UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

PhD Thesis

Klemen Lilija

Hand-based Interaction in Mixed Reality

Manipulating, Navigating and Learning

Date: September 30th, 2021

Advisor: Kasper Hornbæk

This thesis has been submitted to the PhD School of The Faculty of Science, University of Copenhagen

Hand-based Interaction in Mixed Reality

Author
Klemen Lilija
Department of Computer Science, University of Copenhagen

Title
Mixed-Reality User Interfaces

Date submitted
September 30th, 2021

Date defended
...

Advisor
Kasper Anders Søren Hornbæk, Professor, University of Copenhagen

Assessment Committee
Daniel Lee Ashbrook, Associate Professor, University of Copenhagen
Stefania Serafin, Professor, Aalborg University
Christian Holz, Assistant Professor, ETH Zurich

Abstract

Mixed reality has promised us better ways of interacting with physical and virtual worlds. Increased productivity, natural communication, on-demand user interfaces, immersive storytelling, and dream-like experiences. Until now, few of these promises have come to fruition or been widely adopted. This is partly because of limitations of the current hardware and partly due to a lack of useful interaction techniques. This thesis focuses on the latter. I study mixed reality interaction techniques that help users manipulate objects, navigate virtual worlds and support them while learning new skills. The techniques are motivated by the wish to involve more of our bodies and expose the spatial capabilities of mixed reality. Since hands are one of the most dexterous parts of our bodies, I based all of the presented techniques on them.

The thesis investigates hand-based interaction techniques for mixed reality through four papers. Paper 1 investigates how to best support free-hand manipulation of occluded objects with the help of augmented reality. We compare four techniques for re-enabling visual guidance and find that a see-through view and a displaced three-dimensional view of the occluded objects perform best. Paper 2 explores the re-configuration of virtual environments for enabling hand-based interactions at a distance. It combines two familiar metaphors (world-in-miniatures and portals) to support a range of techniques, including interaction at scale and occluded interaction. Paper 3 explores navigation of spatial recordings through direct manipulation of objects in the scene. Users can, for example, move an object to its past or future location to navigate to a specific moment in time while staying present and engaged in the recording. Paper 4 investigates a technique for supporting motor skills acquisition. The technique corrects the users' virtual hand movements during training to nudge them towards more accurate movements. The four papers use a mix of quantitative and qualitative methods for evaluating the techniques. We conduct lab (Paper 1) and remote (Paper 4) experiments and use expert reviews (Paper 2) and the think-aloud method (Paper 3). The tasks and applications are grounded in daily-life scenarios to make them understandable for a wide range of participants in our evaluations.

Combined, the papers present several hand-based interaction techniques for improving manipulation, navigation and learning in mixed reality. These techniques allow users to use greater capability of their hands and interact with the virtual and real world around them more fully.

Dansk Resumé

"Mixed reality"-teknologi har lovet os bedre måder at interagere med fysiske og virtuelle verdener. Forøget produktivitet, naturlig kommunikation, on-demand brugergrænseflader, indlevende historiefortælling, og drømmeagtige oplevelser. Indtil nu er få af disse løfter blevet indfriet eller udbredt. Dette er delvist på grund af begrænsninger i den nuværende hardware og delvist på grund af en mangel på nyttige interaktionsteknikker. Denne afhandling fokuserer på sidstnævnte. Jeg studerer mixed reality interaktionsteknikker som hjælper brugere med at manipulere objekter, navigere gennem virtuelle verdener, og støtter dem i at indlære nye færdigheder. Teknikkerne er motiveret af et ønske om at involvere mere af vores kroppe og fremhæve de rumlige egenskaber i mixed reality. Siden hænder er én af de mest fleksible dele af vores kroppe, baserer jeg alle de præsenterede teknikker på brugen af hænder.

Afhandlingen undersøger håndbaserede interaktionsteknikker til mixed reality gennem fire artikler. Artikel 1 undersøger hvordan man bedst kan understøtte frihåndsmanipulation af objekter, som er ude af syne bag andre objekter, gennem brugen af augmented reality. Vi sammenligner fire teknikker til at genetablere visuel kontakt og finder at et gennemsigtigt perspektiv og et forrykket tredimensionelt perspektiv klarer sig bedst. Artikel 2 undersøger omkonfigureringen af virtuelle omgivelser for at tillade interaktion på afstand. Den kombinerer to kendte metaforer (world-in-miniature og portaler) for at understøtte en række teknikker, herunder skaleringsbaseret interaktion og interaktion som foregår ude af syne bag objekter. Artikel 3 undersøger navigation af rumlige optagelser gennem direkte manipulation af objekter i en scene. Brugere kan, for eksempel, flytte et objekt til dets forrige eller fremtidige lokation for at navigere til at specifikt tidspunkt, mens de oplever at være til stede og engageret i optagelsen. Artikel 4 undersøger en teknik til at understøtte indlæring af motorik. Teknikken tilretter brugerens virtuelle håndbevægelser under træning for at skubbe dem mod mere præcise bevægelser. De fire artikler bruger et miks af kvantitative og kvalitative metoder til at evaluere teknikkerne. Vi foretager laboratorie- (Artikel 1) og fjerneksperimenter (Artikel 4) og bruger ekspertevaluering (Artikel 2) samt tænk-højt metoder (Artikel 3). Opgaverne og applikationerne er baseret på hverdagsscenarier for at gøre dem forståelige for et bredt udsnit af deltagere i vores evalueringer.

Samlet præsenterer artiklerne adskillige håndbaserede interaktionsteknikker for at forbedre manipulation, navigation, og indlæring i mixed reality. Disse teknikker tillader brugere at bruge deres hænder bedre og interagere mere fuldbyrdigt med den virtuelle og fysiske verden omkring dem.

Acknowledgements

I would like to thank my advisor Kasper Hornbæk for giving me the opportunity to do curiosity-driven research. By no means did he have to choose me for the open PhD position or be as patient as he was with the scattered ideas I approached. I am happy that he did and am grateful for his help over the course of my PhD.

I would also like to thank my collaborators Henning Pohl, Jess McIntosh, Sebastian Boring and Søren Kyllingsbæk that made the papers of my thesis possible and helped me stay level-headed when the going got tough.

In addition to all of the mentioned above there are several others that made my PhD years as enjoyable as they were and gave me memories and friendships that I will cherish for a lifetime. Jonas Schjerlund, Aske Mottelson, Carlos Tejada, Xiaoyi Wang, Tor-Salve Dalsgaard, Andreea-Anamaira Muresan, Joanna Bergström, Jarrod Knibbe, Daniel Ashbrook and Paul Strohmeier are some of them.

I would never come close to starting or finishing my PhD if it was not for the unconditional support of my family. A thank you to my parents, grandparents, and sister for always giving me peace of mind on whichever road I have chosen. Most importantly, a thank you to Anama and Niko for bearing with me while writing this thesis and for giving me happiness in the world outside of human-computer interaction.

The research presented in this thesis has been funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program, grant agreement 648785.

1. Publications

[69] Klemen Lilija, Henning Pohl, Sebastian Boring, and Kasper Hornbæk. Augmented Reality Views for Occluded Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, New York, NY, USA. Association for Computing Machinery, 2019

[113] Henning Pohl, Klemen Lilija, Jess McIntosh, and Kasper Hornbæk. Poros: Configurable Proxies for Distant Interactions in VR. in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, Yokohama, Japan. Association for Computing Machinery, 2021. DOI: 10.1145/3411764.3445685

[71] Klemen Lilija, Henning Pohl, and Kasper Hornbæk. *Who Put That There? Temporal Navigation of Spatial Recordings by Direct Manipulation*. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2020, pages 1–11

[68] Klemen Lilija, Søren Kyllingsbæk, and Kasper Hornbæk. Correction of Avatar Hand Movements Supports Learning of a Motor Skill. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 1–8, 2021. DOI: 10.1109/VR50410.2021.00069

Part I.

Introduction

2. Thesis Overview

2.1. Structure of the Thesis

This thesis is split into three chapters: introduction, papers, and conclusion. In the introduction (current chapter), I give a brief overview of the thesis and papers I have published during the four years of my PhD. Here I also introduce the topics of *mixed reality* and *interaction techniques*. In the *papers* chapter, I compile the four published papers I authored or co-authored during my PhD. The papers are included exactly as they were published, except for slight changes in formatting to fit the format of this thesis. In the *conclusion*, I discuss the topics that connect the four papers and ideas that were not mentioned in the individual papers. I conclude the thesis by touching again on mixed reality and interaction techniques in general and present a vision for future work.

2.2. Overview of Papers

Within my work, I covered a broad scope of hand-based interaction techniques for mixed reality. *Paper 1* focused on improving the efficiency of manipulating physical objects that are out of sight but within the user's reach. We investigated ways of visualizing the user's hand and occluded objects with the help of augmented reality. *Paper 2* focused on enabling a range of hand-based interactions in virtual reality. We explored proxies that allowed users to reconfigure their environment and reach distant or occluded places through them. *Paper 3* explored temporal navigation of spatial recordings through manipulation of virtual objects in the scene. The users could move through time by moving objects to their past and future states. *Paper 4* explored a technique for guiding users' hands during training by correcting their movement. We investigated if such a technique offers lasting improvements to the trained movement.

2.3. Paper Abstracts

2.3.1. Paper 1: Augmented Reality Views for Occluded Interaction

We rely on our sight when manipulating objects. When objects are occluded, manipulation becomes difficult. Such occluded objects can be shown via augmented reality to re-enable visual guidance. However, it is unclear how to do so to best support object manipulation. We compare four views of occluded objects and their effect on performance and satisfaction across a set of everyday manipulation tasks of varying complexity. The best performing views were a see-through view and a displaced 3D view. The former enabled participants to observe the manipulated object through the occluder, while the latter showed the 3D view of the manipulated object offset from the object's real location. The worst performing view showed remote imagery from a simulated hand-mounted camera. Our results suggest that alignment of virtual objects with their real-world location is less important than an appropriate point-of-view and view stability.

2.3.2. Paper 2: Poros: Configurable Proxies for Distant Interactions in VR

A compelling property of virtual reality is that it allows users to interact with objects as they would in the real world. However, such interactions are limited to space within reach. We present *Poros*, a system that allows

users to rearrange space. After marking a portion of space, the distant *marked space* is mirrored in a nearby *proxy*. Thereby, users can arrange what is within their reachable space, making it easy to interact with multiple distant spaces as well as nearby objects. Proxies themselves become part of the scene and can be moved, rotated, scaled, or anchored to other objects. Furthermore, they can be used in a set of higher-level interactions such as alignment and action duplication. We show how Poros enables a variety of tasks and applications and also validate its effectiveness through an expert evaluation.

2.3.3. Paper 3: Who Put That There? Temporal Navigation of Spatial Recordings by Direct Manipulation

Spatial recordings allow viewers to move within them and freely choose their viewpoint. However, such recordings make it easy to miss events and difficult to follow moving objects when skipping through the recording. To alleviate these problems we present the Who Put That There system that allows users to navigate through time by directly manipulating objects in the scene. By selecting an object, the user can navigate to moments where the object changed. Users can also view trajectories of objects that changed location and directly manipulate them to navigate. We evaluated the system with a set of sensemaking questions in a think-aloud study. Participants understood the system and found it useful for finding events of interest, while being present and engaged in the recording.

2.3.4. Paper 4: Correction of Avatar Hand Movements Supports Learning of a Motor Skill

Learning to move the hands in particular ways is essential in many training and leisure virtual reality applications, yet challenging. Existing techniques that support learning of motor movement in virtual reality rely on external cues such as arrows showing where to move or transparent hands showing the target movement. We propose a technique where the avatar's hand movement is corrected to be closer to the target movement. This embeds guidance in the user's avatar, instead of in external cues, and minimizes visual distraction. Through two experiments, we found that such movement guidance improves the short-term retention of the target movement when compared to a control condition without guidance.

3. Mixed Reality

3.1. What is Mixed Reality?

The term *mixed reality* (MR) describes technologies that can substantially alter users' perception. Such technologies can create an illusion of seeing or touching three-dimensional objects where there are none or an illusion of being elsewhere than in the surrounding physical space. Such illusions are commonly created by monitoring the users' movement and generating stimuli that manipulate the users' senses. For example, an MR headset tracks the user's head and generates the light that would fall on the user's retina if the simulated object was in the room. Similar illusions can be created for other senses by generating auditory, haptic, olfactory and gustatory stimuli.

There are many terms related to MR that were given to similar technologies, systems, techniques and experiences. *Augmented reality*, *augmented virtuality*, *artificial reality*, *virtualized reality*, *virtual reality*, *mediated reality*, *multimediated reality*, *substitutional reality*, *diminished reality*, *modulated reality* and *extended reality* are some of them. These terms often overlap, lack a precise definition, or are conveniently redefined to fit the context in which they are used. Many researchers and practitioners are therefore inclined to use an umbrella term such as *extended reality* (XR) to avoid the messiness of concepts referring to various aspects of computer-enabled experiences. Despite the messiness, the conceptualization of MR (or XR) is not a pointless intellectual exercise. Through the conceptualization, we gain a better understanding of the used technologies and the experiences they enable. Conceptualization can also help us articulate unresolved problems, generate new ideas, and design solutions for existing problems. Virtual reality (VR) and augmented reality (AR) were the first terms that gained wide acceptance. The two terms categorized technologies in ones where the physical world that surrounds the user is combined with the virtual content (AR), and the ones where only the virtual content is essential and the physical surrounding is ignored (VR). These terms did not come into being because of a careful consideration of the semantics of the words and their suitability for building theory, taxonomies or descriptions of a design space. Often it came down to the personal taste of MR pioneers who, for example, liked the world "virtual" more than "synthetic" and tried to avoid terms that are too "nerdy", "techy" or "computery" (e.g., see 1980s interview with Jaron Lanier [50]). These terms were accepted by academic researchers and sparked a lively conversation that helped explore the possibilities of such technologies. In the 1990s, Steve Mann coined the term *mediated reality* [80] through which he highlighted the possibility of altering (e.g., modifying or removing) physical objects instead of only overlaying virtual content on top of them as *augmenting* might suggest. Around the same time, Milgram and Kishino introduced a reality-virtuality (RV) continuum [87]. They described the continuum that allowed to categorize devices and applications that partly or fully replace the user's vision and established a fluid relationship between AR and VR. On one extreme of the continuum is the *real environment* and on the other the *virtual environment*. Everything between, excluding the extremes, is regarded as MR (see Figure 3.1).

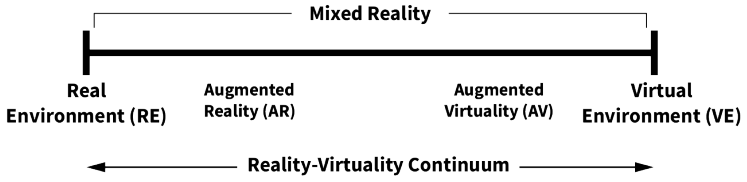

Figure 3.1.: Reality-virtuality continuum as presented by Milgram and Kishino (figure adapted from [87])

AR is placed on the left side of the continuum and VR on the far right. Such conceptualization revealed the space between AR and VR. There, Milgram and Kishino placed *augmented virtuality*, which is a virtual environment that contains elements of the physical environment surrounding the user (where AR is mostly a physical environment containing aspects of virtual). In addition to the RV continuum, Milgram and Kishino described a three-dimensional taxonomy (*extent of world knowledge*, *reproduction fidelity*, and *extent of presence metaphor*) for characterizing the emerging display technologies. Milgram and Kishino's RV continuum and taxonomy helped generate further concepts such as *substiutional reality* (where the virtual environment is adapted to fit the physical world) and articulate open research questions. MR quickly became the most widely reference term in academia for discussing computer-generated realities and was later also adopted in industry[1].

Despite the popularity, RV continuum and MR as defined initially have flaws that can cause confusion. Speicher and Nebeling listed several shortcomings of Milgram and Kishino's MR definition in their recent paper [130]. One of them is the confusion around where VR fits on the RV continuum and if it is part of MR or not. This is why in some papers, we see sentences such as "mixed reality and virtual reality" and also sentences such as "mixed reality (augmented reality and virtual reality)". It is also unclear where on the RV continuum can we place conventional displays, if anywhere. Thorough surveys and interviews, Speicher and Nebeling collected several conflicting definitions of MR that researchers and practitioners use today. To not add to the confusion, they avoided proposing another definition and instead created a framework for discussing MR experiences along seven dimensions. Skarbez et al. took a different approach and offered a revised definition of Milgram and Kishino's RV continuum and taxonomy [127]. With their revision, they answered the main criticism of RV concept and developed it further by introducing the distinction between *external virtual environments* and *matrix-like virtual environments* (see Figure 3.2). They also revised the taxonomy by proposing dimensions that consider the observer (extent of world knowledge, immersion, and coherence), which made the taxonomy more suitable for characterizing experiences and not just display technologies.

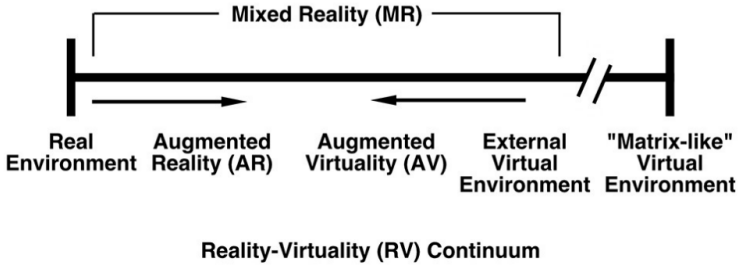

Figure 3.2.: Revised reality-virtuality continuum as presented by Skarbez et al. (figure from [127])

The revised MR definition by Skarbez et al. is the one used throughout this thesis. Therefore when discussing MR, both AR and VR are included. The reason for not using an umbrella term such as XR [2] is to keep the concepts that have proven useful and refine them instead of dismissing them because of their shortcomings. This does not mean that other concepts should not be entertained. The ones that I find especially interesting for offering an alternative perspective on MR are concepts that are free of the notion of *reality*. Marshall and Tennent point at such concepts in *The Limitations of Reality* [81] by stating that we should focus on fundamental descriptions of how our senses are manipulated to explore the potential of current and future hardware fully. Such descriptions can, for example, be seen in Sutherland's description of the first MR system, which is highly lucid without any mention of the word *reality* or *virtual* [137]. More fundamental descriptions can also be helpful outside of describing the technicalities of interactive systems but also for describing experiences (e.g., by borrowing concepts from cognitive science).

[1] https://docs.microsoft.com/de-de/windows/mixed-reality/discover/mixed-reality
[2] The term XR is also not confusion-free [79].

3.2. Mixed Reality Research Topics

MR research covers a wide range of topics ranging from technical ones focusing on improvement of MR hardware (e.g., optics, integrated circuits and display), to philosophical ones that, for example, focus on ethics of MR. Topics that concern the user interactions more directly deal with simulation sickness, ergonomics, depth perception, quality of the rendering, tracking latency, haptics, spatial sound, immersion, presence, embodiment, effects of long-duration VR sessions, and many more. LaViola et al. textbook gives a good overview of topics and open research questions[65].

This thesis focuses on a subset of MR research concerned with interaction techniques for accomplishing basic manipulation and navigation tasks. The techniques are motivated by using the capacity of MR to involve more of our bodies when interacting with the physical and virtual world around us. I specifically focused on hand-based interaction techniques since hands are one of the most dexterous parts of our bodies[3] and can be tracked reasonably well with current MR technologies. Besides being able to exercise a high degree of control over our hands, we are already highly skilled in using them from daily life interactions. I aimed to leverage this knowledge when designing the interaction techniques presented in this thesis.

[3]The largest area in primary motor cortex is used for hands.

4. Interaction Techniques

Interaction techniques are methods for accomplishing tasks via a user interface. MR provides a more extensive design space for techniques than conventional computing. It can simulate traditional techniques and enable new ones that are only possible because of the capabilities of MR (e.g., its spatiality and ability to manipulate human sensory-motor loop). This section focuses on MR interaction techniques for three fundamental tasks – manipulation, navigation, and learning. These tasks are fundamental because they are often needed for carrying out higher-level tasks and achieve more complex goals (e.g., sensemaking of past events). Despite being fundamental, the interaction techniques for these tasks are not yet perfected, and we are far from exploring the full capability that MR technologies offer. Below I give a brief overview of open issues in research of interaction techniques for manipulation, navigation, and learning.

4.1. Manipulation

Manipulation of real and virtual objects is a fundamental task that helps us achieve more complex goals [65]. For example, to win in a game of virtual table tennis, we need to be good at manipulating the table tennis racket. Early on, researchers spent a lot of energy on making MR manipulation efficient and easily learnable. Many MR manipulation techniques were inspired by how humans interact with physical objects – mostly by grasping them. Such techniques leverage the users' knowledge from their daily-life interactions. However, it is difficult to recreate real world interactions in a virtual world accurately.

The user's hands and fingers need high-quality tracking to be able to recreate real world like interactions. Hand tracking solutions have recently become practical enough to be included in low-cost commercial headsets[1]. Most of these setups are based on inside-out tracking that works with the help of depth cameras on the headset. Such setups have enabled more accessible ways to conducting research of hand-based interaction techniques by avoiding the need for specialized equipment. Furthermore, the proliferation of low-cost headsets enabled easier recruiting of experienced VR users and even allowed reaching them in their homes (e.g., via remote studies like the one conducted in Paper 4). Despite the advancement of hand tracking methods based on depth cameras, precise low-latency setups that can handle occlusion are still impractical and reserved for laboratories with expensive motion capture systems.

Lack of haptic feedback is another issue that distinguishes manipulation of virtual objects from manipulation of physical objects. Researchers have tried to address this with haptic gloves, controllers, ultrasonic devices, electrical muscle stimulation, and passive props. Each of these has their advantages and disadvantages and are being extensively researched. One limiting trend of haptic research (or any MR research focused on one modality) is that it is often investigated in isolation from vision and sound. The coupling haptic feedback to different modalities is rarely investigated since it requires more complex study designs and experimental setups.

Accuracy and precision of manipulation are also influenced by the human body itself. Mid-air interactions with 6 DOF can be very cumbersome, especially when considering the lack of haptic feedback and the gorilla arm effect[2]. Inventing MR manipulation techniques that are as usable as, for example, mouse-based interactions in CAD workspaces is still needed. Mendes et al. suggest that a promising research direction is to continue exploring manipulation techniques that can limit the DOF on users' demand[85].

[1] Oculus Quest 1 and 2 can track the users' hands by using depth cameras on the front of the headset.

[2] Tiredness that comes from constant motion of arms when interacting in mid-air.

Virtual environments can offer more than what is possible in the real world. For example, the users' avatars can be of arbitrary size, and the objects they manipulate can be modified with disregard for the laws of physics. The users could change the size, location, or shape of an object in an instant. This puts a limit on what can be accomplished by interactions that are based on the real world. The interaction techniques that depart from what is possible in real world are sometimes called magical. They, for example, allow the users to cast a ray from their finger and teleport themselves closer to the object they wish to manipulate [ref]. Research of magical interactions (e.g., Paper 2) is essential for exploring the full capability of MR and often goes beyond metaphors base on daily-life interactions.

Both real-world inspired techniques and magical techniques often require training before users can use them effectively. This is a challenge that has been under-explored in MR research. Most MR manipulation techniques are based on metaphors that are easy to understand and quick to learn. Such techniques can be great, however, focusing only on them leaves out a large design space of techniques that might require extensive training and, in exchange, offer benefits that easily learnable techniques do not.[3].

Apart from manipulating virtual objects, MR can also support the users when manipulating physical objects, or, using physical objects within the MR application (e.g., as a passive prop). There are many taxonomies and surveys that focus on manipulation of virtual objects (e.g., [65, 85], however, few include MR manipulations of physical objects. Similar, MR techniques for manipulating physical objects are also much less explored than virtual ones.

An issue related to many of the above is properly evaluating existing and newly created manipulation techniques. One of the difficulties here is how to standardize the evaluation to fairly compare the techniques among each other. To deal with this issue researchers, have proposed standards, testbeds, and general guidelines (e.g., see [10]).

I focused on several problems related to manipulation listed above. Paper 1 focused on physical object manipulation supported by mixed reality (occluded interaction). Paper 2 addresses magical interactions in VR: distant interaction, small and large scale interaction as well as occluded interaction. It presents a system based on proxies and marks which behave similarly to worlds-in-miniature and portals. This system allows non-destructive rearrangement of the space, enabling high-fidelity free-hand interaction through arbitrary distances and sizes. Paper 3 and 4 address MR manipulation indirectly. Paper 4 focuses on supporting users' during motor learning. The technique presented could also be used during learning of new manipulation techniques. Paper 3 couples manipulation of objects in the world to moving through time (temporal navigation).

4.2. Navigation

Navigation is a fundamental task that is most often studied in the context of moving through a physical or virtual space [65]. For example, meeting a friend on the other side of the city or finding a treasure in a VR game requires navigation. Apart from navigating through space, the user can also navigate through time. Such temporal navigation is often used for conventional recordings (e.g., YouTube videos) and does not have a real-world counterpart.

Navigation is usually analyzed by splitting it into two components – wayfinding and travel. With wayfinding, the user determines, plans, and follows a route while travel is the physical or virtual movement on that route. Travel tasks can be of different types, and we generally distinguish between search tasks (where the user has a target location in mind) and exploration (where the user is browsing the environment).

One of the main issues for spatial navigation in MR is that the virtual world can be infinite while the real world the user is situated in is constrained. To be able to travel through virtual worlds, researchers have designed techniques that, for example, allow the users to teleport or redirect their walking.

[3] Learning how to type on a keyboard, hand-write or to play a music instrument also requires extensive training.

The main issues and opportunities for temporal navigation are less clear since they have not been studied as extensively as spatial navigation. This is likely due to two reasons 1) many MR applications imitate the real-world where temporal navigation is not possible, and 2) there is little prior research on temporal navigation from disciplines outside of computer science (in contrast, spatial navigation has been extensively studied in psychology). Due to this MR, applications that could benefit from temporal navigation are under-explored, problems and opportunities not articulated, and the design space uncharted. Paper 3 focused on this space and explored temporal navigation techniques for spatial recordings. The techniques leaned heavily on direct manipulation of objects which was the basis for navigating through time.

4.3. Learning

Learning happens on many levels and is intertwined with almost every human activity. For example, to play a game of VR basketball, the user might need to learn several things – the rules of basketball, the basic moves for handling, passing, and shooting a virtual ball, learning how to cooperate with other players, as well as how to play strategically and tactically. This thesis focuses on only a small sub-part of motor learning – learning of hand movements. Being able to move the hands in particular way is a basis for most manipulation and navigation techniques in MR.

In MR motor learning is usually supported with augmented feedback – an enhanced feedback that occurs during training [126]. Such augmented feedback can be concurrent (e.g., arrows during the movement) or terminal (e.g., visualization of already executed movement). One of the open issues in motor learning is understanding what type of feedback works best and in what situations. Another open question is what visualizations are best for supporting the users during training without confusing or overwhelming them. One opportunity that MR offers is the ability to manipulate where and how the users see their bodies. The dissociation between, for example, where the hand is and where the virtual hand is shown can make the user redirect the movement towards the virtual hand (e.g., as in self-avatar follower effect [35]) or away from it (e.g., as in [5]). Paper 4 focused on concurrent augmented feedback embedded in the avatar's movement. The paper investigated if such a technique supports learning of complex hand movements and provides lasting improvement of the trained movement.

Part II.

Papers

5. Augmented Reality Views for Occluded Interaction

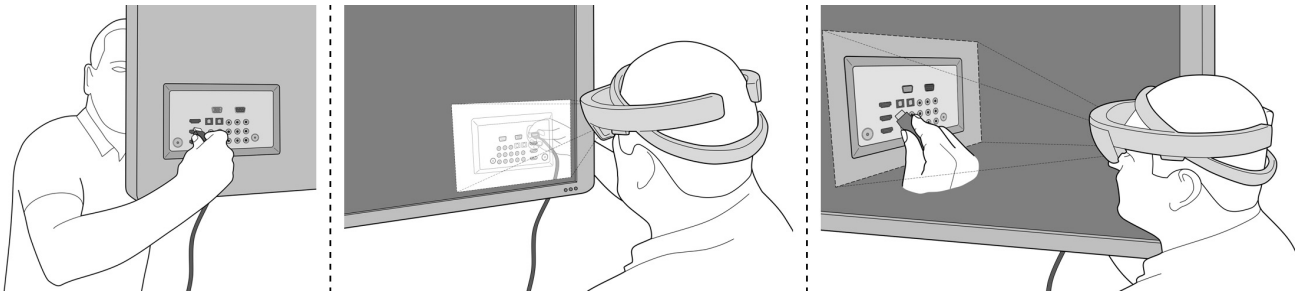

Figure 5.1.: In some situations, we need to manipulate objects out of our sight. We investigate how different views of occluded objects support users during manipulation tasks. An example of such a task is plugging in an HDMI cable. While the port is normally out of sight, see-through view (middle) and displaced 3D view (right) enable visual feedback during interactions.

5.1. Abstract

We rely on our sight when manipulating objects. When objects are occluded, manipulation becomes difficult. Such occluded objects can be shown via augmented reality to re-enable visual guidance. However, it is unclear how to do so to best support object manipulation. We compare four views of occluded objects and their effect on performance and satisfaction across a set of everyday manipulation tasks of varying complexity. The best performing views were a see-through view and a displaced 3D view. The former enabled participants to observe the manipulated object through the occluder, while the latter showed the 3D view of the manipulated object offset from the object's real location. The worst performing view showed remote imagery from a simulated hand-mounted camera. Our results suggest that alignment of virtual objects with their real-world location is less important than an appropriate point-of-view and view stability.

5.2. Introduction

Our hands can reach places and manipulate objects out of direct sight. This allows us to fish for keys under a car seat, scratch the back of our head, tighten a hidden engine bolt, or plug an HDMI cable in a port on the back of a TV (see Figure 7.1). During these interactions, users cannot rely on eye-hand coordination to accurately guide their reaching and grasping. Instead, they have to use their sense of proprioception, tactile feedback, and past knowledge of the object's shape, position, and orientation.

When the user's direct view of an object is occluded, the object may be observed from some other perspective. For example, endoscopic inspection cameras are commonly used in construction to provide a view inside of walls or confined spaces. Similarly, finger [125, 131, 159] and body-mounted [52, 63] cameras have been used to provide remote perspectives. We envision camera and sensor technology to shrink further, allowing systems

to collect real-time visual data from anywhere the user wants to interact. With such data, occluded objects can be rendered into the user's visual field through augmented reality (AR) headsets. However, how to best do that is unclear. Often, remote imagery is shown to users on a dedicated monitor or in a picture-in-picture (PIP) view. In contrast, work on see-through AR [90] assumes that keeping the remote view stable, in a position and orientation corresponding to remote imagery's real location, is beneficial. However, depending on the task, other types of views could be preferred.

We compared four views of occluded objects (with an additional baseline, where the object is not seen at all) to empirically identify the trade-offs with respect to performance and subjective satisfaction across a set of manipulation tasks of varying complexity. The views included two variants of PIP (a static and a dynamic, hand-mounted camera), a see-through view, and a displaced three-dimensional (3D) view. The views were implemented using virtual models of the occluded objects and virtual cameras, giving us full flexibility in manipulating the camera position and view content.

We contribute a set of views designed to support interaction with occluded objects, and empirical data on the performance and user satisfaction with those views. Our results show that the displaced 3D view performs on par with the see-through view and is often the preferred choice for occluded interaction. The worst performing view showed remote imagery from a virtual hand-mounted camera. The above two results suggests that alignment of remote imagery and its real-world location is less important than an appropriate point-of-view and view stability.

5.3. Related Work

We build upon related work on use of remote cameras, ways of presenting the remote views, and perceptual adaptation to discrepancies between vision and proprioception.

5.3.1. Remote Cameras

In many situations, users require imagery of something that is not directly viewable; remote cameras provide this. For example, remote cameras in video surveillance systems enable the security officer to monitor multiple rooms. Similarly, rear view cameras in cars enable drivers to see the car's surroundings.

Cameras can also be placed on a user's body. Yang et al. explored the design space enabled by a finger-worn RGB camera [159]. They proposed using the device as an extension of the user's sight. Stearns et al. explored how a finger-mounted camera can aid users with impaired vision when reading [131]. Horvath et al. instead mapped the visual information from a finger-worn remote camera to haptic information [42]. Kurata et al. proposed a shoulder-mounted camera system for remote collaboration, where remote collaborators can control the camera's angle [63]. Depending on the application, the placement of wearable cameras can vary. Mayol-Cuevas et al. provided a model for choosing the camera's location on the body, based on the field of view and the resilience to the wearer's motion [83].

Remote cameras can also be mounted on tools. For example, endoscopic cameras are used for inspection in construction as well as by surgeons. However, endoscopic cameras are not easy to control as they move in counter-intuitive ways. This has inspired work on improving control of the remote camera's movement and ways of presenting the camera's view to the user [92, 116, 151].

