Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Towards a Particle Accelerator Metaverse: Mixed Reality Interfaces for Operations of Complex Physics Facilities

Dissertation to achieve a doctoral degree at the Faculty of Mathematics, Informatics, and Natural Sciences at Universität Hamburg

Ke Li
Fachbereich Informatik
Fakultät für Mathematik, Informatik und Naturwissenschaften

October 2024

Supervisor:	Prof. Dr. Frank Steinicke
Supervisor:	Dr. Reinhard Bacher
Supervisor:	Prof. Dr. Wim Leemans
Supervision Panel Chair:	Prof. Dr. Thomas Ludwig
External Reviewer:	Prof. Dr. Yoshifumi Kitamura
Chair of Examination Commission:	Prof. Dr. Jianwei Zhang
Deputy Chair of Examination Commission:	Prof. Dr. Walid Maalej
Date of Thesis Defense:	October 24, 2024

Declaration

I hereby declare, on oath, that I have written the present dissertation on my own and have not used other than the acknowledged resources and aids.

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Hamburg, July 16 2024	Ke Li, Unterschrift

Declaration

The formatting and structuring of this dissertation are made through the adaptation of the novathesis LaTeX template [139]. Additionally, following the ACM guidelines [1], I used OpenAI's ChatGPT for rephrasing sentences and correcting grammar to improve the clarity and readability of my writing. ChatGPT was also employed as a search aid to obtain direct links to important academic papers, ensuring that all referenced works were sourced and reviewed independently by me. The assistance from ChatGPT helps enhance the overall quality and coherence of the thesis, while all original research and analysis remain my own. Detailed documentation of the usage and a list of typical prompts employed can be found in Appendix A.2.

Die Formatierung und Strukturierung dieser Dissertation erfolgte durch die Anpassung der novathesis LaTeX-Vorlage [139]. Zusätzlich habe ich gemäß den Richtlinien der ACM OpenAIs ChatGPT verwendet, um Sätze umzuformulieren und die Grammatik zu korrigieren, um die Klarheit und Lesbarkeit meiner Arbeit zu verbessern. ChatGPT wurde auch als Suchhilfe eingesetzt, um direkte Links zu wichtigen akademischen Artikeln zu erhalten, wobei alle zitierten Werke unabhängig von mir bezogen und geprüft wurden. Die Unterstützung durch ChatGPT trägt zur Verbesserung der Gesamtqualität und Kohärenz der Dissertation bei, während alle ursprünglichen Forschungen und Analysen von mir selbst stammen. Eine ausführliche Dokumentation der Nutzung sowie eine Liste typischer verwendeter Aufforderungen sind im Anhang A.2 zu finden.

[1] https://www.acm.org/publications/policies/frequently-asked-questions

To my future self,
Never give up your love and passion.

Acknowledgements

This dissertation would not have been possible without the support and guidance of many individuals and institutes. Firstly, I express my deep gratitude to my supervisors, Prof. Frank Steinicke, Dr. Reinhard Bacher, and Prof. Wim Leemans. Their unwavering trust in me, by allowing me to implement my research ideas while providing guidance and support, has been invaluable. Collaborating with such esteemed scientists has been both an honor and a pleasure. They are not only exceptional mentors but also incredibly kind, consistently offering opportunities for me to promote myself and my work. I want to thank the members of my dissertation committee, Prof.Kitamura, Prof.Jianwei Zhang, and Prof.Walid Maalej for their time in reviewing this dissertation and their strong support in helping me complete the doctoral procedures.

I am also grateful for the support of the Helmholtz Graduate School for the Structure of Matter (DASHH), which provided me with a generous scholarship, allowing me to fully focus on my research. My heartfelt thanks go to my colleagues from DESY and the HCI group, whose collaborative efforts and expertise across various academic fields and operational functionalities of particle accelerators were crucial for this project. Dr. Susanne Schmidt taught me the methodologies of HCI and consistently supported me in statistical analysis. Dr. Aradhana Choudhuri, Dr. Ingmar Hartl, and Dr. Tino Lang introduced me to the application domain of MR for laser safety and assisted with field studies and empirical user evaluations at high-power laser laboratories. Tim Rolff provided invaluable assistance with C++ programming and computer graphics. André Dehne and Julien Branlard offered support with field studies at the accelerator and the operation of the MARWIN robot. Additionally, the members of the HCI group always lifted my spirits and provided encouragement whenever I felt stuck.

On a personal note, I would like to acknowledge the unconditional love of my friends and family. First of all, my deepest gratitude goes to my family in China for believing in me and supporting my decision to pursue a scientific career abroad. To my extended family in Germany, your warmth and love have provided me with a second home, greatly contributing to my ability to focus and persevere. I am immensely grateful to my friends Ankita, Abhik, Adrià, Carlitos, Jeanette, Jordan, James, Lusitos, Linghua, Max, Rana, Shaan, Ying, Yifan... for continuously adding joy to my life, broadening my worldview, and being my unwavering support system. A special thanks to my friend Lo-I for spending hours listening to me and offering constructive feedback. Your talent and aspirations have inspired me for a decade, and I have no doubt you will become a great philosopher of our time. Finally, to my fiancé, Henrik, you add so much positivity and joy to my life every day. Thank you for understanding my restless mind and sharing my passion for science, music, and art. You are the best thing that could have happened to me, and I can't wait to spend my life with you.

Abstract

Large-scale physics facilities such as particle accelerators and high-power lasers are important laboratory equipment for industrial processes and scientific research. However, they also present some of the most demanding human-computer interaction (HCI) challenges in facility control and maintenance. In recent years, the rise of mixed reality (MR) technology started to transform everyday HCI from conventional 2D user interfaces (UIs) and monitors to 3D spatial UI and immersive head-mounted displays (HMDs). Accompanied by the rapid advancement of computer vision, artificial intelligence (AI), and robotics, the availability of commercial HMDs already provides a glimpse into our digital future of the "*metaverse*", where users can seamlessly perceive and interact with digital content and perform their daily activities such as working, learning, and entertainment in an interconnected immersive MR space. In this dissertation, we present the design, implementation, and evaluation of several novel MR interfaces for operating complex physics facilities by developing novel sensor fusion, immersive neural rendering, and immersive robot teleoperation techniques. As a first step towards a "*particle accelerator metaverse*", this dissertation explored how MR can be used to improve HCI for users working in extreme, complex, and hazardous conditions such as high-energy laser facilities and particle accelerator tunnels while contributing to the advancement of MR technology in general.

The first contribution of the thesis introduces "**Taming Cyclops**", a novel application domain of using video see-through head-mounted displays (VST-HMD) as laser safety goggles at advanced optics laboratories such as the accelerator injector laser facilities. We developed a high-resolution stereoscopic VST-HMD and conducted an empirical study at the laser science and technology (LS-FA) laboratories at Deutsches Elektronen-Synchrotron DESY, where we evaluated the usability, perceived safety, advantages, and limitations of using stereoscopic VST-HMDs as laser safety goggles. Our system and user study results confirm that the complex and hazardous working conditions at high-energy laser laboratories could be significantly improved with MR technology.

The second contribution of the thesis introduces "**Mixed Reality Tunneling**", a novel sensor fusion technique to balance the trade-off between limited render performance and high visual quality of conventional stereoscopic VST-HMD design. By merging a color

video stream from an external stereoscopic camera with the low-resolution VST that is commonly integrated into some of today's standalone virtual reality (VR) headsets, we create a perceptual high-resolution and wide field of view VST-HMD prototype that could improve the usability of future VST-HMD at advanced optics laboratories. A controlled user experiment confirms that MR tunneling leads to higher overall usability, less perceived motion sickness, and a higher sense of presence.

The third contribution of the thesis presents "**Magic NeRF Lens**", an interactive framework that supports online facility inspection and maintenance planning in immersive VR to improve the efficiency of particle accelerator operation by reducing the duration of onsite maintenance windows. By using neural radiance fields (NeRF) models to replicate, store, and visualize the appearance of complex accelerator facilities, "Magic NeRF Lens" enables real-time volumetric rendering of complex accelerators at one-to-one real-world scale as well as dynamic 3D interaction with these neural 3D representations. To overcome the performance bottleneck of VR NeRF rendering, we present two novel interactive data fusion techniques that can merge a NeRF model with its' corresponding CAD model through the MR tunneling effect and contextual 3D NeRF drawing interaction. We demonstrate the effectiveness of our framework and methods through a technical benchmark, a visual search user study, and expert reviews at the European X-Ray Free Electron Laser(EuXFEL) linear accelerator (LINAC).

In the last contribution of the thesis, we introduce "**Reality Fusion**", an MR system that enables robust robot teleoperation at complex particle accelerator tunnels. Our system can localize, stream, project, and merge a typical onboard depth sensor with a photorealistic, high resolution, high framerate, and wide FoV rendering of the complex remote environment represented as 3D Gaussian splats (3DGS). Our framework enables robust egocentric and exocentric robot teleoperation in immersive VR, with the 3D Gaussian splats effectively extending spatial information of a depth sensor with limited FoV and balancing the trade-off between data streaming costs and data visual quality. We evaluated our framework through a user study, expert reviews, and real-world testing at the EuXFEL LINAC tunnel. We demonstrate that reality fusion leads to significantly better user performance, situation awareness, and user preferences.

Zusammenfassung

Großtechnische physikalische Anlagen wie Teilchenbeschleuniger und Hochleistungslaser sind wichtige Laborausrüstungen für industrielle Prozesse und wissenschaftliche Forschung. Allerdings stellen sie auch einige der anspruchsvollsten Herausforderungen in der Mensch-Computer-Interaktion (HCI) bei der Steuerung und Wartung der Anlagen dar. In den letzten Jahren hat der Aufstieg der Mixed-Reality-(MR-)Technologie begonnen, die alltägliche HCI von herkömmlichen 2D-Benutzeroberflächen (UIs) und Monitoren zu 3D-räumlichen UIs und immersiven, am Kopf montierten Displays (HMDs) zu transformieren. Begleitet von den rasanten Fortschritten in der Computer Vision, künstlichen Intelligenz (KI) und Robotik bieten kommerziell erhältliche HMDs bereits einen Einblick in unsere digitale Zukunft des "Metaverse", in der Nutzer nahtlos digitale Inhalte wahrnehmen und damit interagieren sowie ihre täglichen Aktivitäten wie Arbeiten, Lernen und Unterhaltung in einem vernetzten immersiven MR-Raum ausführen können. In dieser Dissertation präsentieren wir das Design, die Implementierung und die Bewertung mehrerer neuartiger MR-Schnittstellen zum Betreiben komplexer physikalischer Anlagen durch die Entwicklung neuartiger Sensorfusionen, immersiver neuronaler Wiedergabe und immersiver Robotik-Teleoperations-Techniken. Als erster Schritt in Richtung eines "Teilchenbeschleuniger-Metaversums"untersucht diese Dissertation, wie MR verwendet werden kann, um die HCI für Benutzer zu verbessern, die unter extremen, komplexen und gefährlichen Bedingungen arbeiten, wie beispielsweise in Hochenergie-Laseranlagen und Teilchenbeschleunigertunneln, und gleichzeitig zur allgemeinen Weiterentwicklung der MR-Technologie beizutragen.

Der erste Beitrag der Dissertation stellt "**Taming Cyclops**"vor, ein neuartiges Anwendungsgebiet der Nutzung von Video-See-Through-Head-Mounted Displays (VST-HMD) als Laserschutzbrillen in fortschrittlichen Optiklabors, wie zum Beispiel in den Beschleuniger Injektor Laseranlagen. Wir haben ein hochauflösendes stereoskopisches VST-HMD entwickelt und eine empirische Studie an den Laserwissenschafts und technologielabors (LS-FA) am Deutschen Elektronen-Synchrotron (DESY) durchgeführt, in der wir die Benutzerfreundlichkeit, die wahrgenommene Sicherheit, die Vorteile und die Einschränkungen der Verwendung stereoskopischer VST-HMDs als Laserschutzbrillen evaluiert

haben. Unsere System- und Benutzerstudienergebnisse bestätigen, dass die komplexen und gefährlichen Arbeitsbedingungen in Hochenergie-Laserlaboren mit MR-Technologie erheblich verbessert werden können.

Der zweite Beitrag der Dissertation stellt "**Mixed Reality Tunneling**"vor, eine neuartige Sensorfusionstechnik, die einen Ausgleich zwischen begrenzter Rendereffizienz und hoher visueller Qualität des konventionellen stereoskopischen VST-HMD-Designs herstellt. Durch die Verschmelzung eines Farbbildstroms von einer externen stereoskopischen Kamera mit der niedrigen Auflösung des VST, das in einigen der heutigen Standalone-Virtual-Reality-(VR-)Headsets integriert ist, schaffen wir einen wahrnehmungsbasierten, hochauflösenden und weitwinkligen VST-HMD-Prototypen, der die Benutzerfreundlichkeit zukünftiger VST-HMD in fortschrittlichen Optiklabors verbessern könnte. Ein kontrolliertes Benutzerexperiment bestätigt, dass MR-Tunneling zu einer höheren Gesamtnutzerfreundlichkeit, weniger wahrgenommener Bewegungskrankheit und einem höheren Präsenzgefühl führt.

Der dritte Beitrag der Dissertation präsentiert "**Magic NeRF Lens**", ein interaktives Framework, das die Online-Inspektion und Wartungsplanung in immersivem VR unterstützt, um die Effizienz des Teilchenbeschleunigerbetriebs durch Verkürzung der Onsite-Wartungsfenster zu verbessern. Unter Verwendung von Neural Radiance Fields (NeRF)-Modellen zur Replikation, Speicherung und Visualisierung des Erscheinungsbildes komplexer Beschleunigeranlagen ermöglicht "Magic NeRF Lens"die Echtzeit-Volumenwiedergabe komplexer Beschleuniger im Maßstab 1:1 der realen Welt sowie die dynamische 3D-Interaktion mit diesen neuronalen 3D-Darstellungen. Um den Leistungsengpass von VR-NeRF-Rendering zu überwinden, präsentieren wir zwei neuartige interaktive Datenfusionstechniken, die ein NeRF-Modell mit seinem entsprechenden CAD-Modell durch den MR-Tunneling-Effekt und die kontextuelle 3D-NeRF-Zeicheninteraktion verbinden können. Wir demonstrieren die Effektivität unseres Frameworks und unserer Methoden durch einen technischen Vergleichsmassstab, eine visuelle Suchstudie und Expertenbewertungen am Europäischen Röntgen-Freie-Elektronen-Laser (EuXFEL) Linearbeschleuniger.

Im letzten Beitrag der Dissertation stellen wir "**Reality Fusion**"vor, ein MR-System, das eine robuste Robotik-Teleoperation in komplexen Teilchenbeschleunigertunneln ermöglicht. Unser System kann einen typischen Tiefensensor an Bord lokalisieren, streamen, projizieren und mit einem fotorealistischen, hochauflösenden, hochfrequenten und weitwinkligen Rendering der komplexen entfernten Umgebung, dargestellt als 3D-Gaussiansplats (3DGS), verschmelzen. Unser Framework ermöglicht robuste egozentrische und exozentrische Robotik-Teleoperation in immersiver VR, wobei die 3D-Gaussiansplats die räumlichen Informatio-nen eines Tiefensensors mit begrenztem FoV effektiv erweitern und den Kompromiss zwischen Datenübertragungskosten und Datenvisualisierungsqualität ausgleichen. Wir haben unser Framework durch eine Benutzerstudie, Expertenbewertungen und reale Tests im EuXFEL-LINAC-Tunnel evaluiert. Wir zeigen, dass Reality Fusion zu einer signifikant besseren Benutzerleistung, einem besseren Situationsbewusstsein und besseren Benutzerpräferenzen führt.

Contents

List of Figures

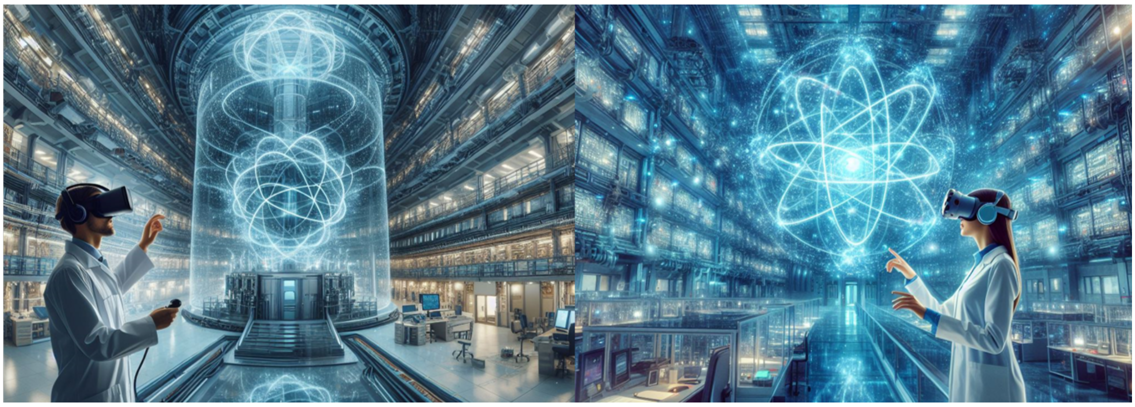
List of Tables

Part I

Introduction

1

Introduction



How can mixed reality technology be used to improve human-computer interaction for users working in extreme, complex, and hazardous conditions such as high-energy laser facilities and particle accelerator tunnels?

