259] L. Qian, M. Unberath, K. Yu, et al. “Towards Virtual Monitors for Image Guided Interventions-Real-Time Streaming to Optical See-Through Head-Mounted Displays”. In: *arXiv preprint arXiv:1710.00808* (2017) (cit. on p. 98).

List of Figures

List of Tables