Instead of showing a 2D RGB view of the scene, several papers explored the use of depth, infrared (IR), and stereoscopic cameras. For example, *Room2Room* used Microsoft's Kinect to capture a 3D point cloud of a remote participant for a telepresence application [105]. Several cameras can be combined for fully instrumented rooms, allowing for real-time volumetric capture of the scene [47].

5.3.2. Presenting the Views

Data from remote camera can be displayed in many ways. In diminished reality, the scene is altered to remove, hide, and see through objects and thus reveal the area of interest (see Mori et al. for a survey [90]). For example, Sugimoto et al. presented a method for removing a robotic arm from the remote camera view, enabling the operator to see more of the work area [135]. Researchers also investigated best ways to visualize hidden objects to keep the spatial understanding and enable depth judgments [30, 74, 164].

Commonly, the visual information is shown on dedicated displays or as a picture-in-picture view in a head-mounted display. Alternatively, AR can be used to combine the user's view with the camera data. Colley et al. use handheld projection to reveal the physical space on the other side of the wall [25]. The *Room2Room* used projection to render a volumetric capture into the user's space [105]. Similarly, Krempien et al. used projection to augment a surgeon's view of a patient with medical imagery [60]. Furthermore, such blending of virtual and real can enable new interaction techniques such as *The Virtual Mirror* [12] and transparency-enabling gestures as presented in *Limpid Desk* [43].

The most flexible way for presenting remote imagery is via virtual reality. An example of this is recent work by Lindlbauer and Wilson, who explored the concept of *remixed reality*, whereas user's viewpoint can be moved to an arbitrary location, while the scene's objects can be copied, moved, and removed on demand [73]. While many of the above ways of presenting the remote visual information work well for exploration tasks, they are not well suited for tasks that also require manipulation of the observed.

5.3.3. Perceptual Adaptation

Manipulating an object while viewing it from a uncommon perspective can cause a mismatch between the visual feedback, proprioception, and tactile feedback. Users can adapt quicker to some discrepancies than to others. A number of studies investigated pointing errors during the displacement of vision induced by prism spectacles [39, 44, 118, 150]. They found that perceptual adaptation is rapid, with pointing errors drastically decreasing within a few trials.

The adaptation is slower for the left-right reversal of the visual field. In Seklyama et al.'s study, the participants wore prism spectacles inducing left-right reversal of visual field and needed a month to adapt well enough to be able to ride a bicycle [124]. Interestingly, fMRI scans show that a new visuomotor representation emerges at about the same time. Subjects are then able to switch between the old and new mappings for eye-hand coordination.

Arsenault and Ware investigated the influence of a distorted perspective and haptic feedback on rapid interaction with virtual objects [4]. They found that accurate perspective and haptic feedback improve performance in fish tank VR. Ware and Rose [149] conducted a series of experiments to investigate the differences between virtual and real object rotations. They suggest that manipulation of virtual object is easier when the hand is in approximately the same location. Pucihar et al. investigated the perceptual issues related to rendering of AR content from the device perspective (versus user perspective) [26].

Remote imagery of interaction with occluded objects introduces even stronger perceptual mismatch than some of the prism spectacle experiments and thus likely comes with even larger adaptation costs. There are few studies investigating the influence of displacement, view stability, and point-of-view in dexterous manipulation tasks with haptic feedback.

5.3.4. Summary

Previous studies investigated application of remote cameras and ways of presenting the captured visual information to support exploration and visual tasks. However, few of those studies touched upon supporting manual interaction. Similarly, there are many studies on perceptual adaptation that give insight into separate aspects

of manual interaction (e.g., eye-hand coordination [124]). However, it is unclear how to unite these findings to design views that best support manual interaction with occluded objects.

5.4. Types of Occluded Interaction

We define occluded interaction as interaction where users manipulate objects that are partially or fully occluded. Users can self-occlude parts of the scene with their body (e.g., fingers covering part of the touch screen). Similarly, tools can occlude as well (e.g., a handheld drill covering part of the drilling area). In many situations, the occluder is a separate object, located between the user and the target. For example, a couch occludes what is underneath for people sitting on it. Not only do such objects block the sight, they also constrain the users' movement and make it hard to *reach* what is occluded. For example, tightening a bolt at the back of a sink requires reaching around it. While self-occlusion and tool-occlusion can often be remedied, occlusion by the environment or larger objects is generally too costly or impossible to remove.

Our everyday life is full of tasks that require occluded interaction, and complexity of these interactions varies. For example, a relatively simple task is pressing a switch hidden under a table. However, occluded interaction can also involve complex movements (e.g., a mechanic working on a part in the back of an engine compartment).

To structure the space of occluded interaction, we looked into task taxonomies [21, 31, 75, 117, 160]. From this we built a selection of tasks that is representative of the range of movement complexities and constraints [31]. Furthermore, we selected tasks that commonly occur in everyday life (see *A taxonomy of everyday grasps in action* [75]).

All tasks we identified as representative for occluded interaction require acquisition of an object, followed by manipulation. These tasks are:

Pressing, where users toggle an object, such as a light switch, power button, or door bell.

Rotating, where users grasp an object and then twist it (with one DoF of rotation) to the desired orientation. Examples of such objects are radiator valves, thermostats, and dimmer switches.

Dragging, where users grab an object and move it (constrained to one DoF) to the desired position. Common examples are chain locks, mounted at the back of doors.

Plugging, where users have to slide one object into another. This requires matching orientation as well as position. A common example is plugging of a USB cable or stick at the back of TV or computer.

Placing, is a task where one object is put onto another. This requires positional alignment, but can also include orientation constraints. For example, hanging a key on a hook, or an umbrella onto a door handle.

5.5. Visualizing Occluded Objects

Removing occlusion (i.e., bringing back the object into the user's view) to enable manual interaction in the occluded area can be done in several ways. The simplest approach is having 2D cameras, placed *statically* within an environment, showing the user an unmodified view of the occluded scene. In this case, the camera maintains its spatial relation to the occluded objects, whereas the perspective the user sees depends on the camera's position and orientation. If placed wrong, the user's hand or used tool might eventually become an occluder. Some of the techniques from diminished reality deal with this problem by removing the object of occlusion from the scene (e.g., [135]). Another issue that may arise by using static cameras to reveal the occluded scene is the mirror-effect. For example, when the camera is placed on the opposite side of the user (i.e., an angle of 180°), moving one's hand to the left in the user's frame of reference, results in movement to

the right in the camera's view. This can impede manual interactions in which a user relies only on a remote camera's view.

To avoid the constraints of statically placed cameras, cameras can be placed on the user's body (e.g., on fingers as in [131]) or onto a tool controlled by the user (e.g., [135, 151]). Such *dynamic* cameras enable users to control the camera's orientation and location, allowing the users to select a perspective relevant to the task at hand. Some techniques from augmented reality allow for arbitrary viewpoint positioning (e.g., [73, 91]). Being able to control the remote camera's perspective can alleviate the frame-of-reference issues, while at the same time causing a new set of problems related to viewpoint instability and counter-intuitive control of the remote view.

With the advent of AR headsets as well as depth cameras, showing occluded objects is no longer restricted to two dimensions. One option is to simply remove the occluder, as in diminished reality, while keeping the user's frame-of-reference. That is, the occluded object becomes visible through the occluder in a *see-through* fashion (e.g., [7]). This has the advantage that the spatial relation between the object (which is now visible), the user, and the interacting hand remains the same, which should allow for interactions as if the occluder would not be present. One issue here, however, is that the object might occlude itself. This is the case when interaction occurs on the side of the occluded object, facing away from the user (e.g., user is frontally facing a TV while searching for a slot on it's backside, as in Figure 7.1). In this case the irrelevant parts of the occluded object could occlude the relevant parts, or the interacting hand. In cases when virtual models of the scene are available, this can be alleviated by varying the amount of transparency depending on relevance and depth of the objects.

The 3D virtual representation of the occluded object can also be shown to the user, as if the user would stand in front of it, independently of the actual location of the physical object. Such *cloned 3D* view can be positioned in a way to reveal the most relevant perspective to the user and/or enable the most ergonomic positioning of user. The cloned virtual object maintains its 3D properties and appearance. The user can move freely in front of that view, for example, to explore the sides of the occluded object or to get a different perspective, once the interacting hand is present. However, as the view is offset from the object's actual location we expect difficulties as noted in the related work on perceptual adaptation (see Section 5.3.3).

While each of these views mitigates the actual problem of occlusion, we expect that their advantages and disadvantages would render some of them more useful than others. We predict that the see-through view will perform the best on all the occluded interaction tasks, especially on the ones that involve complex manipulation and orientation of handheld objects (i.e. placing and plugging). The cloned 3D view should perform slightly worse than see-through view because of the view's displacement from the area of manipulation. However, cloned 3D should still perform better than the static 2D view as the latter does not provide depth cues nor does allow adjustment of the viewing angle. Our last assumption is that the dynamic camera view will perform the worst. While it allows adjustment of the viewpoint by moving the camera, it does so on expense of camera instability. We believe that its benefits will be outweighed by the drawbacks.

5.6. Evaluating Occluded Interaction Views

To investigate the views and assumptions mentioned in the previous section, we designed an experiment in which we compared five views of occluded objects across a set of five everyday manipulation tasks.

5.6.1. Tasks

We base the experiment tasks on the task categories described in Section 5.4. Within each category we chose a familiar object encountered in everyday life. The tasks and their corresponding objects (see Figure 5.2) were as follows:

Pressing a Light Switch: Required either pressing (changing the state of) the left button, right button or both of the buttons on a two-button light switch.

Rotating a Dial: Required participants to rotate a dial knob with a small arrow indicating its orientation. The starting position was always at 6 o'clock, and participants had to rotate the knob to either the 9 o'clock, 12 o'clock or 3 o'clock position.

Dragging a Slider: Required participants to position a slider at a specified position. Similar to the dial, the starting position was constant and the participants were instructed to positioned the slider to either 25 %, 50 % or 75 % of the full length of the slider's rail.

Plugging in an HDMI Stick: Required the participants to plug a HDMI stick into one of four vertically arranged HDMI slots. The participants were instructed which port to use at the beginning of the task.

Placing a Key Fob: Required the participants to hang a key fob onto one of three hooks. The participants were instructed to either hang the item onto the left, middle, or right hook.

Figure 5.2.: During the study participants performed five different tasks (left to right): *pressing* one of two light switches, *rotating* a dial, *dragging* a slider, *plugging* an HDMI stick into one of four ports, and *placing* a key fob onto one of three hooks.

5.6.2. Views

The experiment compared five views. Four of the views were already mentioned in the Section 5.5. We added a baseline condition in which participants did not receive additional visual help. All views were implemented using exact virtual copies of their corresponding physical objects. This allowed us full control over rendering of the objects into the participants' view. The views were:

No Visualization: Here, the participants did not receive any visual help from the AR headset.

Static Camera: In this view, we showed a virtual remote camera view, rendered as a picture-in-picture (PIP) in the participant's visual field via an AR headset (see Figure 5.3a). The PIP followed the participant's head movement similarly to the windows in operating system of Microsoft Hololens. The remote virtual camera (60 FOV, 16×9) was positioned at a static location, 30cm from the object of manipulation.

Dynamic Camera: In this view, we showed a virtual remote camera view, rendered as a PIP, similarly to the static camera view (see Figure 5.3b). The virtual remote camera (90 FOV, 16×9) was attached either to the tip of the participant's index finger or the tip of the handheld object (i.e., HDMI device or the key fob).

Cloned 3D: In this view, we rendered the 3D models of the occluded objects at a static location in the proximity of the participant. More specifically, the virtual model of the occluded object was rotated 70° around the vertical axis and moved 65 cm away from the object's physical location (see Figure 5.3c), to the left of the participants.

See-Through: Here, the 3D model of the occluded object is rendered at its actual position. The participants get an impression of seeing through the occluder (see Figure 5.3a).

In all of the views, we also either rendered the fingertips of the participants' index finger and thumb (in *Pressing*, *Rotating* and *Dragging* task), or the tracked handheld object (in *Plugging* and *Placing* task).

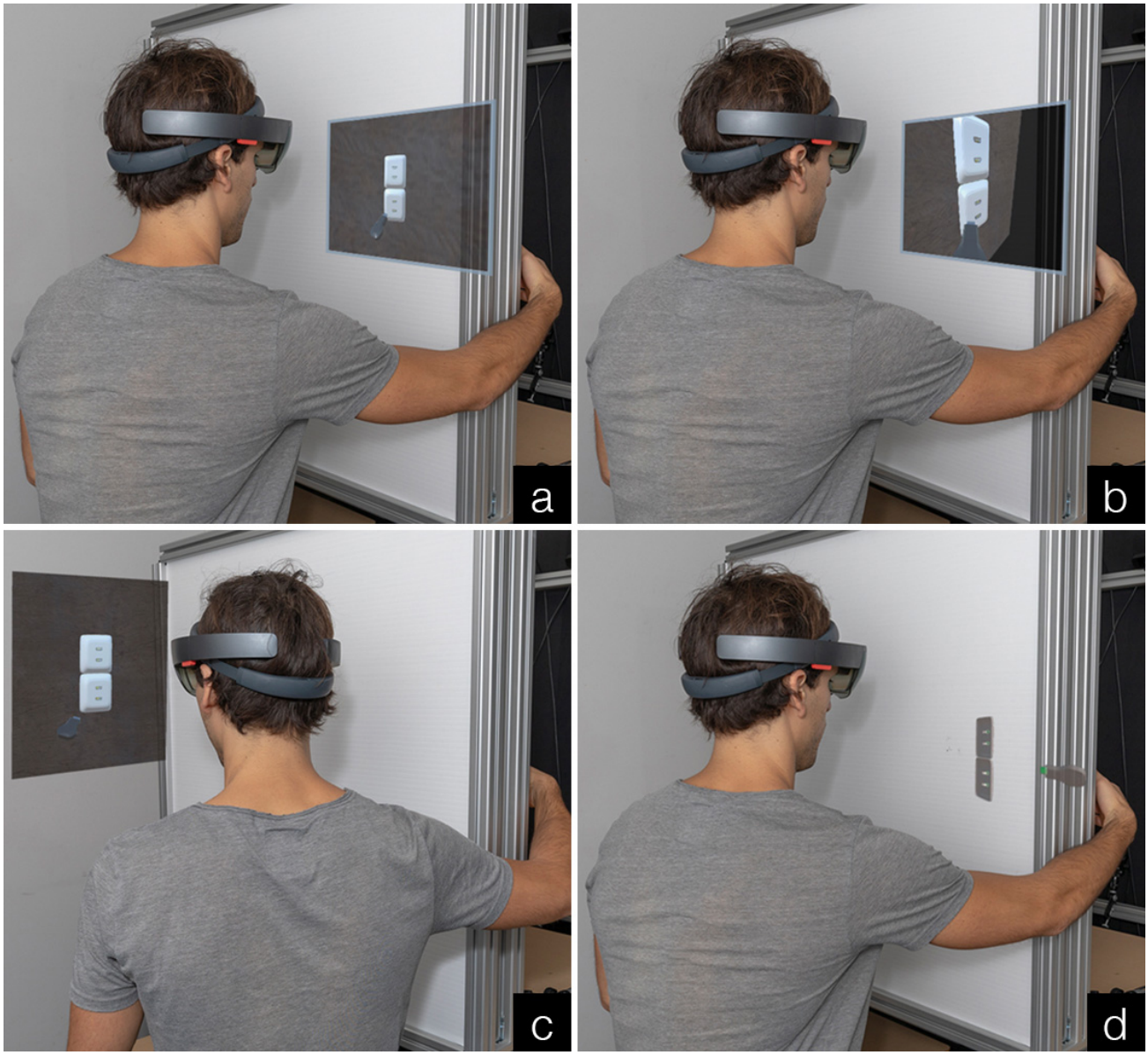

Figure 5.3.: Four of the views used in the experiment: A) Static camera, b) Dynamic camera, c) Cloned 3D and d) See-through view. No visualization view is not shown, as it does not render anything in user's field of view. Frame e) shows the used apparatus.

5.6.3. Apparatus

We used a 1x1x1 meter aluminum frame as a base for mounting the task objects, the tracking setup, and the panel that occluded the participants' view (see Figure 5.4). The task objects were mounted into a wooden panel placed at 30° angle to the cube's surface facing the participants.

Participants were seated in front of the cube, with a numpad placed on the left of them. The enter key on the numpad served as a trigger to start and end the trial. To track the participants' hands we used an OptiTrack setup (eight Flex 13 cameras, 1280×1024 pixels, 120 fps). In the *pressing*, *rotating* and *sliding* task, we used

OptiTrack markers placed on the index finger and thumb. In the two tasks that involved a handheld object, we placed the OptiTrack markers on the object itself (see Figure 5.2).

Figure 5.4.: During the study, participants were seated in front of a 1 m^3 frame. They reached in from the right side and interacted with objects placed on a wall, that was tilted at a 30 ° angle. Movement inside the frame was tracked with an Optitrack setup and participants received visual feedback in a HoloLens headset.

We used Microsoft HoloLens mixed-reality headset (2.3M Holographic resolution, 30°H and 17.5°V FOV) to display the views. The HoloLens was connected wirelessly to a host computer which captured the movement data and tracked the objects' state via a Teensy 3.2 microcontroller. To align the coordinate systems of the HoloLens and OptiTrack, we used a one-time calibration procedure in which we placed a HoloLens's spatial anchor at the origin of OptiTrack's coordinate system.

5.6.4. Design

We used a repeated-measures within-subjects factorial design. Independent variables were **Task** (*Pressing*, *Rotating*, *Dragging*, *Plugging* and *Placing*), and **View** (*None*, *See-Through*, *Cloned 3D*, *Static*, and *Dynamic*).

We split the tasks into two blocks and counter-balanced across participants both the tasks within the block and the blocks itself. Tasks were split in blocks to minimize the tasks preparation time, as each of the blocks required a distinct hand tracking setup. One block contained tasks which used finger tracking (*Pressing*, *Rotating* and *Dragging*), while another contained tasks with a tracked handheld object (*Plugging* and *Placing*).

For each *Task*, participants used each of the *Views*. We used one practice repetition and two timed repetitions per *Task*. The *Task* goal was randomized (e.g., which hook to place the key fob onto) and the *Views* were presented in random order within each repetition. Each participant completed a total of 75 trials.

5.6.5. Measures

We collected three performance measures for each trial: (1) how long it took the participants to start the manipulation of the task objects, (2) the duration of manipulation, and (3) the error of the manipulation.

The temporal measures were calculated from the moment participants' hand entered the cube. The start of the manipulation for the pressing, rotating and dragging tasks, was the time till the first change of the object's state was detected (e.g., when the dial was rotated). For the plugging and placing conditions, this was the time till the first plugging or placement of the object. The duration of manipulation is measured from the time participants' hand entered the cube till the last state change (e.g., last rotation of the dial, or last placement of the key fob). Finally, manipulation error in the dragging and rotating conditions, was measured as the difference between the set position and the target position. The pressing, plugging, and placing tasks only allow for discrete state changes and thus error was registered accordingly. Error was recorded at the end of each trial, so intermittent wrong states were not penalized.

5.6.6. Participants

We recruited 24 participants (8 female, age 19–71, M = 33.6, SD = 11.9) via mailing lists and social media. All participants were right-handed, four participants wore glasses. When asked to rate their experience with augmented reality on a 1–5 scale, 13 participants stated no experience. Five participants each rated their experience as 2 and 3, and only one participant had higher than average (4) experience. For participation in the study, participants received gifts worth about $30.

5.6.7. Procedure

First, we explained the purpose of the study to the participants and presented an overview of the tasks and objects. Once participants understood the tasks and how to manipulate the objects, the experimenter introduced the HoloLens and the views. After being familiarized with the five views, the experiment began.

Each of the five tasks started with a practice trial for each of the views. Participant received instruction shown via the HoloLens (e.g., "See-through view: Rotate the dial to 12 o'clock."). The trial started once the participant pressed the enter key on a numpad placed in their proximity, and ended once they pressed the key again; this also initiated new instructions. Between the trials the experimenter reset the object to its starting state, and participant could start the next trial as described above. In the practice trials, participants were encouraged to take their time and ask the experimenter for any clarifications they needed.

After the practice trials, participants continued with the timed trials going through all the views twice in a randomized order. In the timed trials participants were instructed to complete the trials as fast and accurate as possible. Once they completed the task, participants had a short break during which they removed the HoloLens and were asked to fill out a questionnaire.

The questionnaire asked the participants to rate the view with respect to four Likert-scale questions: (Q1) whether they liked that visualization, (Q2) whether they could easily manipulate the object, (Q3) whether they felt supported by the visualization, and (Q4) whether the visualization allowed them to easily check the object's state. We also asked participants to describe advantages and disadvantages of the views, as well as challenges particular to the task and further comments. Participants were provided with a sheet showing the names and images of the five views as seen through the HoloLens.

After completing all tasks, participants filled out a final questionnaire where they provided an overall rating for each view and additional comments. The experiment took 60 to 90 min, depending on the participants' pace of going through the trials and how much time they used on the in-between the tasks questionnaires.

5.7. Results

We separate the results into three sections: (1) analysis of the performance measures collected during the trials, (2) analysis of the ratings from the questionnaires, and (3) thematic analysis of the participants' comments. The overall results of the analysis show that the see-through and cloned 3D view preformed the best, with the latter being the preferred choice of participants. The worst performing and perceived view was the dynamic camera view.

5.7.1. Differences in Performance

To determine differences in performance, we analyze the 1200 timed trials, sans invalid ones. Trials are invalid if: (1) the participant did not interact with the object, or (2) the experimenter did not properly reset the setup for that trial. This was the case for 19 trials (i.e., 1.6 % of the trials).

Because each task required a different kind of interaction, measures are not directly comparable. For example, the average manipulation duration ranged from 3.7 s (pressing task) to 9.1 s (plugging task). To better show the differences per view, we hence normalized the data for visualization and show relative performance measures.

For statistical analysis of differences we used repeated measures two-way ANOVAs. We report on main effects of view and interaction effects, but not on effects of tasks. All post-hoc tests used permutational paired t-tests with Holm-Bonferroni correction and 1000 permutations. We ran post-hoc tests to compare the views, but not the tasks or interactions.

Time till Start of Manipulation

As shown in Figure 5.5, the manipulation delay differed between the views. It took participants more than 1 s longer than average to start manipulation when they used the dynamic camera view. On the other hand, with the see-through view they started manipulation almost a second earlier on average.

Figure 5.5 also shows how this delay differed depending on the task. This highlights interaction effects, such as the see-through view being particularly beneficial in the plugging task. Similarly, no visualization fared worse for all tasks but the pressing task, where participants only had to press a light switch.

We found a main effect of view ($F(4, 92) = 12.0, p < 0.001, \eta^2 = 0.06$) as well as an interaction effect ($F(16, 368) = 2.7, p < 0.001, \eta^2 = 0.04$). Post-hoc testing showed significant differences between the dynamic camera view and all other views ($p < 0.05$), but not with the no visualization condition ($p = 0.2$). Furthermore, the see-through view was significantly different from no visualization ($p < 0.05$).

Duration of Manipulation

For the duration of manipulation we also see differences between the views (see Figure 5.6). Here the dynamic camera view performed badly, resulting in task durations longer by an average of more than 2 s. Having no visualization also impacted performance, albeit not as much as the dynamic camera view.

Interaction effects between task and view are also visible in Figure 5.6. For example, the cloned 3D view worked especially well for the rotating task, while dragging and rotating tasks are much harder with the dynamic camera view than placing tasks.

We found a main effect of view ($F(4, 92) = 24.9, p < 0.001, \eta^2 = 0.11$), as well as a significant interaction effect ($F(16, 368) = 2.9, p < 0.001, \eta^2 = 0.05$). Post-hoc testing showed significant differences between the dynamic camera view and all other views ($p < 0.05$).

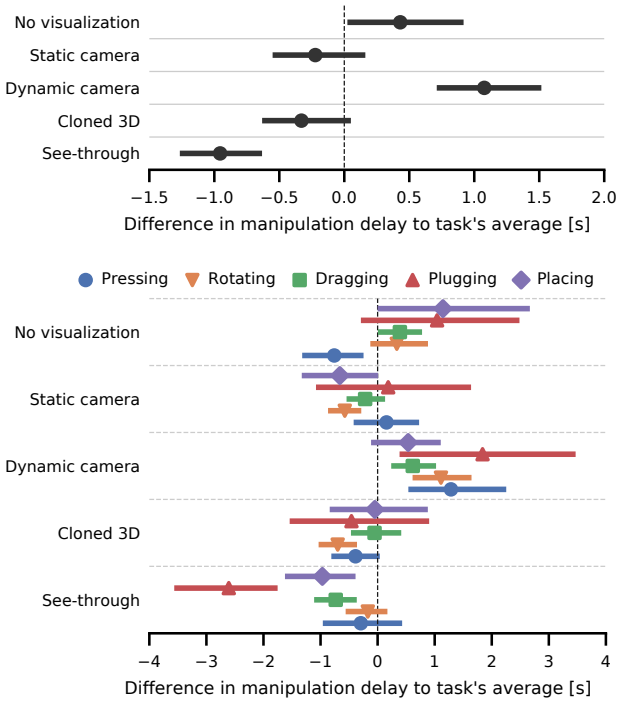

Figure 5.5.: Delay till start of manipulation per view (top) and per view and object (bottom). Error bars show bootstrapped 95 % confidence intervals.

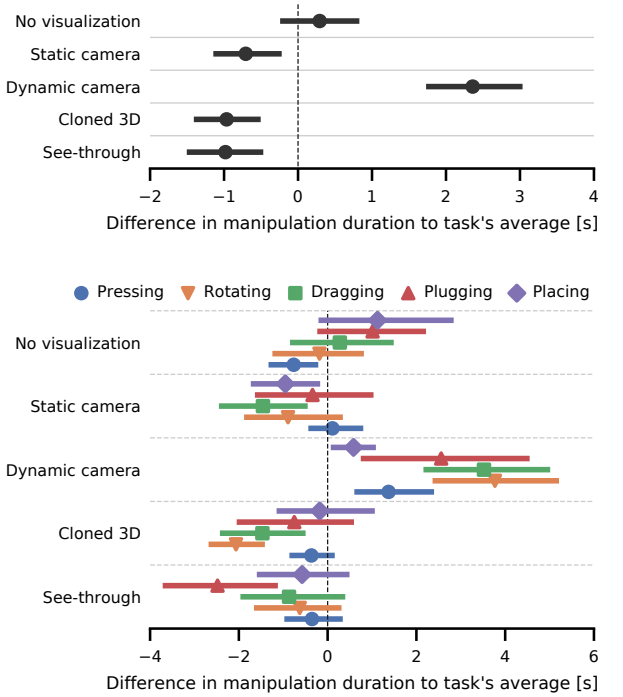

Figure 5.6.: Duration of manipulation per view (top) and per view and object (bottom). Error bars show bootstrapped 95 % confidence intervals

Manipulation Error

Finally, we investigated how precisely participants were able to manipulate the objects. In the dragging and rotating conditions, error is measured as the difference between the set position and the target position. The pressing, plugging, and placing tasks only allow for discrete state changes and we only compare the presence or absence of error. Error is recorded at the end of each trial, so intermittent wrong states are not penalized.

We separately analyze the error for dragging and rotating (relative) from the other three conditions (absolute). Figure 5.7 shows how the view influenced the two kinds of manipulation error. As can be seen, in the no visualization condition participants were more prone to errors than in other conditions. While participants could still use tactile and proprioceptive information, the lack of visual feedback made accurate manipulation challenging.

Figure 5.7 also shows the interactions between task and error. For example, with the see-through view the rotating task resulted in less error than the dragging task. With no visualization, this relationship is inverted. For the former, we found a main effect of view ($F(4, 92) = 3.8, p < 0.001, \eta^2 = 0.05$). However, there was no interaction effect ($F(4, 92) = 2.2, p = 0.07, \eta^2 = 0.03$). Post-hoc testing showed significant differences between the cloned 3D view no visualization ($p < 0.05$).

For the absolute error conditions, we also found a significant effect of view ($F(4, 92) = 2.5, p < 0.05, \eta^2 = 0.02$), as well as an interaction effect ($F(8, 184) = 2.0, p < 0.05, \eta^2 = 0.05$). However, post-hoc testing showed no significant differences in any comparison of two views.

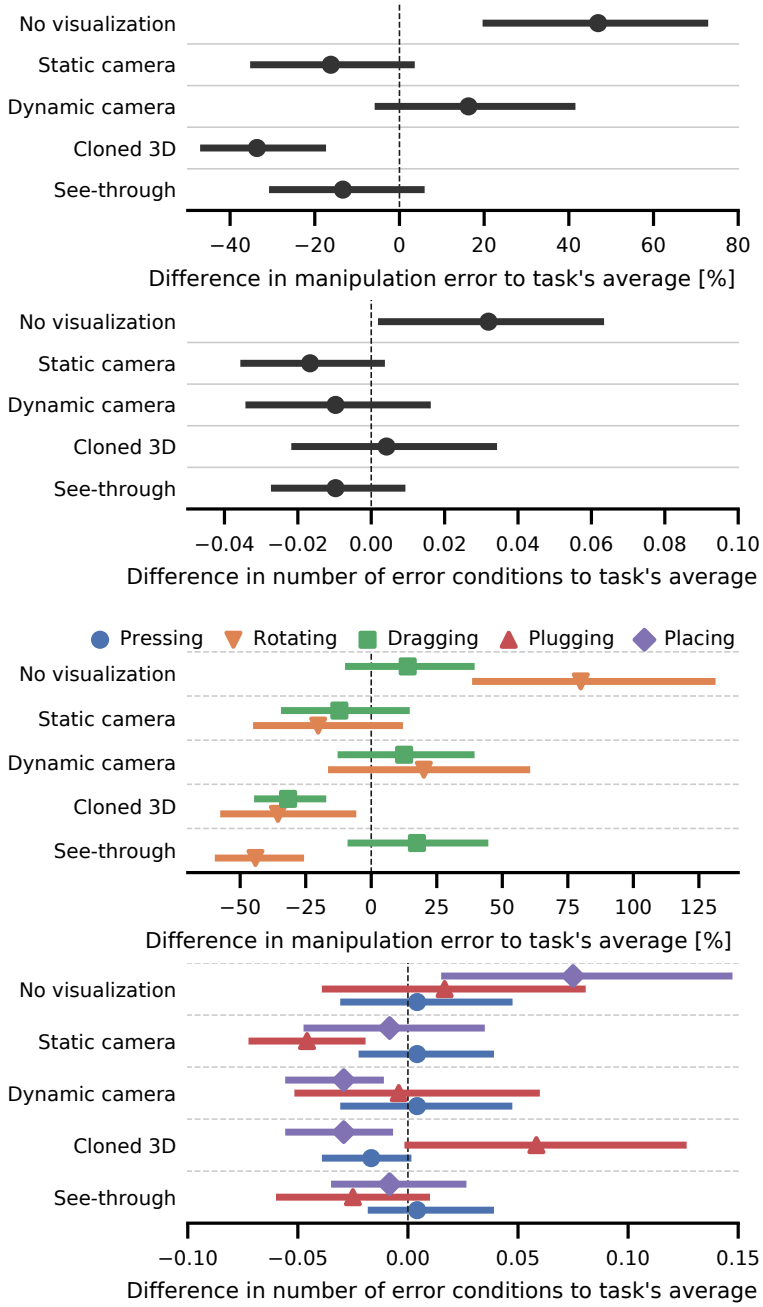

Figure 5.7.: Relative error and absolute errors per view (top two) and per view and object (bottom two). Error bars show bootstrapped 95 % confidence intervals.

5.7.2. Differences in Ratings

Asked for their overall rating of each view at the end of the study, participants gave favorable ratings to all but the dynamic camera view and the no visualization condition (see Figure 5.8). A Friedman test confirmed the significant effect of view on overall rating; $\chi^2(4) = 54.644, p < 0.001$. Post-hoc testing with a pairwise Wilcoxon signed rank test with Holm-Bonferroni correction showed significant differences between the cloned 3D view and no visualization ($p < 0.001$), as well as the dynamic ($p < 0.001$), and static ($p < 0.05$) camera views. The see-through view also differed from the dynamic camera view ($p < 0.001$) and no visualization ($p < 0.001$). Finally, the static camera view was significantly different from the dynamic camera view ($p < 0.01$) and no visualization ($p < 0.01$). In addition to the overall preferences, we took a closer look at the ratings the participants provided after each block. There we noted relatively consistent ratings for all of the views except the dynamic view. The latter was rated more highly than average in all four questions after participants did the placing task.