This teaser image was created using Microsoft Designer with the following prompt: "Images depicting scientists working with particle accelerators utilizing mixed reality technology."

1.1 Motivation

In recent years, an increasing number of complex physics facilities have been built around the world. These facilities, including particle accelerators, nuclear power plants, and lithography machines, are critical infrastructures for scientific research and industrial processes. As they are often complex systems built after decades of planning and implemented at enormous economic cost, scientists and engineers today are actively seeking methods to effectively maintain existing facilities to maximize their operational lifetimes and economically upgrade them to new operation standards. However, the operation of these complex physics facilities presents various demanding human-computer interaction (HCI) challenges in facility control and maintenance. For example, a particle accelerator could involve thousands of components that require frequent inspection and maintenance.

However, it must also operate continuously for thousands of hours per year. During this time, on-site human access is not possible due to radiation hazards, and any unexpected interruptions to operation will result in high energy and setup costs [42]. As a result, many facility inspection and maintenance planning tasks rely on an effective digital interface for operators to retrieve, visualize, understand, and interact with information about complex facilities. In fact, how to enable users to intuitively control and explore ever-increasingly complex digital information is one of the most important research topics in nowadays' HCI and data science research [15].

In the past, mixed reality (MR) technology has demonstrated enormous potential for improving HCI in various industrial processes [54]. For example, the possibility for users to spatially interact with digital content and thoroughly inspect a 3D model in an immersive telepresence environment can significantly improve users' productivity, engagement, and workflows in various industrial tasks [39]. MR technology also enables the development of spatial user interfaces (UIs) for robot teleoperation that allows remote control of robots with augmented visual cues or immersive visualization of remote environments, resulting in potentially significant improvement in operators' situation awareness, task performance, and trust in robots compared to teleoperation using traditional 2D UIs and desktop displays [248]. Furthermore, MR head-mounted displays (HMD) found important applications in specialized domains such as industrial welding halls [114] and laser laboratories [181], where HMDs are not only just an assistive visualization tool but also a robust eye protection measure to meet the health and safety standards of unusual working conditions. The immense application space of MR and the availability of high-quality commercial MR HMDs such as the Meta Quest Pro, Apple Vision Pro, and the Microsoft Hololens already started transforming modern HCI into a "*metaverse*" [174], where physical reality and virtual reality (VR) converges, and artificial intelligence (AI) and human intelligence comes together through an interconnected immersive space.

Despite its significance, developing a "*metaverse*" for operations of complex physics facilities requires scientific research to gain a deep understanding of human factors in empirical application settings while addressing the technical limitations of existing MR solutions. Although, many previous works already investigated applying MR technology to industrial processes through proof-of-concept prototypes and simulated tasks, research on adapting MR for real-world tasks at extreme, complex, and hazardous environments such as high-energy laser facilities and particle accelerator tunnels is sparse, leaving the human factors associated with using MR in performing complex real-world tasks unknown and unstudied. For example, although Quercioli [181, 180] has proposed the concept of using a monoscopic video see-through head-mounted display (VST-HMD) as laser safety goggles, previous work does not include any empirical user studies, leaving the usability, advantages, and limitations of using VST-HMDs at actual laser laboratories largely unknown. On the other hand, building a practical "*particle accelerator metaverse*" requires technical innovation in HMD design, computer graphics, and robotics for MR technology to meet the complex real-world application requirements. For example, while

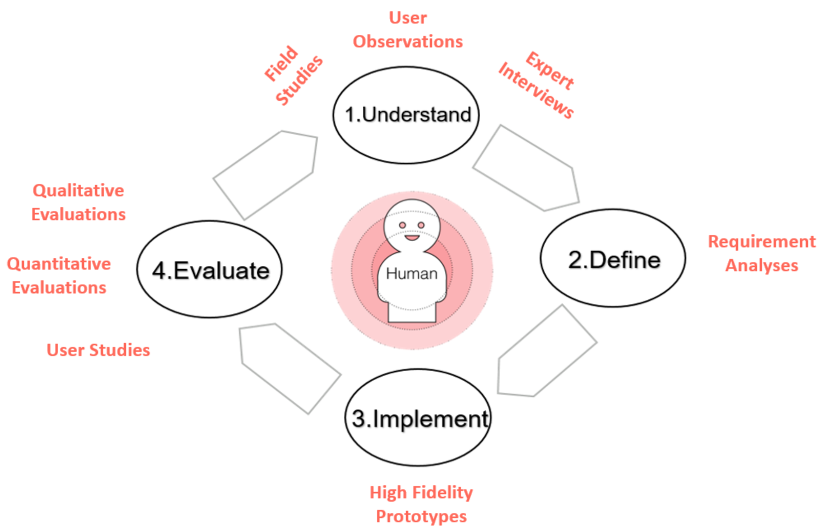


Figure 1.1: Illustration of the iterative human-centered design (HCD) [103] process with its four key steps ranging from 1. understand user behaviors, 2. define prototype requirements, 3. prototype implementation, and 4. system evaluation with user studies.

"*digital twins*" play a central role in industrial MR applications, conventional 3D reconstruction methods based on RGBD sensors [252], photogrammetry [183], or LiDAR scanners [18] can only reconstruct an industrial facility with limited completeness and accuracy, which often fall short in capturing the intricate geometries and detailed appearances that are necessary for an immersive and accurate virtual representation of complex real-world environments. Thus, bridging these existing technological gaps is also imperative for realizing the full potential of a " *particle accelerator metaverse*".

As a step forward in addressing the aforementioned research and technological gaps, this dissertation presents the design, implementation, and evaluation of several novel MR systems tailored for supporting the operations of complex physics facilities. While contemporary MR technology includes a diverse range of components, spanning from haptics and audio to visual displays [204], the technical contributions of this dissertation primarily focus on the advancement of immersive MR visual systems. Traditionally, visual systems are considered to be the main element of MR [155] as they appeal to the core stimulus for users during MR experiences and interactions, especially in immersive environments. Therefore, the contributions from this dissertation target addressing various technical and practical challenges for MR visual systems, including MR visual displays, MR graphical processing and rendering algorithms, as well as 3D UIs and interaction designs. Specifically, we investigate novel sensor fusion concepts to improve MR visual display designs at advanced optics laboratories, integrate the latest developments in neural rendering for interactive photorealistic virtual facility inspection, and explore multi-modal data fusion techniques to provide users with real-time volumetric visual feedback for immersive robot teleoperation tasks at particle accelerator tunnels. As

Figure 1.1 illustrates, all of our technical development adapts a human-centered design (HCD) approach [103] to thoroughly understand the application domains and analyze user requirements through field studies, user observations, and expert interviews at particle accelerator and high power laser facilities at the Deutsche Elektronen-Synchrotron DESY. In addition, all systems and technologies are evaluated through qualitative and quantitative user studies to validate their usability in real-world application scenarios, quantify users' perceptions, and identify the systems' limitations to motivate future work. In summary, this dissertation aims to bridge the gap between theoretical MR concepts and practical applications within the context of complex physics facilities while contributing to the technical development of practical immersive MR visual systems in general.

1.2 Research Questions

Particle accelerators are some of the most complex machines built by humans and their operations involve a large amount of topics. The overall research goal of this dissertation is to provide a deeper insight into how MR technology can be used to improve HCI for users working in extreme, complex, and hazardous conditions such as high-energy laser facilities and particle accelerator tunnels. However, due to the limited scope of a doctoral project, this dissertation does not intend to provide an exhaustive list of MR applications at particle accelerators. Rather, it identifies and focuses on three key use cases where MR demonstrates significant potential and has direct impacts on improving the efficiency and workflows of particle accelerator operations: **i)** using MR HMD as laser safety goggles, **ii)** immersive photorealistic visualizations for virtual facility inspection, and **iii)** teleoperation of robots at particle accelerator tunnels.

1.2.1 Head-Mounted Displays as Laser Safety Goggles

High-power lasers play a crucial role in the development and operation of different types of particle accelerators. For example, they are essential for generating intense pulses that drive the acceleration of charged particles in the development of plasma wakefield accelerators [76]. For larger particle accelerators, laser-based injection systems are used to create short, intense bursts of electrons [241]. As these laser sources can rapidly cause permanent damage to human eyes, research and development (R&D) work with those lasers typically requires wearing personal protective equipment (PPE), such as laser safety goggles as eye protectors. Currently, laser safety goggles are based on optical spectral filters, which block spectral bands where hazardous laser radiation is emitted. Such laser safety goggles can filter up to 99% of the visible spectrum, limiting the vision of researchers working in hazardous and complex laboratory environments. In theory, MR HMDs with video see-through (VST) capabilities could be used as eye protectors without reducing users' visibility of the environment since they can be constructed such that all laser and ambient light is blocked from the human eye [181]. However, there has been no comprehensive work that investigates the practicability and the human factors of such an eye protection method at an actual advanced optics laboratory. As a result, the usability, advantages, and limitations of VST-HMD for laser safety remain largely unknown. To bridge the gap between theoretical concepts with practical applications of using MR for laser eye safety, the first part of the thesis investigates the following research question:

- **RQ1**: *Can a stereoscopic VST-HMD be used as laser safety goggles at advanced optics laboratories? What are its current usability, advantages, and limitations?*

For advanced optics laboratories, VST-HMDs offer significant potential not only in providing eye protection from lasers but also in improving complex optical system operation workflows. However, many existing VST-HMD designs still exhibit several limitations that mitigate their practicality for certain tasks such as fine-motor assembly work like optical fiber splicing or those requiring locomotion within the environment [133]. For example, a high-resolution, wide field of view (FoV), colorful stereoscopic VST-HMD such as the Varjo-XR3 [230] might require tethered connection to an expensive and non-portable computer station. On the other hand, portable VST-HMDs such as those that come with standalone virtual reality (VR) headsets might only provide grayscale low-resolution VST. In the second part of the thesis, we explore how to enhance the VST functionality of a standalone HMD by developing novel sensor fusion techniques. The following research question guided our technical development:

- **RQ2**: *How to balance the trade-off between limited render performance and high visual quality of a stereoscopic untethered VST-HMD?*

1.2.2 Virtual Facility Inspection

One major advantage of MR is its ability to provide immersive and interactive 3D visualizations of complex systems. At particle accelerator facilities where human onsite visits to the facilities are limited due to safety hazards and operation constraints [42, 58], immersive visualization systems that accurately represent the complex facility conditions and fully immersive users in inaccessible remote environments are crucial for online facility inspection and maintenance planning [39]. However, conventional 3D representations such as meshes and point clouds often provide only limited realism in modeling complex geometries of real-world conditions. The recent innovation of implicit 3D volumetric representations using neural radiance fields (NeRF) can generate photorealistic 3D reconstructions with a relatively small amount of input images [154], and, thus offers a new approach to establishing visualization frameworks for virtual facility inspection. Although, there are vast amount of recent work on NeRF [158, 25], prior efforts have mainly concentrated on demonstrating proof-of-concept experiment results, rather than delivering a user-friendly toolkit for visualizing and interacting with real-world NeRF models, especially in immersive environments. The third part of the thesis investigated the following research question to apply the latest NeRF 3D modeling and rendering methods for virtual visual inspection of complex accelerator facilities in immersive MR:

- **RQ3**: *How to develop a user-friendly toolkit for visualizing and interacting with real-world NeRF models in immersive MR, particularly for virtual inspection of complex physics facilities?*
- **RQ4**: *How does such a neural rendering system perform for real-world immersive MR applications? What are users' perceptions of the overall usability, spatial presence, and their task performance during virtual facility inspection tasks?*

1.2.3 Visual System for Immersive Robot Teleoperation

Robot telepresence and teleoperation are important means for humans to transfer their intentions to a robot surrogate in a remote environment inaccessible to humans at particle accelerator tunnels to monitor and provide real-time measurement of the facilities [215, 58]. According to Adamides et.al [5], one central aspect of an efficient teleoperation system is its capability to provide operators with a high level of situation awareness of the robot's surroundings. Previous research demonstrated that teleoperation through an immersive HMD can significantly enhance users' situation awareness and task performance compared to traditional 2D display [72, 249, 117, 238, 217]. However, a typical onboard stereo vision sensor can only provide (a potentially high-quality) visual display with a limited field of view (FoV). Such limited 2D visual feedback typically impedes spatial awareness and a sense of presence in the remote environment [72]. Although several previous works investigated using a multi-camera setup [216] or streaming videos from an omnidirectional camera [249], these solutions introduce a significant increase in data streaming latency, leading to undesirable effect of VR motion sickness [249] and delay in robot control and intervention [216]. Therefore, the final part of the thesis investigates the following research question in search of a solution to balance the trade-off between the visual quality of volumetric data and its' processing and streaming latency.

- **RQ5**: *How to provide low-latency and high-quality volumetric visual feedback to the operators during robot teleoperation tasks at particle accelerator tunnels while ensuring that operators have a high level of situation awareness of the complex remote environments?*

1.3 Outline

The descriptions of motivation and research questions above conclude Part I of this thesis. The remainder is structured as follows.

Part II describes the fundamentals of this thesis and introduces the general background knowledge of particle accelerators, mixed reality, and photorealistic 3D scene representations. Specifically, Chapter 2 describes typical processes and challenges in particle accelerator operations such as facility inspection, maintenance, and robotic intervention. Chapter 3 introduces relevant concepts in MR, such as VST-HMD, the human visual system, cybersickness, immersion, and presence. Chapter 4 introduces the topic of 3D reconstruction and 3D representations and describes the theoretical background for NeRF and 3DGS.

Part III reports the details of our technical and scientific contributions in developing MR interfaces for the "*particle accelerator metaverse*". Chapter 5 introduces "*Taming Cyclops*", an empirical evaluation of the effectiveness of existing VST-HMD devices for laser safety at advanced optics in answering **RQ1**. Chapter 6 presents "*Mixed Reality Tunneling*", a novel sensor fusion for improving standalone VST-HMD designs in answering **RQ2**. Chapter 7 presents "*Magic NeRF Lens*", an interactive toolkit that supports immersive photorealistic visualizations of complex physics facilities leveraging the recent advancement of NeRF in answering **RQ3** and **RQ4**. Chapter 8 presents "*Reality Fusion*", a volumetric data fusion method that enables robust immersive real-time robot teleoperation at particle accelerator tunnels in answering **RQ5**. In Chapter 9, we present a collection of exploratory contributions that are related to the main research questions of the dissertation, including the initial development of a multi-modal robot teleoperation framework and the conceptual design of a NeRF system for cross-reality version control at R&D optical workbench.

Finally, Part IV discusses the findings of the introduced research questions and concludes the thesis with an outlook for future work.

1.4 Publications

The main contributions of this dissertation have been published in the following journals and conferences. These contributions are fully integrated into this dissertation.

1.4.1 Main Authorship

The following publications were mainly authored by myself, while my co-authors contributed to various aspects such as system implementation, writing specific sections of the paper, or providing supervision.

1.4.1.1 Journal Paper

- **Ke Li**, Susanne Schmidt, Tim Rolff, Reinhard Bacher, Wim Leemans, Frank Steinicke. (2024). Magic NeRF Lens: Interactive Fusion of Neural Radiance Fields for Virtual Facility Inspection. Frontiers in Virtual Reality. [DOI: `https://doi.org/10.3389/frvir.2024.1377245`]

1.4.1.2 Conference Papers

- **Ke Li**, Aradhana Choudhuri, Susanne Schmidt, Tino Lang, Reinhard Bacher, Ingmar Hartl, Wim Leemans, Frank Steinicke. (2022). Stereoscopic Video See-Through

Head-Mounted Displays for Laser Safety: An Empirical Evaluation at Advanced Optics Laboratories. 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) [DOI: https://doi.org/10.1109/ISMAR55827.2022.00025]

- **Ke Li**, Susanne Schmidt, Reinhard Bacher, Wim Leemans, Frank Steinicke. (2022). Mixed Reality Tunneling Effects for Stereoscopic Untethered Video-See-Through Head-Mounted Displays. 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) [DOI: https://doi.org/10.1109/ISMAR55827.2022.00018]

- **Ke Li**, Reinhard Bacher, Susanne Schmidt, Wim Leemans, Frank Steinicke. (2024). Reality Fusion: Robust Real-time Immersive Mobile Robot Teleoperation with Volumetric Visual Data Fusion. 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (Accepted/To Appear, Oral Presentation)

1.4.1.3 Workshop Papers

- **Ke Li**, Reinhard Bacher, Wim Leemans, Frank Steinicke. (2022). Towards Robust Exocentric Mobile Robot Tele-Operation in Mixed Reality. ACM Human-Robot Interaction (HRI) Workshop on Virtual, Augmented and Mixed Reality for Human-Robot Interaction (VAM-HRI). [https://openreview.net/pdf?id=HYIes841hJc]

1.4.1.4 Others

- **Ke Li**, Tim Rolff, Reinhard Bacher, Frank Steinicke. (2023). RealityGit: Cross Reality Version Control of R&D Optical Workbench. 2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). [DOI: https://doi.org/10.1109/ISMAR-Adjunct60411.2023.00178] [1] (Interactive Demo & Poster)

- **Ke Li**, Tim Rolff, Susanne Schmidt, Reinhard Bacher, Wim Leemans, Frank Steinicke. (2023). Interacting with Neural Radiance Fields in Immersive Virtual Reality. Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems. [DOI: https://doi.org/10.1145/3544549.3583920] (Interactive Demo & Poster)

- **Ke Li*, Tim Rolff***[2], Susanne Schmidt, Reinhard Bacher, Simnoe Frintrop, Wim Leemans, Frank Steinicke. (2023). Bringing Instant Neural Graphics Primitives to Immersive Virtual Reality. 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). [DOI: https://doi.org/10.1109/VRW58643.2023.00212] (Interactive Demo & Poster)

[1]This work received the honorable mentioned award at the student design competition at ISMAR 2023
[2]* indicate equal contributions

- **Ke Li**, Aradhana Choudhuri, Susanne Schmidt, Tino Lang, Reinhard Bacher, Ingmar Hartl, Wim Leemans, Frank Steinicke. (2023). Mixed Reality for Laser Safety at Advanced Optics Laboratories. 2023 International Laser Safety Conference (ILSC) [https://www.edit.fis.uni-hamburg.de/ws/files/30357120/ILSC_2023_Final_Draft.pdf] (Invited Talk)

- **Ke Li**, Aradhana Choudhuri, Susanne Schmidt, Reinhard Bacher, Ingmar Hartl, Wim Leemans, Frank Steinicke. (2022). Taming Cyclops: Mixed Reality Head-Mounted Displays as Laser Safety Goggles for Advanced Optics Laboratories. 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). [DOI: https://doi.org/10.1109/VRW55335.2022.00123] (Poster)

1.4.2 Co-Authorship

I was further involved in the following publications, with contributions such as system implementation and drafting sections of the paper. However, the following publication is not integrated as a part of this dissertation.

1.4.2.1 Conference Papers

- Tim Rolff, Susanne Schmidt, **Ke Li**, Frank Steinicke, Simone Frintrop. (2023). VRS-NeRF: Accelerating Neural Radiance Field Rendering with Variable Rate Shading. 2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). [DOI: https://doi.org/10.1109/ISMAR59233.2023.00039]

1.4.2.2 Others

- Tim Rolff, **Ke Li**, Julia Hertel, Susanne Schmidt, Simone Frintrop, Frank Steinicke. (2023). Interactive VRS-NeRF: Lightning-fast Neural Radiance Field Rendering for Virtual Reality. 2023 ACM Symposium on Spatial User Interaction (SUI). [DOI: https://doi.org/10.1145/3607822.3618020] (Demo)

Part II

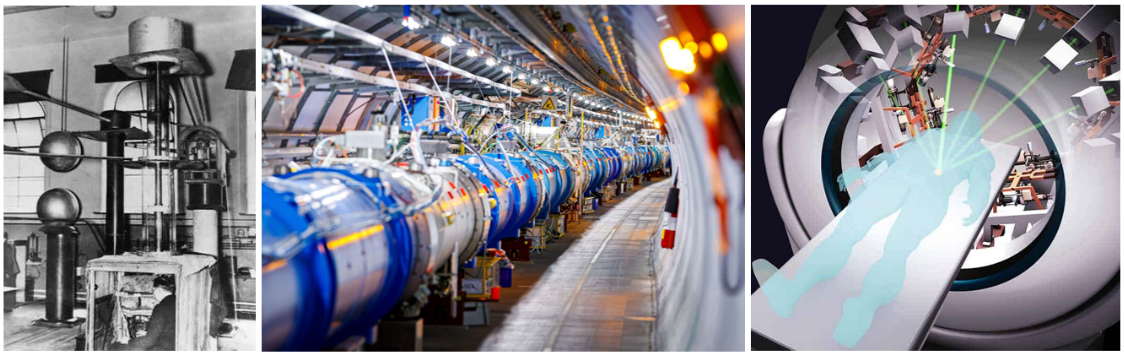
Fundamental

2

Particle Accelerator Operations

"The system stability, robustness, and maintenance should not limit machine availability or delay commissioning." —-The European X-Ray Free-Electron Laser Technical Design Report, 2007

In the last few decades, hundreds of particle accelerators have been built worldwide, ranging from small-scale accelerators used in laboratory experiments to large-scale facilities like the LHC at CERN. As some of the most ambitious instruments built by humans, these facilities play a central role in today's fundamental scientific research and industrial processes. While particle accelerators offer enormous opportunities for particle physics, material science, and photon science research, their operations involve controlling and maintaining some of the most complex machines built by humans. In this chapter, we provide a general background of particle accelerators (Section 2.1) and introduce the key challenges of their operations (Section 2.2). In addition, we provide insights into the current strategies of HCI utilized to manage these facilities and identify their constraints and shortcomings (Section 2.2). Finally, we discuss the current implementations of robotic intervention methods for remote facility inspection and maintenance tasks at particle accelerators and provide an overview of previous UI technology for robot teleoperation and controls at hazardous accelerator environments (Section 2.3).



(a). Early version of particle accelerator | (b). Modern accelerator facility for particle physics research | (c). Modern particle accelerator for cancer therapy

Figure 2.1: Evolution of particle accelerators: (a) one of the first controllable particle accelerators developed by Cockcroft and Walton [43], and (b) the LHC accelerator at CERN, demonstrating the scale of a part of the modern particle accelerator facility [43], and (c) the applications of modern particle accelerators for cancer treatment [227].

2.1 Background on Particle Accelerator

A particle accelerator is a physics instrument designed to accelerate charged particles such as electrons, protons, or irons to controllable physical states. In 1932, the first controllable particle accelerator was developed by Cockcroft and Walton, where the resulting charged beam can be manually manipulated in terms of various physics parameters, such as energy level, intensity, and energy spread for the first time [43]. As particle accelerator research advances, modern particle accelerators have evolved into facilities with many different scales, particle sources, and operation principles. As Figure 2.1 (a) shows, the initial design of the particle accelerator is relatively compact and utilizes relatively simple voltage multiplication techniques to accelerate charged particles to high energy. Since then, particle accelerators have not only grown in scale, allowing the generation of particle beams with much higher energy and intensity but also have grown in robustness and operability, making it possible to accurately focus particle beams to targets at nano-scale precision.

The possibility to robustly accelerate and control charged particles makes these facilities not only important for fundamental research but also for various medical and industrial processes. One of the most well-known functionalities is high-energy physics research. Figure 2.1 (b) shows a section of the particle accelerator tunnel of the Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN). The LHC, which is housed in a 27 km long ring tunnel, usually accelerates protons to almost the speed of light and creates highly energetic proton-to-proton collisions during which processes new particles can be generated and detected [43]. To date, the LHC is the largest accelerator facility in the world and is well-known for its groundbreaking discovery of the Higgs boson, which validates the theoretical foundation for the Standard Model and provides important insights into the history of development and the physical laws of the universe [2]. As particle accelerators can produce electromagnetic radiation with special physical properties, in recent decades, they have been widely used in photon science research as well. For example, electrons after accelerating through a circular path close to the speed of light emit *synchrotron radiation*, which is electromagnetic radiation ranging from infrared to X-ray with high tunability in terms of intensity, wavelength, and temporal resolution [202]. As Figure 2.1 (c) shows, the possibility to accurately control such radiation enables potential novel cancer treatment technology, such as synchrotron radiation therapy (SRT). Compared to standard external beam radiation therapy (EBRT) using a linac accelerator, in SRT, highly collimated beams of synchrotron-generated X-rays can deliver radiation to tumors with higher precision and target larger volume, in the meantime, maximizing the damage to cancer cells and minimizing harm to surrounding healthy tissues [98]. Highly coherent and intense synchrotron X-ray radiation also presents laser-like properties. Traditionally, developing photon-based X-ray lasers is particularly difficult due to the challenges of constructing laser cavities using optical mirrors, which are typically transparent to short-wavelength photons like X-rays. Therefore, coherent X-ray synchrotron radiation, often referred to as X-ray free-electron lasers (XFELs), presents unparalleled possibilities for capturing atomic-scale images of material and biological structures.

The enormous range of significant applications and the broad future potential of using particle accelerators as a general physics instrument has driven a tremendous amount of investment from the government and private sectors. Some of the most well-known publically funded scientific organizations such as CERN and DESY are established with thousands of scientists dedicated to designing, constructing, and operating a range of accelerator facilities. However, as the particle accelerator research progresses, particle accelerator facilities' development and operation conditions are becoming ever-increasingly complex. For large-scale accelerators such as the LHC and EuXFEL, there are always demands for obtaining charged particles at more and more extreme physical conditions. For example, for the EuXFEL, although the electrons are already accelerated close to the speed of light, obtaining more energetic beams can enable X-ray generation at shorter wavelength, higher intensity, and more coherency, making these XFEL possible to resolve structures at even finer scale and details. However, upgrading and building larger-scale accelerators also indicate higher energy consumption, construction costs, and operational hazards such as high voltage, strong magnetic field, and high radiation conditions. As more and more increasingly complex accelerators are built around the world, the key questions nowadays for accelerator research are no longer restricted to investigating new designs and methods in developing more sophisticated and controllable facilities, but also to how to efficiently and economically maintain existing ones while upgrading them to new operation standards. This dissertation contributes to the research and development of future facility management and control strategies to increase the overall facility availability and improve operation efficiency by developing MR UIs to support virtual facility inspection, immersive robot teleoperation, and laser eye safety.

2.2 Key Challenges in Accelerator Operations

In this section, we introduce the key challenges in acceleration operations. The content of this section is primarily based on both previously published technical reports and onsite field studies, expert interviews, and user observations at the EuXFEL facilities. Although the operation challenges of different accelerator facilities may vary based on the accelerator scale and physical principles, this section aims to introduce some of the common challenges using the operation of the EuXFEL as an example. Moreover, it is also important to note that accelerator operation can cover a broad range of topics, ranging from hardware design to automatic fault analysis and recovery. This section aims to provide background information on accelerator operation workflows in HCI areas relevant to this dissertation. For a comprehensive detailed introduction to accelerator design and operations, we refer the reader to the EuXFEL Technical Design Report [3].

2.2.1 Complex Facility Management and Control

The EuXFEL includes a large-scale, linear superconducting particle accelerator (LINAC) located in Schenefeld, Hamburg. The entire facility is 3.4 kilometers (km) long, starting on the DESY campus in the northwest part of the city of Hamburg, and ending in the neighboring Federal State of Schleswig-Holstein, south of the city of Schenefeld, where the experimental hall is located [242]. Many public images of the particle accelerator typically focus on the large-scale linear accelerator tunnels, through which the charged particles travel and gain energy. However, as Figure 2.2 shows, operations and control of particle accelerators consist of numerous interconnected subsystems. The EuXFEL accelerator starts with high-power laser laboratories where injector laser systems are developed and the injector laser beam is sent to the main accelerator injector to generate electron bunches and provide the initial energy of the electrons before the particles are sent to the main accelerator. In the main accelerator, the electron bunches travel through a 1.7 km long superconducting linear accelerator tunnel where electrons gain even higher energy. The superconducting cavities, which are obtained by cooling niobium (Nb) metal with liquid helium to below -271°, can transmit energy to the electron bunches with minimum loss from electrical resistance [242]. The high-energy particles are then compressed and reshaped by a bunch compressor and an X-ray laser beam is created through the self-amplified spontaneous emission (SASE) process within undulator setups. The resulting light pulses are transferred to different photon beamlines in the experiment hall, where the XFEL beam is further characterized for different experiment setups and requirements.

All of the systems, from accelerator injector lasers to photon beamlines, need to operate collectively to deliver high-quality particle beams with high controllability. To achieve this, a complex central control system is designed which handles more than 10 million control system parameters, generating $130 Terabyte/day (TB/d)$ uncompressed data and $30 TB/d$ temporarily archived data [240]. Multiple computer screens organize and display essential

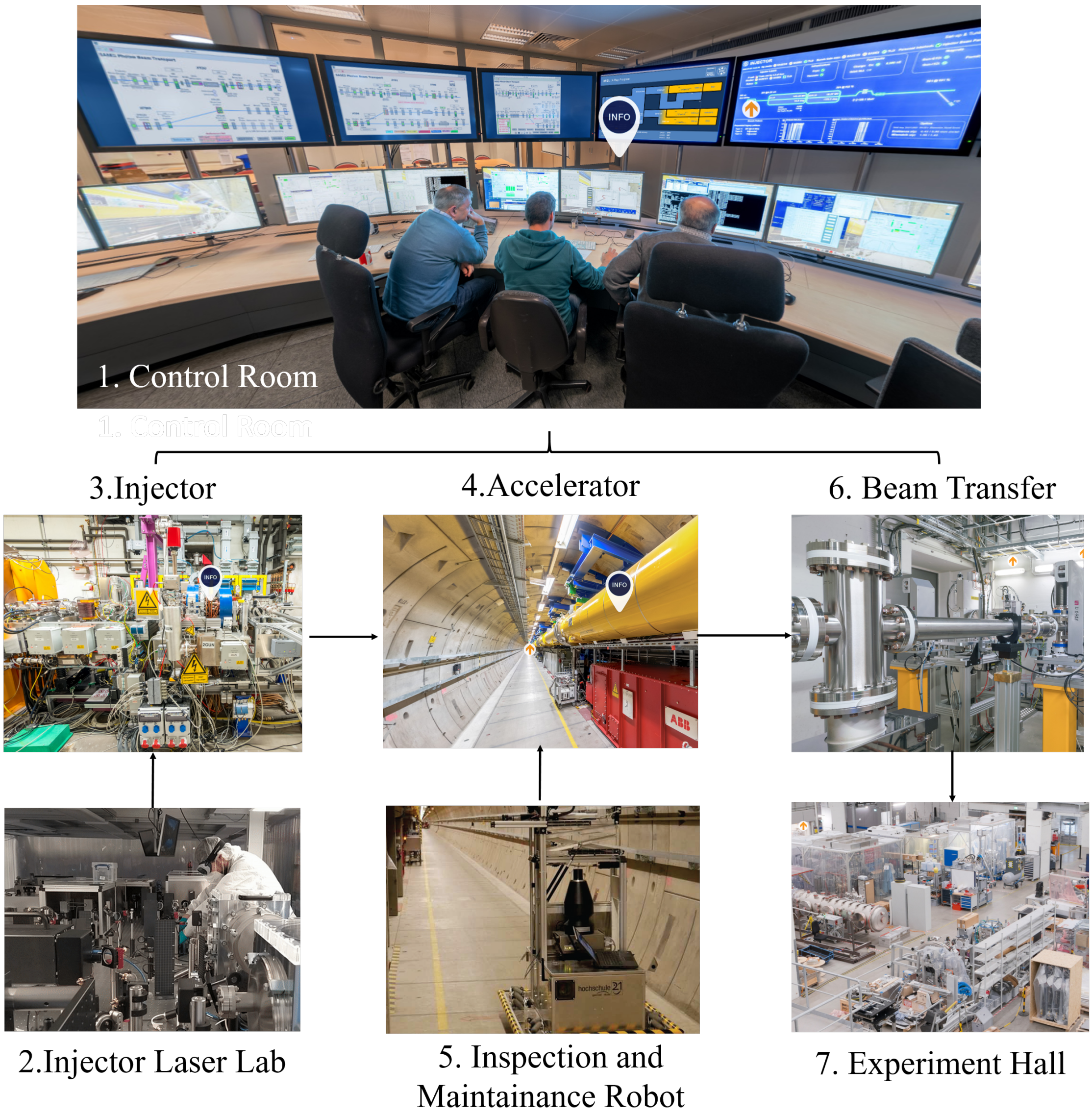


Figure 2.2: An overview of the operation of the EuXFEL linear accelerator and their subsystems, ranging from the injector to beam transfer systems (image 1, 3,4,5, and 7 are taken from the EuXFEL virtual tour website. Image 5 is taken from [58].)

information for providing an overview of the facility status, such as the states of the power supplies and the energy level of the produced photon beams. Operators have access to specialized control panels and programs that present and visualize real-time details of the different subsystems, such as the beam profile of the injector laser, the stability of the electron beam, or the overview of each RF station. Although the facility implements various automation for operation and setups [3], a small crew of operators still need to take shifts in monitoring the facility 24 hours per day, 7 days per week during its runtime to maintain optimum operating conditions and to respond to warnings from machine protection systems, detect potential system failures, and record needs for component replacements [3]. Moreover, every year, more than 2,700 scientists from 33 countries visit accelerator facilities at DESY to conduct experiments at different photon beamlines [26]. Control system experts also collaborate with beamline scientists and beamline users to adjust various parameters of the accelerators to fine-tune photon beams to desired physical states for different experiments.