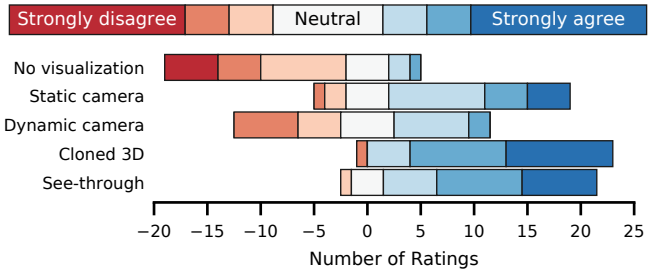

Figure 5.8.: At the end of the study, participants provided an overall rating of each view, indicating on a 7-point Likert scale how well the view supported them in the tasks. Shown here are the number of ratings for each level of the scale, stacked horizontally to highlight overall trends.

For statistical analysis of the factors influencing these ratings, we used cumulative link mixed models that enabled regression on the ordinal rating data in repeated measure designs. We fit two models to the data: one including interaction effects between view and object, and one without these effects. Both models included the user id as random effect variable. A nested model ANOVA showed that the interaction between view and object is significant ($p < 0.001$). Analysis of the main effects on the model with Chi-squared tests showed no significant effect of object ($p = 0.2$), but a significant effect of view ($p < 0.001$).

5.7.3. Qualitative Differences

The questionnaire given to participants between the tasks and at the end of the experiment contained a set of open-ended questions. We did a thematic analysis on the participants' responses, Table 5.1 shows the reoccurring themes and the number of participants mentioning the topic.

Static Camera

Participants liked the static view as it provided an overview of the area of manipulation and enabled seeing the state of the manipulated object.

Participants were split over how supportive the view was for manipulation. Five participants found the view helpful for movement and "easier orientation of objects" (P1), while five had troubles "to determine distance and details" (10) and "finding the correct angle" when manipulating the occluded objects.

Four participants complained about the distance of the static camera (P3: "The static camera is too far away."). This is a limitation of the fixed camera perspective as it cannot at both provide a good overview and a detailed view.

Dynamic Camera

Participants disliked dynamic camera view for a number of reasons. Participants complained about difficulties of seeing the relevant parts of their interaction. P5 mentioned that "dynamic view was annoying because the camera kept shifting and got in the way of getting a good view." A number of participants also mentioned self-occlusion as a problem, because either the handheld object or "the fingers obscured the view" (P11).

The unstable camera perspective "was confusing to adjust to" (P3), and impeded interaction with occluded objects. This is supported by the qualitative results, which showed that dynamic camera view took participants the longest to finish the manipulation tasks (see Figure 5.6).

When using dynamic camera view in task where the virtual camera was showing the perspective of the handheld objects, three participants drew parallels with the point-of-views they use in games. P9 said "it was like playing a 1st person shooter" and P7 that "it's like driving a spaceship.".

Surprisingly, the dynamic camera view was less disliked in the placing task (Figure 5.8). Three participants mentioned that in the placing task the dynamic camera view was "not horrible anymore" (P23). Task-dependence of view preference was also expressed by P10, saying that "it's funny how different views are helpful for different tasks."

Judging by the participants responses the handheld object-mounted camera was more helpful than the finger-mounted camera because of more relevant and stable perspective.

Cloned 3D

The cloned 3D view was perceived as intuitive (P23: "It was intuitive, I didn't have to translate the experience to make sense of what to do next."), natural and real. One of the participants event tried reaching for the virtual object to manipulate it before realizing that the physical object is at a different location. Three participants explicitly mentioned that the "cloned 3D view was advantageous because it helped (the) depth perception" (P24).

While the cloned 3D view was intuitive for most, five participants mentioned difficulties when using cloned 3D view. P23 mentioned that "it was confusing, like trying to do everything mirrored." and P2 mentioned that "the angle (of the object) was hard to adjust" during the plugging task.

Two participants made an observation mentioning that they saw the offset of virtual object from the real object's location as a benefit (P19: "With the cloned view the fact that you were looking off to the side made it easy to abstract movement from visualization")

See-through

Similar to the cloned 3D, the see-through view was also perceived as intuitive by ten participants. After using it in a task for the first time P22 exclaimed "this was scary easy."

The see-through view was perceived as giving a good overview of the area of interaction and good at supporting object manipulation. P5 said that "it enables your brain to relax since you see it as you would."

Contrary to the above, some participants found the see-through view confusing. P12 called it "an odd perspective" and P9 mentioned that "seeing the object from behind makes it difficult". Seeing the object from behind can also cause confusion between left and right, as expressed by four participants. For example, P19 turned the dial to nine instead of three in one of the practice trials and after noticing it exclaimed: "Oh yeah, it's because I see it from behind."

Table 5.1.: Benefits and drawbacks of the views as expressed by the participants.

View	Benefits	Drawbacks
Static camera	good overview (10) view's stability (2)	distant view (4)
Dynamic camera	game-like (3)	hard to manipulate (10) bad overview (8) confusing (8) self-occlusion (7) unstable view (4)
Cloned 3D	natural, intuitive, real (10) easy to manipulate (9) good overview (9) depth perception (3)	confusing (4)
See-through	natural, intuitive, real (8) easy to manipulate (11) good overview (8) depth perception (2)	confusing (5)

5.8. Discussion

We compared four views of occluded objects to identify the best support for occluded interaction. We found few differences between the views in performance (i.e., manipulation duration and manipulation error). However, the dynamic camera view performed the worst. With the cloned 3D and see-through views participants completed the tasks the fastest; these two views were also rated the highest on subjective satisfaction. Static camera view also supported occluded interaction well. In light of these results, we discuss three factors that contributed to these results: *point-of-view*, *view stability* and *view displacement*.

Point-of-View: The cloned 3D and see-through view were rated as the most liked and supportive. Many participants mentioned that those views felt natural and intuitive and that they showed the relevant part of the interaction. This was partly due to participants being able to choose their point-of-view by moving their head, just as we do in non-occluded interactions. This was not the case with the static camera view, thus some participants complained about the camera's view being too far away. Such limitations result from a fixed viewpoint, as it cannot provide both a good overview and a detailed view.

View Stability: Being able to change the point-of-view brings several benefits. However, if it comes at the expense of view stability, the drawbacks can quickly outweigh the benefits. Poor view stability was most noticeable in the dynamic camera view. While the participants could change the point-of-view, they were often confused by not knowing what will be shown next or how their hand movements affects the viewpoint. Issues with view stability can be only partly mitigated by smoothing the camera's movement. Good view stability requires intuitive mapping between the user's and viewpoint's movement (e.g., as in cloned 3D view).

View Displacement: Another factor that should have affected the performance of the view and users' satisfaction is view displacement. Past works suggest that view displacement negatively affects performance in manipulation tasks (e.g., [149]). Taking this into account, the cloned 3D view should have performed worse than see-through, as it was rotated and offset from the actual location of the occluded objects. We assume that the negative effect of the displacement is not noticeable in our study because of the use of everyday tasks. This allowed participants to rely on their tactile sense and past knowledge of familiar objects. The discrepancy between our results and past findings warrant further studies investigating the influence of view displacement on complex manipulation tasks, in which participants can rely on a spectrum of senses and skills used in their daily lives.

These factors are important, not only when considering the future research on interactions with occluded objects, but any research that deals with out-of-sight interaction. For example, dexterous input for AR headsets that involves hand movement out of user's direct sight.

5.8.1. Limitations

There are a few study limitations related to experiment design decisions. First, we conducted the experiment in a controlled setting to limit the external influences when collecting performance measures. This means that occluded interactions in the experiment only approximate the ones from real-life situations. However, we believe that the use of everyday tasks and objects, even if only in an artificial setting, is generalizable to many real-life situations.

Second, we chose one specific configuration and appearance for each of the view. A different appearance (e.g., transparency in see-through view) or configuration (e.g., the static camera placed at a different angle) might have influenced the results. Evaluating these aspects would require comparison of view variations, which would have extended our experiment duration (90 min per participant) even further.

Third, all the views were simulated by modeling the occluded objects and the use of virtual cameras. This gave us flexibility at expense of realism. We believe this was not a major issue considering the comments of participants' expressing how real the objects looked.

Last, participants had only a short practice run with each of the views. It is possible that with extensive training or longitudinal use participants could have adapted to more uncommon views.

5.8.2. Future Work

While our study shows that cloned 3D and see-through views support occluded interaction well, there are several open questions and directions to explore.

It is unclear what parts of an occluded scene and of the user's body should be rendered to best support occluded interaction. In our study, participants saw only the fingertips of their index finger and thumb, or the handheld object. Despite the minimalistic rendering of the hand's location, we did not note any complaints about showing too little. On the contrary, we had comments about self-occlusion interfering with the interaction. A systematic investigation into what hand features to render to best support eye-hand coordination would help make more informed decisions when designing the views for occluded interaction.

When considering the appearance of the occluded objects, investigating more extreme visual alternations can reveal new ways of supporting manipulation of occluded objects. For example, exploring planar abstractions [84], or hybrid visualization of see-through and cloned 3D view might offer additional benefits. Such visual alternations would also require unrealistic mapping between user's movement and visual feedback, as for example explored by Teather and Stuerzlinger [140].

5.9. Conclusion

We investigated how well different views can support manual interaction with objects that users cannot directly see. We evaluated the system in a lab study where we varied views and tasks in a controlled manner. We found out that see-through and cloned 3D view perform the best, with the latter one being preferred by the participants. The worst performing and most disliked view simulated remote imagery from a hand-mounted camera. We believe that alignment of remote imagery of the manipulated objects and their real-world location is less important than suggested by previous work. Furthermore, our results highlight the importance of view stability and an appropriate point-of-view.

6. Poros: Configurable Proxies for Distant Interactions in VR

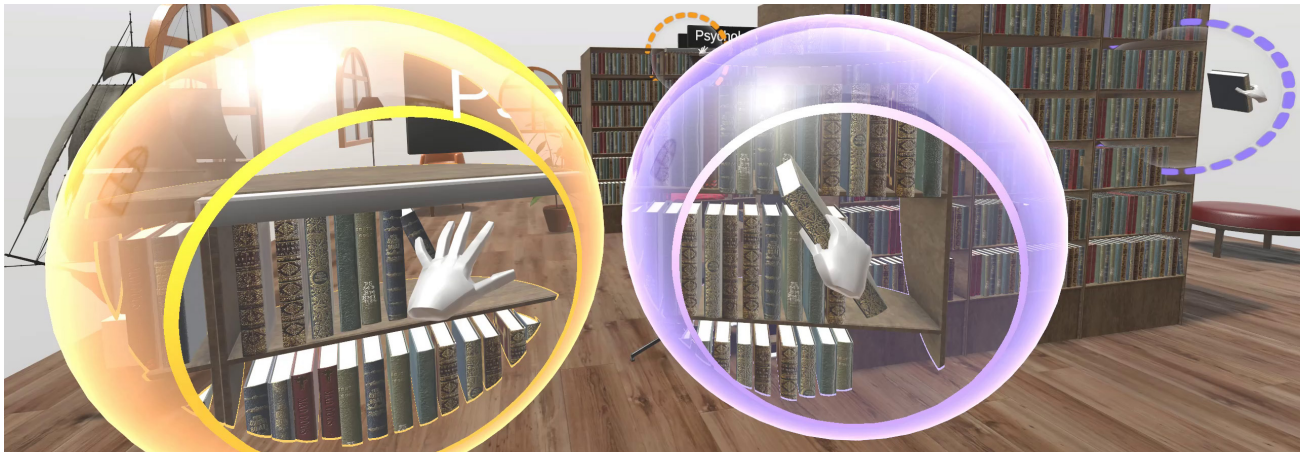

Figure 6.1.: Poros enables users to bring portions of distant spaces closer so that they can interact with and across them. Shown here are two proxies, linked to marked spaces (shown in the same color) around two different bookshelves. The user is about to move a book from one space to the other. In addition to direct interactions through them, users can move and arrange proxies, as well as perform operations on them, such as merging and aligning.

6.1. Abstract

A compelling property of virtual reality is that it allows users to interact with objects as they would in the real world. However, such interactions are limited to space within reach. We present *Poros*, a system that allows users to rearrange space. After marking a portion of space, the distant *marked space* is mirrored in a nearby *proxy*. Thereby, users can arrange what is within their reachable space, making it easy to interact with multiple distant spaces as well as nearby objects. Proxies themselves become part of the scene and can be moved, rotated, scaled, or anchored to other objects. Furthermore, they can be used in a set of higher-level interactions such as alignment and action duplication. We show how Poros enables a variety of tasks and applications and also validate its effectiveness through an expert evaluation.

6.2. Introduction

Virtual reality (VR) enables high levels of immersion [103] but at a cost: immersive interaction is often not efficient interaction. For example, reaching an object at the other end of a room requires physical effort as the user first needs to walk to the object. This issue is exacerbated when users need to be at different locations in quick succession. Locomotion techniques, such as teleportation, can help but incur a cost every time the user

switches position (e.g., due to disorientation [18]). An alternative approach is to enable users to directly act at a distance.

Several techniques have been proposed to enable more efficient distant interactions in VR. For example, users' reach can be extended (e.g., by extending hands [115] or raycasting) or the distant space can be brought closer (e.g., by portals [58]). Both of these approaches have unresolved problems. Extending the user's reach decreases the precision of interaction due to angular error. Bringing the distant space closer can interfere with interaction by distorting the visual space [22], creating inconsistency of hand mapping [89], making it difficult to notice the boundary between the space that is brought closer and the scene, or by occluding the scene as portals do [58]. Worlds in miniature [133] avoid some of these problems by enabling easy access to other parts of the scene while existing outside of the scene itself. However, in their current forms worlds in miniature only afford limited manipulation support, such as not allowing users to create and arrange them as they see fit.

Building on worlds in miniature and portals, we propose *Poros*, where portions of space can be marked and linked to proxies. The proxies can be brought close to the user and allow all interactions as if they were next to the marked space, effectively acting as surrogates [64]. This approach does not distort the visual space, keeps the 1:1 hand mapping, allows users to interact simultaneously with both distant spaces and nearby objects, makes the boundaries between multiple spaces visible, and allows multiple levels of indirection. As shown in Figure 6.1, this can, for example, be used to move objects efficiently between two distant spaces.

Proxies do not just enable direct interaction with distant spaces. They become part of the scene and can be transformed, arranged, or anchored to objects. For example, a user can create a proxy that is linked to a toolbox and anchor it to themselves. In this way, the user can always access a tool (e.g., a ruler) through the self-anchored proxy. Further interactions are possible once multiple proxies have been created. For example, users can perform operations between proxies to align them, highlight shared objects, or duplicate actions by linking multiple marks to a proxy. To demonstrate the usefulness of Poros, we show several examples of interactions with distant spaces, interactions at different scales, multiple perspectives, and multi-proxy interactions. We also validate that Poros can be easily understood and used effectively through an expert evaluation.

6.2.1. Contributions

Our paper makes the following contributions:

- We present a VR system, *Poros*, that enables users to create proxies to distant spaces and interact through them—all using direct hand-based manipulations.
- We show a range of operations on single, pairs, and groups of proxies enabled by having proxies as objects in the scene.
- We demonstrate that by combining operations, Poros enables many interaction techniques, including: occluded object interactions, interacting on different scales, multi-perspective manipulation, replicated interactions, space searching, and inter-space alignment.
- We validate Poros' effectiveness with an expert evaluation.

6.3. Related Work

Our paper is informed by previous work on enabling interaction with distant targets in VR. In particular, interaction techniques for distant reaching (e.g., ray-casting), ways of warping the virtual space (e.g., Erg-O [89]), and the creation of meta-spaces (e.g., overviews). Each of these three approaches has shortcomings that our technique addresses.

6.3.1. Interaction Techniques for Distant Targets

There are a large number of selection techniques for VR [3]. When selecting distant targets, two classic examples are ray-casting and arm-extension [17]. An example of the latter is *go-go*, where the user's arm grows non-linearly as it moves away from their body [115]. The PRECIOUS technique solves one of the problems of ray-casting: disambiguation between close targets [86]. If a selection is ambiguous, the user is moved closer to the candidate targets, makes a selection there, and afterward moved back to the original position. Another approach to disambiguation was proposed by Pierce et al. who used hand gestures to, for example, put a frame around the desired object [110]. Similarly, gaze information can be used for selecting distant objects which can then be manipulated as if they are close [108]. Another approach for improving ray-casting is to "bend" the ray towards potential targets [132] or through user control [98], easing the selection of dense, occluded, or distant targets. Common issues with these techniques are that (1) they distort or move the user's body, potentially impacting body ownership or disorienting the user [18], and that (2) interaction no longer occurs directly through the user's hands, which has similar repercussions. With Poros, users' bodies are not altered. Instead of extending their reach, moving them, or introducing pointer-like constructs, we move parts of space, bringing objects of interest into reach.

6.3.2. Warping and Moving Space for Interaction

Another approach to bringing targets closer is to warp the whole space. For instance, Chae et al. presented an augmented reality (AR) technique where users can shrink a room along one axis to bring distant objects closer [22]. Also for AR, Sandor et al. built a system that distorts space to show points of interest that are out of view or occluded [121]. In VR, Mine et al. scale the world as users grab distant objects for manipulation [88]. In Elmqvist's *BalloonProbe* technique only objects are warped instead of the whole space [29]. By repelling objects away from each other, occluded ones can be accessed more easily. In general, warping or distorting the space requires users to adapt to new, likely unfamiliar, spaces. This is exacerbated in non-linearly warped spaces, where movement and interaction can be particularly difficult.

Another method to warp space is the use of portals. With portals, arbitrary locations in space can be linked—stepping through one portal instantly moves one to the linked location. This can be used to shorten distances, but also to break the spatial consistency of a virtual world [58]. Stoev and Schmalstieg named this *through-the-lens* interaction and discussed how it can be used for a range of tools [134]. Portals can also be used to avoid the use of teleporting within a virtual environment [76]. With *PhotoPortals*, Kunert et al. proposed the use of portals for easier collaboration [62]. Users can capture views of the scene, manipulate them, and share them with others. Their portals also include a view mode where a whole cuboid slice of the captured space is shown. *SpaceTime* [155] also focuses on supporting collaboration. Containers in *SpaceTime* could be seen as portals to another place as they only allow teleportation to the remote content contained in it and not direct manipulation through it. But *SpaceTime* also breaks the spatio-temporal consistency of the world, enabling objects within containers to exist as clones, or time-shifted versions of the original ones. Similar to version control software, this enables collaborators to work in parallel and resolve conflicts later. However, it is unclear if the conflicts that arise by allowing this can be easily handled by direct manipulation in VR.

We take a similar view of distant spaces as *PhotoPortals* and *SpaceTime*, but instead of collaboration we focus on distant interactions and interactions built atop the combination of multiple spaces. Furthermore, we maintain the spatio-temporal consistency of the scene as spaces and objects are only made accessible, not cloned.

6.3.3. Interaction with Meta-Spaces

Another alternative for extending the reach of users are meta-spaces, such as Stoakley et al.'s *Worlds in Miniature* [133]. In addition to a first-person view, users see an overview of the world in which they can interact as in the main view. Furthermore, the miniature can be used for navigation [102]; in some systems, several miniatures

can be available. Instead of functioning as an additional view, the whole world can be turned into a miniature by scaling the user up, which then allows them to move faster through the scene [1]. The main world and the world in miniature can also be distributed between users, enabling collaboration across scales [112].

Worlds in miniature for large or complex spaces can be hard to use and hence Trueba et al. presented several improvements to the technique that clip and filter what is shown [145]. Another version of worlds in miniature is Bluff's *Miniature Metaworld* [14]. In contrast to the work above, his miniature is situated within the original scene. Users can manipulate objects in either the original or miniature view, but also move objects between them. An alternative to showing whole worlds in miniature is to only present distant landmarks to users and allow them to teleport to these [109].

Instead of replicating space, Pierce et al. explored how to replicate individual objects for interaction [111]. Their *Voodoo Dolls* technique enables users to grab distant objects via an image plane technique and then manipulate them in their hands. To show context, nearby objects can be placed in the other hand to, for example, allow for placing an object on a shelf.

In Poros, we also apply the worlds in miniature concept. As with metaworlds, our "miniatures" become part of the original scene. With Poros we improve upon worlds in miniature in several ways, enabling users to (1) create instances on demand and at a distance, (2) manipulate the bounds, orientation, and location of worlds in miniatures already in the scene, (3) anchor worlds in miniature to scene objects or themselves, making them dynamic, (4) merge and split worlds in miniature to replicate spaces and actions, (5) perform operations on worlds in miniature like aligning them or searching them, and (6) gradually peek into worlds in miniature, controlling how much of the view is taken over by them.

6.4. Poros for Interaction at a Distance

The primary goal of Poros is to enable manual interaction with distant objects. The core idea for this is to allow users to bring parts of space closer to themselves and directly interact inside and across those parts. Based on this capability, the second goal of Poros is to enable users to exploit the introduced indirection in order to make common tasks easier. More specifically, Poros allows users to:

- Mark distant spaces and bring them close in the form of *proxies*. Users can then perform **interactions through proxies**, to allow direct interaction with these distant spaces. Users can reach into proxies to manipulate their contents and peek into them to inspect a distant location.
- **Manipulate proxies** as they are first-class citizens in the scene. Users can scale, transform, minimize, align, clone, and otherwise manipulate proxies. Thereby proxies offer more interaction possibilities than worlds in miniature and portals. Users can leverage these possibilities to configure and optimize their workspace for the task at hand. Users can also anchor proxies to other objects and avatars to make these workspaces mobile.
- **Manipulate marks** to adjust what part of space a proxy shows. Marks can also be anchored to objects, which allows for tracking and manipulation of moving objects through proxies.
- Perform **abstract operations on proxies** that change them according to their or other proxies' content. This includes several alignment operations, as well as content-sensitive highlighting, and merging. Making use of these operations allows users to perform complex tasks that would otherwise require many actions or substantial movement.

Figure 6.2 illustrates the basic components of Poros: (1) users place *marks* in the scene to denote spaces they want to link to, which results in (2) *proxies* close to them that they then can interact with. Proxies are an exact mirror of a subset of the scene; any change to either is reflected in the other. Note that most commonly one proxy is linked to one mark, however, this association can also be one-to-many or many-to-one (as a result of merging

and cloning operations, described in Section 6.4.6). In contrast to portals, proxies are not two-dimensional gates to another place, they are three-dimensional replications.

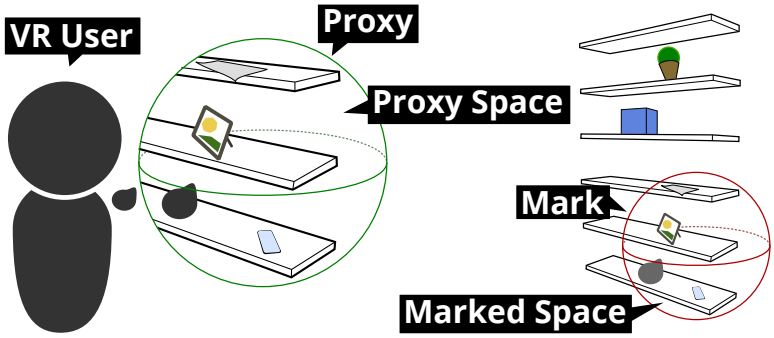

Figure 6.2.: In Poros, users create marks around space they want to bring close and subsequently called. The enclosed space is called the marked space. Through marking, a proxy is also created next to the user. It contains the proxy space, which is an identical mirror of the marked space.

6.4.1. Setup

We implemented Poros using the Unity game engine. We used an HTC Vive headset for output, with a Leap Motion controller attached for hand tracking. This setup allows users to walk around to explore the scene and use their hands to interact with the scene. The use of hand tracking, instead of controllers, was motivated by three considerations: (a) to heighten the sense of immersion, (b) to allow for complex and high-dimensional control, and (c) to leverage already acquired knowledge of object manipulation from real life.

For the visual replication inside proxies, we extended the built-in shaders and added a custom render feature to Unity's scriptable render pipeline. That changes the pipeline to add an additional render pass to opaque as well as transparent objects. On top of the default rendering, each object is then rendered another time for each proxy in the scene. For each active proxy that entails culling, modifying transformation matrices, setting global shader clipping data, and drawing the contained objects. To replicate hands, we wrap the Leap provider to transform hand data before passing it on to the Leap interaction system.

Poros is available as open source software[1] so others can extend and try it themselves.

6.4.2. Basic Properties of Proxies

In Poros, proxies and marked spaces are always spherical. This reduces the complexity for the users when creating or editing them. Conceptually, however, proxies and marks could be of any shape. For example, the interaction technique from *TunnelSlice* would be suitable for marking cubic volumes [66].

Figure 6.3 shows how proxies and marks are rendered inside a VR scene. Both use a fresnel effect for shading as well as further highlighting where they intersect the scene. This makes for a translucent and ephemeral appearance that limits scene occlusion. A circle with a marching-ants effect circumscribes each mark. Proxies and marks are color-coded to show which ones are connected. As users approach a proxy with their hands or head, the colored shell opens up (see, e.g., Figure 6.4) in order to provide a clearer view of the contained space.

Proxies display all the content within their linked marked space. If an object only intersects a marked space, it is clipped to the space's boundary. A proxy's content is scaled according to the relative size of the marked and proxy space; users may change that using operations on the proxy (explained later). If a marked space is twice as large as a proxy, its content is shown at half the size inside the proxy. When entering the proxy, the

[1] Available at https://github.com/henningpohl/poros

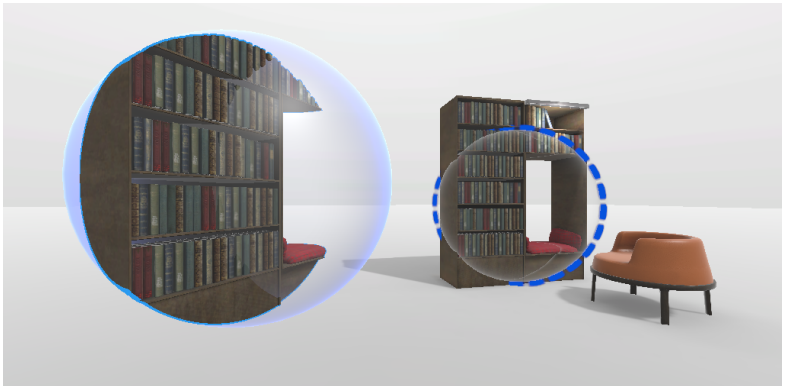

Figure 6.3.: Proxy spaces are rendered inside colored bubbles, marked spaces are shown in a fainter style. A dashed outline with a marching ants effect around marks is colored to hint at their connected proxy space.

Figure 6.4.: Proxies and marks can exist at different scales. Here the marked space is much larger and the user's hand hence scaled up inside of it.

user's hands always stays at the size it had outside of proxy (see Figure 6.4). Such 1:1 mapping is important for accurate motor control during entry and exit as well as for the manipulation of objects in the proxy. When the hands are inside of a proxy they are also rendered in each of the marked spaces the proxy links to. While the user's hands always stay at the same scale, the replicated hands in the mark are scaled. If a marked space is twice as large as a proxy, so are the replicated hands rendered in it (see Figure 6.4).

Just as proxy spaces show the visual content of linked mark spaces, they also bring their sounds closer. If a book falls down in a different room, this would normally not be audible to the user. However, if the book hits the ground within a marked space whose proxy is close to the user, that sound is also played at the corresponding location in the proxy space. Only sounds that occur within a marked space are audible in proxies; just as we use a hard boundary for visuals, we also apply this boundary for audio sources.

6.4.3. Creating Marks and Proxies

In Poros, users can mark a space through manual interaction; this process simultaneously creates a proxy. As shown in Figure 6.5, users start this creation process by making a hand gesture where the index fingers point towards each other. This creates a proxy between the fingers, and a mark in the distance; similar to aperture-based selection methods [32].

By moving their hands, users can continuously translate and scale the proxy, which is constrained to fit between

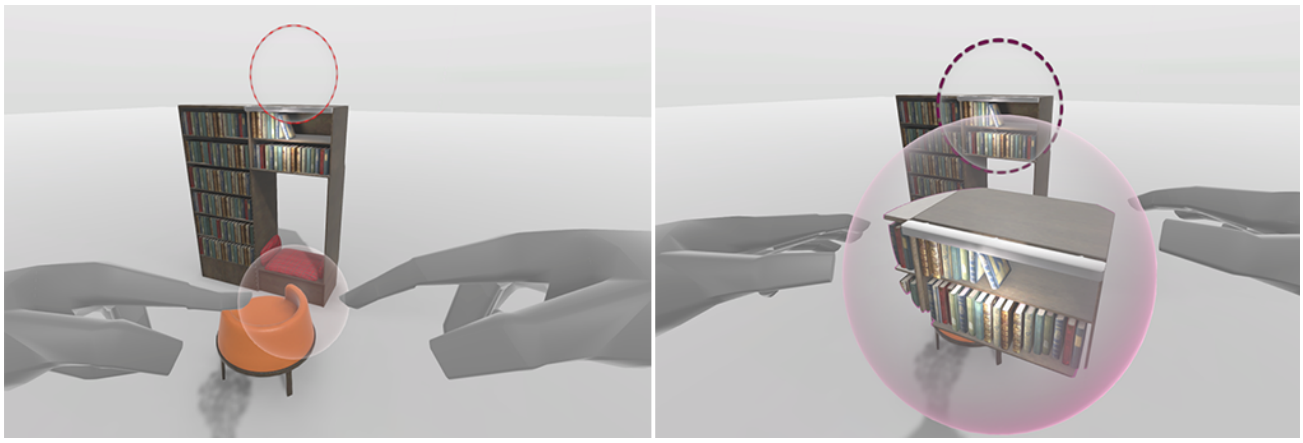

Figure 6.5.: Proxies and marks are created using a bimanual pointing hand gesture. Users can adjust the size and position of the mark inside the scene by moving their hands.

users' index fingers. This ensures that everything in the proxy (and thus also the marked space) is within reaching distance. As the proxy is sized, the marked space scales proportionally. Furthermore, the distance to the proxy changes proportionally to the distance between the user's body and hands. These proportions can be adjusted, but we have found that an exponentially scaling proportion allows for the greatest flexibility. Through this mechanism, users are able to create marked spaces far away from them. The mechanism means that the proxy and mark are dependent on each other, and finer adjustments should be done as a subsequent action afterward. We chose this approach as it fits the direct manipulation approach used in Poros. Alternatively, one could use techniques such as ray interaction [154] or clutching [9] to place the marked space.

6.4.4. Interactions Through Proxies

The main interaction afforded by proxies is direct manipulation of distant objects. Any movement or action within a proxy is handled exactly as if it occurred in the linked marked space.

Interacting Inside a Proxy Space

When users reach into a proxy space, their hands are effectively transported to the marked space. Consequently, they see their hands twice: within the proxy and within the scene. This extended reach allows users to grab and move objects far away from them but also to push buttons or operate other mechanisms. In Poros, we treat the proxy and the marked space as identical and thus objects are shown twice but do not exist twice. The objects only exist in the scene, that is, at the marked space. This also extends to events triggered via interaction—a button press inside a proxy only triggers one button press event.

The scale and orientation of a proxy space are independent of the rest of the scene. Hence, objects can appear much smaller or larger in the proxy space but also can be seen from other perspectives (see Figure 6.6). Allowing users to freely pick their desired scale and orientation enables interactions that are hard or impossible to do otherwise. For example, a marked space that encompasses an entire room can allow for an easy rearranging of furniture.

Moving Objects into and out of a Proxy Space

Interaction with distant objects is not limited to the inside of proxies. A common need, for example, is to use these objects elsewhere in the scene. To enable this, objects can be moved into and out of proxy spaces (see

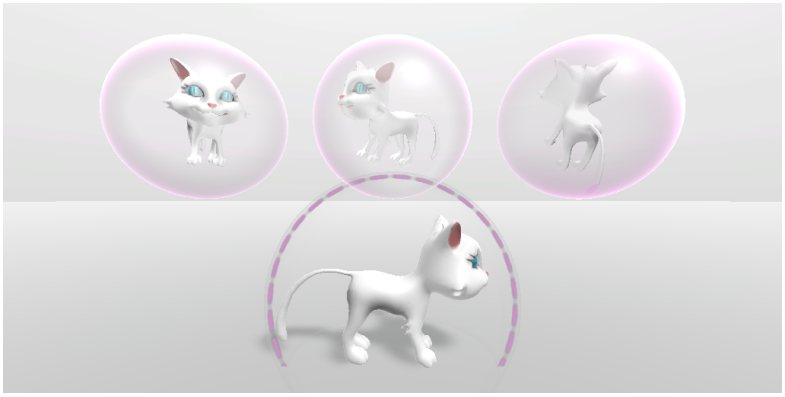

Figure 6.6.: Proxies allow for different views on the scene. Shown here are three proxies linked to the same marked space. Each is rotated in a different way, showing a cat from multiple perspectives.

Figure 6.7). When objects enter a proxy they exist only in the scene location, that is, at the marked space. Similarly, when the user takes an object out of the proxy, it only exists in the user's hand and stops existing at the distant location. As the proxy is only a representation of the actual space, an object never exists in two places at once and it always exists at its actual scale.