Despite the complexity of accelerator facilities and their controls, existing UIs of control systems are developed largely based on the principles of function-orientated design, where the software system focuses on delivering proper functionalities rather than optimizing user experiences and needs [148]. Although such design approaches are widely adopted for developing industrial software [17], the lack of consideration of user behaviors and preferences often results in difficulty for users to understand and learn the software, thereby, decreasing user engagement, and increases the efforts and training needed for onboarding new users [103]. For example, retrieving specific parameters of a component could require users to search through multiple layers of drop-down menus and perform multiple checkbox selections, which could be overwhelming and lead to cognitive overload, resulting in lower task performance and reducing the overall system usability [4]. Conventional 2D data visualization designs also present challenges in supporting users with exploring and interacting with large-scale multi-dimensional data [34] such as an extended range of historical operation records, limiting users' performance in discovering anomalies and patterns with existing visualization tools. Moreover, traditional 2D UIs primarily depend on inputs from keyboards and mice, which are not the optimal interaction modality for users at the accelerator tunnels or laser laboratories, who need to access the control system while both of their hands are occupied performing maintenance and implementation tasks [17].

Changing how users interact, retrieve, and perceive the accelerator control system has the potential to significantly improve accelerator control and operation workflows, thereby, increasing the stability and availability of the complex facility [17]. In recent years, human-centered design approaches have set new standards for software engineering, so that the best practices for developing systems with high usability priorities understand requirements from the perspectives of human factors over specific features and functionalities. For example, the introduction of novel input modalities such as speech control can reduce hand-based inputs when interacting with complex control systems [17]. Adaptation of

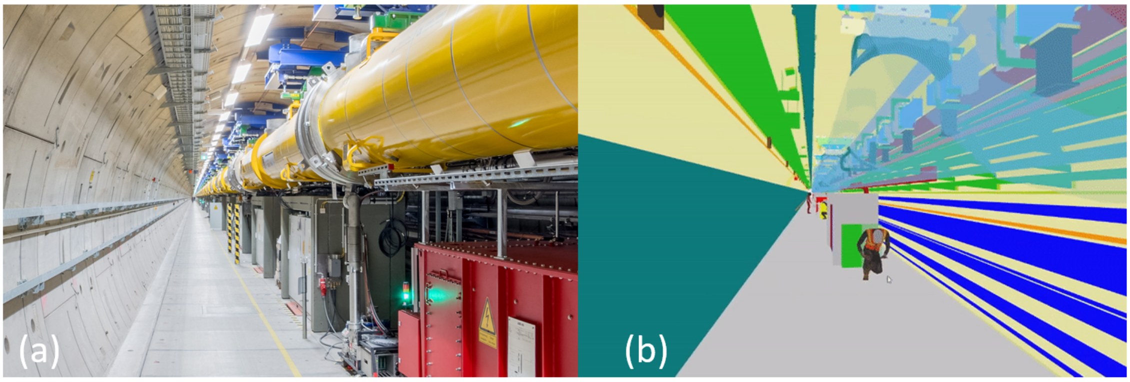


Figure 2.3: (a) Visual comparison between real-world accelerator implementation and (b) its CAD Design for virtual facility inspection.

3D UIs enables more intuitive HCI where the characteristics of the tasks are essentially spatial, leading users to use their natural skills and instincts, reducing the mental demands of acquiring new skills for interacting with UIs [31]. As accelerator facilities increase in complexity and scale, research investigating and adapting novel input modalities and 3D UIs is becoming more and more relevant for particle accelerator operation and control systems [17]. This dissertation contributes to developing novel accelerator control systems, transforming them from conventional 2D UIs built with conservative function-oriented design to novel 3D UIs built through HCD, enabling multi-modal user inputs, intuitive user interactions, and improved operation workflows of particle accelerators.

2.2.2 Online Inspection and Maintenance

As the accelerator facility is built with enormous costs and its operation consumes an enormous amount of energy, its operation principles aim for the highest possible efficiency, with minimum downtime to maintain the stability of the beam quality and maximize availability to continuously supply users with photon beams [64]. For example, the EuXFEL facility is scheduled to run for more than 5,000 hours per year continuously with only a few weeks of scheduled maintenance and a shutdown period over the year. However, during accelerator runtime, access to the accelerator tunnels is prohibited due to various hazardous conditions, such as high radiation, high voltage, and high magnetic field exposure, which could be lethal or harmful to humans. Moreover, unexpected human intervention requests to the accelerator tunnels during scheduled runtime can be costly, as unplanned switch-off of the facility leads to loss of scheduled beam time, disturbance of the beam quality due to system interruptions, and extra energy consumption and setup costs required to calibrate and fine-tune the facility to a desired state. Even during the facility shutdown period where radiation exposure is reduced to below the safety threshold, onsite human intervention involves numerous safety procedures, making the accelerator tunnels largely inaccessible to personnel without special training.

On the other hand, accelerator facilities are made of hundreds of thousands of specialized hardware systems that need regular inspection and maintenance. Many inspection and maintenance tasks, such as monitoring the water cooling system, calibration of beam monitor equipment, and maintaining the delicate RF cavities need to be completed within the limited annual scheduled shutdown period. However, as suggested in the EuXFEL Design Report, "*the system stability, robustness, and maintenance should not limit machine availability or delay commissioning*" [3]. Therefore, maintaining the availability of the facility requires not only robust algorithms for monitoring thousands of control system parameters, redundancy for fault tolerance and recovery mechanism, and extremely reliable hardware [3], but also effective online facility inspection tools to help operators plan, manage, and execute various maintenance activities when onsite access is not possible.

Visual systems are of central importance for online inspection and maintenance activities for operators to understand the complex remote environment [17]. Typical workflows for facility inspection often involve viewing computer-aided design (CAD) models to examine the facility structures and components in detail. As Figure 2.3 shows, although CAD models are a common type of "digital twin" representation in industrial applications [111, 140], they often only represent the initial design of the facility or equipment and are not always complete or up-to-date to match the structures and appearances of the real-world conditions. Nonetheless, CAD models are an indispensable form of 3D visualization as they can include semantic information about each component and can reveal their internal structures, offering important insights into the general functionality, organization, and assembly of complex accelerator systems. In contrast to CAD models, virtual facility inspection also adapts other types of spatial data, such as 3D meshes generated from photogrammetry [192] or Light Detection and Ranging (LiDAR) point cloud scans [184] for creating updated visualizations and 3D reconstruction. However, these methods offer limited capability to duplicate the visual appearance of complex physical realities as well. Intricate geometries, such as cable layout, and reflective surfaces such as equipment with metallic materials that are prevalent in particle accelerators could lead to measurement errors [184], limiting the accuracy of the 3D reconstruction, and reducing the usability of these 3D representations for various online visual inspection tasks. A part of this dissertation contributes to the advancement of visualization tools for virtual facility inspection and maintenance by leveraging novel 3D representations through neural networks and volumetric rendering [154], aimed at enhancing operators' spatial understanding and helping them to form an accurate mental model of the remote environment.

2.3 Robot Interventions

To maximize the availability of accelerators, robot intervention presents enormous potential [42, 58, 159]. Robot intervention can be more desirable and flexible than human onsite intervention as robots can perform tasks in radioactive and hazardous conditions without

(a). Various Robot Platforms at CERN

(b). Conventional Robot Teleoperation Interface

(c). Immersive Mixed Reality Robot Teleoperation at CERN

Figure 2.4: Overview of robot platform and teleoperation framework implemented at CERN, with (a) showing an overview of existing robot systems [42], (b) showing the conventional robot teleoperation interfaces [217], and (c) showing recent implementation of a MR robot teleoperation interfaces at CERN [217]

shutting down the accelerator facility [42]. Moreover, robot intervention presents the advantages of possible autonomous operation, which can greatly reduce human efforts in robot programming and motion planning. In the pursuit of more reliable and efficient operation of large-scale particle accelerators, various robot platforms have been developed to support operators in performing remote maintenance activities. For example, as Figure 2.2 shows, The MARWIN mobile robot was designed for inspecting and maintaining the EuXFEL accelerator tunnel. It assists operators in detecting radiation hot spots from malfunctioning cavities when they are pushed to their limits. Additionally, it measures residual radiation levels after the facility is shut down to ensure the safety of human entry [58]. As Figure 2.4 (a) shows, the LHC at CERN, robots with various functionalities are deployed, ranging from lightweight mobile robots for surveillance and patrol to large-scale platforms with manipulators and end-effectors for more complex tasks such as handling radioactive components [215].

While intelligent robot systems that can make optimal and correct decisions autonomously without manual control are the ultimate goal of robot development at particle accelerators, there is currently no proof of operation safety of existing autonomous robot technology [11]. In hazardous environments, unexpected robot behaviors need to be fully avoided, as a single failure or mistake by an autonomous system could lead to more damage to the facility than the intended maintenance activities. As a result, existing robot operation and control at particle accelerators largely adapt fully manual or semi-autonomous control strategies [42]. However, accurately communicating human intentions to a maintenance robot which often has complex configuration and redundancy hardware setups is challenging [42]. For example, in particle accelerator tunnels, the wireless signal can be restricted or even interrupted due to radiation interference, making transmission of the robot's feedback delayed and unstable. For the control of the MARWIN robot at the EuXFEL, the reception of the robot's visual feedback can take 2-3 seconds after sending the operator's command, making instantaneous and smooth teleoperation technically challenging. Moreover, the accelerator environment also presents various non-robot-friendly conditions. For instance, the robot's hardware and software could be interfered with by the so-called "bit-flip" phenomenon when traveling through a certain area of the tunnel with high radiation. The high radiation and high energy of the charged particles can trigger random interference of the robot's software and memory, leading to possible unknowns and instability of the robot's system [62]. In addition, the design and construction of existing large-scale accelerator facilities rarely take into consideration of possible robot interventions. As a result, robots might need to visit narrow gaps between components to complete their missions or have the capability of moving through uneven and irregular floor layouts while performing autonomous parking and accurate self-localization. These constraints further lower the tolerance of mistakes and increase the uncertainty of their controls, leading to more demanding accuracy requirements for robot controls and motion planning.

While advanced hardware and software solutions are increasingly integrated for handling errors and avoiding a single point of failure (SPOF) [58], under the unique operation conditions and technical limitations of autonomous systems, robot teleoperation at particle accelerators also require the development of intuitive and efficient human-robot interaction (HRI) interfaces to enable timely and accurate remote control of the robots. The robot research and development team at CERN has pioneered the experimentation of various novel HRI interfaces [142]. As shown in Figure 2.4 (b), traditional HRI interfaces are largely based on conventional desktop display with keyboard and mouse as inputs. While visual information is crucial for operators to understand the robot's surroundings, using only 2D displays and 2D visual feedback such as videos and images can limit operators' spatial understanding and situation awareness for motion control tasks that are inherently spatial. To overcome this problem, MR robot teleoperation interfaces were developed where 3D data such as real-time point cloud from an RGBD camera is streamed and displayed in complementary to 2D videos, as shown in Figure 2.4 (c). MR robot teleoperation can provide the operators with a sense of "presence" through which they can flexibly observe the robot's behaviors from any perspective as if they are "co-located" with the robot [10]. Multi-modal user input such as hand tracking, body tracking, and eye gaze tracking in immersive MR also enables more natural and intuitive HRI. For example, motion planning and programming of 6 degrees of freedom (DoF) robotic manipulator can be greatly simplified by using the user's tracked arm gestures and its inverse kinematic mapping as inputs [206]. Central to the efficiency and usability of an MR HRI interface is the quality of the immersive visual system which operators rely on to understand the spatial layout of the robot's environment. In Chapter 8, we demonstrate the possibility of teleoperating robots at particle accelerator tunnels by developing a photorealistic visual system that enables both low latency spatial visual feedback and enhanced situation awareness of the remote environment.

3

Mixed Reality Visual Displays

*"With appropriate programming, such a **display** could literally be the Wonderland into which Alice walked." – Ivan Sutherland, the Ultimate Display, 1965.*

The teaser image was created using Microsoft Designer with the following prompt: "An image illustrates mixed reality visual displays."

An MR interface comprises various technical components, from visual displays and tracking devices to sensor fusion techniques, graphical processing pipelines, and 3D UIs. Among them, MR visual displays are one of the most critical parts of an MR interface, as they are the primary component that mediate users' vision sensing and perception during an immersive experience. How to develop an MR visual display capable of delivering flexible, high resolution, high framerate, and wide field of view immersive MR experiences with a light-weight form factor has been an open research question since the development of the first MR visual system by Ivan Sutherland in 1986 [214]. This chapter introduces the fundamental concepts and theories related to developing MR visual displays. We first define MR (Section 3.1), then establish its connection to the "metaverse" (Section 3.2). Afterward, we provide an overview of MR HMD (Section 3.3), which serves as the foundation for MR applications designed to improve accelerator operation workflows in environments such as advanced optics laboratories. Finally, we discuss relevant human factors associated with immersive MR visual displays, highlighting their influences on

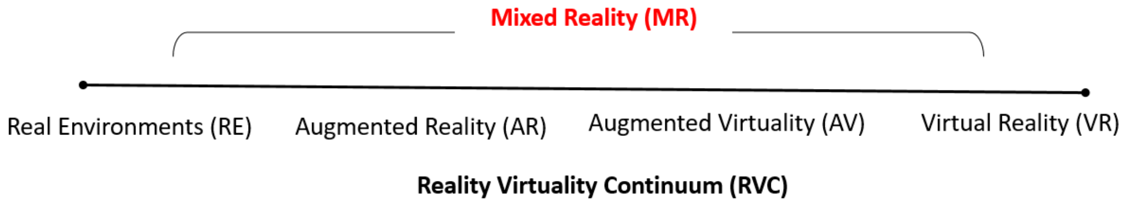


Figure 3.1: Reality-Virtuality Continuum Diagram, adapted from [155].

user experiences while introducing psychological metrics used in this dissertation for quantifying users' perception and performance (Section 3.4). [1]

3.1 What is Mixed Reality?

In today's diverse landscape of extended reality (XR) technology, many terminologies seek to define the wide variety of immersive experiences enabled by the innovation of different digital visual displays and computer graphics techniques. The standard definition of XR technology typically originates from Milgram et al.'s reality-virtuality continuum (RVC) [155], which categorizes various immersive experiences on a spectrum ranging from the real environment to fully virtual ones. Based on the degree of immersion of users in a fully computer-generated synthetic digital space, there are typically three types of distinguishable visual displays:

- **Real Environment (RE)**: As one end of the RVC spectrum, a RE represents the physical space of the reality in which we live without any digital augmentation.
- **Augmented Reality (AR)**: In an AR experience, users perceive the real environment augmented with virtual information, for example, through the overlay of a UI.
- **Augmented Virtuality (AV)**: In an AV experience, users perceive the virtual environment (VE) that is augmented with real-world information, for example, through the integration of real-world physical objects, people, or data streams into the VE.
- **Virtual Reality (VR)**: In a VR experience, users' perceptions are only exposed to synthetically generated content, typically, using a VR HMD through which users are fully immersed in an artificial 3D world and are visually excluded from their real environments.

However, these discrete categorizations of XR technology often fail to encapsulate the complex interconnection between the real world and the digital world, Milgram instead

[1] Some part of Section 3.3.4 and Section 3.4.1 was already published in the related work section of the following paper: **Ke Li**, Susanne Schmidt, Reinhard Bacher, Wim Leemans, Frank Steinicke. (2022). Mixed Reality Tunneling Effects for Stereoscopic Untethered Video-See-Through Head-Mounted Displays. 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) [DOI: https://doi.org/10.1109/ISMAR55827.2022.00018] [132]

defines **Mixed Reality**, which refers to a spectrum of experiences that blend elements of the physical and digital worlds to unify various concepts in XR. The wide acceptance of referring to immersive technology generally as MR suggests the important insight that real-world applications of XR are unlikely to isolate users into a truly virtual environment. As pointed out by Skarbez et al., the conventional VR experiences we have nowadays can also be considered as a subset of MR technology, as virtual contents might situated based on real-world experiences [198]. Such a categorization is supported by the fact that VR immersive experiences, even though displayed through an HMD that fully covers users' eyes, could still make users' perceptions largely grounded in the physical realities they know, rather than fully replaced by the virtual ones that are indistinguishable from the physical realities.

Another emerging concept closely related to MR is **Cross Reality** (CR). Different from MR, CR emphasizes the concurrent use of, or transitions between, different visual displays or experiences in the RVC, including VR, AV, AR, and physical reality [1]. The distinctive feature of CR is user collaboration despite using different manifests of the experiences [80]. For example, a CR system allows interactions between VR users and non-VR users, connects users working on 2D displays to those who are immersed in an MR environment, and coordinates transitions between different immersive experiences in the VC [80]. While this dissertation primarily focuses on developing MR interfaces, CR systems are crucial for the real-world adoption of MR interfaces. Different application domains will use various forms of MR experiences, and it is unlikely that traditional HCI methods, such as 2D displays, mice, and keyboards, will be entirely replaced by emerging immersive technologies. In Chapter 9, we will explore the potential applications of CR at particle accelerators, and present the initial design of a CR system for the "*particle accelerator metaverse*".