In most aspects, this behaves exactly as when objects are moved in the rest of the scene. However, with proxy spaces, a few situations arise that require special handling. One is dropping an object in a proxy that only links to an empty volume of space. If a user drops an object in such a proxy it will fall through the bottom of the marked space and come to rest at the distant location.

Another situation that needs special handling is moving objects to and from the proxy spaces of different scales. Because proxy spaces can be magnified or shrunk in comparison to the rest of the scene, objects also appear larger or smaller respectively, even though their actual size is unchanged. For example, when a marked space spans an entire room, the furniture appears much smaller inside a proxy space than it actually is. This brings benefits, such as being able to grasp objects that would otherwise be too large to grasp (e.g., shelves). However, it also requires transitioning to the object's actual scale when they are taken in and out of a proxy or moved between proxies. We handle this by keeping object sizes visually constant while they are moved to and from proxies. Once the user lets go of an object, it snaps back to its actual size. For example, if the furniture is taken out of the proxy described above, it appears small at first, but grows to full size as it is placed down. The same

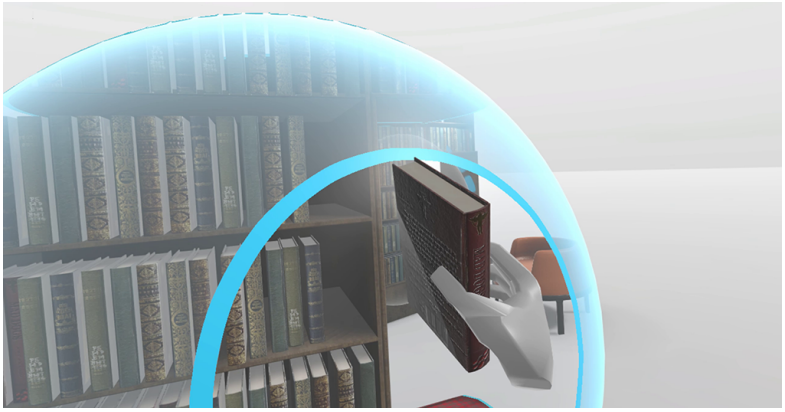

Figure 6.7.: Users can interact with objects inside of proxies, but also move objects into and out of them, like this book. If the proxy was scaled, objects taken out of it soon afterwards shrink or grow to their actual size.

happens when objects are being moved between proxies.

In addition to objects, users can also move *themselves* inside a proxy space (see Figure 6.8). For Poros, we restricted users to sticking their head into proxies, in order to "peek" at a distant space. When peeking, the proxy space expands, giving the user a wider view of the marked space. This allows for inspection of a distant space, with easy transition back to a normal perspective (i.e., by taking the head out of the proxy). When the user's head is fully inside a proxy, the whole view is identical to the one a user would have at the mark's location in the room—not clipped anymore as when looking at the proxy space from outside. Furthermore, if the proxy is visible from the mark's location then the user can see themselves within it.

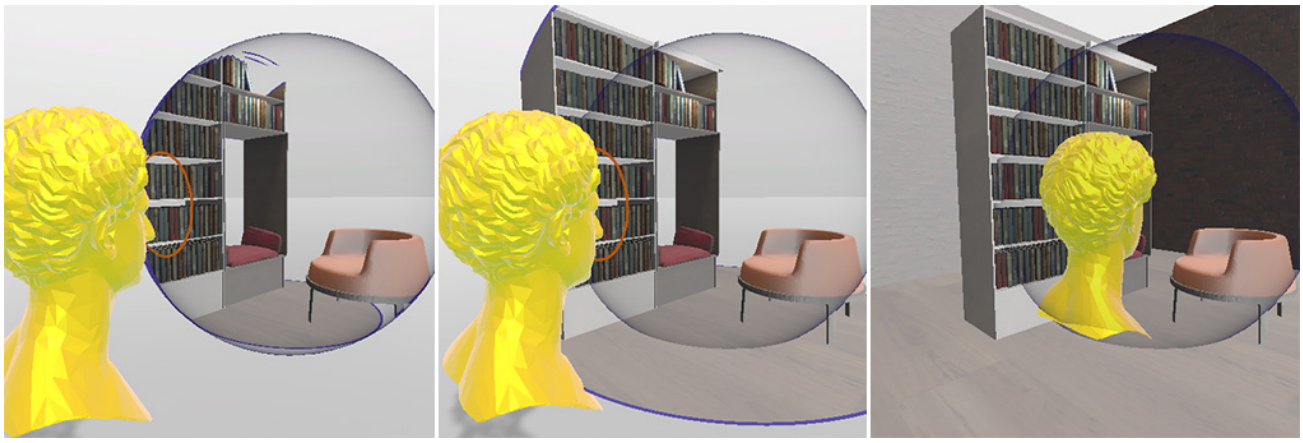

Figure 6.8.: Proxy spaces initially only show what is contained inside. As the user moves into a proxy space, the shown volume expands exponentially until the user is finally immersed in the linked marked space.

6.4.5. Manipulation of Proxies and Marks

As proxies become part of the scene, they also become available for interaction. Users can manipulate proxies, as well as the marked spaces linked to them with manual interactions. This allows users to, for example, arrange proxies, adjust mark locations, or take proxies along as they move around.

Proxy Manipulation

Just like other objects in the scene, proxies can be manipulated. In Poros we allow for translation, scaling, and rotation of proxies (see Figure 6.9 for an example of translation). These manipulations can be performed by grabbing the shell of the proxy with one or two hands respectively.

However, to prevent accidental changes to a proxy, we require a mode switch to enable manipulation. For this, users have to briefly rest their hand on a proxy's shell. Upon switching to manipulation mode, proxies no longer open up for the user, and appear more opaque. When users move away or stop interacting, proxies return to the default mode for easy access to their contents.

Sometimes a static arrangement of proxies is not sufficient. For example, users might want to take a proxy along when walking. To address this, we allow users to *anchor* proxies to other elements of the scene, including to themselves. To trigger anchoring the user rests one of their hands on a proxy while grabbing and pulling with the other. While the proxy stays in place, a tether emerges and follows the pulling hand. Users can then drag this tether to other objects for anchoring. To anchor a proxy to themselves, users drop the tether at a target that appears near their waist. An anchor can be disengaged by starting the anchoring process again, then releasing in open space.

Figure 6.9.: By resting their hand on the outside of a proxy, users can activate a manipulation mode. The proxy then turns more opaque and allows users to translate (shown here), scale, and rotate it with their hands.

While anchoring enables users to take proxies with them, proxies are not always needed and could obstruct other parts of the scene if always close. For this reason, we also provide a way for users to temporarily store away and bring along proxies (see Figure 6.10). When scaling a proxy below a certain size, it is instead minimized and flies to a storage location on the user's left forearm. Where anchoring to oneself keeps proxies in view, this alternative storage option allows users to move around without being obstructed by a proxy. Once users need to get access to a proxy again, they can restore it by lifting their forearm and selecting the desired proxy on it. The proxy then returns to its original size and appears in front of the user.

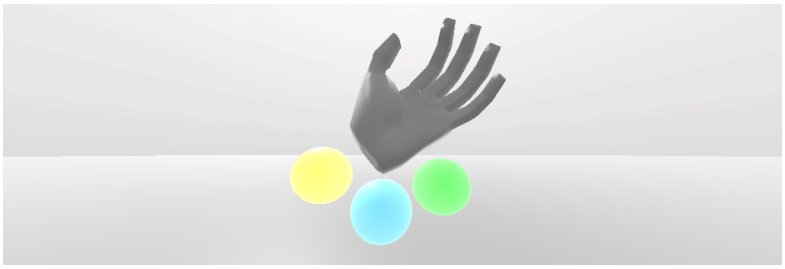

Figure 6.10.: Users can minimize proxies, which then attach to their wrist. This allows users to take them along. Holding the wrist up and touching a minimized proxy returns it to its original size.

Mark Manipulation

In addition to proxies themselves, we enable users to change the boundaries of the marked spaces they are linked to (see Figure 6.11). Marked spaces can be moved and resized, changing what part of the scene they encompass, but not the scene itself. As with proxies, mark manipulation needs to be activated. Users have to place their hand on the inside boundary of a proxy (i.e., when their hand is already in the marked space). This makes a handle appear which acts as a joystick for the mark. When grabbing it with their left hand, users can translate the marked space, while grabbing with their right hand controls its scale.

Marked spaces can be anchored to objects (see Figure 6.12) and users just the same as proxies. This can be useful, for example, if a distant object is moving. Users can anchor a marked space by pinching on the manipulation handle, dragging an emerging tether to an object, and releasing the pinch on it. Self-anchoring and detaching work as in proxy anchoring.

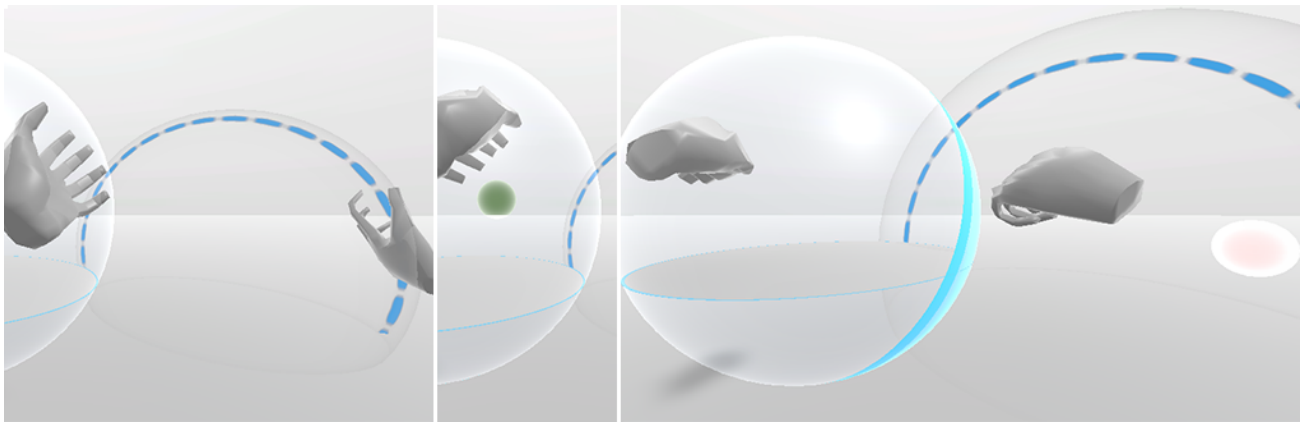

Figure 6.11.: Resting a hand in the inside of a proxy activates mark manipulation. A handle appears that can be used to translate and scale the mark. Shown here is how grabbing it with the left hand and dragging outwards expanded the marked space.

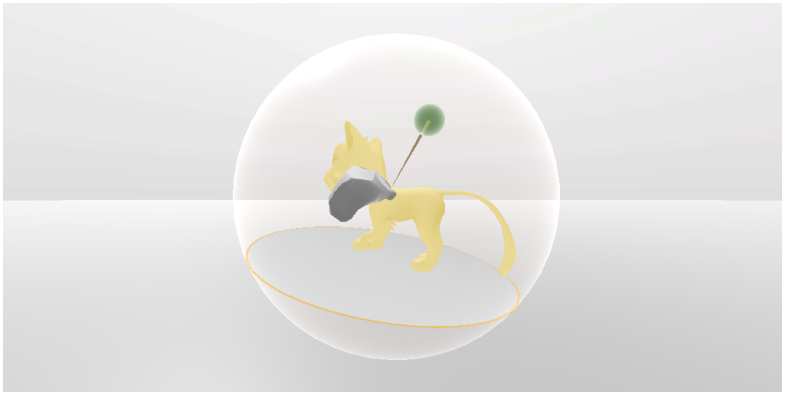

Figure 6.12.: Users can anchor marked spaces to objects within, like this cat. When in manipulation mode, pinching on the handle reveals a tether that can be dragged onto the object to anchor to. As the cat walks around, the marked space will now follow it.

6.4.6. Abstract Operations on Proxies

Poros also supports abstract operations on one proxy as well as multiple proxies. These can speed up simple manipulations (e.g., aligning of proxies) or enable interactions that would otherwise not be possible (e.g., highlighting of overlapping objects).

Triggering Operations

Operations are triggered by a hierarchical crossing menu, accessible through pinching. When pinching the surface of a proxy and dragging away, the menu shown in Figure 6.13 appears in the dragging direction. Moving through a menu item selects that item, or triggers the next menu within the hierarchy. The release position of the pinch can be used as parameter for single proxy operations, and to pick the second proxy in multi-proxy operations.

Single Proxy Operations

Poros supports four operations on single proxies: cloning, splitting, aligning to scene, and resetting.

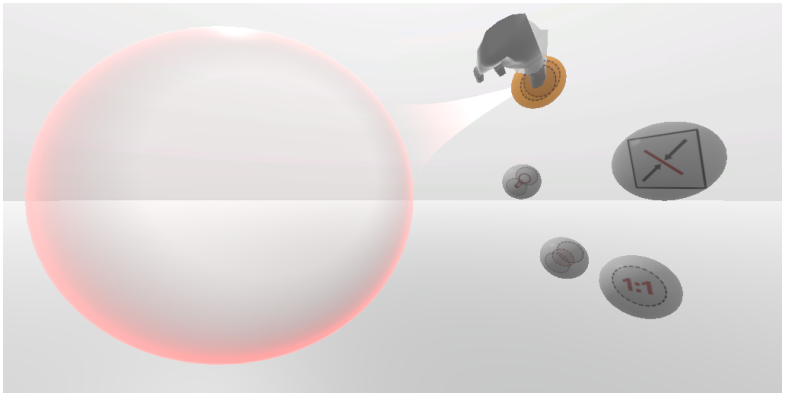

Figure 6.13.: Pinching the surface of a proxy and dragging away reveals a crossing menu with several operations to chose from. Menu items and submenus are selected by dragging through them. For operations with a target location or object, users continue dragging and release on them.

Cloning allows users to create another proxy space that is linked to the same marked space as the proxy the operation was used on. This can be useful to, for example, create multiple views on the same space in order to see an object from different angles.

Splitting is an operation only available on proxies that have previously been merged (see below). Such proxies are linked to multiple marked spaces. With this operation, the proxy splits up into separate proxies—one per linked marked space.

Aligning to the scene rotates a proxy to make it best fit in with the scene around it. For example, a proxy space containing a cabinet, when close to other furniture, would align with it. This operation makes use of the prevalent orientations within a proxy space, as well as the scene around it. We derive orientations from each object's representation in the physics simulation (e.g., as a box). Resulting from this are two sets of world-space direction vectors. We then use the Kabsch algorithm [46] to find the rotation matrix that best aligns the proxy space's content with the surroundings.

Resetting, reverts any orientation and scale differences between a proxy space and the linked marked space. Afterwards, the scale within the proxy is equal to the rest of the scene and everything within is facing the same direction it does in the scene.

Multi-Proxy Operations

Four additional operations work on multiple proxies at once: aligning, convenience aligning, merging, and highlighting. Users activate them with the same menu as above and, after command selection, then drag to the other proxy to include in the operation.

Aligning and **convenience aligning** both use the mechanism for *aligning to the scene* described above. However, instead of aligning to the scene, the former makes groups of proxies face in the same direction (e.g., having a row of cabinets all face forward). The latter takes into account the position of the user to make transfer of objects from one proxy to the other more convenient. In the case of two bookcases, for example, moving books between them is easier if both are tilted towards the user, instead of just facing forward. As object data can be ambiguous with respect to an object's main orientation, we currently manually annotate them with a preferred orientation (e.g., the forward direction for a bookcase).

Merging results in one proxy that is linked to multiple marked spaces. The spaces overlap and are all shown concurrently. While this can, at times, be visually confusing, it does allow for a set of advanced interactions. The user's hands in the proxy are mirrored in each marked space, all moving in unison and all with the ability

to interact with the scene. Hence, interaction in a proxy that is linked to multiple marked spaces is replicated across them. For example, pushing down on a button in one also results in pushes in all other spaces (if the buttons are aligned). This feature can be used to move multiple objects at once or trigger multiple actions at once. Such replication has the most potential where a scene contains multiple instances of the same interactive object. For example, consider turning on multiple machines, opening a row of windows, or playing with several slot machines at once.

Highlighting helps users search proxy spaces. Whatever is contained in one proxy defines a lens for another one. For example, one proxy encompassing just one egg can be used to find all the eggs in a second, much larger, proxy. In our current implementation, objects are tagged manually and search is then performed within these annotations. When highlighting, floating exclamation marks are temporarily attached to every matching instance. Users can then see the highlighting within the searched proxy, but also as part of the scene.

6.5. Example Applications

Poros enables novel applications but also allows for easy implementation of a range of existing interactions. In this section, we describe several examples of such applications and uses. See Figure 6.14 for an overview of these examples.

6.5.1. Occluded Object Interactions

Sometimes what we want to interact with is hidden behind other objects or otherwise hard to get to. For example, imagine a scenario where the user desires to turn on a PC that is under a table (e.g., as in [70]). Since the PC is hard to reach, the user can mark the power button and create a proxy at a more suitable location. For example, the user can place the proxy containing the power button on the table, while anchoring the mark to the PC (Figure 6.14-A). After rearranging the space the user can now easily access the PC's power button when at the table, even if the PC is moved.

6.5.2. Large and small scale interaction

Direct manipulation in VR breaks down if objects are much larger or smaller than the user. Poros allows the users to interact at a comfortable scale. By adjusting the size of the proxy and mark, users can make the objects within the proxy appear smaller or larger. For example, the users can easily rearrange furniture and other large objects by creating a mark spanning an entire room and linking it to a small proxy in front of them (see Figure 6.14-B). This functionality is similar to worlds in miniature [14, 133] and the *Voodoo Dolls* technique [109], while giving the user additional freedom over the scale and perspective. Alternatively, if the user is interacting with a tiny object then the mark and the proxy can be adjusted so that the objects within it appear larger. For example, the user can magnify a part of a book to read the small print (see Figure 6.14-C). This functionality supports low vision users similar to the tools proposed in *SeeingVR* [162], while keeping the magnified space fully interactive. When reading the book with small print, for example, users can flip pages and continue reading on the next one.

Users can also poke their head into such scaled proxies. This expands the proxy and allows the users to have a wider look at a distant marked space—effectively being teleported there while keeping the scale as it was in the proxy. This allows the user to feel like a giant or a dwarf within the scene and allowing perspectives and manipulations that are not possible at a normal scale [155, 162].

6.5.3. Observing Spaces and Oneself

Users often need to monitor or observe out of sight places. For example, consider a security guard monitoring a building. Such tasks commonly involve more than one space to be monitored, such as when several rooms

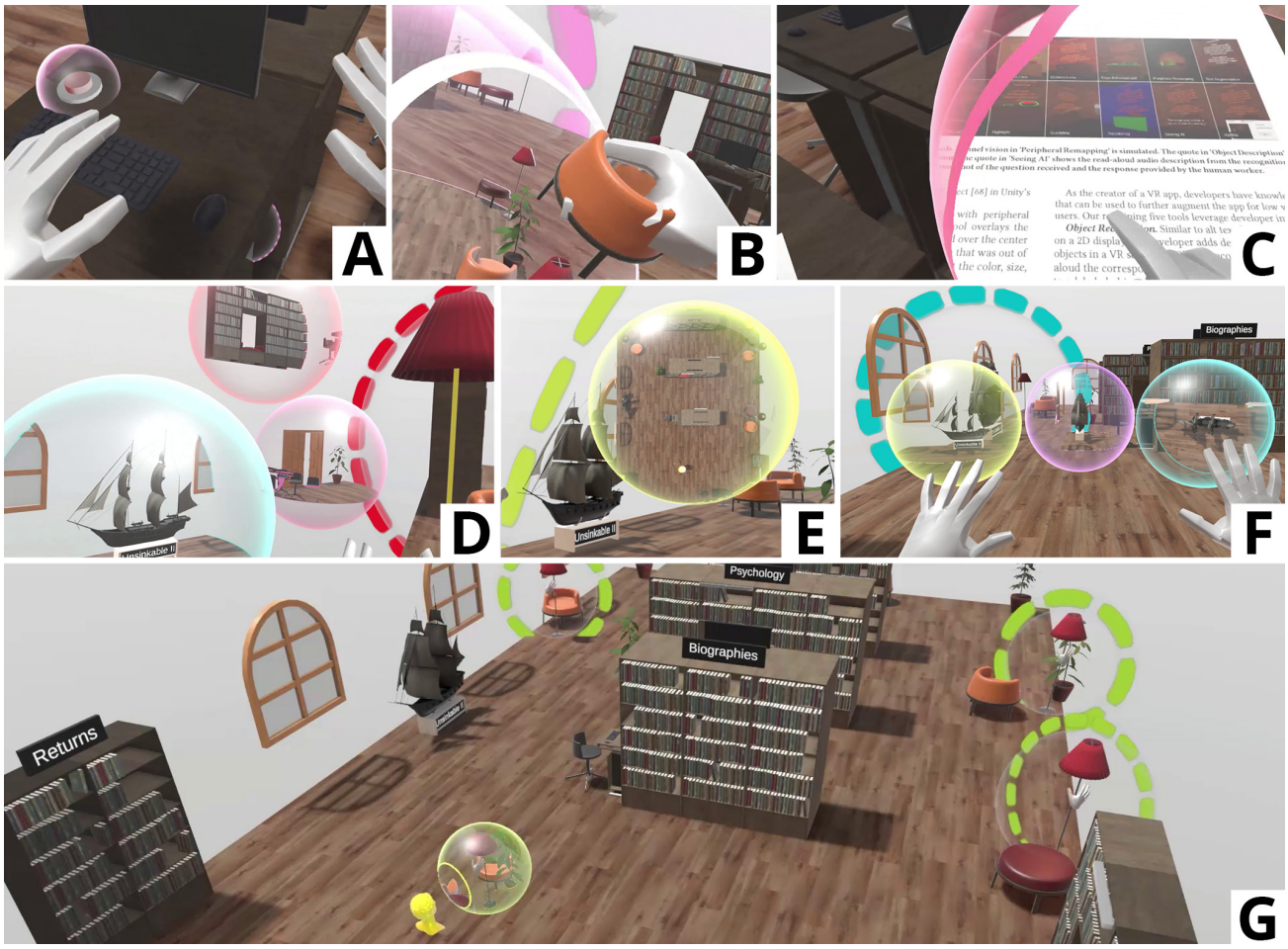

Figure 6.14.: We build a range of examples with Poros. Shown here are (A) re-configuration of space so the PC's power button is on the table, (B) moving furniture, (C) reading small print on books, (D) monitoring different parts of a room, (E) setting up a mini-map that includes oneself, (F) inspecting a sailing ship from multiple perspectives, and (G) replicating an action across multiple spaces to turn off several lamps at once.

are watched at once. Poros enables users to easily build ensembles of proxies to watch several distant spaces simultaneously. Figure 6.14-D shows a proxy configuration that allows the user to monitor different parts of a library. As the space within the proxy is fully interactive the user can quickly act on the remote space when needed (e.g., to close an open window). Users can also anchor the proxies to themselves to take them along as they move around the scene.

The ability to anchor proxies and marks also enables monitoring of moving objects in the scene. For example, when anchoring a mark to a cat, this allows the user to constantly monitor what the cat is doing and to manipulate its surrounding (e.g., cleaning up a mess the cat made). To observe themselves the users can anchor a mark to themselves For example, if the user anchors a room-sized mark and a top-down view proxy to themselves, they essentially create a mini-map centered around them (see Figure 6.14-E). Such anchoring could also be used to show users what is behind them all the time.

6.5.4. Multi-Perspective Object Manipulation

When manipulating objects, only seeing them from one perspective can hide important details and limit interaction. To help with this, many VR painting and modeling application enable change of perspective to more easily draw or manipulate objects [2][3]. Furthermore, modeling software like Blender[4] allows multiple-perspectives of the same object. With Poros, users can easily achieve both. To change the perspective they can transform the proxy, and to create a multi-perspective view on the object they can duplicate proxies and manipulate those to achieve the desired arrangement of perspectives. Figure 6.14-F shows a multi-perspective view of a sailing ship, allowing the user to quickly interact from multiple perspectives.

6.5.5. Replicated Interactions

Some tasks require repetition, such as filling the bowls of several pets or opening all windows in a room. To facilitate such tasks, Poros allows for the replication of interactions. After merging proxies, one proxy space is connected to multiple marked spaces and hence any user action is replicated across them. As an example, Figure 6.14-G shows how one proxy is linked to several marks around a series of lamps. Users can reach into the proxy and pull on the cord switch to turn on all lamps at once. For this to work, alignment of the different cords is necessary and thus users could first trigger an *align within* operation on the proxy. In aligned spaces, switching on one lamp is no different than switching on any number of lamps. Similarly, users could move multiple objects at once, or fill the bowls mentioned above.

6.5.6. Searching Spaces

To manipulate objects, users first need to find them, which can be time-consuming in complex scenes. For example, consider a kitchen, library, or archive, which all contain many similar objects. The highlighting operation in Poros can help users find the objects they want to interact with. Proxies here act similar to Perlin and Fox's *portal filters* [106]. In Poros, however, the view is not just changed inside a proxy, but instead the scene itself shows the highlighting. This allows users to also see results in the larger context of the scene.

6.5.7. Organizing Shelves

In scenes that contain many items, organizing those can be an important task. For example, a user could desire to move all shirts to the other side of a store. When many objects need to be moved between places far apart, this necessitates a substantial amount of locomotion from the user. With Poros, users can create a workspace that is optimized for this task. For example, a user would create two proxies, linked to two different shelves. The shelves can be far apart, but the proxies can be arranged next to each other. By making use of the *aligning for convenience* operation, users can then also have the two proxies rotate to make moving items from one to the other easier. This results in a setup where no locomotion is required and users can move objects with little effort. Furthermore, this could be combined with anchoring to take a proxy along or place items on a moving target.

6.6. Expert Evaluation

Poros enables a variety of interactions, therefore there is no single performance metric that would have covered them all. Furthermore, there are distinct trade-offs for different tasks, making comparison across the wide set of scenarios problematic. Thus, rather than conducting an empirical evaluation comparing a component of Poros to

[2] Tilt Brush, https://www.tiltbrush.com/
[3] Blocks, https://arvr.google.com/blocks/
[4] https://www.blender.org/

another technique, we opted for an expert evaluation focused on conceptual understanding, breadth, and overall experience. Hence, a limitation of this study is that we are not able to draw any conclusions about comparisons to possible alternative techniques in each scenario. However, the experts do hint towards comparable techniques which could be useful for future empirical evaluations.

We invited nine experienced VR developers and researchers (with at least 2 and on average 5.7 years of VR experience) to our lab. They watched a video introducing Poros beforehand and then each had 30 minutes to work with the system in four scenarios: (1) large and (2) small scale interactions (per Section 6.5.2), (3) proxy and mark creation and manipulation (as in the first part of Section 6.5.7), and (4) proxy duplication (Section 6.5.4). The scenarios were identical to the examples described earlier and exposed the experts to a wide range of uses.

Throughout the evaluation, we asked the participants to talk about their thoughts and actions. At the end, we also interviewed them using open-ended questions. We audio recorded each evaluation and transcribed the recordings. For the analysis we group participant quotes by themes and report on each theme below.

6.6.1. Concept

Generally, most users "got the concept pretty quickly" (P1). As P2 noted, "when [the proxy] is placed there, and you can just sort of put your head in it, then I think it is very easy to understand what is going on. Like that aspect of it is very good. I didn't really need an introduction" (P2). Some likened it to other concepts: "was it Super Mario or Crash Bandicoot that has spheres and then you go to those portals, like the old old PlayStation games" (P7).

The participants could also conceptually distinguish Poros from teleportation: "So I see it as like an alternative to teleporting around" (P2). They also saw clear advantages of Poros over teleportation "In that way [in reference to a task which requires moving between places frequently], if you teleport around, it's a lot of teleportation" (P1). Furthermore, P8 identified another difference to locomotion techniques: "and you can move large objects, which you wouldn't be able to, if you use locomotion to get to that thing".

From a short demonstration, participants were able to easily understand the concept visually, and also see the differences and advantages when compared to more traditional locomotion techniques such as teleportation.

6.6.2. Learnability and Consistency Across Interactions

Some experts noted that the "controls require a fair amount of memorization" (P1), but that "once you get familiar with it, I think it's very nice to go from one space to another space" (P5).

Some of the interactions were perceived to be natural: "it's so one to one what is going on so you don't have to learn it" (P2). On the other hand, experts felt that the interactions were inconsistent. For instance, P6 stated that "the biggest thing [usability flaw] was when you create the sphere you do this, but then you want to change its size afterwards you have to do something else".

P3 suggested manipulating the two spaces in the same way (by grabbing them with fists) especially for smaller marked spaces: "So I think for small scale things, that would be super useful to be able to just pick it up, and then stand and look and move it around. If I'm inspecting something very small, instead of having the controls inside". Hence, by making the interactions more consistent, learnability could be improved.

6.6.3. Additional Feedback

Participants noted that they "miss extra feedback. Like some sort of a haptic or sound when you do something" (P5). P8 also mentioned haptic feedback that could ease the difficulty in knowing when the hands are in contact

with objects: "One thing I'm missing is like haptic feedback when you touch the spheres, so when you know that, okay, now it's touching outside — or inside". Future versions of Poros should include more feedback to let the user know when actions are performed or which mode they are in. Ideally, haptics would help the user to know if they are touching the spheres.

6.6.4. Design Choices

In this implementation, the hand size from the users perspective is kept constant. One user noted that, while this is sometimes beneficial, "it's nice to have a large hand if you want to grab a couch, for instance" (P1) this may make some interactions difficult and proposes to have "the option to scale it because if I read a paper, for instance, it's nice to you know, grab the paper". Further exploration is necessary to understand the situations where certain hand sizes are beneficial and also to allow user control over this.

Some found it difficult to position the marks to exactly where they needed to be during creation. P5 mentioned "I think the marking goes too fast. Can it be more precise, a little bit slower?" P2 described an alternative to the gesture driven movement, but then suggested he "struggles with moving it and scaling it and all that was more a matter of hand tracking." Creation and manipulation seems difficult in certain situations and is difficult to cater for all, but there may be ways to let the user control parameters of this, or to change it automatically based on context.

6.6.5. Utility and Applications

Several experts commented on potential applications of Poros. One expert mentioned that it could be used as a "very detailed inventory system" (P2) that could replace traditional menu based inventories with icons, with spatial references to objects. This suggestion was echoed by P4: "most training apps use menus that you have in a hand that kind of gives you the power to do things or, or tools that are ideal for this specific use case". They specifically recalled a training app for fire investigations which could be fitting: "where you can be multiple people and you can walk around in a scanned, burned out building and place evidence, numbers and take photos. It's kind of like those tools".

P3 found parallels to real world interactions with the small-scale effect: "you can do fine grained manipulation, but with larger body movements, so I think it would be useful for that. ... in that sense, it's not that different from what they do with like, robotic surgery work".

P1 mentioned how it could be useful to clean up a room and return items quickly to where they belong: " I clean it all up and that takes forever because everything is scattered and they go to different places. So I can definitely see like if I had it [Poros], it's a bit like putting six boxes in front of me books in this one, Lego in that one, dirty clothes in that one...". Therefore, they used a series of boxes as a real life analogy of Poros, for moving objects to pre-designated spaces.

Although we did not investigate collaborate tasks, several experts commented on this: "You probably also use it collaboratively so that you're able to all of you share the perspective together, so I think especially for any challenge where we collaborate on something very, very small" (P2). P7 recalled a project where this could facilitate it either collaboratively/socially: "some people in Minecraft are trying to build the whole world. And if people are going to visit different parts of the world, different cities, then they can have a sphere and then they can just have a kind of multi tasking window in that virtual 3d space there and then see what other people are doing in another part of the world".

We finally had comments about VR replacing desktop interfaces: "As we move further towards virtual reality becoming more and more used ... I can see this working as a really good workplace, maybe just as a desktop" (P9). There, Poros could be helpful "like a 3D room as some kind of menu and then being able to interact easily with everything in the room".

6.6.6. Summary

The expert evaluation showed that the concept was easy to grasp, clearly different from locomotion techniques, and perceived useful across a range of applications. Having established that the system was conceptually understood well and that these tasks can be completed, we can now directly compare Poros empirically (or variations of it) to other techniques for interacting with distant objects. Regarding usability, experts mentioned a need for higher consistency between actions and additional feedback for interactions. There is further exploration with respect to either intelligently adjusting parameters of hand size, creation and manipulation based on context, or to enable user control of these.

6.7. Discussion

Poros tackles an inherent limitation of spatial interfaces: Interaction at a distance is cumbersome. Users commonly need to traverse space and can only be in one location at once. With Poros, users can rearrange space for a given task, ensuring that what they need is within reach, wherever it is. They can glance at one or more distant locations and move objects between them. Furthermore, proxies can provide more powerful tools than just access to distant space. What they encapsulate can be operated on—used to filter, scale, or align spaces. As we have shown, this enables a range of applications that are complicated or impossible with existing technologies.

Whether setting up the kind of workspace enabled by Poros is beneficial depends on the nature of the user's activity. For one-off interactions, or when users want to be at another location for a longer time, setting up a collection of proxies might be too costly. While there is a setup cost when using Poros, subsequent interactions with distant objects are essentially "free". This is in contrast to common locomotion techniques, where no setup is necessary, but each movement incurs a cost (e.g., due to a need to reorient). Just as selection/manipulation and travel techniques are generally considered distinct [19], we consider Poros and the latter complementary—each suited to distinct situations.

Poros is designed for VR, but also builds upon work in 2D user interfaces where surrogates [64] are commonly employed to modify "distant" objects. For example, with wall-sized interfaces users also need to walk to interact and hence techniques like *Frisbee* [51] have been developed. Conceptually, this is the 2D analogy to Poros: a local "telescope" through which one can see and interact with elements at a remote destination. One of the benefits of surrogates is that they allow for non-destructive re-configuration of space. For example, when creating a proxy to a bookshelf, the original scene is still kept intact for later interactions and other users.

Enabling users to configure their workspace to fit a task, Poros brings to VR what is common in other forms of interactive computing. As we have proxies that can be arranged to make moving books easier, desktop user interfaces make it easy to place windows next to each other, in order to allow for convenient drag and drop operations, and context switching. Poros hence shows that VR space can be made just as malleable as other interactive spaces.

6.7.1. Future Work

There are a range of possible extensions to Poros that could be explored. For example, we only allow users to observe themselves in marked spaces, yet not to manipulate their avatar (e.g., picking themselves up to move to a different location). While we enable users to replicate an action in several linked spaces, this currently requires good alignment. Instead, a future version could adjust the hands in each linked space to their specific context. For example, opening a window in one marked space could snap the hand to window nearby window handles in all the other spaces, alleviating the need for good alignment. In Poros, we have also decided to prevent object duplication. However, conceptually, a proxy space linked to multiple marked spaces would offer an opportunity to do just that. Placing a book into such a proxy could result in a different copy coming into existence in each linked marked space.

In this implementation, we have chosen to define spaces spherically. A more complex system could explore using shapes that adapt intelligently to content, either based on the outline of objects or perhaps based on context. Alternatively, users could be given finer control over the shape of marks and proxies.

6.8. Conclusion

We have described Poros, where users can manipulate the space around them through proxies and also perform a variety of interactions on these. Through combining these interactions, we enable many applications with varying complexities. These range from interactions that other systems have enabled individually, to new interactions such as replicating actions across multiple spaces, aligning spaces, and interacting through different perspectives. We have shown several examples of how Poros can be used for interactions that are hard to do or impossible with existing techniques. The source code for Poros is available and we hope to inspire follow-up techniques. With more work moving into VR, techniques like Poros will be important to allow users to reconfigure their workspaces get the most out of their use of VR.

7. Who Put That There? Temporal Navigation of Spatial Recordings by Direct Manipulation

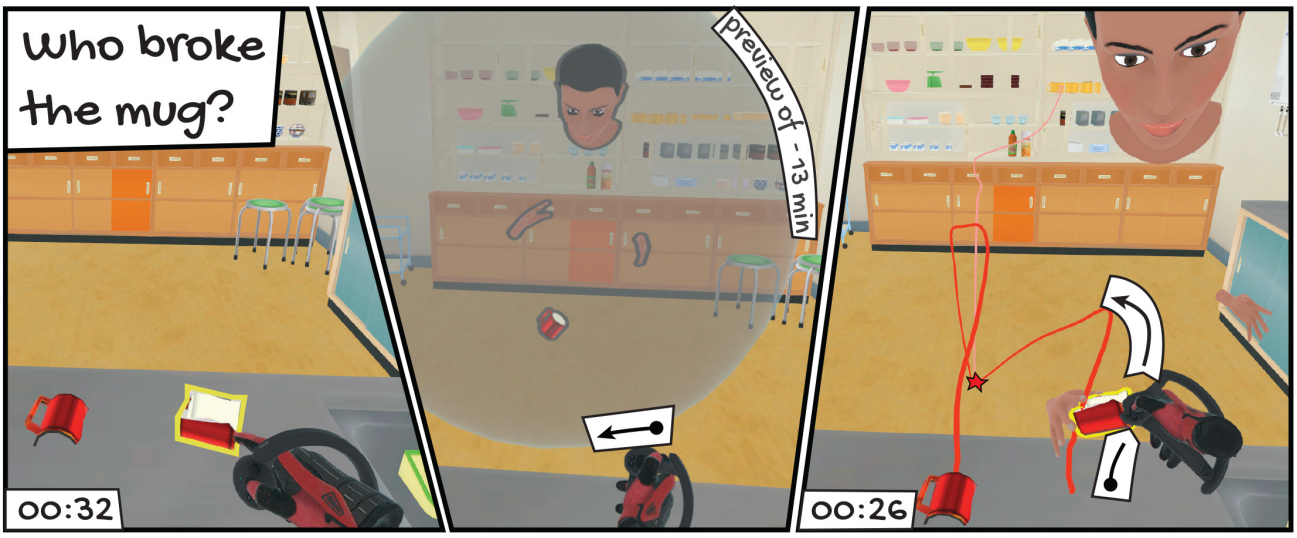

Figure 7.1.: The viewer is in a spatial recording of a virtual kitchen and would like to find out who broke the mug on the kitchen counter (left). By touching one of the mug pieces and stepping through its changes, the viewer previews the accident which happened 13 minutes ago (middle). To see the full story of the mug, the viewer investigates its trajectory and drags the broken piece along it to navigate to the moment it was put on the kitchen counter, 26 minutes into the recording. The *Who Put That There* system enables viewers to preview and navigate by an object's changes.

7.1. Abstract

Spatial recordings allow viewers to move within them and freely choose their viewpoint. However, such recordings make it easy to miss events and difficult to follow moving objects when skipping through the recording. To alleviate these problems we present the *Who Put That There* system that allows users to navigate through time by directly manipulating objects in the scene. By selecting an object, the user can navigate to moments where the object changed. Users can also view trajectories of objects that changed location and directly manipulate them to navigate. We evaluated the system with a set of sensemaking questions in a think-aloud study. Participants understood the system and found it useful for finding events of interest, while being present and engaged in the recording.

7.2. Introduction

Spatial recordings include virtual reality (VR) captures, volumetric recordings, motion captures, and 3D video game replays. In contrast to traditional video, the viewer is *in* a spatial recording and can move within it.

Currently, more and more spatial recordings are being created. For example, modern 3D video games such as Fortnite come with the ability to record spatial recordings[1], which can be shared with others. Media companies have also started experimenting with production of VR movies[2] and volumetric captures[3]. Furthermore, future instrumented rooms [73, 99] and augmented reality (AR) devices will likely increase the amount of spatial recordings.

Spatial recordings are viewed differently than traditional video. For example, while videos impose a fixed viewpoint, spatial recordings allow viewers to choose their own view and location within them. This gives viewers more control over the experience, but also puts a burden on them to use this freedom effectively. While spatial recordings are consumed differently, the main controls for actively exploring them are the same as for traditional video: a timeline as well as play, stop, fast-forward, and rewind. These controls limit how well viewers can interact with spatial recordings and can cause issues, such as missing of events, motion sickness, or a break of immersion. Further, limiting interaction to the control of *time* complicates exploration of a recording. Navigating by control of time requires skipping to find the moment when an interesting change occurred in the scene. This is especially burdensome for long recordings where it is easy to miss a relevant moment and where the resolution of the timeline is limited.

We propose to structure navigation through spatial recordings along the *changes of objects* therein. Viewers can explore and navigate recordings by directly interacting with objects. Instead of scrubbing through time, viewers can preview an object's changes and go to the moment a change occurred. Continuous changes (e.g., change of location) are previewed in the form of trajectories, while discrete changes (e.g., shape change) are previewed as an animated loop of the change. Figure 7.1 shows an example of such previews. Navigating by manipulation of objects places control *into* the scene and ensures that viewers do not miss events of interest. This enables more efficient access to interesting parts of a recording and supports sensemaking by allowing users to infer relations among objects. For example, a viewer of a spatially recorded basketball game can inspect the ball's trajectory to see if a player touched it before it went out of bounds. In a spatial recording of a game the viewer might wonder why, when and how a house was build. By stepping through house's changes the viewer can see or infer the answers.

We exemplify the concept via the *Who Put That There* temporal navigation system for spatial VR recordings. The system contains a set of direct manipulation techniques which allow users to step through objects' changes and navigate by objects' trajectories. The system also allows scrubbing through time by using a timeline, as well as rewind, fast-forward, play and pause of the recording by using conventional media controls. We evaluated the system in a think-aloud study using a spatial recording of a virtual kitchen and a set of sensemaking questions. The results confirmed the usefulness of the system and provided insights for designing future object-based temporal navigation systems.

7.3. Related Work

Earlier work on spatial recordings focused on ways of capturing them [48, 73, 99, 104] and less so on systems and techniques that improve the viewing experience. In the next subsections we first present conventional ways to temporally navigate recordings and their application to spatial recordings. Then we present the related work on direct manipulation techniques for temporal navigation of recordings.

7.3.1. Conventional Temporal Navigation

Conventional temporal navigation systems typically consist of a timeline as well as play, stop, fast-forward, and rewind controls. These help users find events of interest or get a quick overview of what happened in the

[1] Fortnite replay system, https://www.epicgames.com/fortnite/en-US/news/fortnite-battle-royale-replay-system
[2] Disney's *Cycles*, https://doi.org/10.1145/3214745.3214818
[3] Zero Days VR, https://www.zerodaysvr.com/

video. Such tools can be complimented with automatically generated video summaries (e.g., a teaser) and video abstractions (e.g., a storyboard). Borgo et al. referred to these techniques as *video visualization* and provided a good overview of them in his recent report [16]. It is unclear how the techniques and research findings related to viewing of conventional videos apply to spatial recordings. For example, a study of fast-forward speeds for conventional videos showed that a speed of 1:64 works well for the viewers to still comprehend changes in the scene [152]. However, using such fast-forward speed in spatial recording might make the viewers nauseous or make them miss important events in their periphery. Similarly, conventional techniques for automatic video abstraction are not easily transferable to spatial recordings.

7.3.2. Conventional Temporal Navigation for Spatial Recordings

Current user interfaces for temporal navigation of spatial recording typically use the same controls as the conventional video user interfaces (e.g., [94, 95, 142]). Using conventional tools to navigate spatial recordings can cause issues such as missing of events, motion sickness and breaking of immersion. To combat the motion sickness caused by speeding through the recording, Nguyen et al. proposed a technique which shrinks users' field of view (FOV) dynamically, depending on the motion in a recording [95]. While this technique does mitigate the motion sickness, it might break the immersion or make it more likely to miss events of interest. To avoid missing of events Liu et al. introduced view-dependent video textures [77] which loop a part of the scene until the viewer turns to a right direction to notice an important event. This technique is more suitable for passive viewing than active exploration of spatial recordings. Furthermore, it requires that the producer marks the important events in advance. To avoid braking of the immersion (e.g., while scrubbing the timeline) researchers have investigated more natural ways to interact with recordings. One way to accomplish this is by using gestures for temporal navigation [55, 78, 107, 120]. However, these gestures are often detached from the scene and mostly act as a substitute for play, pause, and rewind buttons.

7.3.3. Temporal Navigation by Direct Manipulation

In 1999, Satou and colleagues introduced the idea of using direct manipulation to temporally navigate videos [122]. They placed polygonal lines over a video to trace objects' trajectories. These lines could then be used as spatio-temporal sliders. For example, by clicking on tennis ball trajectory, the video frame changed to the time when the tennis ball was at the clicked location. The idea to couple the temporal control to video's spatial information gained traction a few years later when researchers investigated how to automate the technique [27, 34, 49, 53, 96]. They developed methods for extracting objects' trajectories, dragging of objects in the video, and ways of dealing with a moving background and change of camera's perspective.

Many of these challenges are specific to using objects' trajectories in a traditional video. Since a conventional video recording is a 2D representation of 3D information, the spatial information is skewed. For example, the distance of an object can only be gauged by its size. Furthermore, the enforced static viewpoint limits the possible interactions as the users cannot move to, for example, look behind an occluder or to follow an object's trajectory out of the current view.

There have been few explorations of direct manipulation techniques in spatial recordings. We are only aware of work by Kuhlen and colleagues, exploring navigation of scientific visualization [61, 153, 163]. Those systems were designed to navigate blood cell simulations, the workings of an impeller and similar. Therefore the interface was designed to assist detailed observation by controlling animation speed, zooming into sub-spaces and marking the spatio-temporal areas of interest. While such techniques incorporate aspects of object-based navigation, there were designed for a specialized use case. Furthermore, the techniques focused only on movement based changes (i.e., trajectories) without considering changes of appearance, configuration, topology and interactions between objects. We believe that temporal navigation by a variety of objects' changes is beneficial for a multitude of spatial recordings, not only for scientific visualizations. Next, we present our interaction concept and a system that exemplifies it.

7.4. Direct Manipulation for Spatial Recordings

Consider a spatial recording of a capture the flag match. A common way to watch it is to follow a player from the beginning of the recording to the end. However, not all of the recording is equally interesting. Rather, there are pivotal moments in a match, such as when a flag is stolen or a player enters the enemy compound. We posit that structuring and navigating recording by changes is especially useful for spatial recordings. Viewers are able to explore and navigate through a recording by directly interacting with changing objects, instead of having to use a timeline.

7.4.1. Concept

Figure 7.2 illustrates the concept. The fundamental idea is that viewers of spatial recordings are interested in finding moments of change. Instead of focusing on manipulation of time, we make *changing objects* and their *direct manipulation* central to viewing and navigation.

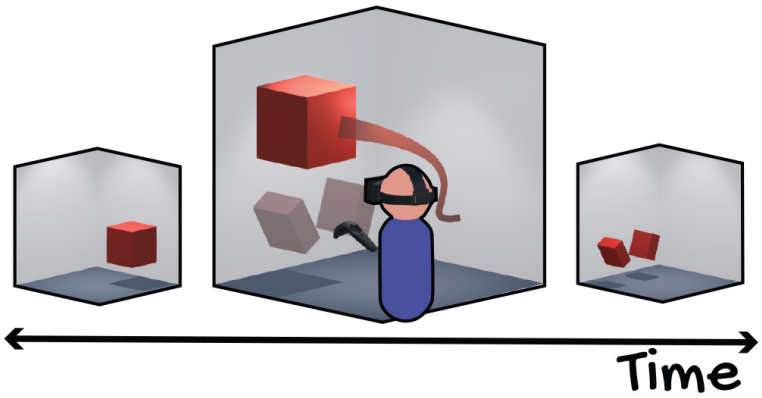

Figure 7.2.: Our concept for interacting with spatial recordings builds on two ideas: (1) The user is interested in changes (e.g., movement of the red cube). (2) Those changes are shown in the scene as previews, for instance as a trajectory showing the cubes movement throughout the recording or as an insert showing that an object has split. Both of these may be used to navigate to the corresponding moment in time.

Change concerns the properties of objects or people (we use the former as a placeholder for both). For example, objects can move, change size or color, split, disappear, change internally or change relation to another object. Depending on the specific form of spatial recording and the nature of the scene, different properties are relevant for the viewer. For example, in a football match, the players' movement, passes, strikes, and penalties can be crucial. While in a poker game the change of emotion or relation between two gazes could interest the viewer. Moreover, changes can be brief, constituting an event, such as a goal being scored, or they can happen over a longer duration of time (e.g., a player dribbling a ball from one side of the field to the other). For spatial recordings, such changes may be given as part of the recording (e.g., objects' events in a VR application). They may also be derived from the recording, for instance by activity-recognition algorithms [67, 136, 141], or through techniques for finding moments of interest (i.e., keyframes) in a recording [146]. Changes structure spatial recordings into meaningful sequences and moments. In comparison to timeline navigation, the focus on changes in many situations aligns better to viewers' questions and interests. With timeline controls, changes are secondary to the movement through time; users need to scrub the timeline while looking around the scene to see the changes occurring.

The second key idea is that a viewer can *directly manipulate* the objects of change to view the spatial recording. For instance, a user might select an object to inspect how it changed over time. Figure 7.2 shows a user selecting a cube to *preview* its changes; the trace of the cube's movement and the moment it split in half. The previewed

changes are shown directly in the scene, keeping the context of the spatial recording. How such changes are shown depends on their nature. If an object moved, this can be visualized as a trajectory (as in [27]). If an object changed shape, the changed object can be shown at the location of the change. With *navigation*, users can directly manipulate objects to move between episodes of change by acting upon a preview and being transported to that moment in the spatial recording. When changes are gradual (e.g., movement, scaling, color change) users can navigate via small units of change. For example, an object's trajectory can act as a non-linear time slider on which users can scrub. If the change is momentary (e.g., object being cut in half), then users are transported to the beginning or end of the change. The change-to-time mapping hence can be discrete or continuous, depending on the nature of the change itself.

In comparison to timeline navigation, direct manipulation of objects places the navigation directly in the spatial recording. For example, viewers can select an object and jump to the moment in time it was last moved. Instead of manipulating *time*, viewers manipulate *objects* for the same effect.

7.4.2. Benefits

Navigating spatial recordings by direct manipulation of objects within the scene has a number of benefits.

Easy Mapping of Intentions to Actions

Embedding navigation actions in a scene makes it easier to find out how to navigate by making the actions visible and concrete. When navigating, users are interested in changes within a scene. To navigate they manipulate the scene. This means that the input and output vocabularies are similar which allows users to easily transfer their intentions into actions. For example, if a user is interested in putting *that* over *there* then pointing to *that* and then *there* is an easily discoverable and executable action [15]. Similarly, when viewers are interested in who put *that* there they can point to it and find out.

Changes are Visible In-Scene

Previewing changes in the scene guides the users' view during navigation. For example, a viewer watching a spatial recording of a theater performance might wonder how a prop entered the stage. Instead of scrubbing through the timeline, the viewer can directly query the object. Once the trace of the object is revealed, the viewers can use it to temporally navigate by manipulating the prop's position, knowing in advance where to focus their gaze. This helps users avoid missing events of interest (e.g., where the prop came from) and being overwhelmed by irrelevant changes when scrubbing the timeline.

Integrated Coarse and Fine-Grained Navigation

Our concept integrates coarse and fine-grained navigation. The temporal resolution of a timeline control is constant and limited by its width. This limitation does not apply to navigation via objects' changes. For example, consider a viewer who is interested in the movement of a tiger in a ten-hour spatial recording. The tiger might be asleep for the first five hours, then wake up and walk around slowly for a bit, sprint to the boundary of the enclosure, and then go back to sleep for another four hours. Viewers scrubbing through such a recording could easily miss the sprinting tiger. Yet, when navigating by changes of the tiger's state, this is easily found. At the same time, fine-grained navigation is supported. With our concept, the tiger's movement trace could be shown as preview and allow users to scrub *along the trace*. Hence, temporal scrubbing is replaced by spatial scrubbing (with each position linked to a point in time). While timelines are limited in space, such trajectories can be much longer. They naturally have higher resolution when more change is happening (e.g., movement).

Sensemaking

Our concept supports sensemaking by allowing key questions (e.g., who, what, when) to be answered in ways timelines cannot support. Users can inspect the subjects in question and receive quick answers (e.g., "who left the burger on the table?"). Changes happening to objects are more likely to relate to viewers' questions and interests. With a timeline, questions are secondary to control of time. Users need to look around the scene while scrubbing through time to find the moment or absence of change.

Navigating a recording by using objects' changes also allows for easy discovery of causal sequences. For example, consider a recording where you see a broken coffee mug (as in Figure 7.1). With direct manipulation, the viewer can navigate to the moments the mug was used (i.e., touched, moved, filled or broken). At that point, the viewer might notice other things in the recording, such as a person interacting with the coffee drinker. The viewer could then follow that person to go back in time to when they entered the room, uncovering what they came in for (and why they disturbed the coffee drinker).

7.5. The Who Put That There System

We designed a system that enables users to temporally navigate spatial VR recordings by direct manipulation of objects within the scene. The *Who Put That There* system tracks changes of location, size, appearance (e.g., a pot getting dirty), configuration (e.g., a stove being turned on), and change of shape or topology (e.g., an object breaking), as well as interactions with an avatar (i.e., object being thrown). Users can select an object and preview its changes in two ways: (1) by stepping through notable changes, or (2) by showing the object's trajectory, indicating the change in location. Users can then navigate via the previewed changes by either selecting one of them, or by scrubbing the object's trajectory. The system also includes a timeline and conventional media controls to play, pause, fast-forward, and rewind the spatial recording.

7.5.1. Apparatus

We implemented the *Who Put That There* system using *Unity3D*. We used a *HTC Vive Pro* with *Valve Index Controllers* for all testing and evaluation of the system. These controllers allowed us to implement direct grasping interactions with objects. They also contain triggers and a thumbstick, which we mapped system commands to.

7.5.2. Stepping Through Objects' Changes

To access the moments where an object changed, the users can select objects within the scene and step through their notable changes (see Figure 7.3). The changes we define as notable are: object going from static to moving or the other way around, object being grasped or released, change of configuration (e.g., microwave being opened or closed), appearance (e.g., a plate getting dirty), and shape or topology (e.g., burger being bitten in). Users can select an object by hovering the controller over it and then use the thumbstick to preview what happened to the object before or after their current time in the spatial recording. For example, Figure 7.3 shows what happened to the burger before and after being put on the kitchen counter. Once the preview is active the user can continue stepping through all the changes by using the thumbstick.

The preview is rendered in transparent sphere to separate it from the rest of the scene. It shows the selected moment of change as a looped animation. The objects outside the preview sphere stay static at the current time in the recording. To further distinguish the previewed objects, we render a gray outline around the previewed ones. Scale and position are retained and viewers can move to observe the preview from different angles. This preserves the context and supports viewers' sensemaking activities.

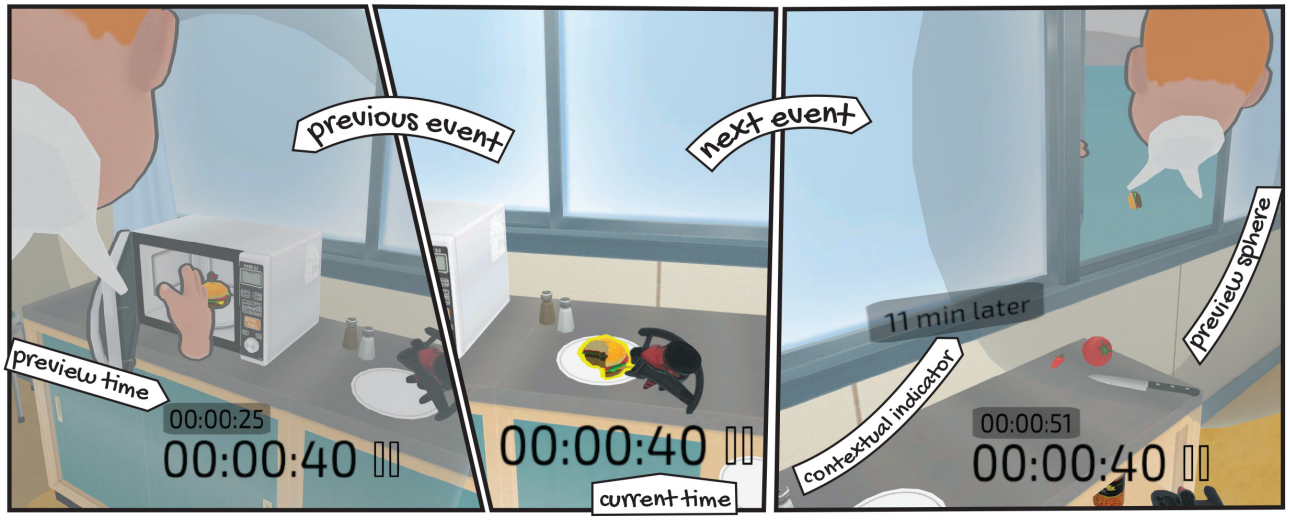

Figure 7.3.: The stepping through objects' changes technique allows users to preview the moments the object had changed. The user can select an object (middle - burger) and step through previous (left - burger being taken out of the microwave) or next changes (right - burger being thrown out of the window) relative to the current point in the recording (00:00:40). The preview is shown within a sphere around the queried object. Objects outside of the sphere are at the current point in the recording, while the objects within the sphere are shown as they are at the previewed moment (00:00:25 and 00:00:51 into the recording). When stepping through changes a contextual indicator floats towards the shown preview while showing the relative time distance to it.

Some objects stay static throughout the whole recording. To identify those that did change we included a *reveal objects* functionality which flashes the objects within the scene that changed at any point throughout the recording. Further, when stepping through an object's changes we show a contextual indicator (similar to [23]) which directs the user to the previewed change and indicates the relative time to it.

Skipping through previews allows viewers to quickly skim through all of an object's changes without changing the whole scene around them. Essentially, viewers can peek into the past or future while staying grounded at their current position in the recording. However, we also allow viewers to fully transition to a previewed moment by selecting it. The spatial recording then skips to that moment in time and hence updates the rest of the scene. Where previewing allows viewers to select interesting moments, transitioning to these moments enables a broader set of activities. For example, in a spatial recording of a kitchen a viewer might find the moment the knife was last used. To see exactly how it was used and what else was happening in the other parts of the kitchen they can navigate to that moment and inspect the rest of the scene. For example, they might discover that other people were in the kitchen as well, or that the pan was already on the stove when the vegetables were still being cut.

7.5.3. Scrubbing Objects' trajectory

While stepping through changes works well for events, it is limiting when change occurs over longer duration. For example, when an object is carried around, this results in many location changes. For this case, we developed a trajectory-based navigation technique. To toggle an object's trajectory, viewers select the object and, as shown in Figure 7.4, add the trajectory to the current scene. Viewers can enable multiple objects' trajectories at the same time. Trajectories are color-coded per object to help distinguish them and to identify the object they belong to.

Figure 7.4.: The trajectory scrubbing techniques allows viewers to trace the path of an object and discover events along its history. The user can scrub the trajectory by dragging the object (pot) on it and navigate to the moment the object was at specific location (e.g., stove). To find all the trajectories in a region of space the user can trigger a trajectory sphere which reveals trajectories that passed through it (e.g., trajectory of canned beans leading above the pot).

Trajectories enable a time-agnostic view of an object's history. Hence, they allow viewers to answer a variety of questions about an object, such as: (1) where it came from, (2) whether it was used at a specific location, or (3) whether it covered an area with its movement. For example, users could be wondering where the pot was taken from or if the whole floor was mopped.

To navigate the spatial recording by using a trajectory the viewers are able to scrub along it or jump to any points on the trajectory (see Figure 7.4). By doing this the user moves to the time in the spatial recording when the object was at the selected position. The trajectory here acts as a form of timeline, yet differs from conventional timeline in many ways: (1) it is anchored in the world instead of being an overlay, (2) constant movement along a trajectory generally results in non-linear movement through time, instead of a fixed time-step, (3) their resolution is dependent on the movement speed of the object instead of the total duration of the recording, and (4) they are specific to one object instead of the whole recording. Trajectories hence take up little space when an object was static, yet allow for fine-grained navigation when objects were moving around a lot or at high speed.

Navigating by scrubbing an object's trajectory concurrently updates the whole spatial recording. In other words, viewers always see a consistent world, in contrast to previewing discrete changes by stepping through an object's changes. Further, trajectories give a snapshot of objects' whole story and help viewers notice events they would not with a more limited preview, or when skipping through the recording. This is valuable for sensemaking as it allows viewers to discover what happened alongside the selected object's changes. For example, as shown in Figure 7.4, they may discover that the canned beans were put in the pot but not the tomato.

The techniques we discussed so far only work on visible objects. To show trajectories of absent objects we implemented a trajectory search technique which allows users to find trajectories that passed through a region of space at any point in the recording. The users can enable a *trajectory sphere* which is anchored to one of their controllers. The sphere then acts as a magic lens revealing all the objects' trajectories that passed through it. Users can move around with the sphere to scan the environment and uncover trajectories of interest. When the user disables the sphere the revealed trajectories disappear. To keep a revealed trajectory, the user needs to enable it while the *trajectory sphere* is active. Once the trajectory is enabled they can use it for scrubbing or

jumping to any point on it. In addition to facilitating exploration the *trajectory sphere* allows users to inspect and enable objects' trajectories from a distance.

7.5.4. Timeline Scrubbing

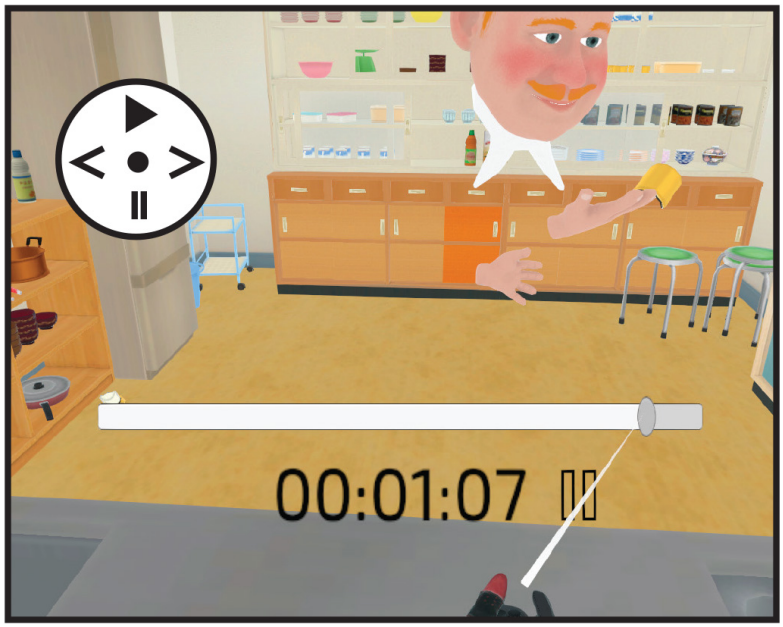

Figure 7.5.: In addition to navigation by direct manipulation of objects, the system also includes a timeline and media controls. The viewers can scrub the timeline via a raycast from the controller to the desired point in time, and use the thumbstick to rewind, fast-forward, play or pause. The current state of the replay system is indicated by an icon next to the time.

In addition to our direct manipulation techniques, we implemented a conventional timeline control (shown in Figure 7.5). When activated the timeline appeared at the bottom of the user's view and was view-stabilized. Users can scrub the timeline via a raycasting pointer. While our direct interaction techniques allow users to quickly jump to and explore moments of change, conventional timeline controls are still useful. For example, when the user wants to jump to a specific point in time or to the end or beginning of a spatial recording.

7.5.5. Media Controls

To give users the basic control over the spatial recording we included the conventional media controls. The user can rewind, fast-forward, play and pause the recording. These controls enable users to fine-tune the playback while finding or viewing the moment of interest.

7.6. Evaluation

To learn how users experience object-based navigation, we evaluated the *Who Put That There* system in a think-aloud study. In particular, we were interested in whether the system was easy to use, understandable, and useful. Participants were given a spatial recording and a set of sensemaking tasks. To complete the tasks, participants had to view the relevant parts of the recording. We encouraged them to verbalize their thoughts while they carried out the tasks. After they completed the tasks, we conducted an open-ended interview.

7.6.1. Spatial Recording

Figure 7.6.: We evaluated the "Who Put That There" system in a 22-minute spatial recording of a kitchen. The blue trajectories show all the objects' movement through the recording and the red trajectories the two avatars' movement.

We produced a 22-minute spatial VR recording of a kitchen containing two avatars. The virtual kitchen was 6x4 meters in size and contained two kitchen counters, two shelves, a fridge, microwave, sink, stove, as well as an assortment of kitchen and food items (plates, cups, boxes, cans, vegetables, drinks, etc.). The spatial recording contained a variety of events that commonly occur in real kitchen. For example, in the recording the two avatars arranged kitchen items, cooked, ate, cleaned, as well as dealt with kitchen accidents such as the breaking of a plate. Figure 7.6 shows a top-down view of the kitchen with the trajectories of all objects and avatars.

7.6.2. Sensemaking Task

To motivate a concrete use of our system, we asked participants to answer a set of sensemaking questions while viewing the spatial recording. We had two types of questions varying in complexity. One type started with a "*What*" referring to an object (e.g., "What happened to the beer can?"). Participants could answer these by observing the object in question while viewing the recording (e.g., the beer was taken off the shelf and put on the counter by one of the avatars, then poured in a glass by another avatar). The *"What"* questions were 1) *what happened with the blue beer can*, 2) *what happened with the orange pot* and 3) *what happened with the red mug*. Questions of the second type started with "*Why*" and were asking for a reason an event had happened (e.g., "Why was the sliding door left open?"). The participants could not answer these questions by solely focusing on the question's subject (i.e., sliding door), but had to understand the context around an event (e.g., one of the avatars broke a plate while preparing a meal, she then hid the broken pieces in the closet leaving the sliding door slightly open before rushing out of the kitchen). The *"Why"* questions were 4) *why was the burger thrown out of the window* 5) *why was the trash bin kicked over* and 6) *why was the sliding door left open*.