In summary, it is safe to conclude that MR is a space where physical realities and digital realities converge and will continue to take the central spot when human-machine interaction transits from conventional 2D UIs to natural, intuitive, and immersive 3D realms. If not otherwise specified, this dissertation adapts the MR definition derived from the RVC as initially proposed by Milgram et al [155], where the term "Mixed Reality" refers to a collection of immersive technology, including immersive experiences facilitated by traditional VR, AR, and AV displays.

3.2 What is a "*Metaverse*" ?

The term "*metaverse*" was first introduced by Neal Stephenson in his science fiction novel "Snow Crash" in 1992, which describes a virtual environment that is a successor of the internet with users representing themselves and interacting with each other as virtual avatars [174]. Nowadays, the advancement of XR technology makes various concepts of the "metaverse" no longer just science fiction. Recently, major technology companies such as Meta, Microsoft, and Apple started to release a series of HMDs and applications

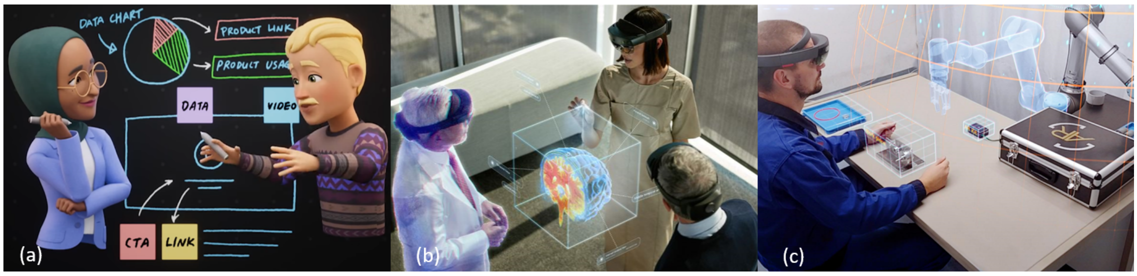


Figure 3.2: MR and the Metaverse, with (a) showing a screenshot of the Meta Horizon Workroom, an application allowing collaborations and interactions of remote users in a virtual environment [91], and (b) showing a screenshot of the conceptual design of the Microsoft Mesh application for co-located metaverse experiences in immersive MR [100], and (c) showing humans collaborating with other forms of agents (e.g. a robot) in the metaverse [77].

that have already taken the first step in enabling users to play, learn, and work in an interconnected immersive 3D space [14, 151]. For example, as shown in Figure 3.2 (a), the Meta Horizon Workroom application enables teams that can not be physically together to have meetings in a virtual environment as if they are in the same physical room, with each user's body gesture and facial expressions tracked and transferred to their virtual avatar [91]. While the initial concepts primarily focus on building the "metaverse" in a pure VR space, real-world applications of the "metaverse" are at its core facilitated by a variety of MR experiences. Figure 3.2 (b) shows a screenshot of the conceptual design of the Microsoft Mesh application, where virtual users can co-locate with physical users by "teleporting" into their 3D space in the physical world through MR [100]. Co-located users can also work together in the shared MR metaverse space for collaborative sense-making through interactions with persistent virtual content [213].

Although public attention on the "metaverse" typically focuses on shared experiences between humans, it is important to notice that the metaverse ultimately refers to the unifying connectivity between a variety of agents, regardless of whether they are humans [100], intelligent virtual agents [118], or physical robots [216]. As shown in Figure 3.2 (c), an important part of the "metaverse" research also covers the interaction between humans and virtual intelligent agents [118], as well as between humans and physical robots mediated through immersive displays [216]. This dissertation explores several technical aspects of the "metaverse" for real-world industrial applications and presents the first steps toward a "particle accelerator metaverse" where complex physics facilities can be inspected, operated, and maintained in immersive MR. Moreover, we investigate a key aspect of creating shared experiences within the "metaverse" by developing systems for visualizing and interacting with immersive, photorealistic digital twin environments, laying the foundation for the "particle accelerator metaverse" to be a dynamic space for a wide range of simulation, training, and maintenance activities.

3.3 Head Mounted Displays

At the center of MR technology for the "metaverse" are MR visual systems that enable users to perceive and interact with immersive 3D content. In this section, we provide an overview of MR HMD (Section 3.3.1), highlighting their enhanced capabilities in visually blending the digital and the physical world [155]. Moreover, we introduce OST-HMD (Section 3.3.2) and VST-HMD (Section 3.3.3), reviewing the technical challenges of their development and presenting their advantages and limitations. In addition, we discuss the fovea-peripheral characteristics of the human vision system, highlighting the foveated rendering technique in enhancing rendering performance for immersive HMD (Section 3.3.4). Finally, we provide an overview of vergence-accommodation conflicts (VAC) and discuss the potential drawbacks of the current single-focal design in HMDs (Section 3.3.5).

3.3.1 MR HMD Overview

Nowadays, mainstream MR interfaces are mostly based on HMD, where display screens are integrated into eyeglasses and goggles [214]. Compared to other types of immersive technology, such as projection-based, Cave automatic virtual environments (CAVEs)-style systems with displays positioned at a fixed location in the form of room-scale screens [53], display screens of HMDs are often placed in front of users' eyes, offering more flexible and versatile setups where users can integrate individual preferences and customizations in their immersive experiences. Therefore, HMD finds important applications in many domains such as medical visualization, medical surgery [51], human-robot interactions [216], and industrial facility inspection and maintenance [54].

As HMDs are gaining increasing popularity, they might become the next generation of personal computing devices [66]. The enormous potential of HMDs is exemplified by the race of major technology companies such as Meta, Microsoft, and Apple, in defining and developing HMDs with increasingly lightweight form factors, higher resolution, and more intuitive and natural HCI modalities such as voice input, eye tracking, gesture controls, and face tracking [151]. While the early development of commercially available HMDs provides support for only one of the distinct immersive experiences in AR, VR, or VR, in recent years, the release of major HMDs increasingly emphasized enabling MR displays, which further blurs the boundaries of conventional AR, VR, and AV experiences. For example, previous iterations of HMDs such as Oculus Quest 2 [150] and the HTC Vive Pro headsets primarily support distinct VR experiences, where users perceive no direct visual stimulus from their real environment. Nowadays, conventional VR headsets have evolved into MR HMDs through the integration of VST capability, where users can not only interact with a fully synthetic environment, but also perceive their real environments through stereoscopic video captured by integrated visual sensors on the headset. Such hybrid HMD designs create flexibility to transit users between immersive experiences in the VC [49] and broaden the application domains of MR HMDs [133]. Such a trend enables versatile MR experiences that smoothly transition users between their real-world environment and a virtual one, exemplified in the development of AR HMDs as well. For example, the recent release of the Magic Leap 2 headset [144], implements a dynamic dimming technology, which allows users to dynamically control the optical opacity of the display to enable brighter digital content in a broad range of ambient light conditions, thereby, offers the possibility to adjust the level of immersion in the virtual environment [145]. Therefore, as predicted by Milgram, modern immersive HMDs are gradually evolving into general "MR HMDs" [155], emphasizing their enhanced capabilities and flexibility in visually blending the digital and the physical world.

3.3.2 Optical See-through Head Mounted Displays

One of the most prominent features of an MR HMD is the capability of users to perceive the real-world environment while interacting with virtual content. Depending on how the real-world environment is captured and displayed to users, MR HMDs are generally categorized as video see-through head-mounted displays (VST-HMDs) [105] and optical see-through head-mounted displays (OST-HMDs) [137].

OST-HMD enables the optical superposition of digital information onto users' direct view of the physical world [94]. Dewen et al. distinguished three common optical solutions for OST-HMDs based on the underlying imaging principles [60]. Early concepts of OST-HMDs were primarily based on macro optics design, where virtual contents are reflected from (Light Emitting Diode) LED displays to users' eyes based on basic reflection and refraction principles determined by Snell's laws [60]. While macro optics systems are easy to manufacture, they often result in bulky setups, degrading user experiences, and reducing the headset's flexibility. Given the need for more compact, lightweight, and ergonomic designs, the exploration of micro optics and nano optics solutions remains a focal point in the ongoing research for OST-HMD development. Current mainstream OST-HMD such as Microsoft Hololens [153] and Magic Leap [144] headsets already adopt various micro optics techniques using diffractive waveguides, which can guide the display light from small micro-led displays to users' eyes through a thin glass plate [126]. With the advancement of nano-fabrication technology in achieving unprecedented resolution and precision, the size and weight of an OST device can be further reduced by utilizing nano-optics, which can compress multiple components of a complex optical system into a thin layer of metalenses, potentially making future OST devices possible at the scale of an eye-glasses or even contact-lenses [109].

One main factor contributing to OST's importance for MR HMD is the possibility of having a non-obstructed view of the real-world environment with the optical clarity of natural human vision. However, designing and developing an OST-HMD that can be adapted for everyday uses not only faces common computational constraints of HMD for running complex spatial computing algorithms on a resource-limited mobile device, it also needs to overcome more demanding optical system design challenges to deliver a

Figure 3.3: Commercial video see-through head-mounted displays and optical see-through head-mounted displays.

cost-efficient solution [60]. For example, even though diffractive waveguides can deliver holographic images with a more compact form factor, higher clarity, and fewer aberrations compared to conventional macro optics solutions, their' fabrication introduces high manufacturing costs [185]. Moreover, imaging through planer diffractive waveguides is often restricted to a narrower FoV compared to conventional magnification-based HMDs [60]. As the development of OST-HMDs is still in the preliminary stage, there is a recent trend of commercial MR HMD development focusing on using VST for immersive MR experiences.

3.3.3 Video See-through Head Mounted Displays

As shown in Figure 3.3, VST technology enables users to perceive the real-world environment through images captured by digital cameras. Early VST technology creates immersive MR experiences by mounting an external stereoscopic camera on a VR display such as the custom-made VST solution using the ZED Mini [247] camera and the Oculus Rift headset [165], with both the camera feed and the VR display processed by an additional computation workstation. Users' head poses are recorded by the integrated tracking system of the VR headset and the stereoscopic video from the external camera is processed, streamed, and updated in the VR displays accordingly [247]. In recent years, designs and configurations of VST-HMDs have largely evolved with the advancement of computer vision, computer graphics, and sensor fusion technology. Nowadays, high-performance, standalone VST systems such as the Apple Vision Pro can achieve VST with $3,660 \times 3,142$ pixel (px) per eye display resolution, around 100° horizontal field of view (HFOV), and $12ms$ motion to photon latency have recently become commercially available. Compared to OST-HMD [137], current VST-HMDs offer MR experiences with a wider FoV and, therefore, can support MR experiences with a higher level of immersion. Moreover, the optical system of a VST-HMD often employs the design of a VR headset, which primarily consists of a simple magnification setup that projects images from a small liquid crystal device (LCD) to human eyes [246], resulting in lower costs in optical fabrication and manufacturing than OST-HMDs. While OST and VST technology both have their strengths and limitations, the system development of this dissertation is primarily based on HMDs with VST capabilities to create MR experiences with a higher level of immersion and flexibility in switching immersive experiences within the VC. It is also important to point out that HMDs with high-quality VST capabilities such as Apple Vision Pro [14] and Meta Quest Pro [151] have only begun to be commercially available after late 2023 and early 2024. As previous iterations of mainstream HMDs such as the Meta Quest 2 headset [150] only support low-resolution and grayscale VST, in Chapter 5 and Chapter 6, we developed custom-made HMDs to create high-resolution VST solutions that were not commercially available at the time of the experiments and user studies.

A primary challenge in VST-HMD development is to create a visual display system that enables perception which resembles and even extends human's natural vision using digital displays. As human vision is highly sensitive to visual variations and artifacts, especially when exposed to a fully immersive environment [56], user experiences with a VST system largely depend on the quality of the integrated visual display. For example, while traditional 2D displays in a television or monitor typically suffice with a refresh rate of 48–60 Hz, a recent study demonstrates that human eyes can perceive visual flicker artifacts at rates over 500 Hz when a display includes high-frequency spatial edges [56]. Even though nowadays' advanced VST-HMDs have a refresh rate of up to 120 Hz per eye, the discrepancy between human's natural vision and the capability of the digital display may introduce subtle unwanted effects that negatively influence users' experiences [207]. Another important factor for VST-HMD design is to address the inherent latency of the video feed, which can cause uncomfortable registration errors between the captured real-world videos and the rendering of these videos in the virtual environment [20]. Human eyes can detect motion to photon latency that is higher than $17ms$ [6], and an immersive VR experience typically needs to have a latency at least lower than $50ms$ to feel responsive [207]. Higher latency can increase the risk of users experiencing cybersickness, an adverse effect of MR usage that leads to symptoms such as nausea, fatigue, and blurred vision [146]. As a result, a significant amount of computational resources of VST-HMD is dedicated to real-time video processing and rendering to minimize the delay between the movement of the user's head and the update of the digital display reflecting the user's action. In Chapter 5, we will present a VST-HMD design that could potentially increase the perceived framerate and resolution of a VST-HMD while reducing the perceived latency through sensor fusion.

3.3.4 Foveated Rendering

Foveated rendering is an important technique in MR HMD for reducing the rendering workload by matching the rendering quality with the characteristics of the human visual system [176, 175]. As shown in Figure 3.4 (a), human vision begins with detecting light by two types of photoreceptor cells within the retina [96]. In the foveal region of the human eye, which represents only 5.2° of the human visual field [212], the cone photoreceptor

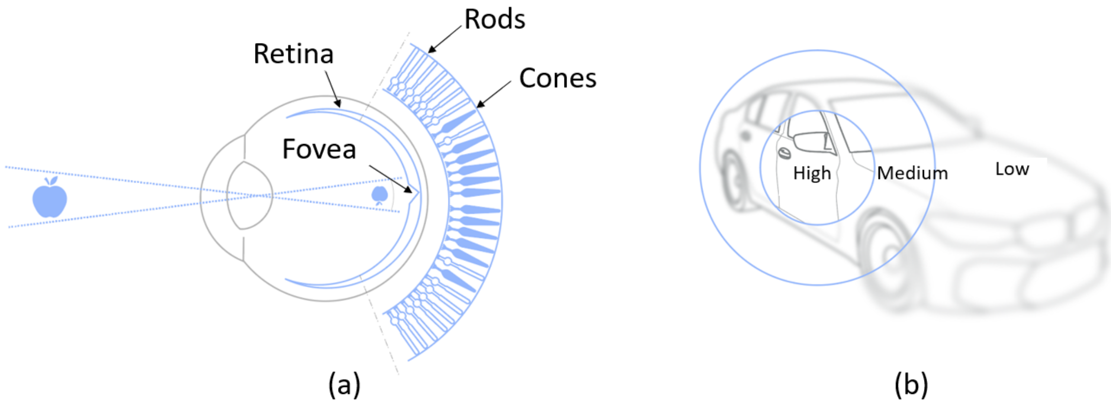


Figure 3.4: Subfigure (a) illustrates the uneven distribution of the rod and cone photoreceptor in the retina of the human eyes, and subfigure (b) illustrates the resulting foveated rendering technique that matches the rendering quality with the characteristics of the human visual system (image adapted from [65]).

dominates, giving this region high visual acuity and helping to form a sharp, high-resolution perception of the environment. On the contrary, in the peripheral area of the human eye, which spans from 60° to nearly 180° in horizontal diameter in the human visual field [212], the rod receptor dominates, making this region particularly sensitive to motion detection and low light vision [96] while forming only a blurry low-resolution perception due to the lack of rod receptors. As shown in 3.4 (b), such uneven distribution of rods and cones in the retina leads to distinct regions in human vision that differ in their sensitivity to displayed visual quality and can be exploited in immersive graphics systems to save computational resources by rendering progressively less detail outside the eye fixation region [176]. By integrating modern eye-tracking technology into an immersive HMD, an MR visual system can precisely determine the user's eye fixation point and adjust the rendering quality in real-time. This process can reduce the number of rendered pixel quads by up to 70 % without noticeable visual degradation for the user [12, 176]. In Chapter 6, we will explore similar aspects of the foveal and peripheral vision for sensor fusion in HMD. There, high-resolution, colorful VST is presented in the foveal region for enhanced visual acuity, while low-resolution, grayscale, but high-framerate VST is shown in the peripheral region for enhancing motion detection acuity. Additionally, in Chapter 7, we will discuss an immersive visualization design that leverages the foveal-peripheral characteristics of human eyes, which enables more efficient real-time immersive NeRF rendering through data fusion.