7.6.3. Study Design

We split the think-aloud study into three blocks, each of them containing one *"What"* question and one *"Why"* question. In the first block, the participants navigated temporally only by using *media controls* and by *stepping through objects' changes*. In the second block the participants used only *media controls* and *objects' trajectories*. In the third block the participants used only *media controls* and *timeline*. We split the functionality of the system into three blocks to ensure that participants used all the techniques during the study, and could not rely only on those they were most comfortable with. Furthermore, limiting the available interactions allowed us to shorten the training phase and keep the study duration under an hour. We randomly assigned the questions to the blocks per participant. Participants went through a training phase before each of the blocks (e.g., training for *media controls* and the *objects' trajectories* before starting the first block).

7.6.4. Protocol

We recruited 11 participants (4 male, age 23–48, M = 31.9, SD = 6.7) with limited VR experience. We first introduced the participants to the VR setup and the concept of spatial recordings. After, we gave them an overview of the study and introduced them to the think-aloud protocol. To help them familiarize themselves with the protocol we engaged the participants in a simple think-aloud math task. Once the general introduction was over, the participants went through a training phase for the first block, where they were shown how to use the *media controls* and the *stepping through objects' changes* to navigate a spatial recording. For this we used a training recording and in-application controller hints to guide the participants to the right controller inputs.

Once the participants understood the techniques and were comfortable with using them, they were put in the main spatial recording and started with the first *"What"* question. During the task we encouraged participants to keep talking with verbal prompts as well as a large "Keep talking" sign on one of the walls in the virtual kitchen. After they answered the first question they moved on to the *"Why"* question. When they finished with both of the questions they were offered to take a break or to continue with the next block. The procedure for the second and third block was similar except for the techniques being used (i.e., in the second block they used *media controls* and *objects' trajectories*). At the end of the study we conducted an open ended interview to gain additional insight into their comments and the perceived advantages and disadvantages of the individual techniques. On average, the study took 60 minutes to complete.

7.6.5. Data Collection and Analysis

The experimenter took notes of participants' interactions and comments during the think-aloud study. Furthermore, we recorded the participants' comments by using the microphone of the headset and screen captured the participants' view while they were interacting with the spatial recording. We analyzed the collected data by following the approach of Braun and Clarke [20]. We first identified meaningful bits in the data and coded them with shorthands, focusing on participants language, as well as the temporal navigation concepts described in the previous sections. From the codes we identified broad topics and reviewed them, omitting the ones irrelevant to our research focus and merging the related ones into larger themes.

7.7. Results

Three topics emerged from the thematic analysis of the think-aloud study and the end-interviews: *being there*, *making sense*, and *having fun*. All participants were able to understand and use all the functionality of the system to answer the questions. We note the observations directly related to usability in the *usability* subsection and discuss the improvements to the system in the discussion section.

7.7.1. Being There

Depending on the used technique, the participants expressed different level of involvement with the spatial recording. P1 stated that when *stepping through objects' changes* and dragging on *objects' trajectories* he felt more "like being there". Similarly, P6 compared scrubbing the global timeline to using the objects' trajectories, and said that the first one is "more like watching a movie" while the second one is "more like playing a game". The physicality of walking and manipulating objects gave the impression of being in an interactive world even when the possible manipulations were constrained to the recorded ones. However, this physicality also disturbed some participants as they would rather interact with objects at a distance and with less effort.

The *Timeline* was experienced as less interactive. Most often, the participants keep the view "pretty steady when using" the timeline (P1). A common pattern was to lock their view on the region of interest, start scrubbing from the beginning of the timeline, overshoot, and then correct to close in on the event of interest (e.g., a beer can being picked up). P4 mentioned that such scrubbing felt like waiting.

7.7.2. Making Sense

When using objects' trajectories to complete the sensemaking tasks participants mentioned greater understanding of what had happened and the context around it. P5 compared using the timeline to reading the ending of a book to find out what happened, and using the objects' trajectories to reading the full book and seeing how it happened. P11 experienced the timeline as efficient, however, it gave her the feeling of "not seeing the whole story" and "missing of events" even when she was sure she missed none. Similarly, P6 experienced trajectories as giving her more knowledge and allow for better reasoning when trying to interpret them.

The additional knowledge encoded in the trajectories could also confuse participants, especially when the trajectories were entangled or were of similar color (P10). On the other hand, many participants were quick to understand the associations between objects and trajectories. P8 mentioned it was easy to notice when an object broke or to which object the trajectory belongs just by the color. Stepping through objects' changes also facilitated active exploration. Participants were quick to inspect changes of objects in question as well as the ones that might be related. For example, when asked to find out "Why the burger was thrown out of the window?" they inspected the fridge and the tomato, to see if those two objects played a role in it.

7.7.3. Having Fun

Most participants mentioned that using the object-based navigation was fun, interesting and "cool". Part of this is the novelty effect which the timeline lacks. P9 mentioned that while the timeline was "definitely the simplest" it was also "the most primitive and not as much fun". However, participants mentioned more than just novelty when interacting with objects' trajectories. P2 especially liked the *trajectory sphere* and felt like it gave her "some kind of superpower". P5 drew parallels to "seeing the code behind the matrix" and "realizing the canvas" on which the interactions happen.

7.7.4. Usability

Participants were quick to learn and use the full functionality of the system, despite most of them having little or no prior VR experience. Object-based techniques required more explanation than timeline and media controls, as participants were familiar with the latter two from traditional video recordings. Except for minor issues, such as miss-clicks because of difficulty using the thumbstick, the participants did not experience any major usability problems when using object-based techniques. Similarly, participants had no troubles when using the media controls, while a few experienced problems with precision when using the timeline. However, in general all participant were able to use all of the systems functionality to navigate the recordings and answer the sensemaking question.

7.8. Discussion

We have presented a concept for interaction which helps users navigate spatial recordings, such as recordings of VR experiences, volumetric captures, motion captures, and 3D video games. The concept departs from the idea that viewers of spatial recordings are interested in finding moments of change. Therefore we structure the navigation of spatial recordings by objects' changes. Viewers are able to directly manipulate the objects to preview changes and navigate through time. We illustrated the concept with the *Who Put That There* system, which enables such navigation for spatial VR recordings. We have argued that this way of interacting with spatial recordings helps users navigate in a meaningful manner while being closely integrated with the content of the recording. Next, we discuss the limitations, advance over previous work, potential extensions and future application scenarios.

7.8.1. Limitations

There are limitations to the presented concept, the *Who Put That There* system that exemplifies it, and the evaluation of the system. First, direct manipulation techniques might not be suitable for all scenarios. Our concept requires active engagement with the scene which can become exhausting or cumbersome in case of complex queries. Second, the *Who Put That There* system is only one instance of the presented concept. It does not track, preview and allow navigation by all possible changes that could happen to objects and avatars. Furthermore, the system does not support filtering of the tracked changes or compound operators for multiple objects. Third, the evaluation has limited generalizability as we only tested one scenario and recording duration. For a more comprehensive evaluation of the usability of the specific techniques, a comparison of them in different scenarios would be needed.

7.8.2. Advances over Previous Work

Current interfaces for navigating spatial recordings mainly use timelines, both in commercial interfaces (e.g., the Fortnite Replay System[4], or VR Player[5]) and in research (e.g., Vremiere [95], or [120]). As argued, timeline scrubbing has drawbacks when used for spatial recordings. For example, it makes it easy to miss events and difficult to follow moving objects, it can cause nausea, and can lower immersion. Timelines are excellent for questions about time, but less useful for making sense of spatial recordings when users ask questions about patterns (where is this object usually), causality (why is this here), and agency (who put this here).

We draw inspiration from previous work on direct manipulation of video content [27, 34, 49, 96, 122] and expanded on their concepts. Systems like the one by Dragicevic [27] and Wolter [153] focused on changes in the form of movement. We allow for any type of change and examplify the concept in a *Who Put That There* system, allowing navigation by change of appearance, configuration, and shape, among others. Furthermore, we include previewing and navigating by discrete as well as continuous changes.

We believe that our focus on sensemaking is unique. Earlier work has outlined how people relate to past in real [147] and virtual worlds [93]. It appears that relations to past are rarely solely about time, and are often focused around objects, activities and events. We choose *change* as the unit that connects them; partly because it is easy to integrate it and operate with it in a temporal navigation system. Furthermore, we demonstrate the usefulness of the our system for sensemaking with a qualitative study identifying additional qualities of interaction such as increased sense of presence.

[4] https://www.epicgames.com/fortnite/en-US/news/fortnite-battle-royale-replay-system
[5] http://www.vrplayer.com/

7.8.3. Potential Extensions

The *Who Put That There* system could be extended in several ways. First, support for additional kinds of changes in spatial recordings could be added. For instance, objects might change their proximity to a specific area (e.g., knife leaving the kitchen area), or relative location to another object or an avatar (e.g., a cap being worn backwards instead of forwards). The system could also support person related changes such as change of emotion and other internal states.

Second, objects' changes could be previewed and controlled in several additional ways. For example, changes could be visualized as multicolored sculptures (e.g., similar as in [161]) to be able to preview continuous changes of appearance, shape and size. Another improvement would be to encode more information in trajectories to show their direction, time differences between its points, or interactions between multiple trajectories. They could also visually highlight notable changes such as change of configuration. The previews could also be controlled in more ways. For example, the user could scale the object by pulling it apart and navigate by change of size (e.g., of a plant growing).

Third, many spatial recordings contain a notion of viewpoint or viewpoints; this is the case for VR, AR, and many game recordings. The *Who Put That There* system currently does not handle switching between viewpoints and viewers need to move to a new location on their own. Future systems could enable changing of viewpoints as well as stepping into viewpoints of an avatar. The latter would require ways to mitigating motion sickness, however, if possible this could further increase the sense of "being there".

Fourth, in some cases changes in objects occur over large scales or far away from the viewers location. For example, consider a spatial recording that spans multiple rooms or buildings. To preview changes within such large space, the system could include a mini-map with abstracted preview and teleport transitions to navigate to them.

7.8.4. Further Application Scenarios

Using direct manipulation within spatial recordings to control time is a concept applicable to various settings. However, in our study we have focused on a daily life scenario demonstrated in VR. We envision similar interactions to be possible in AR in the near future. Apart from daily life, there are several other scenarios where the concept could be useful. Game replay systems could be complimented with navigating by changing objects. For example, in a capture the flag game the player could use a trajectory sphere to reveal the flag. Spatial sports recordings such as basketball games could benefit by keeping the viewer immersed while navigating through the highlights. Training material for tasks (e.g., engine assembly) that come in the form of spatial recordings (e.g., AR instructions [101], VR simulations [33], or educational application [6]) could be enhanced to allow trainees to navigate through instructions step by step, and to allow the instructors to quickly review their performance.

7.9. Conclusion

We proposed structuring the navigation through spatial recordings by objects' changes. The changing objects can be directly manipulated to preview and navigate to moments of interest. We examplify the concept with the *Who Put That There* system, which among other, allows stepping through an object's changes (e.g, change of appearance, configuration, grasp, size), and scrubbing on objects' trajectories. We evaluated the system with a sensemaking task and demonstrated its usefulness. Furthermore we identified qualities such as increased sense of presence, engagement and understanding of activity in the recording.

[6] Labster, https://www.labster.com/

8. Correction of Avatar Hand Movements Supports Learning of a Motor Skill

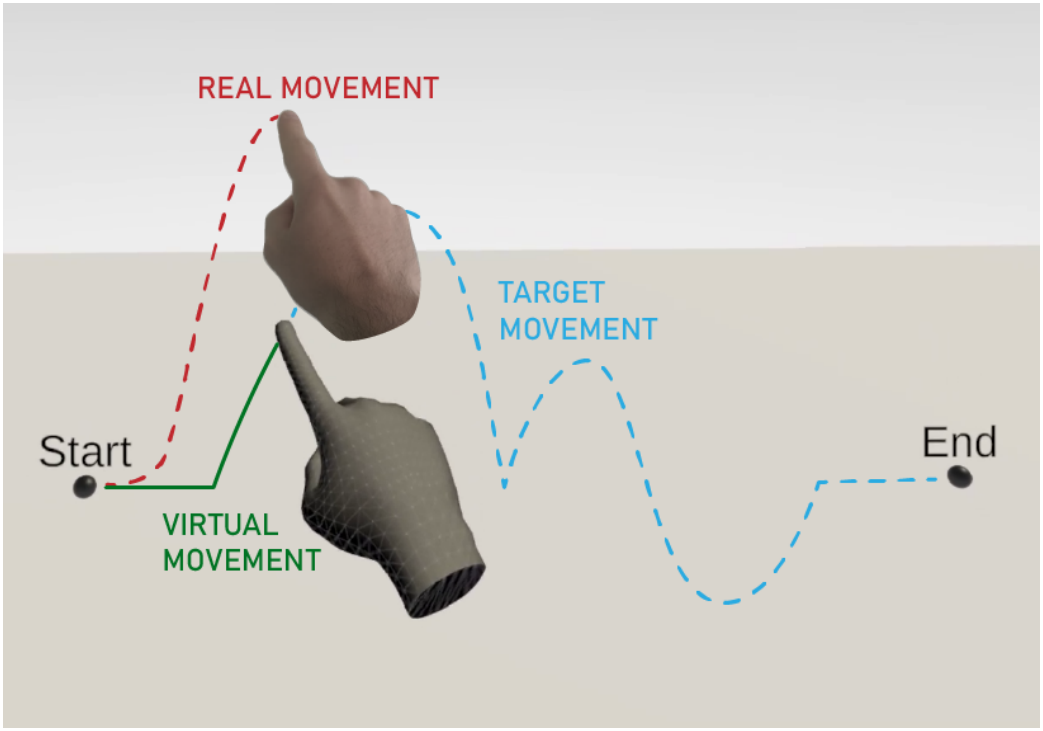

Figure 8.1.: The participant's virtual hand movement was corrected to be closer to the target movement than the hand's real movement. The blue line shows the path the participant needed to trace with the index finger, the red line shows the participant's actual movement, and the green line shows the movement of the virtual hand that the participant saw in virtual reality. None of the paths nor the participant's real hand was visible during the experiment.

8.1. Abstract

Learning to move the hands in particular ways is essential in many training and leisure virtual reality applications, yet challenging. Existing techniques that support learning of motor movement in virtual reality rely on external cues such as arrows showing where to move or transparent hands showing the target movement. We propose a technique where the avatar's hand movement is corrected to be closer to the target movement. This embeds guidance in the user's avatar, instead of in external cues and minimizes visual distraction. Through two experiments, we found that such movement guidance improves the short-term retention of the target movement when compared to a control condition without guidance.

8.2. Introduction

Sports, games, artistic endeavors, and gestural communication all benefit from precise hand movements. However, learning complex movements is difficult and we often benefit from some type of guidance. Guidance such as playback of the target movement, or arrows rendered in virtual reality (VR), can help the user learn a movement. However, guidance can also be ineffective or distracting if the cues hide content in a scene or visually interfere with it. Further, the guidance might be confusing, or the users might get used to it to such an extent that they perform poorly once the guidance is removed. Such over-reliance on guidance can negatively affect the use of VR applications since users cannot train indefinitely. Thus, it is crucial to design appropriate guidance to benefit the user.

Many techniques have been proposed for supporting the learning of hand movements in VR. The majority of them provide augmented feedback around the user's avatar. Such feedback includes, an extra pair of hands showing where the user should move next [28], a mirrored view of the user with cues overlaid on it [2], or abstract cues such as arrows guiding a user towards a target.

We investigate a technique that embeds the guidance directly in the avatar's movement. The technique corrects the user's virtual movement when it differs from the target movement (Figure 8.1). The user's virtual hand location and pose are corrected to be closer to the target movement. Such correction introduces a misalignment between the user's actual hand and its representation in virtual reality. This misalignment then serves as a guide towards the target movement. Such guidance immediately improves the user's virtual performance, avoids introducing external visual cues, and might be used to subtly impact the user's movement during learning.

We compare different degrees of virtual movement correction to a control condition and a conventional guidance technique. In two experiments, we found that the virtual movement correction improved the short-term retention of the trained movements. The technique outperformed the control condition in which the users did not receive any guidance and performed equally well as a conventional VR guidance technique (viz., ghosted hand). Our results suggest that the correction of virtual movement is a viable method for supporting hand movement learning. The applications that could especially benefit from it are those where minimizing visual distraction is critical (e.g., communication with gestures), as well as applications that require good virtual performance during training.

8.3. Related Work

Researchers have explored several VR techniques for providing guidance during a motor task. Application domains vary from learning VR painting [143], Tai-Chi practice [38], conducting [28], tennis and table tennis playing [45, 144], engine assembly [40], caligraphy [158], rehabilitation and physiotherapy [129, 139], as well as augmenting social interactions [119]. The review paper by Sigrist et al. gave a comprehensive overview of different feedback types that the user can receive during motor learning [126]. Apart from categorizing the types of augmented feedback, they discussed evaluation methods, motor learning theories, and the impact of different feedback strategies. According to their categorization, our technique falls within training with concurrent visual feedback in a complex task scenario. In the next section, we list the techniques and systems that are related to our work.

8.3.1. Techniques and Systems

Most existing techniques that provide visual feedback during training place the guiding cues in the world around the user. For example, *EGuide* rendered an extra pair of hands in the user's periphery for egocentric guidance of the user [28]. Similarly, the *Just Follow Me* system guided the user during calligraphy by placing an additional translucent brush (i.e., ghost brush) in the virtual environment [158]. Systems such as *YouMove* [2],

MotionMA [148], *SleeveAR* [129], and *Physio@Home* [139] placed cues in the world from an allocentric perspective. They provided a mirrored view of the user with added cues to serve as guides. The cues varied from realistic representations of the user [157], abstract representation of the user (e.g., rendering of a skeleton in *YouMove* [2]), to use of abstract symbols (e.g., arrows and circles in *SleeveAR* [129]).

Previous work has also explored placing cues on the user's body and manipulation of the user's avatar to help execute and learn motor movement. *LightGuide* projected visualizations on users' real hands to guide their movement [128]. Visualizations such as 2D and 3D arrows on the hand's surface helped users execute simple mid-air movements. Work on intermanual skill transfer [100, 156] looked at the effects of manipulating the user's avatar on motor learning. For example, Xiao et al. [156] showed that practicing with one hand while seeing the same movement in the other hand, helps improve performance for the unpracticed hand. Similarly, Ossmy and Mukamel showed that seeing the movement in immobile hand as well as yoking it to passively follow the practicing hand further improves the motor learning [100]. Such visual manipulations of the user's avatar are similar to the ones we apply to the practicing hand.

8.3.2. Problem and Evaluation

The problem domain we focus on is how to best support motor learning by providing visual guidance during training. Sigrist et al. characterized motor learning as a lasting increase of performance *after* training, that is, once the guidance is removed. To evaluate the effectiveness of different guidance techniques, they suggested testing the short- and long-term retention of the practiced movement. Much of the related work performed such an evaluation (e.g., [2, 144, 158], while some did not (e.g., [13, 28, 128]). The latter techniques therefore fall in a different problem domain or need further evaluation to establish if they are suitable for motor learning.

8.3.3. Theoretical Background

Apart from specific techniques, previous work also investigated the general influence of avatar manipulations. Gonzalez-Franco et al. [35] reported the *self-avatar follower effect*. They found that, under certain conditions, participants followed their avatar when it did not overlay with their physical body, without being explicitly instructed to do so. Cohn et al. continued their work in *SnapMove* in which they mapped arbitrary physical reaching movement to a single virtual movement and reported on drift of participant's hand toward the avatar's movement [24]. The opposite effect was observed already in the 1960s by Nielsen [97]. In his experiment, the participants drifted *away* from their illusory hand during a motor task. This effect has been frequently exploited in *motion retargeting* work [5, 11, 56, 57]. For example, Azmandian et al. [5] shifted the user's virtual arm and hand during the reaching movement to redirect the user's reach towards a physical prop. Similarly, Bergström et al. [11] warped the user's fingers during the reaching movement to resize the users grasp.

These contrary reactions to visual manipulation could potentially be explained by predictive coding models where the conflicts are minimized at different hierarchical levels [72]. That is, normally the conflicts between proprioception and visual feedback are minimized. However, in case of a visual task that does not allow this, the users decrease their reliance on proprioception and increase their reliance on the visual feedback. In such a task, the minimization of visual conflicts becomes more important than the minimization of proprioceptive conflicts. We designed our hand guidance technique to produce a self-avatar follower effect and *support* the minimization of the virtual and real hand movement's misalignment.

8.4. Correction of Virtual Hand Movement

When learning a new hand movement, the user first builds up a motor program [126]. Learning is often initiated by showing the user a visual demonstration of the target movement. The movement program is then later refined through training. Initially, users make many errors and often need feedback to correct their movement. Visual

guidance provided during training can help in this phase of learning. It can prevent cognitive overload, remind the users of the next movement in a sequence, and help them notice and avoid errors.

The key idea of our approach is to correct the virtual representation of the user's hand movement to be closer to the target movement. The correction depends on the size of the movement error. When the user's hand location and pose is far removed from the target one, the correction is large; when the subject is doing well the correction is smaller.

The implementation of our approach has two parts, we first *predict* where the user's hand should be at any moment during the movement, and second, we *correct* the hand's virtual representation.

Depending on the type of the target movement, the prediction can either be simple or complex. If the target movement requires to move a hand from one point to another in a straight line, only moving forward (e.g., as in *SnapMove* [24]), then it is relatively simple to predict where the hand is supposed to be at any point during the movement. In contrast, if the target movement is poorly defined and requires intricate hand movements (e.g., sign language), the prediction is more complicated. For our study, we selected a target movement that requires a high degree of precision while being simple to predict. The participants had to trace an invisible line in a single movement going from left to right (see Figure 8.1 for an example target movement). We predicted where the hand should be, by using the participant's hand location on the left-to-right axis and projected it on the target movement.

Once we can predict with high accuracy where the user's hand needs to be, we can correct its virtual representation. We can either fully correct the virtual hand by moving it to the predicted location and pose, or we can partially correct it by moving it somewhere between the predicted and actual location. In our initial study, we used three variations of movement correction to investigate if they support learning of a motor movement.

8.5. Study 1

We conducted an exploratory study to investigate the usefulness of virtual movement correction for learning of hand movement patterns. We were interested if such guidance improves the short-term retention of the trained movement.

8.5.1. Participants

We recruited 30 participants via an online Oculus Quest community[1]. The study was administered remotely by using participants' personal Oculus Quest headsets. We discarded data from participants that experienced poor hand tracking quality. We used jitter experienced during the measured repetitions as a proxy for the quality of hand tracking. If a participant experienced more than ten virtual hand moves larger than 20 cm from frame to frame during the experiment then their data was discarded. With this criteria, we discarded the data of 8 participants. The average age of the 22 participants left (all male) was 31.5 years ($SD = 9.3$).

8.5.2. Hand Guidance Techniques

We had three conditions of virtual movement correction: *interpolation50*, *interpolation75* and *snapping*. Additionally, we had a *ghost* condition (Figure 8.2) where the participant's movement was guided by a ghosted hand (mimicking the state-of-the-art in movement guidance [28]), and a *control* condition in which the participants did not receive any guidance.

The five experimental conditions were implemented as follows:

[1] https://www.reddit.com/r/oculus/

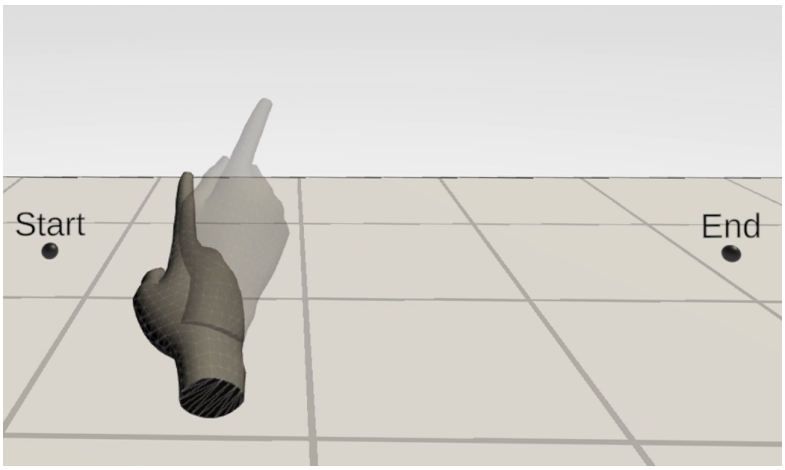

Figure 8.2.: A transparent hand was guiding the participants through the target movement in the *ghost* condition.

control: Participant's virtual hand was always rendered at the hand's actual location. Participants did not receive any guidance during training.

ghost: The participant's virtual hand was rendered at its actual location, while a ghosted hand (i.e., an additional translucent hand) indicated where to move next (Figure 8.2). The ghosted hand therefore moved with the same speed as the participant's hand while always showing the next location and pose in the target movement.

interpolation50: The participant's virtual hand was rendered 50% between the target location (i.e., where the hand should be) and the actual location of the participant's hand. Similarly, the virtual hand's pose was interpolated to be mid-way between the target pose and the participant's actual hand pose.

interpolation75: The participant's virtual hand was rendered 75% of the way between the target location and the actual location of the participants' hand. The virtual hand's pose was also interpolated to be three-quarters of the way between the target pose and the participant's actual pose.

snapping: The participants' virtual hand was rendered at (i.e., snapped to) the target location and in the target pose throughout the task.

All conditions were set in an identical virtual environment with minimal visual distractions. The virtual floor was textured with a grid pattern to aid the depth perception (see Figure 8.2).

8.5.3. Design

We used a within-subject design and a Graeco-Latin square to balance the five conditions and the five target movements the participants were instructed to repeat (see Figure 8.3 for target paths).

8.5.4. Procedure

Each experimental condition consisted of a training block and a test block. The participants started with the training block in which they needed to replicate a target movement in mid-air as many times as possible within 40 seconds. The target movement was shown to them before the training by an animated hand performing the movement three times. The index finger tip of the animated hand traced one of the target paths (see Figure 8.3 for the target paths); the target path was not visible during the animation. Once the participants started the training, they were guided by using one of the five techniques described in Section 8.5.2. After completing the training block, the participants were given a Likert-scale question about the task difficulty. Once they answered the question, they started the test block (i.e., short-term retention test), where they were instructed to repeat the

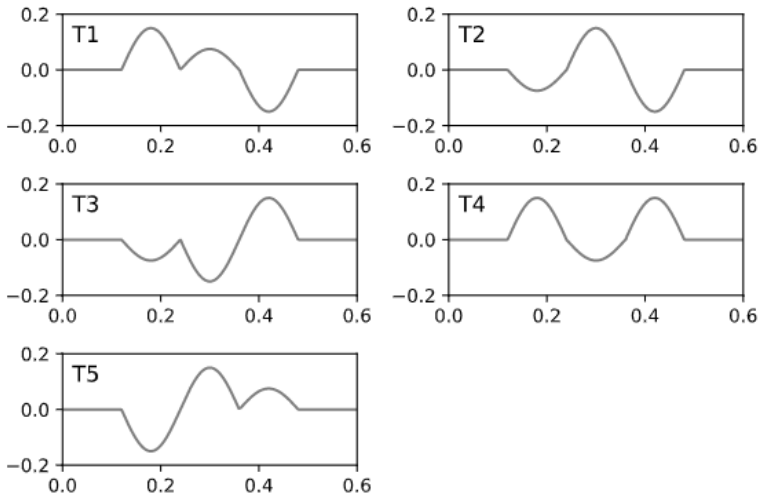

Figure 8.3.: The five target paths the participants needed to trace mid-air with their index finger in Study 1. Several paths were randomly generated in a plane from half-period sine waves of two different amplitudes – 7.5 and 15 cm. Each path was assembled from three wave segments and a straight segment at the beginning and end of the path. Among the generated paths, we selected five of them of approximately equal difficulty through pilot testing. The paths were never directly shown to the participants.

movement from the training block as many times as possible within 20 seconds. Participants were not shown the target movement again before starting the test block, nor did they receive any guidance during the test block. They could only rely on what they learned in the training block.

This procedure was the same for each condition; each participant went through it a total of five times. Once the participants completed the experiment, the study application sent the logged movement data to a remote server. Finally, they were asked to fill a post-study questionnaire and received a Steam[2] game worth 15$ as a reward.

8.6. Study 1: Results

The main performance metric used to compare the conditions was the accuracy of the executed movements. We used mean squared error (MSE) as a proxy for accuracy and compared the absolute performance in the training block, the absolute performance in the test block (i.e., short-term retention test), and the performance increase from the training block to the test block. We also investigated how performance changed over time, the perceived task difficulty, and compared the total number of repetitions the participants executed per condition. For the analysis, we used linear mixed effects models (LMM) due to the advantages over more commonly used ANOVA [6, 8, 82].

8.6.1. Number of Repetitions

The LMM analysis did not reveal any significant differences in the number of executed repetitions among the conditions for neither the training block ($F(4, 84) = 2.167, p = 0.08$) nor the test block ($F(4, 84) = 2.167, p = 0.08$). In the training block, the participants on average executed 12.15 repetitions ($SD = 3.13$), while in the test block the mean number of repetitions was 5.81 ($SD = 1.57$). Figure 8.4 shows the mean number of repetitions per condition for the training and test block.

[2] https://store.steampowered.com/

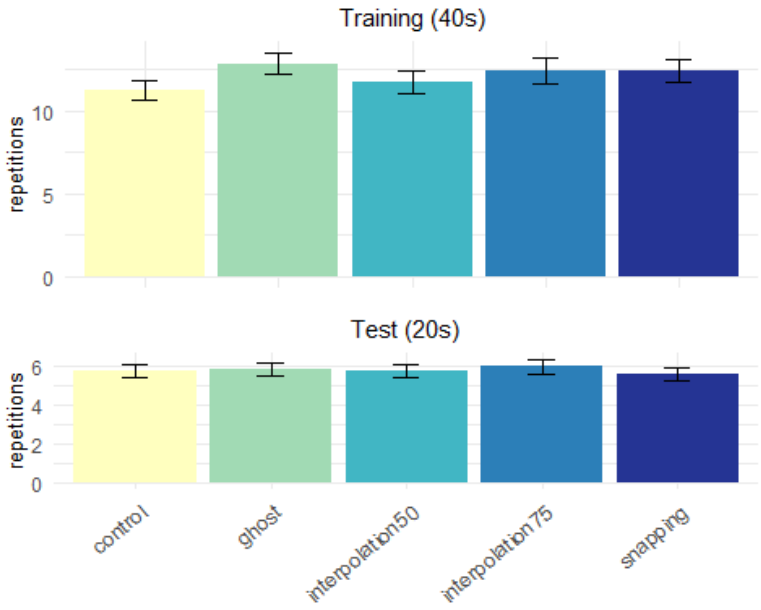

Figure 8.4.: Mean number of repetitions per condition for the training and test block in Study 1. The error bars represent the standard error of the mean.

Lack of differences in the number of repetitions indicates that participants executed the mid-air movements with similar speed, independent of the guidance technique used in the training block.

8.6.2. Movement Accuracy

We compared the accuracy of the executed movements between the conditions to see if any of the guidance techniques were better at supporting learning of the target movements. The accuracy measure we used was the mean squared error (MSE) of the executed movement over the target movement. We calculated the error by squaring the difference between the index finger's logged position and the target position (i.e., where on the target path the index finger should be) for each point on the target path. To calculate the MSE of a repetition, we summed all the errors and divided the sum with the number of them. To calculate the MSE for a specific condition we averaged the MSE across all the repetitions for a participant and then averaged again across all participants. Figure 8.6 shows the MSE for each of the conditions for the training and test block, and the performance improvement (i.e., decrease in MSE) from the training to test block.

To see if there are significant differences between conditions, we conducted an LMM analysis. The analysis of absolute accuracy in the training block did not reveal a main effect of condition on MSE ($F(4, 84) = 1.313, p = 0.272$). Thus, participants did not significantly benefit from any guidance during the training.

The analysis of absolute accuracy in the test block found a main effect of condition on MSE ($F(4, 84) = 3.931, p < 0.01$). A post hoc pairwise comparison test with Bonferroni adjustment showed significantly lower MSE for the *snapping* condition when compared to the *control* condition ($p = 0.046$).

The analysis of accuracy improvement from the training to test block found a main effect of condition on MSE improvement ($F(4, 84) = 5.169, p < 0.001$). The post hoc pairwise comparison with Bonferroni adjustment found that accuracy in *snapping* and *interpolation75* conditions significantly improved ($p = 0.005$ and $p = 0.012$, respectively) when compared to the *control* condition.

To better understand the size of the movement errors, Figure 8.5 shows a movement plot and MSE of a select participant in the *snapping* condition. The participant's accuracy in the training block was poor and good in the

test block.

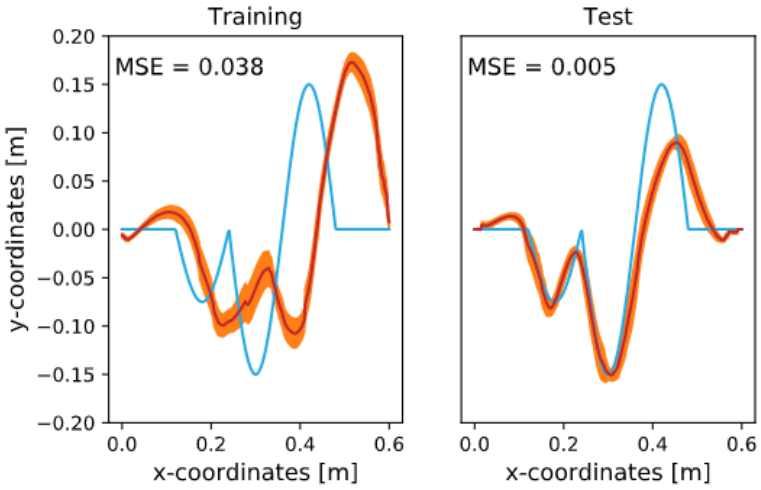

Figure 8.5.: Logged movement in the y-x plane for the *snapping* condition for one of the participants in Study 1. The blue line shows the target path while the red line shows the mean trajectory of the participant's index finger; the orange interval indicates the standard error of the mean.

8.6.3. Perceived Difficulty

After the training in each experimental condition, the participants answered a Likert-scale question on task difficulty. The LMM analysis did not reveal any significant differences in perceived difficulty among conditions ($F(4, 84) = 0.773, p = 0.546$).

8.6.4. Learning Effect

To investigate the learning effect, we used the correlation of MSE and time for the training and test block. The Pearson correlation coefficients indicate a weak positive correlation to no correlation in MSE over time (see Table 8.1). Lack of negative correlation indicates that MSE did not decrease with time during the training block, meaning that participants did not get more accurate during training. Furthermore, the Pearson correlation coefficients suggest that participants were less accurate with time in most conditions (e.g., in the *control* condition in the training block).

	control	ghost	inter50	inter75	snapping
Training	r = 0.30 p < 0.001	r = 0.20 p < 0.001	r = 0.16 p = 0.01	r = 0.09 p = 0.15	r = 0.08 p = 0.20
Test	r = 0.22 p = 0.01	r = 0.14 p = 0.11	r = 0.03 p = 0.71	r = 0.20 p = 0.02	r = 0.12 p = 0.16

Table 8.1.: Pearson correlation coefficients and p-values for time and MSE for the training and test block of Study 1.

8.6.5. Summary

The results of Study 1 suggest that correcting participants' virtual movement helps learning of a target movement. Training with *snapping* guidance increased the accuracy of participants in the short-term retention test (i.e., the test block). Similarly, *snapping* and *interpolation75* improved the most from the training to test block.

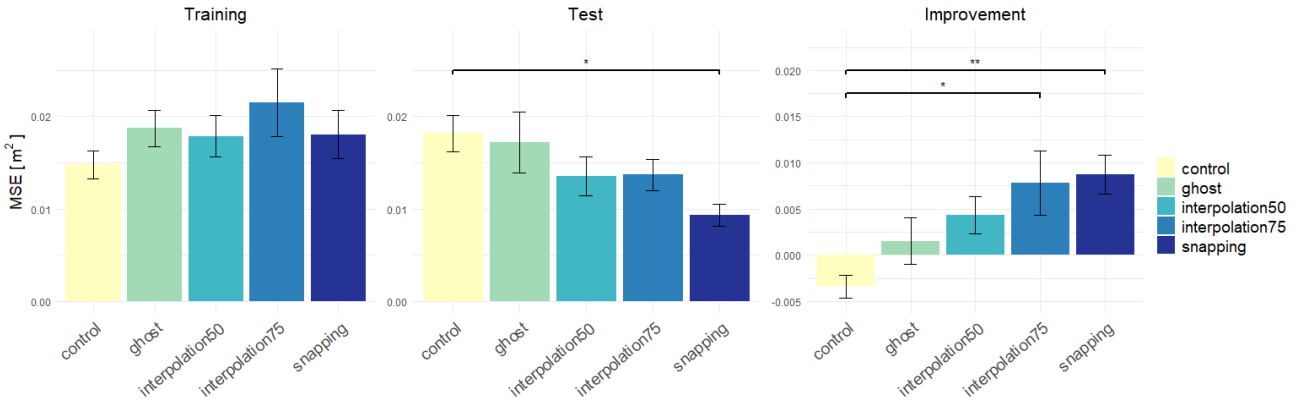

Figure 8.6.: Mean MSE across participants for training and test block, and the relative improvement in MSE from the training to test block in Study 1.

While *interpolation50* also performed well, the analysis did not show significant difference when compared to other conditions. Full correction of virtual movement (i.e., *snapping*) was better at supporting learning than partial correction (i.e., interpolation).

We expected to see a clear increase or decrease in performance from *interpolation50*, *interpolation75* to *snapping*, however, there was no clear relationship between the three conditions. A more complex approach than interpolation may be needed for varying the degree of virtual movement correction.

We did not notice any clear patterns of improvement within the blocks. The correlation analysis does not indicate an improvement in accuracy *during* the training or test, even when there is a large improvement in accuracy *from* the training block to the test block. To confirm the findings from Study 1, we decided to replicate it while keeping only the best performing correction condition (i.e., *snapping*).

8.7. Study 2

Study 2 replicated the initial study while removing the *interpolation50* and *interpolation75* conditions. With replication, we intended to confirm the benefits of the virtual movement correction for short-term retention of the target movement and to confirm the surprising lack of improvement *during* training. We kept only one of the virtual movement correction conditions (*snapping*) as we were more interested in the general effects than the effects of specific variations (e.g., degree of correction) of the virtual movement correction technique. Furthermore, having only one virtual movement correction condition avoided any potential training effect across conditions. Our hypotheses for Study 2 were that virtual movement correction (i.e., *snapping*) will outperform the *control* and the *ghost* condition in the short-term retention test, and that *snapping* will show the largest performance improvement from the training to test block. In other words, except for *snapping* significantly outperforming *ghost*, we expected to replicate the results from the Study 1.

8.7.1. Participants

We recruited participants over online Oculus Quest communities until we had 36 validated data sets (1 female, 35 male, age $M = 28.7$, $SD = 8.23$). To validate the data sets, we used the same criteria as in Study 1 and discarded the data from three participants that experienced hand tracking problems during the tasks. The drastic reduction of the rejected data when compared to the the Study 1 was due to the added screening test that the participants had to take before starting the experiment. In the screening test, the participant's hand tracking quality was evaluated in a simple target selection task. If the participant's hand tracking quality was under a

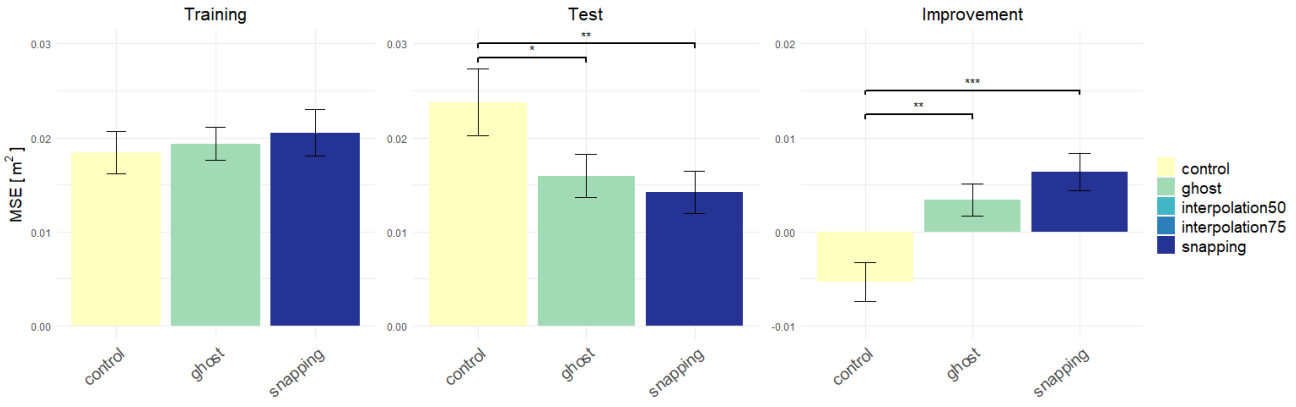

Figure 8.7.: Mean MSE across participants for training and test block, and the relative improvement in MSE from the training to test block in Study 2.

predetermined threshold, then the study application did not allow them to proceed to the experiment. In such a case, the participants were encouraged to find a room with better lighting conditions and try the screening test again. We conducted Study 2 two months after the initial study and did not limit the participation to only those who did not take part in Study 1. We assumed that after such a time span any carryover effect from Study 1 would be negligible. In the post-study questionnaire only one participant stated that he took part in Study 1.

8.7.2. Design

In Study 2 we used *control*, *ghost* and *snapping* conditions, implemented as in Study 1. We selected the three target paths from Study 1 with the least deviation from their overall mean MSE. The selected target paths were T1, T4 and T5, as seen in Figure 8.3. We used a within-subject design and balanced the conditions and target paths by using all 36 combinations and order permutations.

8.7.3. Procedure

Except for the screening test and fewer conditions, the study procedure was identical to the one of Study 1. The study took approximately 20 minutes and the participants were given a Steam game worth 15$ as a reward for their participation.

8.8. Study 2: Results

We conducted an identical LMM analysis to Study 1, using MSE as our main performance metric, while also looking into the learning effect and the number of executed repetitions per condition.

8.8.1. Number of Repetitions

As in Study 1, there were no significant differences ($F(2, 70) = 0.236, p = 0.79$) in the number of repetitions between the conditions in the test block ($M = 6.15, SD = 2.05$). We found a main effect of condition on the number of repetitions in the training block ($F(2, 70) = 9.555, p < 0.001$). Tukey's post hoc test revealed that in the training block participants executed significantly more repetitions ($p = 0.001$) in the *ghost* condition ($M = 13.72, SD = 3.54$) than in the *control* condition ($M = 11.58, SD = 3.5$). There were no interaction

effects between the *snapping* and *ghost*, and *snapping* and *control* condition. Figure 8.8 shows the mean number of repetitions per condition for the training and test block.

8.8.2. Movement Accuracy

Study 2 confirmed our hypothesis of *snapping* performing better than the *control* condition, and rejected our hypothesis of *snapping* outperforming *ghost* as they did not differ significantly (see Figure 8.7).

The analysis with LMM revealed a main effect of condition on MSE in the test block ($F(2, 70) = 6.082, p < 0.01$). Tukey's post hoc test showed that *snapping* ($p = 0.005$) and *ghost* ($p = 0.025$) were significantly more accurate than the *control* condition (Figure 8.7, middle). We also found a main effect of condition on improvement in accuracy from training to test block ($F(2, 70) = 11.647, p < 0.001$). Tukey's post hoc test showed that the improvement in accuracy from training to test block was significantly larger for *snapping* ($p < 0.0001$) and *ghost* ($p < 0.01$) when compared to the *control* condition (Figure 8.7, right). In the *control* condition, the participants were less accurate in the test block than in the training block.

We did not find any significant differences between the conditions in the training phase ($F(2, 70) = 0.428, p = 0.654$), indicating that guidance used in *snapping* and *ghost* did not impact the accuracy during training (Figure 8.7, left).

8.8.3. Perceived Difficulty

As in Study 1, the LMM analysis of the Likert-scale answers on task difficulty did not find a main effect of *condition* ($F(2, 70) = 0.843, p = 0.435$). This means that all the conditions were judged as equally difficult, suggesting that the participants were not negatively impacted by the misalignment between the virtual and the real hand that the *snapping* introduced.

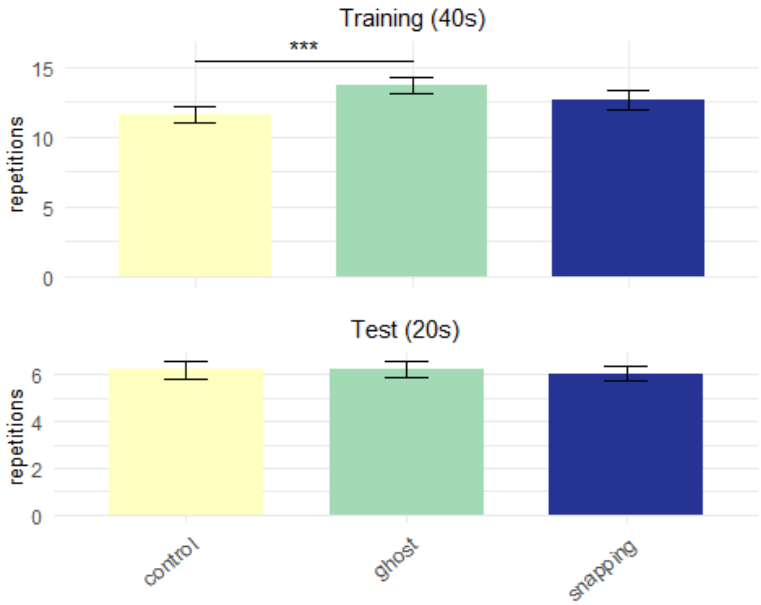

Figure 8.8.: Mean number of repetitions per condition for the training and test block in Study 2. The error bars represent the standard error of the mean.

8.8.4. Learning Effect

Similarly as in Study 1, we found weak to no correlations between MSE and time, which indicates no improvement in accuracy during the training block. Table 8.2 shows Pearson correlation coefficients and p-values for each of the conditions per block.

	control	ghost	snapping
Training	r = 0.21 $p < 0.001$	r = 0.01 p = 0.05	r = 0.12 $p < 0.01$
Test	r = 0.08 p = 0.23	r = 0.07 p = 0.29	r = 0.11 p = 0.11

Table 8.2.: Pearson correlation coefficients and p-values for time and MSE for the training and test block of Study 2.

8.8.5. Summary

Study 2 confirmed the benefits of virtual movement correction for the learning of hand movement while not finding any significant differences between the *snapping* and *ghost* conditions. Furthermore, Study 2 confirmed the lack of clear improvement in accuracy during the training.

8.9. Discussion

We have demonstrated that training in which the participant's virtual movement is corrected, significantly improves performance in the short-term retention test when compared to training without the correction. The effectiveness of such training was comparable to training with a conventional VR guidance technique (i.e., ghosted hand).

The results of Study 1 show that full correction of virtual movement (i.e., *snapping*) supported learning better than partial correction (i.e., *interpolation50* and *interpolation75*). Nevertheless, there is a noticeable positive trend in improvement from the training block to the test block for interpolation conditions (Figure 8.6). The benefit of only partially correcting the virtual movement is that it introduces less misalignment between the location of participant's real and virtual hand than the full correction of movement. For example, *interpolation50* displaced the participant's virtual hand only halfway towards the target location compared to *snapping*. In our task, the smaller misalignment of the interpolation conditions did not outweigh the benefits of full correction. We speculate that *snapping* might have supported learning better because it showed the participants the exact target movement. Seeing the exact movement could be especially crucial in the early phases of motor learning, when the participant is not yet familiar with the movement. In later phases of learning, perhaps the partial correction of movement would be more beneficial.

Surprisingly, we did not observe an improvement in accuracy over time *during* training. This could be due to many reasons. For instance, participants might have become progressively more tired and consequently less accurate with time; they might have forgotten the movement's exact details when training without guidance; or they might have felt rushed towards the end by the training block countdown timer. Furthermore, we noticed that two participants had little regard for accuracy during the training block of the *snapping* condition. They executed the movement almost in a straight line from the marked starting point to the ending point. These participants still benefited from training with *snapping* and performed better in the test block than in the *control* condition. We did not instruct the participants in any specific strategy they should or should not take during the task. Therefore the performance of the mentioned two participants was valid and analyzed the same as others.

The field of human-computer interaction (HCI) only rarely discusses the importance, or lack thereof, of immediate improvement during training for motor learning. One exception is a paper by Kirsh, which describes

how *marking*, a simplified or abstracted form of movement, is used in dance during practice [54]. The dancers execute a partial or a substitute movement while imagining the full movement. Kirsh listed general benefits of marking, such as requiring less energy, and benefits of it over pure mental simulation in which there is no movement at all. The performance of two participants, that were moving their hand almost in a straight line during training in *snapping* condition, could be interpreted as *marking* the movement. Such *marking* could have additional benefits when done in VR, since the users can see the full practiced movement instead of just imagining it as in Kirsh's description of marking practice [54].

The self-avatar follower effect [35] is another mechanism that might have influenced our results. The effect suggests that users will start following their avatar's movement to minimize the misalignment between the virtual and the real hand. The effect differs from *marking* and might be at work during training in the *snapping* condition for most of the participants. Participants were as accurate at executing the target movement in the *snapping* condition as they were in the *ghost* condition (Figure 8.6 left, and 8.7 left). This is interesting when considering that participant's virtual performance (i.e., the movement the participant saw in VR) in the *snapping* condition would be perfectly accurate irrespective of trying to align the physical hand with their virtual hand or not. The fact that the participants did align their hand with their avatar's hand might therefore be described as self-avatar follower effect. The participants' accuracy did not improve over time during the training (see Table 8.1 and 8.2), however, this could be explained by the self-avatar follower effect appearing immediately after starting the training. This speculation is also supported by results of immediate avatar correction (when contrasted to gradual) in the work of Gonzalez-Franco et al. [35]

8.10. Limitations and Future work

It is difficult to predict how well our findings generalize to motor tasks of different complexity and difficulty. Based on the review paper by Sigrist et al. [126], our experimental task falls within the complex type, as it has several degrees of freedom and cannot be mastered in a single practice session. The participants needed to trace an invisible path with their index finger, moving and rotating their hand and fingers freely. The target paths were multi-segment curves generated in a plane, however, to trace them accurately the participants needed to move their hand within a 3D space. Movements of similar complexity could, for example, be used as a control gesture in VR applications or for drawing in mid-air. However, it is unclear how well the virtual movement correction would work in motor tasks that, for example, required dexterous movement of all the fingers, higher degree of accuracy, or memorization of longer movement sequences. Future work should therefore investigate motor tasks of different complexity. One application we find especially interesting is learning of sign language. A sing correction system, conceptually similar to one used for text input correction, might prove immediately useful for communication while possibly have lasting long term improvements.

Apart from investigating other motor task of different complexities, future work should also collect data over longer periods of time. It is unclear how virtual movement correction affects long-term retention of the practiced movement as well as what are the benefits of such training in later stages of motor learning. In their review paper of augmented feedback, Sigrists et al. [126] suggested that concurrent feedback is most efficient in early phases of motor learning. Investigating how to integrate correction techniques with other types of feedback (e.g., terminal feedback or feedback in other modalities) is therefore another interesting research direction.

8.11. Conclusion

We investigated learning of a motor skill with a help of VR hand guidance technique. The technique corrected the participant's virtual hand movement to be more accurate than the the participant's actual movement. In the first experiment, we compared three levels of correction (*interpolation50*, *interpolation75* and *snapping*) to a control condition and a conventional hand guidance technique (viz., *ghost*). In the second experiment, we replicated

the initial experiment with only one of the correction conditions (viz., *snapping*). The two experiments showed that the correction of virtual movement supported motor learning by improving the short-term retention of the practiced movement. Apart from supporting motor learning, the investigated technique has other benefits, such as immediate improvement of the virtual performance during training. By embedding the guidance in the user's avatar it also minimizes visual distractions. We connect the technique to past work on manipulating visual feedback for the purpose of user guidance and interpret the results of our study through *self-avatar follower effect* [35] and *marking* [54] theory.

Part III.

Conclusion

9. Discussion

This thesis has set out to advance hand-based interaction techniques for MR. It covered the space broadly by investigating manipulation, navigation, and learning. The techniques were designed to use the capabilities of our hands more fully when interacting with the world around us. Moreover, the techniques aimed to expose the unique capabilities of MR instead of focusing on elements that have already proven useful in desktop computing (e.g., simulating a desktop-like workspace with a keyboard and WIMP interface).

9.1. Key Results

MR enables new ways of seeing the real world and virtual worlds from an egocentric perspective. It can expand our vision to see through physical and virtual occluders (Paper 1 and Paper 2), rearrange the space we are in (Paper 2), help us see the past and future events (Paper 3), and modify how we see our movement (Paper 4). Seeing the world differently also enables us to move differently. MR can enable manipulation of objects that would otherwise be hard to see or grasp (Paper 1 and 2), it can help us trace movements by dragging objects on their spatio-temporal trajectories (Paper 3), and it can subtly modify our movement without us realizing it (Paper 4).

The bulk of my thesis has focused on manipulation (Paper 1 and 2), navigation through manipulation of objects (Paper 2 and 3), and learning of motor skills that could support manipulation (Paper 4). A common theme explored in most of the papers is where the users' virtual hand is shown in relation to the real hand. Techniques in Paper 1 showed the user's virtual hand (i.e., fingers) at the location of the physical hand (i.e., see-through view) or displaced from the user's physical hand (i.e., displaced 3D view). Poros (Paper 2) simultaneously showed the user's virtual hand at its physical location and at a distant location (i.e., marked location) when the user's hand was in a proxy. The technique in Paper 4 remapped the user's virtual hand movement so that it was closer to the target movement. Allowing such alternations to the user's body representation is a unique capability of MR that has been explored in previous works. For example, the Go-Go technique remapped the user virtual hand to reach further when extended [115]. Portals allowed user's to reach go into them and travel to a distant location [59]. Hand redirection techniques enabled use of limited amount of physical props for haptic feedback in VR [5].

My thesis furthers this line of research by introducing new concepts (e.g., proxy and its corresponding marks), techniques (e.g., correction of avatars movement), empirical findings (e.g., usefulness of displaced view for manipulation of physical objects), and applications (e.g., occluded interaction). The concept presented in Paper 2 points at design space in which not only the space but also the user's hands can be arbitrary distributed (i.e., this is possible with Poros by reaching into a proxy that is connected to multiple marks). Recent related work in this space showed that representing a hand by many virtual hands can improve performance of object selection [123].

An alternative to duplicating the user's hands is to avoid extra representation of them. For example, an extra pair of translucent hands is often used for guiding users during training (sometimes called ghosted hands). Such guidance can be embedded in a single-hand representation as shown in Paper 4. This research direction has been explored in related work that investigated the integration of two persons into one avatar [37, 138]. When relating this research direction to Paper 4, it is interesting to consider replacing the error minimization algorithm with another person that nudging the user towards target movement. Such help could be embedded in the user's avatar with techniques similar to the one in Paper 4.

Apart from exploring how to show the user's hands, I have also explored ways of showing the objects the users interacts with. Paper 1 compared different ways of showing the occluded objects (including at their actual location as well as displaced), Paper 2 explored showing the objects through surrogates, and Paper 3 explored showing the objects at their past and future location. The latter coupled manipulation of objects to manipulation of time and allowed users to engage in temporal navigation directly. While Paper 3 did this in VR it is possible to imagine that such visualizations would to some extent work in AR by manipulating real objects or their virtual counterparts. Such technique could serve as an artificial memory since one of the major challenges there is the retrieval of memories (see SenseCam for potential applications of artificial memory [41]).

The motivation to use greater capability of our hands also implies exploring techniques that allow mastery and not just ease of use. Like writing, drawing and playing a musical instrument requires training, the future MR techniques that push the boundaries of what our bodies and technologies are capable of will also need training. This thesis has only investigated one technique for subtly supporting training of hand movements (Paper 4). Subtle techniques might not always be appropriate [114] or possible therefore some MR techniques will require extensive training. Human-computer interaction research rarely designs techniques that need extensive training since they would require longitudinal studies and efforts that most of the research labs cannot afford.

To use the full capacity of our bodies we should not only avoid ignoring parts of our bodies (e.g., hands) but also avoid focusing on only few of our senses. The papers in this thesis mostly focused on vision and ignored other modalities (except for the techniques in Paper 1 which offered rich haptic feedback via interaction with real objects). It happens often that MR interaction techniques focus on only one or two modalities. One of the speculations in Paper 1 was that the displaced 3D view worked well is because of the rich haptic feedback and familiarity with daily life tasks used in the study. The good performance was contrary to earlier related work showing that displaced view can affect manipulation of objects negatively. One speculation here is that involving more modalities could potentially give more benefits than summing the benefits of individual modalities.

It is still unclear which types of the techniques will prevail once MR devices become more widespread. A typical workday in MR might still contain a keyboard and WIMP interface, however, to achieve the full potential of MR it is necessary to avoid only optimizing how we currently interact with computers and envision new ways of seeing, moving, thinking and learning.

9.2. Limitations

This thesis has several limitations related to its scope, methods used and the equipment used. I focused on hand-based MR techniques and explored them broadly by investigating manipulation, navigation, and learning. This in turn did not allow me to focus on one specific task (e.g., selection) or application (e.g., occluded interaction) and optimize the researched techniques for it over several papers. The benefit of my approach was that it allowed me to explore MR's opportunities for hand-based interactions freely. Through this I discovered ways for supporting occluded interaction (Paper 1), interaction at distance (Paper 2), temporal navigation (Paper 3), and learning of motor skills (Paper 4). Focusing on only hand-based techniques limited the solution space for the researched problems but in turn stayed within the bounds of the main motivation of this thesis (involving more of our bodies when interacting with real and virtual around us).

The formative studies in Paper 2 and 3 used tasks and study settings that are difficult to replicate. This makes the techniques difficult to compare the any newly design ones, however, it also avoided creating artificial settings or limiting the scope of the researched concepts. Here I would point to Greenberg and Buxton's paper that mentions the pitfalls researchers can find themselves in when trying to force an empirical study onto a project too early [36].

Contrary to the formative studies, the experiments in Paper 1 and Paper 4 used tasks that can be used to evaluate or compare similar techniques. However, the task there were relatively basic (Paper 1) or abstract (Paper 4). Due to this it might be hard to generalize to more complex scenarios.

There are also a few limitations related to demographics of participants that we recruited for our studies. Some of the participants of Paper 1 and 3 studies were new to AR or VR, which might have had an effect on their performance and rating of the techniques. For the experiments in Paper 4, we recruited only participants that owned a VR headset and were used to being in VR. However, the issue of recruiting through the selected online VR community was that it was composed of mostly male population. Considering that MR technologies can become the next generation computing platforms, we should strive for a representative sample of the general population when developing new techniques.

The hardware we used for the studies also has limitations. The experiment in Paper 1 used HoloLens 1, which display has a limited field of view. The experiments in Paper 4 used Oculus Quest 1 for hand tracking which accuracy and latency lacks behind capabilities of modern motion capture systems. We designed our studies to mitigate these limitations as much as possible (e.g., by using tasks that require movement only in the area that is best covered with tracking).

9.3. Beyond Basic Interaction Techniques

Interactions with the real and virtual can be researched from a staggering number of perspectives. By focusing on low-level tasks and the interaction techniques for accomplishing them, the thinking space becomes less overwhelming. However, low-level tasks are by definition never the end-goal. The fascination that MR offers is not for its ability to allow us to manipulate a virtual cube efficiently but for its ability to manipulate us. To make us see, hear, move, think and feel slightly or wholly different.

Changing how we see and feel can be accomplished without MR. For example, by practicing architecture we can train ourselves to see details on buildings that few can see. Such training would make us walk down the street in a different manner from, for example, a trained psychologist or choreographer. Changing how we see, move and feel can also happen through conversations, music, reading, conceptual understanding, or a simple act of staying still. MR can support all of the above activities and also directly affect perception and movement (e.g., as in Paper 4).

It is not much of a leap to imagine that one day we will be able to build upon basic interaction techniques and create experiences that make us see, move and feel in specific ways when walking down the street. The MR techniques could nudge us to not only move our hands in specific way but also our head and gaze. Employing such techniques could make us see and move similarly to an architect or a choreographer. This might in turn affect what become aware of and eventually think and feel.

Bibliography

[1] Parastoo Abtahi, Mar Gonzalez-Franco, Eyal Ofek, and Anthony Steed. I'm a Giant: Walking in Large Virtual Environments at High Speed Gains. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, Glasgow, Scotland Uk. Association for Computing Machinery, 2019. DOI: 10.1145/3290605.3300752.

[2] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. YouMove: Enhancing Movement Training with an Augmented Reality Mirror. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, St. Andrews, Scotland, United Kingdom. Association for Computing Machinery, 2013. DOI: 10.1145/2501988.2502045.

[3] Ferran Argelaguet and Carlos Andujar. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics*, 37(3), 2013. DOI: 10.1016/j.cag.2012.12.003.

[4] Roland Arsenault and Colin Ware. Eye-Hand Co-Ordination with Force Feedback. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '00*, New York, New York, USA. ACM Press, 2000. DOI: 10.1145/332040.332466.

[5] Mahdi Azmandian, Mark Hancock, Hrvoje Benko, Eyal Ofek, and Andrew D. Wilson. Haptic Retargeting: Dynamic Repurposing of Passive Haptics for Enhanced Virtual Reality Experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, San Jose, California, USA. Association for Computing Machinery, 2016. DOI: 10.1145/2858036.2858226.

[6] R Harald Baayen, Douglas J Davidson, and Douglas M Bates. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4), 2008.

[7] P. Barnum, Y. Sheikh, A. Datta, and T. Kanade. Dynamic seethroughs: Synthesizing hidden views of moving objects. In *2009 8th IEEE International Symposium on Mixed and Augmented Reality*, October 2009. DOI: 10.1109/ISMAR.2009.5336483.

[8] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 2015. DOI: 10.18637/jss.v067.i01.

[9] Hrvoje Benko and Steven Feiner. Balloon selection: A multi-finger technique for accurate low-fatigue 3d selection. In *2007 IEEE symposium on 3D user interfaces*. IEEE, 2007.

[10] Joanna Bergström, Tor-Salve Dalsgaard, Jason Alexander, and Kasper Hornbæk. How to Evaluate Object Selection and Manipulation in VR? Guidelines from 20 Years of Studies. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, Yokohama, Japan. Association for Computing Machinery, 2021. DOI: 10.1145/3411764.3445193.

[11] Joanna Bergström, Aske Mottelson, and Jarrod Knibbe. Resized grasping in vr: Estimating thresholds for object discrimination. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, 2019.

[12] Christoph Bichlmeier, Sandro Michael Heining, Marco Feuerstein, and Nassir Navab. The virtual mirror: a new interaction paradigm for augmented reality environments. *IEEE Transactions on Medical Imaging*, 28(9), 2009.

[13] Frank Biocca, Arthur Tang, Charles Owen, and Fan Xiao. Attention Funnel: Omnidirectional 3D Cursor for Mobile Augmented Reality Platforms. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, Montréal, Québec, Canada. Association for Computing Machinery, 2006. DOI: 10.1145/1124772.1124939.

[14] Andrew Bluff. 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, Brisbane, QLD, Australia. Association for Computing Machinery, 2019. DOI: 10.1145/3359997.3365682.

[15] Richard A. Bolt. "Put-that-there": Voice and Gesture at the Graphics Interface. In *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '80, Seattle, Washington, USA. ACM, 1980. DOI: 10.1145/800250.807503.

[16] Rita Borgo, Min Chen, Ben Daubney, Edward Grundy, Gunther Heidemann, Benjamin Höferlin, Markus Höferlin, Heike Leitte, Daniel Weiskopf, and Xianghua Xie. State of the art report on video-based graphics and video visualization. In *Computer Graphics Forum*, volume 31 of number 8. Wiley Online Library, 2012.

[17] Doug A. Bowman and Larry 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*, I3D '97, Providence, Rhode Island, USA. ACM, 1997. DOI: 10.1145/253284.253301.

[18] Doug A. Bowman, David Koller, and Larry F. Hodges. Travel in Immersive Virtual Environments: An Evaluation of Viewpoint Motion Control Techniques. In *Proceedings of the 1997 Virtual Reality Annual International Symposium (VRAIS '97)*, VRAIS '97, USA. IEEE Computer Society, 1997.

[19] Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola, and Ivan Poupyrev. An Introduction to 3-D User Interface Design. *Presence: Teleoperators and Virtual Environments*, 10(1), 2001. DOI: 10.1162/105474601750182342.

[20] Virginia Braun and Victoria Clarke. Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, and K. J. Sher, editors, *APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological*, pages 57–71. American Psychological Association, 2012.

[21] Ian M Bullock and Aaron M Dollar. Classifying human manipulation behavior. In *Rehabilitation Robotics (ICORR), 2011 IEEE International Conference on*. IEEE, 2011.

[22] Han Joo Chae, Jeong-in Hwang, and Jinwook Seo. Wall-based Space Manipulation Technique for Efficient Placement of Distant Objects in Augmented Reality. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, UIST '18, Berlin, Germany. ACM, 2018. DOI: 10.1145/3242587.3242631.

[23] Luca Chittaro and Stefano Burigat. 3D Location-pointing As a Navigation Aid in Virtual Environments. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '04, Gallipoli, Italy. ACM, 2004. DOI: 10.1145/989863.989910.