3.3.5 Vergence Accommodation Conflicts

While technical metrics such as display resolution, FoV, frame rate, and refresh rate are important parameters that influence the visual quality of an HMD, the visual comforts, user's sense of presence, and user's spatial awareness in an immersive environment are largely determined by the visual display's capability in providing natural depth cues

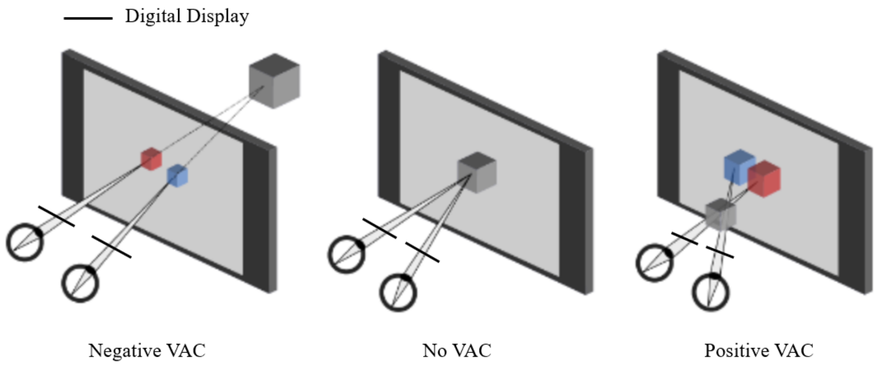


Figure 3.5: Vergence accommodation conflicts in single-focus binocular HMD, with the left image shows a negative VAC effect, where the actual depth of the virtual object is behind the focal plane of the HMD, and the middle image shows the actual depth of the virtual object aligns with the focal distance of the HMD, and the right image shows a positive VAC effect, where the depth of the virtual object is in front of the actual focal plane of the HMD (image adapted from [74]).

[31]. Modern HMDs can simulate a strong sense of 3D perception through stereopsis [143], where binocular vision is enabled through stereoscopic displays that have the disparity of interpupillary distance (IPD) of the user's eyes. However, existing HMDs face a fundamental challenge in simulating natural depth cues due to their single-focal design [90]. Such a design could lead to the adverse effect of VAC, where the virtual image in the HMD can only be focused at a fixed distance. In contrast, the depth of the virtual objects varies with the content, resulting in conflicting information within the convergence-accommodation feedback loops [116, 90]. As shown in Figure 3.5, natural binocular vision relies on the capability of human eyes to move in the opposite direction, known as the vergence process, so that images for each eye can fall on the same areas in the retina for proper stereoscopic sensory fusion in the visual cortex [92]. In the human vision system, vergence occurs simultaneously with accommodation, where the focal length of human eyes can be adjusted accordingly through rapid deformation of the crystalline lens so that objects at different distances to the eyes can remain sharp on the retina [92]. When using single-focal HMDs, accommodation of the visual display could not occur accordingly with the vergence process, leading to a vergence accommodation mismatch which results in distortion of the perceived depth. In fact, VAC is one of the main sources of visual fatigue and discomfort in immersive HMD usage [90].

Although VAC presents in both OST and VST systems, its impact on a user's depth perception and visual comfort is particularly prominent in VST-HMDs, as users' visual system is fully exposed to the digital displays [90]. While there is much active research on varifocal HMDs, their development still needs to overcome numerous technical challenges in eye tracking systems and optical system design for actual implementation in commercial-grade HMDs [90]. In Chapter 5, we discuss how the lack of "true depth perception" can affect a user's task performance in empirical settings where close inspection of fine components is required.

3.4 Human Factors

As pointed out by Bowman et al., *"the term ' **human factors**' refers to the capabilities, characteristics, and limitations of the human user, and includes considerations related to the body (acting), the senses (perceiving), and the brain (thinking)"* [31]. For the development of complex MR interfaces for particle accelerator operation, understanding users' perception and how well users can perform a task is crucial for the safe, efficient, and comfortable use of MR HMDs [31]. In this section, we introduce the background on a few human factors that will be used for the evaluation of the MR interfaces developed in this dissertation, including an overview of cybersickness (Section 3.4.1), immersion and presence (Section 3.4.2), and task load (Section 3.4.3).

3.4.1 Cybersickness

As mentioned in Section 3.3.3, cybersickness is one of the major negative effects of MR, which lead to various symptoms of physical discomfort such as nausea, headache, blurry vision, and eye strain [146]. While these symptoms appear similar to simulator sickness (SS) from a flight simulator or motion sickness that occurs during travel in cars, cybersickness has a distinct profile where symptoms related to disorientation predominately occur [205]. As the risk of experiencing cybersickness increases with the exposure duration to immersive experiences, developing solutions to mitigate and resolve cybersickness is essential for the wide adaptation and extended use of MR HMDs [41]. While many theories were proposed to investigate and understand its causes, the most widely accepted theory for cybersickness is the sensory conflicts theory, which arises from the physiological explanation of motion sickness [167]. The sensory conflict theory states that the causes of cybersickness are due to the mismatch between visual and vestibular senses [160]. In MR HMDs, particularly in immersive VST-HMDs and VR HMDs, the mismatch is mostly induced by the end-to-end display latency, which results in discrepancies between the actual position of the user's head and the displayed position of the virtual image. However, other factors such as limited FoV, low perceived framerate, low refresh rate, and the VAC effect could also induce different levels of cybersickness [41].

Although there is currently no single solution that could resolve all the cybersickness effects at once, several approaches have been developed to compensate for the end-to-end latency in VST-HMDs through a "time warp" effect. The "asynchronous time warp" effect proposed by Van Waveren et al. transforms the video position based on the very latest head tracking information and can significantly reduce the motion-to-photon delay [236]. Freiwald et al. implemented a "camera time warp" effect, a similar reprojection technique for VST-HMD latency compensation [73], and demonstrated that it is possible to lower the frame rate settings of the stereo camera in favor of a higher resolution, without causing a significant increase in discomfort [73]. The time warp technique is widely adopted into existing VR and MR headsets such as the Oculus Quest 2 [163] and the ZED Mini Unity library [208].

One common approach for cybersickness reduction in VR gaming is applying a FoV restrictor (also known as "VR Tunneling"), which limits the optical flow in the peripheral region of the human eyes [125], thereby, visual motion perception is reduced and limits sensory conflicts for stationary users. Abundant studies have demonstrated that although FoV restriction has a positive effect on cybersickness, it can also degrade the user's sense of presence [196] (Section 3.4.2). It has also been demonstrated that a foveal FoV restrictor has no significant effect on cybersickness reduction compared to a fixed FoV restrictor [7]. However, to our knowledge, no previous work has studied whether such FoV restrictors would have similar effects when applied to stereoscopic VST-HMDs. Especially, it is unknown if a similar presence degradation will persist even if the low-latency and low-resolution VST are merged in the restricted peripheral region. In Chapter 6, we will investigate such a tunneling effect in MR for the first time and improve upon the existing VR tunneling effect with VST sensor fusion in the peripheral region.

As cybersickness can have a significant influence on the system usability of an MR application [41], alter the user's sense of presence [237], and affect the user's cognitive and motor functions [115], measuring cybersickness is essential for holistic evaluation of an MR interface. The most widely adopted method for quantifying cybersickness is through the self-report simulator sickness questionnaire (SSQ) developed by Kennedy et al [106]. The SSQ includes 16 items that allow users to record different perceived cybersickness symptoms on a four-point scale (*none*, *slight*, *moderate*, or *severe*). The 16 items are grouped into four categories (*oculomotor*, *discomfort*, *disorientation*, and *nausea*), allowing for an evaluation of cybersickness based on individual subscale as well as the overall total score [106]. A measurement of the user's perceived cybersickness via SSQ is recorded and reported for every MR system development in this dissertation.

3.4.2 Immersion and Presence

Presence and *immersion* are the two most common concepts that are used to describe the unique user experiences of perceiving and interacting with 3D visual displays and UIs [200]. While they are closely related and are often discussed together within similar contexts, they are non-interchangeable concepts that characterize similar aspects of virtual experiences from different perspectives [199]. *Presence* describes the user's subjective psychological response to their engagement and involvement in a computer-generated environment, whereas *immersion* describes the objective level of sensory fidelity a system could provide [200]. For example, the level of visual immersion of an MR HMD refers to how close the system's visual output is to real-world visual stimuli [200]. Technical metrics such as display resolution, FoV, frame rate, and the visual realism of the environment determine the level of immersion of the HMD [30]. A stereoscopic, high resolution, high framerate, wide FoV MR HMD is objectively recognized to provide a higher level of immersion than conventional 2D displays, whereas if a user perceives a higher *sense* of presence can vary based on individual experiences and psychological responses. Nonetheless, various previous studies and user experiments indicate that an increase in the level of immersion tends to increase users' sense of presence as well [30].

Building an MR interface with a higher level of immersion where users can experience a higher sense of presence has various benefits for the operation of complex particle accelerator facilities. For example, a higher immersion enables better spatial understandings and spatial orientation [30], potentially making a virtual inspection of the spatial layout of the complex facility faster and less cognitively demanding. Building a virtual environment with high visual realism can lead to a stronger sense of presence [97], increasing user engagement in simulation, education, and training activities in immersive MR [102]. In Chapter 7 and Chapter 8, we will represent MR interfaces built with photorealistic 3D scene representations, aiming to enhance users' spatial presence in the MR environment, and thereby, improving users' spatial awareness, task engagement, and the system's usability. A typical metric in evaluating users' sense of presence in a virtual environment (VE) is through the Igroup presence questionnaire (IPQ) developed by Schubert et.al [195]. The original IPQ consists of 14 items, with 4 items measuring *spatial presence* (the sense of being physically present in the VE [195]), 4 items measuring user's *involvement* (the attention devoted to the VE and the involvement experienced [195]), 4 items measuring users' *experienced Realism*(the subjective experience of realism in the VE), and 2 additional items assess the general "sense of being there" [195].

3.4.3 Task Load

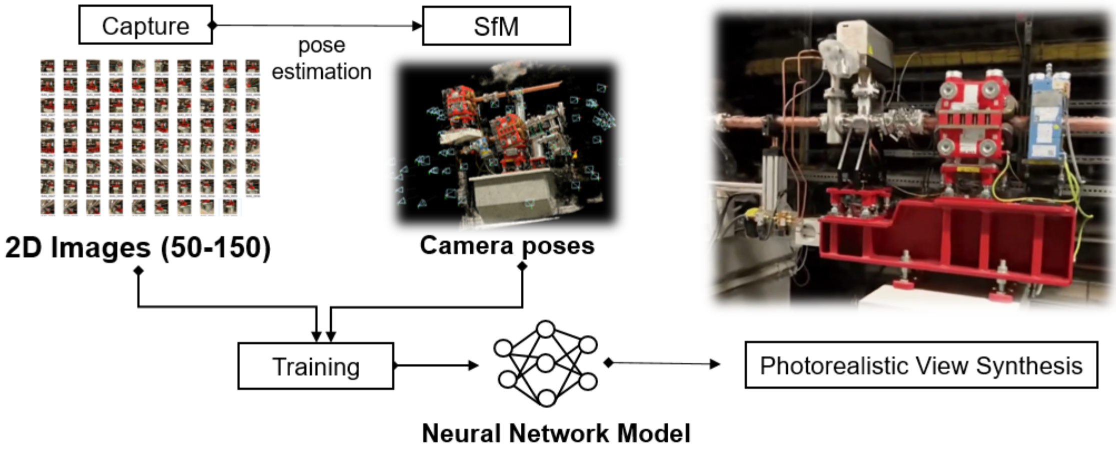
Task load, or workload, is a term that "*represents the cost of accomplishing mission requirements for the human operator*" [87]. Moving HCI from conventional 2D displays to 3D MR HMDs not only introduces changes to users' visual perception, but also could alter users' perception of their task load and performance. For example, while interaction with 3D UIs in MR can be more natural and intuitive, it could lead to a higher amount of physical demand on users, as some tasks in immersive MR might require users to interact with UIs through gesture control with extensive arm movements [108] or navigate inside a VE through natural walking to avoid cybersickness [119]. Adaptation of immersive MR benefits from having the full 3D space for placement of visualization and UI widgets, thereby, opening up wide design space for developing novel user experiences and applications [245]. However, as immersive 3D UIs occupy more open space, they might also introduce additional mental demands on users for spatial memory layout recall at the same time [245]. As a result, analyzing and understanding different aspects of the user's task load is important for the holistic evaluation of an MR interface.

According to the National Aeronautics and Space Administration (NASA) task load index (TLX), users' perceived task load can be represented from six different dimensions:

mental demand, *physical demand*, *temporal demand*, *performance*, *efforts*, and *frustration* [88]. *Mental demand* reflects the amount of mental resources required to perform a set of concurrent tasks [127]. *Physical demand* reflects the amount of physical activity that was required (e.g., pushing, pulling, turning controlling, activating, etc.) [88]. *Temporal demand* is related to the amount of time pressure involved in completing the task, [88]. *Performance* evaluates users' perception of their own performance [88]. *Efforts* indicates users' perceived level of cognitive and physical efforts they need to invest in completing the task [88]. *Frustration* reflects the level of fatigue and mental stress that is introduced by completing the task [88]. Effective design of an immersive MR interface could reduce the user's perceived task load and enable fast, accurate, and reliable task completion [87]. In this dissertation, the NASA-TLX is used as an important evaluation metric to compare different interaction designs and identify their trade-offs and drawbacks.

4

3D Reconstruction & Representations



*"The challenge to computer graphics is to make that virtual world **look real**, sound real, move and respond to interaction in real-time, and even feel real." — Ivan Sutherland, The Ultimate Display, 1965*

As introduced in Section 2.2.2, potential MR applications at particle accelerators such as virtual facility inspection, robot teleoperation, and remote collaboration depend on accurate 3D "digital twins" of the complex physics facilities to provide a high level-of spatial presence and situational awareness. In this chapter, we provide background information about conventional 3D reconstruction methods (Section 4.1). In particular, we give an overview of the recent paradigm shift of adapting implicit neural graphics for reconstructing and representing 3D real-world scenes (Section 4.2). Then, we introduce the fundamentals of NeRF [154] (Section 4.3), 3DGS [107] (Section 4.5), and related work on immersive neural graphics for MR applications [129] (Section 4.4). [1]

[1] Part of Section 4.3 and Section 4.4 were already published in the main and supplementary material of the following paper: **Ke Li**, Susanne Schmidt, Tim Rolff, Reinhard Bacher, Wim Leemans, Frank Steinicke. (2024). Magic NeRF Lens: Interactive Fusion of Neural Radiance Fields for Virtual Facility Inspection. Frontiers in Virtual Reality [DOI: https://doi.org/10.3389/frvir.2024.1377245]

Part of Section 4.5 was included in the following paper: **Ke Li**, Reinhard Bacher, Susanne Schmidt, Wim Leemans, Frank Steinicke. (2024). Reality Fusion: Robust Real-time Immersive Mobile Robot Teleoperation with Volumetric Visual Data Fusion. 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (accepted, to appear)

4.1 Conventional 3D Reconstruction Methods

3D reconstruction is a classic problem in computer vision (CV) to retrieve 3D geometries or appearances of real-world scenes and represent them in digital forms. Based on how 3D information is obtained, 3D reconstruction methods can be classified as active or passive [38]. Active 3D reconstruction such as laser imaging, detection, and ranging (LiDAR) [184], the time-of-flight (ToF) [161], and structured-light [128] methods project lasers, infrared light, or fringe patterns onto the targets to measure their surface geometries through information encoded in the returned optical signal. Active scanners can obtain 3D measurements at sub-millimeter accuracy and are widely adapted in autonomous driving, industrial reverse engineering, and precision engineering [38]. However, they are typically expensive to manufacture and could produce noisy and erroneous results for metallic surfaces with specular reflection due to dependency on the light scattering process from the measurement target [184].

Conversely, passive 3D reconstruction involves techniques that estimate 3D geometries directly from a collection of unstructured 2D images without the need for an additional light source [38]. For example, the structure from motion (SfM) algorithms can extract and match feature points from 2D images taken from different viewpoints, estimate their camera poses, and calculate sparse point clouds [192]. Afterwards, the multi-view stereo (MVS) algorithm can perform dense reconstruction by establishing pixel-wise correspondence between images, resulting in a dense point cloud of the reconstructed scene [193]. Compared to active 3D scanning, such an image-based reconstruction method has comparably low hardware costs and is particularly important for commercial 3D reconstruction applications such as personal avatar generation and customized virtual environment creation for the "metaverse" [83]. However, passive 3D reconstruction methods such as SfM typically struggle with textureless surfaces and materials where distinguishable feature points are sparse and often produce lower measurement accuracy than light-based active 3D scanners. Therefore, conventionally, obtaining an accurate "digital twin" of a real-world scene in the wild requires a hybrid method that involves both active and passive reconstruction, resulting in time-consuming post-processing for data fusion, outliner removals, and texture mapping [183].

4.2 Explicit and Implicit 3D Representations

The recent success of machine learning (ML) and deep learning (DL) in solving ill-posed inverse problems using prior knowledge presents enormous potential for image-based 3D reconstruction problems [83]. One important consideration for learning-based 3D reconstruction is the choice of 3D representations that a neural network should predict [149], as different 3D representations could vary the network architect, training speed, training accuracy, and the degree of supervision required [83].