[24] Brian Cohn, Antonella Maselli, Eyal Ofek, and Mar Gonzalez Franco. SnapMove: Movement Projection Mapping in Virtual Reality. In *IEEE AIVR 2020*, December 2020.

[25] Ashley Colley, Olli Koskenranta, Jani Väyrynen, Leena Ventä-Olkkonen, and Jonna Häkkilä. Windows to other places: exploring solutions for seeing through walls using handheld projection. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*. ACM, 2014.

[26] Klen Čopič Pucihar, Paul Coulton, and Jason Alexander. Evaluating Dual-view Perceptual Issues in Handheld Augmented Reality: Device vs. User Perspective Rendering. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, ICMI '13, Sydney, Australia. ACM, 2013. DOI: 10.1145/2522848.2522885.

[27] Pierre Dragicevic, Gonzalo Ramos, Jacobo Bibliowitcz, Derek Nowrouzezahrai, Ravin Balakrishnan, and Karan Singh. Video Browsing by Direct Manipulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, Florence, Italy. ACM, 2008. DOI: 10.1145/1357054.1357096.

[28] Maximilian Dürr, Rebecca Weber, Ulrike Pfeil, and Harald Reiterer. EGuide: Investigating different Visual Appearances and Guidance Techniques for Egocentric Guidance Visualizations. In *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction*, 2020.

[29] Niklas Elmqvist. BalloonProbe: Reducing Occlusion in 3D Using Interactive Space Distortion. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '05, Monterey, CA, USA. Association for Computing Machinery, 2005. DOI: 10.1145/1101616.1101643.

[30] Mustafa Tolga Eren and Selim Balcisoy. Evaluation of X-ray visualization techniques for vertical depth judgments in underground exploration. *The Visual Computer*, 34(3), 2018.

[31] Thomas Feix, Ian M Bullock, and Aaron M Dollar. Analysis of human grasping behavior: Object characteristics and grasp type. *IEEE transactions on haptics*, 7(3), 2014.

[32] Andrew Forsberg, Kenneth Herndon, and Robert Zeleznik. Aperture based selection for immersive virtual environments. In *ACM Symposium on User Interface Software and Technology*. Citeseer, 1996.

[33] Nirit Gavish, Teresa Gutierrez, Sabine Webel, Jorge Rodriguez, Matteo Peveri, Uli Bockholt, and Franco Tecchia. Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments*, 23(6), 2015. DOI: 10.1080/10494820.2013.815221.

[34] Dan B. Goldman, Chris Gonterman, Brian Curless, David Salesin, and Steven M. Seitz. Video Object Annotation, Navigation, and Composition. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology*, UIST '08, Monterey, CA, USA. ACM, 2008. DOI: 10.1145/1449715.1449719.

[35] M. Gonzalez-Franco, B. Cohn, E. Ofek, D. Burin, and A. Maselli. The Self-Avatar Follower Effect in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2020.

[36] S Greenberg and B Buxton. Usability evaluation considered harmful. In *Proceedings of the 2008 SIDCHI Conference*, 2007.

[37] Takayoshi Hagiwara, Maki Sugimoto, Masahiko Inami, and Michiteru Kitazaki. Shared Body by Action Integration of Two Persons: Body Ownership, Sense of Agency and Task Performance. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2019. DOI: 10.1109/VR.2019.8798222.

[38] Ping-Hsuan Han, Yang-Sheng Chen, Yilun Zhong, Han-Lei Wang, and Yi-Ping Hung. My Tai-Chi Coaches: An Augmented-Learning Tool for Practicing Tai-Chi Chuan. In *Proceedings of the 8th Augmented Human International Conference*, AH '17, Silicon Valley, California, USA. Association for Computing Machinery, 2017. DOI: 10.1145/3041164.3041194.

[39] Richard Held, Aglaia Efstathiou, and Martha Greene. Adaptation to displaced and delayed visual feedback from the hand. *Journal of Experimental Psychology*, 72(6), 1966.

[40] Steven J Henderson and Steven K Feiner. Augmented reality in the psychomotor phase of a procedural task. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 2011.

[41] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, and Ken Wood. SenseCam: A retrospective memory aid. In *International Conference on Ubiquitous Computing*. Springer, 2006.

[42] Samantha Horvath, John Galeotti, Bing Wu, Roberta Klatzky, Mel Siegel, and George Stetten. FingerSight: Fingertip Haptic Sensing of the Visual Environment. *IEEE Journal of Translational Engineering in Health and Medicine*, 2, 2014. DOI: 10.1109/JTEHM.2014.2309343.

[43] Daisuke Iwai and Kosuke Sato. Document search support by making physical documents transparent in projection-based mixed reality. *Virtual reality*, 15(2-3), 2011.

[44] LS Jakobson and Melvyn A Goodale. Trajectories of reaches to prismatically-displaced targets: evidence for "automatic" visuomotor recalibration. *Experimental Brain Research*, 78(3), 1989.

[45] Shixin Jiang and Jun Rekimoto. Mediated-Timescale Learning: Manipulating Timescales in Virtual Reality to Improve Real-World Tennis Forehand Volley. In *26th ACM Symposium on Virtual Reality Software and Technology*, VRST '20, Virtual Event, Canada. Association for Computing Machinery, 2020. DOI: 10.1145/3385956.3422128.

[46] Wolfgang Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5), September 1976. DOI: 10.1107/S0567739476001873.

[47] Bernhard Kainz, Stefan Hauswiesner, Gerhard Reitmayr, Markus Steinberger, Raphael Grasset, Lukas Gruber, Eduardo Veas, Denis Kalkofen, Hartmut Seichter, and Dieter Schmalstieg. OmniKinect: Real-time Dense Volumetric Data Acquisition and Applications. In *Proceedings of the 18th ACM Symposium on Virtual Reality Software and Technology*, VRST '12, Toronto, Ontario, Canada. ACM, 2012. DOI: 10.1145/2407336.2407342.

[48] Bernhard Kainz, Stefan Hauswiesner, Gerhard Reitmayr, Markus Steinberger, Raphael Grasset, Lukas Gruber, Eduardo Veas, Denis Kalkofen, Hartmut Seichter, and Dieter Schmalstieg. Omnikinect: real-time dense volumetric data acquisition and applications. In *Proceedings of the 18th ACM symposium on Virtual reality software and technology*. ACM, 2012.

[49] Thorsten Karrer, Malte Weiss, Eric Lee, and Jan Borchers. DRAGON: A Direct Manipulation Interface for Frame-accurate In-scene Video Navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, Florence, Italy. ACM, 2008. DOI: 10.1145/1357054.1357097.

[50] Kevin Kelly, Adam Heilbrun, and Barbara Stacks. Virtual reality: an interview with Jaron Lanier. *Whole Earth Review*, 64(108-120), 1989.

[51] Azam Khan, George Fitzmaurice, Don Almeida, Nicolas Burtnyk, and Gordon Kurtenbach. A Remote Control Interface for Large Displays. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, UIST '04, Santa Fe, NM, USA. Association for Computing Machinery, 2004. DOI: 10.1145/1029632.1029655.

[52] David Kim, Otmar Hilliges, Shahram Izadi, Alex D. Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. Digits: Freehand 3D Interactions Anywhere Using a Wrist-worn Gloveless Sensor. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, Cambridge, Massachusetts, USA. ACM, 2012. DOI: 10.1145/2380116.2380139.

[53] Don Kimber, Tony Dunnigan, Andreas Girgensohn, Frank Shipman, Thea Turner, and Tao Yang. Trailblazing: Video playback control by direct object manipulation. In *2007 IEEE International Conference on Multimedia and Expo*. IEEE, 2007.

[54] David Kirsh. How marking in dance constitutes thinking with the body, 2011.

[55] Jarrod Knibbe, Sue Ann Seah, and Mike Fraser. VideoHandles: replicating gestures to search through action-camera video. In *Proceedings of the 2nd ACM symposium on Spatial user interaction*. ACM, 2014.

[56] Luv Kohli. Redirected touching: Warping space to remap passive haptics. In *2010 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 2010.

[57] Luv Kohli, Mary C Whitton, and Frederick P Brooks. Redirected Touching: Training and adaptation in warped virtual spaces. In *2013 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 2013.

[58] Ioannis Kotziampasis, Nathan Sidwell, and Alan Chalmers. Portals: Increasing Visibility in Virtual Worlds. In *Proceedings of the 19th Spring Conference on Computer Graphics*, SCCG '03, Budmerice, Slovakia. Association for Computing Machinery, 2003. DOI: 10.1145/984952.984995.

[59] Ioannis Kotziampasis, Nathan Sidwell, and Alan Chalmers. Portals: increasing visibility in virtual worlds. In *Proceedings of the 19th Spring Conference on Computer Graphics*, 2003.

[60] Robert Krempien, Harald Hoppe, Lüder Kahrs, Sascha Daeuber, Oliver Schorr, Georg Eggers, Marc Bischof, Marc W. Munter, Juergen Debus, and Wolfgang Harms. Projector-Based Augmented Reality for Intuitive Intraoperative Guidance in Image-Guided 3D Interstitial Brachytherapy. *International Journal of Radiation Oncology*Biology*Physics*, 70(3), March 2008. DOI: 10.1016/j.ijrobp.2007.10.048.

[61] Torsten W. Kuhlen and Bernd Hentschel. Quo Vadis CAVE: Does Immersive Visualization Still Matter? *IEEE Computer Graphics and Applications*, 34(5), September 2014. DOI: 10.1109/MCG.2014.97.

[62] André Kunert, Alexander Kulik, Stephan Beck, and Bernd Froehlich. Photoportals: Shared References in Space and Time. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, Baltimore, Maryland, USA. Association for Computing Machinery, 2014. DOI: 10.1145/2531602.2531727.

[63] Takeshi Kurata, Nobuchika Sakata, Masakatsu Kourogi, Hideaki Kuzuoka, and Mark Billinghurst. Remote collaboration using a shoulder-worn active camera/laser. In *Wearable Computers, 2004. ISWC 2004. Eighth International Symposium on*, volume 1. IEEE, 2004.

[64] Bum chul Kwon, Waqas Javed, Niklas Elmqvist, and Ji Soo Yi. Direct Manipulation through Surrogate Objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, Vancouver, BC, Canada. Association for Computing Machinery, 2011. DOI: 10.1145/1978942.1979033.

[65] Joseph J LaViola, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. 3D User Interfaces: Theory and Practice, 2017.

[66] H. Lee, S. Noh, and W. Woo. TunnelSlice: Freehand Subspace Acquisition Using an Egocentric Tunnel for Wearable Augmented Reality. *IEEE Transactions on Human-Machine Systems*, 47(1), February 2017. DOI: 10.1109/THMS.2016.2611821.

[67] Jinna Lei, Xiaofeng Ren, and Dieter Fox. Fine-grained Kitchen Activity Recognition Using RGB-D. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, Pittsburgh, Pennsylvania. ACM, 2012. DOI: 10.1145/2370216.2370248.

[68] Klemen Lilija, Søren Kyllingsbæk, and Kasper Hornbæk. Correction of Avatar Hand Movements Supports Learning of a Motor Skill. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, 2021. DOI: 10.1109/VR50410.2021.00069.

[69] Klemen Lilija, Henning Pohl, Sebastian Boring, and Kasper Hornbæk. Augmented Reality Views for Occluded Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA. Association for Computing Machinery, 2019.

[70] Klemen Lilija, Henning Pohl, Sebastian Boring, and Kasper Hornbæk. Augmented Reality Views for Occluded Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, Glasgow, Scotland Uk. Association for Computing Machinery, 2019. DOI: 10.1145/3290605.3300676.

[71] Klemen Lilija, Henning Pohl, and Kasper Hornbæk. *Who Put That There? Temporal Navigation of Spatial Recordings by Direct Manipulation*. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2020, pages 1–11.

[72] Jakub Limanowski and Felix Blankenburg. Minimal self-models and the free energy principle. *Frontiers in human neuroscience*, 7, 2013.

[73] David Lindlbauer and Andy D. Wilson. Remixed Reality: Manipulating Space and Time in Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, New York, New York, USA. ACM Press, 2018. DOI: 10.1145/3173574.3173703.

[74] Fei Liu and Stefan Seipel. Precision study on augmented reality-based visual guidance for facility management tasks. *Automation in Construction*, 90, 2018.

[75] J. Liu, F. Feng, Y. C. Nakamura, and N. S. Pollard. A taxonomy of everyday grasps in action, November 2014. DOI: 10.1109/HUMANOIDS.2014.7041420.

[76] James Liu, Hirav Parekh, Majed Al-Zayer, and Eelke Folmer. Increasing Walking in VR Using Redirected Teleportation. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, UIST '18, Berlin, Germany. Association for Computing Machinery, 2018. DOI: 10.1145/3242587.3242601.

[77] Sean J. Liu, Maneesh Agrawala, Stephen DiVerdi, and Aaron Hertzmann. View-Dependent Video Textures for 360° Video. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, New Orleans, LA, USA. Association for Computing Machinery, 2019. DOI: 10.1145/3332165.3347887.

[78] Andrés Lucero, Dzmitry Aliakseyeu, Kees Overbeeke, and Jean-Bernard Martens. An Interactive Support Tool to Convey the Intended Message in Asynchronous Presentations. In *Proceedings of the International Conference on Advances in Computer Enterntainment Technology*, ACE '09, Athens, Greece. Association for Computing Machinery, 2009. DOI: 10.1145/1690388.1690391.

[79] Steve Mann, Tom Furness, Yu Yuan, Jay Iorio, and Zixin Wang. All reality: Virtual, augmented, mixed (x), mediated (x, y), and multimediated reality. *arXiv preprint arXiv:1804.08386*, 2018.

[80] Steve Mann and Steve Mann Nnlf. Mediated reality, 1994.

[81] Joe Marshall and Paul Tennent. The Limitations of Reality. In *Proceedings of the Halfway to the Future Symposium 2019*, 2019.

[82] Hannes Matuschek, Reinhold Kliegl, Shravan Vasishth, Harald Baayen, and Douglas Bates. Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 2017.

[83] Walterio W Mayol-Cuevas, Ben J Tordoff, and David W Murray. On the choice and placement of wearable vision sensors. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 39(2), 2009.

[84] James Mccrae, Niloy J Mitra, and Karan Singh. Surface perception of planar abstractions. *ACM Transactions on Applied Perception (TAP)*, 10(3), 2013.

[85] Daniel Mendes, Fabio Marco Caputo, Andrea Giachetti, Alfredo Ferreira, and J Jorge. A survey on 3d virtual object manipulation: From the desktop to immersive virtual environments. In *Computer graphics forum*, volume 38 of number 1. Wiley Online Library, 2019.

[86] Daniel Mendes, Daniel Medeiros, Eduardo Cordeiro, Maurıcio Sousa, Alfredo Ferreira, and Joaquim Jorge. PRECIOUS! Out-of-reach selection using iterative refinement in VR. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 2017.

[87] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12), 1994.

[88] Mark R. Mine, Frederick P. Brooks, and Carlo H. Sequin. Moving Objects in Space: Exploiting Proprioception in Virtual-Environment Interaction. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, USA. ACM Press/Addison-Wesley Publishing Co., 1997. DOI: 10.1145/258734.258747.

[89] Roberto A. Montano Murillo, Sriram Subramanian, and Diego Martinez Plasencia. Erg-O: Ergonomic Optimization of Immersive Virtual Environments. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, Québec City, QC, Canada. Association for Computing Machinery, 2017. DOI: 10.1145/3126594.3126605.

[90] Shohei Mori, Sei Ikeda, and Hideo Saito. A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects. *IPSJ Transactions on Computer Vision and Applications*, 9(1), December 2017. DOI: 10.1186/s41074-017-0028-1.

[91] Shohei Mori, Momoko Maezawa, and Hideo Saito. A work area visualization by multi-view camera-based diminished reality. *Multimodal Technologies and Interaction*, 1(3), 2017.

[92] Nassir Navab, Joerg Traub, Tobias Sielhorst, Marco Feuerstein, and Christoph Bichlmeier. Action-and workflow-driven augmented reality for computer-aided medical procedures. *IEEE Computer Graphics and Applications*, 27(5), 2007.

[93] Carman Neustaedter and Elena Fedorovskaya. Capturing and sharing memories in a virtual world. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2009.

[94] Cuong Nguyen, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. CollaVR: Collaborative In-Headset Review for VR Video. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, Québec City, QC, Canada. ACM, 2017. DOI: 10.1145/3126594.3126659.

[95] Cuong Nguyen, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. Vremiere: In-Headset Virtual Reality Video Editing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, Denver, Colorado, USA. ACM, 2017. DOI: 10.1145/3025453.3025675.

[96] Cuong Nguyen, Yuzhen Niu, and Feng Liu. Direct Manipulation Video Navigation in 3D. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, Paris, France. ACM, 2013. DOI: 10.1145/2470654.2466150.

[97] Torsten Ingemann Nielsen. Volition: A new experimental approach. *Scandinavian journal of psychology*, 4(1), 1963.

[98] Alex Olwal and Steven Feiner. The Flexible Pointer: An Interaction Technique for Selection in Augmented and Virtual Reality. In *Proc. UIST'03*, 2003.

[99] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L Davidson, Sameh Khamis, Mingsong Dou, et al. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, 2016.

[100] Ori Ossmy and Roy Mukamel. Neural network underlying intermanual skill transfer in humans. *Cell reports*, 17(11), 2016.

[101] Riccardo Palmarini, John Ahmet Erkoyuncu, Rajkumar Roy, and Hosein Torabmostaedi. A systematic review of augmented reality applications in maintenance. *Robotics and Computer-Integrated Manufacturing*, 49, 2018. DOI: https://doi.org/10.1016/j.rcim.2017.06.002.

[102] Randy Pausch, Tommy Burnette, Dan Brockway, and Michael 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*, SIGGRAPH '95, New York, NY, USA. Association for Computing Machinery, 1995. DOI: 10.1145/218380.218495.

[103] Randy Pausch, Dennis Proffitt, and George Williams. Quantifying Immersion in Virtual Reality. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, USA. ACM Press/Addison-Wesley Publishing Co., 1997. DOI: 10.1145/258734.258744.

[104] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew 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*. ACM, 2016. DOI: 10.1145/2818048.2819965.

[105] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew D 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 - CSCW '16*, New York, New York, USA. ACM Press, 2016. DOI: 10.1145/2818048.2819965.

[106] Ken Perlin and David Fox. Pad: An Alternative Approach to the Computer Interface. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '93, Anaheim, CA. Association for Computing Machinery, 1993. DOI: 10.1145/166117.166125.

[107] Benjamin Petry and Jochen Huber. Towards Effective Interaction with Omnidirectional Videos Using Immersive Virtual Reality Headsets. In *Proceedings of the 6th Augmented Human International Conference*, AH '15, Singapore, Singapore. ACM, 2015. DOI: 10.1145/2735711.2735785.

[108] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. Gaze + Pinch Interaction in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, SUI '17, Brighton, United Kingdom. Association for Computing Machinery, 2017. DOI: 10.1145/3131277.3132180.

[109] J. S. Pierce and R. Pausch. Navigation with place representations and visible landmarks. In *IEEE Virtual Reality 2004*, March 2004. DOI: 10.1109/VR.2004.1310071.

[110] Jeffrey S. Pierce, Andrew S. Forsberg, Matthew J. Conway, Seung Hong, Robert C. Zeleznik, and Mark R. Mine. Image plane interaction techniques in 3D immersive environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics*. ACM, 1997.

[111] Jeffrey S. Pierce, Brian C. Stearns, and Randy Pausch. Voodoo Dolls: Seamless Interaction at Multiple Scales in Virtual Environments. In *Proceedings of the 1999 Symposium on Interactive 3D Graphics*, I3D '99, Atlanta, Georgia, USA. Association for Computing Machinery, 1999. DOI: 10.1145/300523.300540.

[112] Thammathip Piumsomboon, Gun A. Lee, Andrew Irlitti, Barrett Ens, Bruce H. Thomas, and Mark 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*, CHI '19, Glasgow, Scotland Uk. Association for Computing Machinery, 2019. DOI: 10.1145/3290605.3300458.

[113] Henning Pohl, Klemen Lilija, Jess McIntosh, and Kasper Hornbæk. Poros: Configurable Proxies for Distant Interactions in VR. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, Yokohama, Japan. Association for Computing Machinery, 2021. DOI: 10.1145/3411764.3445685.

[114] Henning Pohl, Andreea Muresan, and Kasper Hornbæk. Charting Subtle Interaction in the HCI Literature. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, Glasgow, Scotland Uk. Association for Computing Machinery, 2019. DOI: 10.1145/3290605.3300648.

[115] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao 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.

[116] Rob Reilink, Gart de Bruin, Michel Franken, Massimo A Mariani, Sarthak Misra, and Stefano Stramigioli. Endoscopic camera control by head movements for thoracic surgery. In *Biomedical Robotics and Biomechatronics (BioRob), 2010 3rd IEEE RAS and EMBS International Conference on*. IEEE, 2010.

[117] Alberto Romay, Stefan Kohlbrecher, David C Conner, and Oskar Von Stryk. Achieving versatile manipulation tasks with unknown objects by supervised humanoid robots based on object templates. In *Humanoids*, 2015.

[118] Yves Rossetti, Kazuo Koga, and Tadaaki Mano. Prismatic displacement of vision induces transient changes in the timing of eye-hand coordination. *Perception & Psychophysics*, 54(3), 1993.

[119] Daniel Roth, Gary Bente, Peter Kullmann, David Mal, Chris Felix Purps, Kai Vogeley, and Marc Erich Latoschik. Technologies for Social Augmentations in User-Embodied Virtual Reality. In *25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19, Parramatta, NSW, Australia. Association for Computing Machinery, 2019. DOI: 10.1145/3359996.3364269.

[120] Gustavo Alberto Rovelo Ruiz, Davy Vanacken, Kris Luyten, Francisco Abad, and Emilio Camahort. Multi-viewer Gesture-based Interaction for Omni-directional Video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, Toronto, Ontario, Canada. ACM, 2014. DOI: 10.1145/2556288.2557113.

[121] Christian Sandor, Arindam Dey, Andrew Cunningham, Sebastien Barbier, Ulrich Eck, Donald Urquhart, Michael R. Marner, Graeme Jarvis, and Sang Rhee. Egocentric space-distorting visualizations for rapid environment exploration in mobile mixed reality. In *2010 IEEE Virtual Reality Conference (VR)*, March 2010. DOI: 10.1109/VR.2010.5444815.

[122] Takashi Satou, Haruhiko Kojima, Akihito Akutsu, and Yoshinobu Tonomura. CyberCoaster: Polygonal line shaped slider interface to spatio-temporal media. In *Proceedings of the seventh ACM international conference on Multimedia (Part 2)*. ACM, 1999.

[123] Jonas Schjerlund, Kasper Hornbæk, and Joanna Bergström. Ninja Hands: Using Many Hands to Improve Target Selection in VR. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, Yokohama, Japan. Association for Computing Machinery, 2021. DOI: 10.1145/3411764.3445759.

[124] Kaoru Sekiyama, Satoru Miyauchi, Toshihide Imaruoka, Hiroyuki Egusa, and Takara Tashiro. Body Image as a Visuomotor Transformation Device Revealed in Adaptation to Reversed Vision. *Nature*, 407(6802), September 2000. DOI: 10.1038/35030096.

[125] Roy Shilkrot, Jochen Huber, Jürgen Steimle, Suranga Nanayakkara, and Pattie Maes. Digital Digits: A Comprehensive Survey of Finger Augmentation Devices. *ACM Comput. Surv.*, 48(2), November 2015. DOI: 10.1145/2828993.

[126] Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic bulletin & review*, 20(1), 2013.

[127] Richard Skarbez, Missie Smith, and Mary C Whitton. Revisiting Milgram and Kishino's Reality-Virtuality Continuum. *Frontiers in Virtual Reality*, 2, 2021.

[128] Rajinder Sodhi, Hrvoje Benko, and Andrew Wilson. LightGuide: Projected Visualizations for Hand Movement Guidance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Sys-*

tems, CHI '12, Austin, Texas, USA. Association for Computing Machinery, 2012. DOI: 10.1145/2207676.2207702.

[129] Maurıcio Sousa, João Vieira, Daniel Medeiros, Artur Arsenio, and Joaquim Jorge. SleeveAR: Augmented Reality for Rehabilitation Using Realtime Feedback. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, IUI '16, Sonoma, California, USA. Association for Computing Machinery, 2016. DOI: 10.1145/2856767.2856773.

[130] Maximilian Speicher, Brian D Hall, and Michael Nebeling. What is mixed reality? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019.

[131] Lee Stearns, Victor DeSouza, Jessica Yin, Leah Findlater, and Jon E. Froehlich. Augmented Reality Magnification for Low Vision Users with the Microsoft Hololens and a Finger-Worn Camera. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '17*, New York, New York, USA. ACM Press, 2017. DOI: 10.1145/3132525.3134812.

[132] Frank Steinicke, Timo Ropinski, and Klaus Hinrichs. Object selection in virtual environments using an improved virtual pointer metaphor. In *Computer Vision and Graphics*, pages 320–326. Springer, 2006.

[133] Richard Stoakley, Matthew J. Conway, and Randy Pausch. Virtual reality on a WIM: interactive worlds in miniature. In *CHI*, volume 95, 1995.

[134] Stanislav L. Stoev and Dieter Schmalstieg. Application and Taxonomy of Through-the-Lens Techniques. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '02, Hong Kong, China. Association for Computing Machinery, 2002. DOI: 10.1145/585740.585751.

[135] Kazuya Sugimoto, Hiromitsu Fujii, Atsushi Yamashita, and Hajime Asama. Half-diminished reality image using three rgb-d sensors for remote control robots. In *Safety, Security, and Rescue Robotics (SSRR), 2014 IEEE International Symposium on*. IEEE, 2014.

[136] L. Sun, U. Klank, and M. Beetz. EYEWATCHME—3D Hand and object tracking for inside out activity analysis. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, June 2009. DOI: 10.1109/CVPRW.2009.5204358.

[137] Ivan E. Sutherland. A Head-Mounted Three Dimensional Display. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*, AFIPS '68 (Fall, part I), San Francisco, California. Association for Computing Machinery, 1968. DOI: 10.1145/1476589.1476686.

[138] Ryo Takizawa, Takayoshi Hagiwara, Adrien Verhulst, Masaaki Fukuoka, Michiteru Kitazaki, and Maki Sugimoto. Dynamic Shared Limbs: An Adaptive Shared Body Control Method Using EMG Sensors. In *Augmented Humans Conference 2021*, AHs'21, Rovaniemi, Finland. Association for Computing Machinery, 2021. DOI: 10.1145/3458709.3458932.

[139] Richard Tang, Xing-Dong Yang, Scott Bateman, Joaquim Jorge, and Anthony Tang. Physio@Home: Exploring Visual Guidance and Feedback Techniques for Physiotherapy Exercises. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, Seoul, Republic of Korea. Association for Computing Machinery, 2015. DOI: 10.1145/2702123.2702401.

[140] Robert J. Teather and Wolfgang Stuerzlinger. Exaggerated Head Motions for Game Viewpoint Control. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share*, Future Play '08, Toronto, Ontario, Canada. ACM, 2008. DOI: 10.1145/1496984.1497034.

[141] M. Tenorth, J. Bandouch, and M. Beetz. The TUM Kitchen Data Set of everyday manipulation activities for motion tracking and action recognition. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, September 2009. DOI: 10.1109/ICCVW.2009.5457583.

[142] Balasaravanan Thoravi Kumaravel, Cuong Nguyen, Stephen DiVerdi, and Björn Hartmann. TutoriVR: A Video-Based Tutorial System for Design Applications in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, Glasgow, Scotland, UK. Association for Computing Machinery, 2019. DOI: 10.1145/3290605.3300514.

[143] Balasaravanan Thoravi Kumaravel, Cuong Nguyen, Stephen DiVerdi, and Björn Hartmann. TutoriVR: A Video-Based Tutorial System for Design Applications in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, Glasgow, Scotland Uk. Association for Computing Machinery, 2019. DOI: 10.1145/3290605.3300514.

[144] Emanuel Todorov, Reza Shadmehr, and Emilio Bizzi. Augmented feedback presented in a virtual environment accelerates learning of a difficult motor task. *Journal of motor behavior*, 29(2), 1997.

[145] 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*, JVRC'09, Lyon, France. Eurographics Association, 2009.

[146] Ba Tu Truong and Svetha Venkatesh. Video Abstraction: A Systematic Review and Classification. *ACM Trans. Multimedia Comput. Commun. Appl.*, 3(1), February 2007. DOI: 10.1145/1198302.1198305.

[147] Doménique van Gennip, Elise van den Hoven, and Panos Markopoulos. Things That Make Us Reminisce: Everyday Memory Cues As Opportunities for Interaction Design. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, Seoul, Republic of Korea. ACM, 2015. DOI: 10.1145/2702123.2702460.

[148] Eduardo Velloso, Andreas Bulling, and Hans Gellersen. MotionMA: Motion Modelling and Analysis by Demonstration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, Paris, France. Association for Computing Machinery, 2013. DOI: 10.1145/2470654.2466171.

[149] Colin Ware and Jeff Rose. Rotating virtual objects with real handles. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 6(2), 1999.

[150] Robert B Welch and Gerald Goldstein. Prism adaptation and brain damage. *Neuropsychologia*, 10(4), 1972.

[151] Mark Wentink, Paul Breedveld, Dirk W Meijer, and Henk G Stassen. Endoscopic camera rotation: a conceptual solution to improve hand-eye coordination in minimally-invasive surgery. *Minimally Invasive Therapy & Allied Technologies*, 9(2), 2000.

[152] Barbara M. Wildemuth, Gary Marchionini, Meng Yang, Gary Geisler, Todd Wilkens, Anthony Hughes, and Richard Gruss. How Fast is Too Fast?: Evaluating Fast Forward Surrogates for Digital Video. In *Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL '03, Houston, Texas. IEEE Computer Society, 2003.

[153] Marc Wolter, Irene Tedjo-Palczynski, Bernd Hentschel, and Torsten Kuhlen. Spatial input for temporal navigation in scientific visualizations. *IEEE computer graphics and applications*, 29(6), 2009.

[154] Hans Peter Wyss, Roland Blach, and Matthias Bues. iSith-Intersection-based spatial interaction for two hands. In *3D User Interfaces (3DUI'06)*. IEEE, 2006.

[155] Haijun Xia, Sebastian Herscher, Ken Perlin, and Daniel Wigdor. Spacetime: Enabling Fluid Individual and Collaborative Editing in Virtual Reality. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, UIST '18, Berlin, Germany. Association for Computing Machinery, 2018. DOI: 10.1145/3242587.3242597.

[156] Shan Xiao, Xupeng Yeb, Yaqiu Guobl, Boyu Gaoc, and Jinyi Longy. Transfer of Coordination Skill to the Unpracticed Hand in Immersive Environments. In IEEE, 2020.

[157] Shuo Yan, Gangyi Ding, Zheng Guan, Ningxiao Sun, Hongsong Li, and Longfei Zhang. OutsideMe: Augmenting Dancer's External Self-Image by Using A Mixed Reality System. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, Seoul, Republic of Korea. Association for Computing Machinery, 2015. DOI: 10.1145/2702613.2732759.

[158] Ungyeon Yang and Gerard Jounghyun Kim. Implementation and evaluation of "just follow me": An immersive, VR-based, motion-training system. *Presence: Teleoperators & Virtual Environments*, 11(3), 2002.

[159] Xing-Dong Yang, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. Magic Finger: Always-available Input Through Finger Instrumentation. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, Cambridge, Massachusetts, USA. ACM, 2012. DOI: 10.1145/2380116.2380137.

[160] Tsuneo Yoshikawa. Force control of robot manipulators. 1, 2000.

[161] Xiuming Zhang, Tali Dekel, Tianfan Xue, Andrew Owens, Qiurui He, Jiajun Wu, Stefanie Mueller, and William T. Freeman. MoSculp: Interactive Visualization of Shape and Time. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, UIST '18, Berlin, Germany. ACM, 2018. DOI: 10.1145/3242587.3242592.

[162] Yuhang Zhao, Edward Cutrell, Christian Holz, Meredith Ringel Morris, Eyal Ofek, and Andrew D. Wilson. SeeingVR: A Set of Tools to Make Virtual Reality More Accessible to People with Low Vision. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, Glasgow, Scotland Uk. Association for Computing Machinery, 2019. DOI: 10.1145/3290605.3300341.

[163] Daniel Zielasko, Sven Horn, Sebastian Freitag, Benjamin Weyers, and Torsten W. Kuhlen. Evaluation of hands-free HMD-based navigation techniques for immersive data analysis. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, 2016. DOI: 10.1109/3DUI.2016.7460040.

[164] Stefanie Zollmann, Raphael Grasset, Gerhard Reitmayr, and Tobias Langlotz. Image-based X-ray visualization techniques for spatial understanding in Outdoor Augmented Reality. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design*. ACM, 2014.