In CG, a 3D surface can be expressed explicitly or implicitly. Common explicit 3D representations are points, polygonal meshes, and voxels. Voxels are the 3D analogy of 2D pixels, which are commonly used in medical computed tomography (CT) scans and computer animations, to explicitly represent an entire volume. Although such data structure enables direct extension of 2D convolutional neural network (CNN) architecture to 3D, storing 3D scenes at high resolution with voxels is often spatially prohibited as the memory requirement scales cubically with resolution [138]. While point clouds are the common initial output of many conventional 3D reconstruction methods such as LiDAR scan and photogrammetry, they typically need extensive post-processing and refinement to be converted into meshes that are more efficient for rasterization, fast rendering, and interaction in the conventional graphic pipelines [183]. However, polygonal meshes struggle to accurately represent fine details like fur, hair, and clouds [83]. Furthermore, learning-based 3D reconstruction approaches that directly predict these explicit representations also fail to make accurate predictions of fine details due to the spectral bias of a neural network [219] and the lack of high-quality balanced dataset for supervised learning [138].

On the other hand, implicit 3D representations such as signed distance function (SDF) and occupancy grid can represent 3D geometries through mathematical functions [149]. Recent research shows that neural networks can acquire implicit 3D representations through unsupervised learning, which enables learning of 3D structures without training on extensive prior data [138]. For example, NeRF is an implicit neural 3D representation that can synthesize novel views with high resolution and accuracy by training a small multi-layer perceptron (MLP) with a relatively small amount of 2D images [154]. Compared to explicit voxel representations, the neural network model compresses the 3D volume into a small scene function with the color and occupancy of each voxel to be inferred at runtime, enabling storing a large number of scene details with low memory consumption. Furthermore, unlike explicit representations such as point clouds and meshes that can be incomplete due to their discrete nature, 3D representations with NeRF are continuous. This continuity allows for rendering images at high resolution with greater visual completeness, although, often with rendering speed as a trade-off.

It is important to notice that there are novel 3D representations such as 3DGS [107] which fall in between the conventional categorization of implicit and explicit 3D representations. While the initialization of 3DGS is based on explicit sparse point clouds from SfM, each point is represented as a 3D Gaussian with its covariance matrics and spherical harmonic coefficients implicitly defining the appearance of the rendered 3D scene and geometries [107]. Both 3DGS and NeRF are currently state-of-the-art photorealistic 3D scene representations and are suitable for the rapid generation of "digital twins" of complex real-world environments such as accelerator facilities which are difficult to be accurately represented with conventional explicit 3D representations. In Chapter 7 and Chapter 8, we explore the adaptation of these novel scene representations in immersive MR and evaluate their suitability for real-world MR applications such as virtual facility inspection and immersive robot teleoperation.

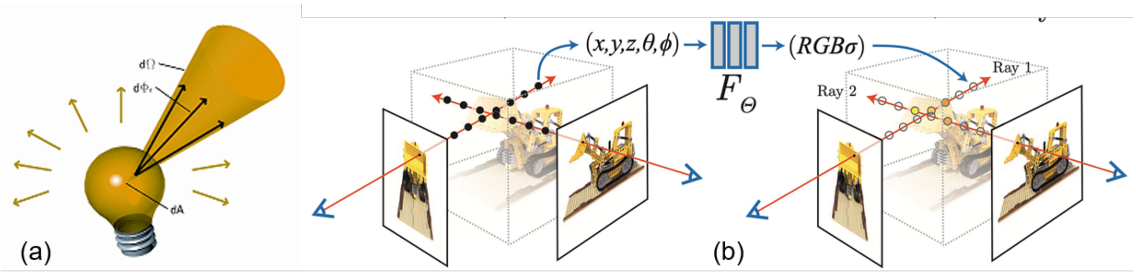


Figure 4.1: Using NeRF as 3D scene representations, with (a) showing radiance as a physical property of surface in emitting light (image adapted from [78]), and (b) showing how 2D images and their camera poses are parameterized using volumetric ray-casting with the position and viewing direction of the 3D points used as input parameters and the color and density of the corresponding voxels as output parameters (image adapted from [154]).

4.3 Neural Radiance Fields

In this section, we provide an overview of the theory of 3D rendering with NeRF, based on which we analyze the performance requirement for VR NeRF experiences.

As shown in Figure 4.1, "radiance" is a physical property that defines the amount of electromagnetic radiation passing through a unit surface [147]. In CG, "radiance" is a parameter widely used to characterize physic-based photorealistic rendering to simulate realistic interaction of lights with different material surfaces [178]. As illustrated by Figure 4.1, the property of radiance is view-dependent. The input parameters of NeRF are the position and viewing direction of the sample point calculated from the given camera poses. The output parameter is the radiance value characterized as the color and volumetric density of a voxel. As suggested by Mildenhall et al. [154], NeRF represents the 3D world with a volumetric scene function F_Θ with learnable parameters Θ by mapping the 3D position $\boldsymbol{p}(x, y, z)$ and the viewing direction $\boldsymbol{d}(\theta, \phi)$ to a color emission vector $\boldsymbol{c}(r, g, b)$ and a volume density float σ:

$$F_\Theta : (x, y, z, \theta, \phi) \mapsto (r, g, b, \sigma). \tag{4.1}$$

To generate a 2D projection from this 3D scene representation, NeRF rendering relies on conventional volumetric ray-casting rendering. A camera ray $\boldsymbol{r}_{(x,y)}$ from $\boldsymbol{o}(x, y)$ with viewing direction $\boldsymbol{d}(\theta, \phi)$ is cast into 3D space for each pixel (x, y): $\boldsymbol{r}_{(x,y)}(t) = o(x, y) + t \cdot d$, where t is within the interval of the user-defined near boundary t_n and far boundary t_f of the ray: $t \in [t_n, t_f]$.

The color value $\boldsymbol{C}(\boldsymbol{r})$ per pixel can be calculated by sampling and accumulating the transmittance along the ray $T(t)$, the volume density function $\sigma(t)$, and the color value $\boldsymbol{c}(t)$:

$$C(\boldsymbol{r}) = \int_{t_n}^{t_f} T(t) \cdot \sigma(\boldsymbol{r}(t)) \cdot c(\boldsymbol{r}(t), d)\, dt. \tag{4.2}$$

As a result, the NeRF rendering performance $P_{(h,w)}$ is bounded by the target render resolution with R_h height, R_w width, the number per ray $N_{\boldsymbol{r}(t)}$, and the average speed of each network query (usually measured in floating point operations per second (FLOPs)) $\overline{F}$ [135]:

$$P_{(h,w)} = R_h \times R_w \times N_{\boldsymbol{r}(t)} \times \overline{F} \tag{4.3}$$

Although the multi-resolution hash encoding data structure of instant-ngp [158] heavily parallelizes the network query process and achieves rendering speedup by significantly reducing $\overline{F}$, another strategy to further optimize rendering performance is to reduce the large number of feed-forward network queries $N_{\boldsymbol{r}(t)}$. One way to achieve this is to reduce the overall render resolution $R_{(h,w)}$, which will inevitably lead to a degradation of the rendering visual quality.

However, for a VR HMD, the per-eye NeRF rendering performance P_{hmd} is determined by the horizontal field of view FoV_h, the vertical field of view FoV_v of the HMD, and the pixel density per degree (PPD), which is the intrinsic parameter of the display hardware:

$$P_{hmd} = FoV_h \times FoV_v \times PPD^2 \times C \times N_{\boldsymbol{r}(t)} \times \overline{F}. \tag{4.4}$$

Note that the constant C is a multiplication factor, given the supersampling requirements in practice for aliasing reduction and anti-aliasing [9].

From Equation 4.4 we could see that reducing the FoV of NeRF rendering in VR could also result in a significant speedup without sacrificing visual quality. Another way to optimize $P_{(h,w)}$ is to reduce the overall sample size $N_{\boldsymbol{r}(t)}$. Abundant previous work has shown that this can be achieved by empty space skipping and early ray termination techniques [134], which were already part of the original instant-ngp implementation [158].

4.4 Related Work on NeRF and VR NeRF

Compared to photogrammetry or conventional RGBD sensors, creating a NeRF often requires fewer image inputs but could produce higher visual quality than photogrammetric point clouds, which tend to be erroneous with limited image feature overlap or uniform textures [252]. Compared to active 3D scanners with sub-millimeter accuracy such as those with structured illumination [250], NeRF can work with "optically uncooperative" surfaces such as metallic or absorbent materials that are common in industrial facilities. While it is also common to combine both image-based approach and active 3D scanning to create a high-quality 3D digital twin [183], it is noteworthy that the post-processing associated with these methods can take weeks to months, during which time the conditions of the accelerator facility may have already changed due to maintenance activities.

As NeRF holds significant potential for various applications, there has been a substantial amount of recent research aimed at enhancing NeRF training [154, 158, 218], rendering

[158, 25, 59], and editing capabilities [124, 84, 101] to make it viable for real-time interactive applications. Recent advancements in NeRF representation and compression, utilizing more efficient data structures such as 4D tensors [44] and multi-resolution hash tables [158], paved the way for real-time NeRF training and rendering [158], making the question of how to complement NeRF rendering with user interface systems to support different application domains becomes more and more relevant. However, much of the prior research in CV for NeRF primarily focused on demonstrating proof-of-concept experiment results, rather than delivering user-friendly toolkits for visualization and interaction with real-world NeRF data in immersive VR applications. For example, popular NeRF visualization toolkits such as NeRF Studio [220] and instant-ngp [158] primarily support scene visualization on 2D desktops. Although these toolkits can potentially enable VR visualization through stereoscopic rendering, they lack further integration into game engines such as Unity, the major platforms for VR application development that can enable more interactive and versatile VR content creation. In Chapter 7, we introduce a visualization framework that integrates a NeRF render plugin that enables interoperability between low-level NeRF inference implementation [158] and high-level game engine, making customized VR NeRF application development more scalable and flexible to a wider range of audiences. Additionally, in line with the open-source approach seen in instant-ngp and NeRF Studio, the framework's development is also open-source to support further research and development of VR NeRF.

Another challenge for VR NeRF development is the enormous network queries required for stereoscopic, high resolution, high frame rate VR rendering [135]. One promising direction is foveated rendering to reduce render resolution in the peripheral region of the human visual field [59]. However, the existing foveated NeRF method requires training and recombining the rendering results of separate networks. Moreover, it did not use the efficient multi-resolution hash coding data structure, which can achieve most of the rendering and performance speedup [59]. Another line of work investigated adapting NeRF into the conventional geometric rasterization render pipeline by converting NeRF models into textured polygons to support photorealistic rendering on mobile devices [45]. However, the surface estimation process can lead to inaccurate results with specular materials and sparse viewpoints, which degrades the visual quality and robustness of NeRF compared to volumetric rendering. In Chapter 7, we introduce a data fusion method that combines the complementary strength of volumetric rendering and rasterization to preserve NeRF render quality at the fovea region of the human's visual field while achieving render volume reduction through multi-modal data fusion without sacrificing users' immersion in the VR environments.

4.5 3D Gaussian Splattings

While accelerating NeRF rendering through foveated rendering [59], data fusion [131], adaptive sampling [158], and efficient data structure [158] present significant steps towards

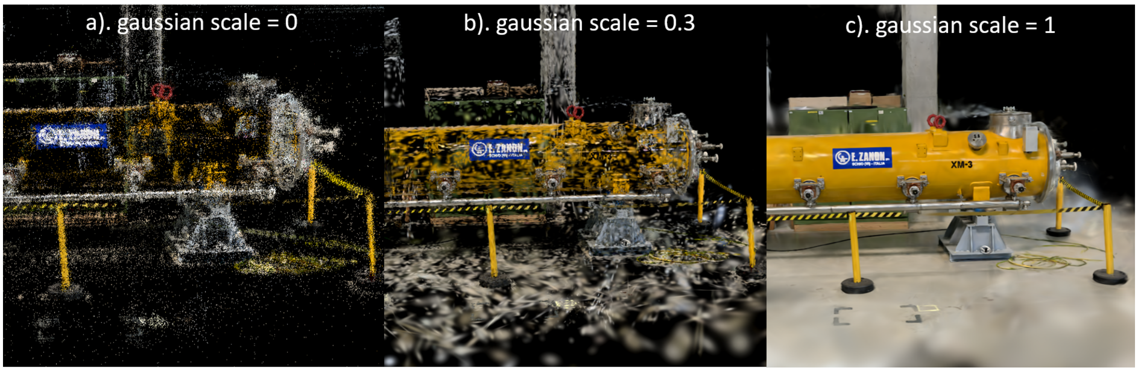


Figure 4.2: Illustration of the splatting process with 3D Gaussians, with (a) showing the initial sparse point clouds obtained from SfM, (b) showing intermediate blending state with the scale of the 3D Gassuains set to 30% of the trained results, and (c) showing the fully blended rendering.

real-time immersive NeRF experiences, representing a 3D scene with a neural network still face the performance bottleneck that makes neural scene representations incompatible for high-resolution rendering on low-resource computational devices such as a mobile phone or a standalone MR headset without large sacrifice in render quality [45]. To address this issue, the introduction of 3DGS presents another paradigm shift in photorealistic 3D scene representations that enables high-quality radiance field rendering without neural networks. As proposed by Kerbl et al, given an initial set of sparse points and camera poses estimated from a set of 2D images of a scene, the opacity and color of a real-world 3D volume can be represented as a set of 3D Gaussians optimized through gradient descent [107]. A density control regulation during training optimizes the number of Gaussians by merging the Gaussians of similar feature groups and generating new Gaussians at areas where detailed geometries and intricate appearances are needed [107]. Each 3D Gaussian represents a part of the 3D volume with its position, rotation, scale, covariance matrix, as well as spherical harmonic coefficients which encode the view-dependent radiance field values. As shown in Figure 4.2, photorealistic rendering is achieved by using a point blending method (a.k.a splatting), which projects 3D Gaussian onto a camera plane and overlays each pixel by blending a summation of radiance contribution using the list of "baked" parameters of 3D Gaussians. As shown in Figure 4.2, view synthesis via 3DGS greatly enhances the visual completeness of the initial SfM sparse point cloud rendering.

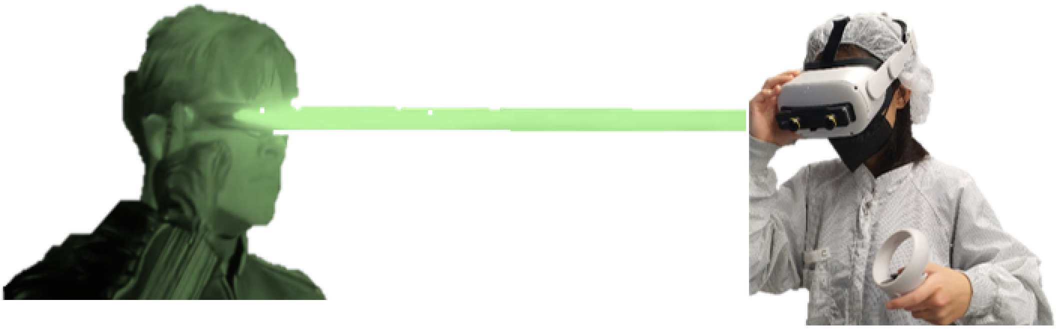
3DGS presents several significant improvements in photorealistic 3D scene representations. Firstly, 3DGS achieves state-of-the-art photorealistic view synthesis quality and rendering performance compared to mainstream NeRF methods such as instant-ngp [158] and Mip-NeRF [25]. Moreover, while 3D Gaussians are optimized using differentiable rendering and gradient descent, the final trained results can be rendered by adapting the traditional rasterization pipelines because the nature of 3DGS as scene representations is partially explicit. This greatly simplifies the integration of 3DGS into conventional game engines [179] and opens the door to efficient photorealistic rendering on low-resource mobile devices without sacrificing the rendering quality as well [27]. In Chapter 8, we present an immersive robot teleoperation system with 3DGS. This system allows real-time, wide FoV, and high-resolution rendering of photorealistic room-scale environments in immersive MR, effectively representing complex remote accelerator facilities to supply robot operators with a high level of situation awareness.

Part III

Mixed Reality Interfaces for the Particle Accelerator Metaverse

5

Taming Cyclops



Cyclops, also known as X-man, is a superhero who can emit high-power laser radiation from his eyes. However, not everyone is equipped with the superpower to protect themselves from dangerous laser radiation. Scientists and researchers at advanced optics laboratories at DESY face the same health and safety challenge. In this chapter, we introduce a novel application of VST HMD for laser eye safety, focusing on answering the following research question:

- **RQ1**: Can a stereoscopic VST-HMD be used as laser safety goggles at advanced optics laboratories? What are its current usability, advantages, and limitations?

The content of this chapter is primarily based on the following publications:

- **Ke Li**, Aradhana Choudhuri, Susanne Schmidt, Tino Lang, Reinhard Bacher, Ingmar Hartl, Wim Leemans, Frank Steinicke. (2022). Stereoscopic Video See-Through Head-Mounted Displays for Laser Safety: An Empirical Evaluation at Advanced Optics Laboratories. 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) [DOI: https://doi.org/10.1109/ISMAR55827.2022.00025]

- **Ke Li**, Aradhana Choudhuri, Susanne Schmidt, Tino Lang, Reinhard Bacher, Ingmar Hartl, Wim Leemans, Frank Steinicke. (2023). Mixed Reality for Laser Safety at Advanced Optics Laboratories. 2023 International Laser Safety Conference (ILSC) [https://www.edit.fis.uni-hamburg.de/ws/files/30357120/ILSC_2023_Final_Draft.pdf] (Invited Talk)

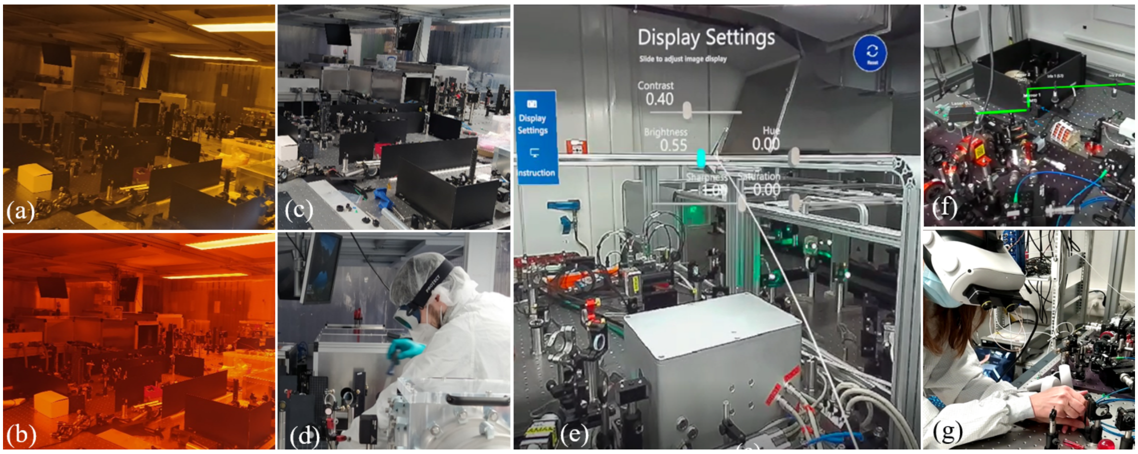


Figure 5.1: (a,b) View of an optics laboratory through two different laser safety goggles with different wavelengths of visible light filtered. (c) Natural view of the laser laboratory. (d) A researcher working with an optical system wearing a conventional laser safety goggle. (e) View from the VST-HMD when observing an optical system (the green light is infrared laser radiation which is not visible through human's natural vision). (f) View from the VST-HMD prototype with a virtual CAD model laying on the optical table. (g) A researcher working with an optical system wearing our VST-HMD prototype.

5.1 Motivation

As introduced in Chapter 2, conducting advanced physics experiments at particle accelerators heavily depends on energetic and multi-spectral lasers. A laser is a universal tool widely used in telecommunication, material processing, scientific research, and industrial processes. Nowadays, an increasing number of science experiments require either high-power wide bandwidth laser sources in the visible spectral range or laser power levels where spectral filters fail to reduce laser light to an eye-safe level. Particularly for research and development work on the laser source itself, encapsulating all hazardous laser radiation is often impossible, as it is typically implemented in industrial processes such as laser welding. Here researchers find it increasingly challenging to implement suitable personal protective equipment, particularly a robust eye protection measure, to cope with the hazardous and extreme working conditions.

The current eye protection method using conventional laser safety goggles based on optical filters can reach its protection limits for today's highest-power laser sources. According to laboratory health and security standards, safe laser operation requires reducing direct laser exposure to human eyes to below the maximum permissible exposure (MPE) threshold [191]. For most types of lasers, the optical spectral filters used by conventional laser safety goggles can provide effective protection. However, many experiments nowadays need to use multiple class 4 lasers, which are the most dangerous and powerful types of laser operating in different optical wavelengths all across the visible spectrum [24]. A typical broadband laser safety goggle can filter up to 96% – 99% of all the visible light [121, 123]. Moreover, there are no conventional laser safety goggles that could provide full-band protection that covers all the laser wavelengths and powers, as it would require filtering all the visible light. As a result, working efficiency at advanced optics laboratories and the possibility of conducting new experiments are limited by the lack of an effective and robust laser eye protection method.

Past research in integrating a VST-HMD into a welding helmet demonstrates the novel functionality of VST-HMDs as eye protectors [8, 172]. Similarly, it has also been proposed that a VST-HMD has the potential to be used as a laser safety goggle [181], especially for advanced optics laboratories [133]. However, to our best knowledge, there is not yet complete work that evaluates the various human factors of using VST-HMDs as laser safety goggles at advanced optics laboratories.

In this chapter, we evaluate a stereoscopic VST-HMD prototype through a user study with 18 participants, including 14 laser experts from DESY. The quantitative evaluation measures the system usability scale (SUS) [21] of the prototype, the perceived VR motion sickness when performing a laser alignment task using a VST-HMD, and the visibility of optics component contamination when viewed through the HMD in comparison to conventional laser safety goggles. The qualitative evaluation focuses on users' preferences, perceived safety, advantages, and limitations of the current prototype via expert reviews and post-study semi-structured interviews. The current prototype reaches an above-average mean SUS score, with users only having slight to no motion sickness symptoms. Many expert participants report seeing the significant potential of using VST-HMDs at optics laboratories, not only in providing eye protection from lasers but also in improving the complex optical system operation workflows. A demo video for this chapter is available here: https://www.youtube.com/watch?v=Qj_CBB8wIVQ

5.2 Related Work

In the following, we provide an overview of industry standards and challenges related to eye protection in advanced optics laboratories before discussing related work of using VST-HMD as an eye protector.

5.2.1 Laser Eye Safety

The increased demand for laser technology also led to an increasing number of laser accidents. In particular, 71% of the reported accidents caused eye injuries, most of which occurred in research and engineering laboratories where high-power lasers must be operated on an optical bench with the laser shield boxes open [69, 244].

Laser light that directly enters a human's eyes can be focused on an extremely small spot on the retina, leading to damages such as corneal burn, aqueous flare, cataract, or even permanent eye blindness [63]. Depending on the wavelength, power, and duration of inappropriate exposures, lasers can cause various degrees of damage to human eyes [141]. A low-power exposure will cause a thermal effect where laser light is absorbed by human eyes, leading only to a rise in temperature [197]. However, inappropriate exposure to high-power laser even only for a very short amount of time could cause an acoustic effect, where mechanical shock waves propagate through human eyes, locally vaporize, and permanently damage the tissues [40]. With exposure to ultraviolet lasers, photo-chemical effects could occur, leading to changes in cell chemistry in the eyes' tissues and even eye cancer [136]. Therefore, it is extremely important to have an effective eye protection measure while working with lasers.

For class 4 lasers, not only direct exposure to the laser beam [191] but also exposure to direct and diffuse reflections can cause permanent eye damage [244]. In laser laboratories, optical components such as mirrors, prisms, and lenses can easily cause spurious reflections of laser beams toward the users' eyes [201]. Moreover, the wavelengths of many lasers used at research and development optics laboratories are outside the visible spectrum, increasing the risk of accidents as users are not able to locate the laser radiation with their natural vision. Therefore, in class 4 laser laboratories, users must wear laser safety goggles at all times, even when they are not directly working at the optical table. Additional measures such as using laser safety windows and laser safety curtains are also commonly seen in optics laboratories.

5.2.2 Challenges of Using Conventional Laser Safety Goggles

The conventional way of laser eye protection using optical filter glass as laser safety goggles has many limitations.

A major limitation is the lower visible light transmission (VLT), which impairs the user's view of the environment and reduces the differentiability of colors. At laboratories where multiple class 4 lasers are used, researchers must wear broadband laser safety goggles that protect a wide range of laser wavelengths. Such safety goggles can filter up to 99% of all the visible light [121]. Although a broad-band laser safety goggle with dielectric coating could have a better VLT of 15% [120], such a goggle is very expensive to produce and is difficult to maintain. The dielectric coating surface applied on the side of the absorbing glass filter which is exposed to the laser radiation is very sensitive to surface contamination, and the goggles can no longer be used after just a minor scratch [47]. In addition, the dielectric coating can also lead to a significant FoV reduction, further decreasing the goggle's usability. In some cases, when multiple high-power lasers at different wavelengths need to be used, there are no filters for goggles available which would reduce all laser radiation to an eye-safe level. Thus, additional protective measures have to be implemented, for example, only allowing open shield laser operation at low power levels.

The second major limitation is that using conventional laser safety goggles requires an expensive, time-consuming, and complex selection process for determining the appropriate goggles for each different experiment. As different experiments could involve different lasers, the wavelength range of the lasers required by the experiment and therefore the required optical density (OD) of the laser goggles can change frequently at research laboratories. Each time a different type of laser is needed for an experiment, a certified laser safety officer (LSO) will need to re-evaluate the laser safety program and purchase the appropriate safety goggles before the experiments can be conducted [191]. This greatly increases the costs of a laser safety program and reduces researchers' productivity in the fast-paced scientific research and development process. Moreover, with such a requirement, human errors in wearing the wrong type of safety goggles pose additional risks to laser eye safety. Moreover, conventional laser safety goggles are expensive to produce and hard to maintain. Laser safety goggles that use reflective dielectric interference coating to filter laser light have the reflection layers usually applied on the laser side of absorbing glass filters in order to either optimize the wavelength blocking range or to further increase the protective effect of the absorbing filter [47]. Such reflective coating is very sensitive to surface contamination and the goggles can no longer be used after even minor damages such as scratches.

5.2.3 Using VST-HMDs as Eye Protectors

Quercioli [181, 180] has proposed the general idea of using a VST-HMD as laser safety goggles using a monoscopic VST-HMD. However, previous work does not include any user evaluation, and therefore the usability, advantages, and limitations of using VST-HMDs at laser laboratories are unknown. Moreover, using a monocular VST-HMD cannot meet the safety and health standards at advanced optics laboratories. With only a monocular view, users do not have sufficient depth perception. Therefore, they cannot accurately and quickly perform fine operations in three dimensions such as fine adjustment of an optical component [75]. This will lead to serious hazards when the laser beam is pointed in the wrong direction (e.g., toward a human), and the operator cannot quickly switch off or change the laser beam path.

5.3 Field Studies and User Observations

In this section, we provide a comprehensive review of the field studies and user observations conducted at the laser science and technology group (FS-LA) at the Photon Science division of DESY, where the optics laboratories have multiple class 4 lasers in operation. One of the laboratories requires researchers to use a type of laser safety goggles with only 6% VLT. Moreover, researchers mention that they have to constantly avoid using certain combinations of high-power lasers, as when the bandwidth of the required laser combination is too wide, the conventional laser safety goggles cannot provide effective protection unless filtering all the visible light. The lack of suitable eye protectors has kept many photon sciences and accelerator science experiments stagnated.

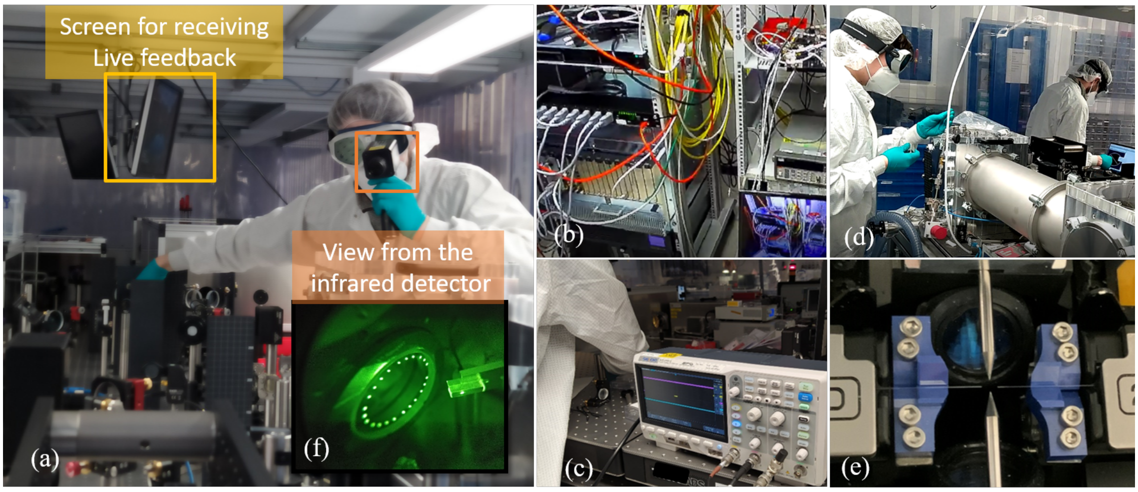


Figure 5.2: Selected images from the field studies. (a) A researcher aligns an infrared laser using an external infrared viewer, with (f) being the image taken from the infrared viewer. The screen for receiving the live feedback of the laser beam profile is placed at an unergonomic position behind the researcher. (b) A cluster of color cables was observed in one of the laser laboratories. (c) An oscilloscope is used in the laboratory with color buttons and signal displays. (d) Researchers collaborating on a task. (e) The optical fiber splicing task, where the fusion of fiberglass with a 125μm diameter is performed manually.

5.3.1 The Laser Safety Procedures

For laser eye safety, all visitors and researchers inside the laboratory need to use a laser safety goggle that only has 11% VLT. Highly reflective objects such as rings and metallic watches must be removed to avoid them redirecting laser light. Other protective measures such as laser safety curtains and interlock systems are used to ensure laboratory safety. In addition, optical setups in the laboratories were covered with a non-transparent safety enclosure to block laser radiation from exiting the optical table. However, when setting up or manipulating an optical system, researchers must open the enclosure shield, and the risk of laser hazards significantly increases.

5.3.2 The Clean Room Laboratory Standards and Requirements

For many experiments, the high-power laser beams will be further focused through optical lenses. The high intensities of the focused beams could damage optical components that are contaminated with dust, dirt, or fingerprints. In serious cases, optical component contamination could also lead to burning of the components and even fire hazards. Therefore, high-power lasers usually are operated in a cleanroom environment. Researchers and visitors need to wear a cleanroom lab coat, shoe covers, and a headcover before entering the experiment space. Before mounting any optical component on the optical table, researchers must check if the components are contaminated. Furthermore, gloves must be worn when manipulating and aligning a laser system to prevent contamination.

5.3.3 Complex Optical System Operation Workflows

One common task at the optics laboratory is laser alignment. For laser light that is beyond the visible wavelength, researchers must use an external light viewer to track and localize the radiation. As Figure 5.2 (a) shows, the researcher must hold an infrared viewer in one hand, and use the other hand to adjust the position and angle of the optical components for directing the primary laser beam to a correct position. As the laser beam must be focused on a very small spot precisely, using an infrared viewer from a far distance is a very user-unfriendly approach. As the infrared viewer can only produce a sharp image at a far distance from the object, the laser alignment of invisible lasers also depends on receiving live beam profile feedback captured by a camera. However, as shown in Figure 5.2 (a), depending on the experimental setups, the computer screens that display the live feedback are sometimes positioned unergonomically behind the operator. In addition, for some setups, there is no possibility to install a camera in the system at the desired observation position. Thus, on some occasions, more than one person is needed to perform experiments, with one person tuning the optical setup based on experience and intuition, and the other checking the live feedback or observing the alignment with an infrared viewer, and informing their colleague when to stop. Figure 5.2 (d) shows two researchers collaborating on a task, where one researcher is at a laptop retrieving and informing the other researcher of the experimental procedures.

Advanced optics laboratories consist of complex optical setups. Correct manipulation of such setups often involves understanding a system with a large number of optical components. Even for experienced users, the operation of complex optics setups relies on lab manuals, CAD drawings of the setups, and instruction videos. In a complex optics laboratory environment, retrieving and interacting with such information based on the conventional HCI methods using a keyboard, a mouse, and a laptop are challenging. Similar to the laser alignment task described before, most optical systems are operated with a combination of software controls and, simultaneously, manual mechanical controls. The unergonomic positioning of screens also leads to a poor overall workflow.

5.3.4 Consequences of Low-visibility and Color-blind Perception

At optics laboratories, users need to carry out high-precision optics alignment tasks with the low VLT broadband laser safety goggle. Such tasks could cause viewing stress even under ordinary room illumination. The color-blind vision makes it difficult for researchers to effectively interpret information that is essential for correctly performing the experiments, such as color visualization in the lab manuals, camera recordings with color-gradient of the laser beam intensity distribution, or signal readings from an oscilloscope, as shown in Figure 5.2 (c). A laser expert describes her experience of color misperception in the laboratory which almost leads to additional hazards in the laboratory:

> " *One time I was working with a black cable in the laboratory. When the lasers*

were switched off for maintenance, I entered the lab without wearing the (laser safety) goggles, but I couldn't find the black cable anymore. It turned out it was a blue cable and I didn't know I was working with a blue cable all the time. "

According to the laser expert, such cognitive errors occur very frequently. Although some preventive measures such as information encoding with QR codes could reduce the dependency on color in the laboratory, these methods are not robust enough to prevent all types of cognitive errors. As Figure 5.1 (a,b) shows, different laser safety goggles could lead to different types of color-blind perception. Removing color dependency will unrealistically require removing all color-related information in the laboratory, and replacing all essential color information with QR code encoding is not an efficient approach in a fast-changing research and development physics laboratory.

One of the most challenging tasks we have observed at the laboratory is optical fiber splicing. Researchers need to wear the conventional laser safety goggles and manually fuse optical fibers with $0.125mm$ outer diameter, as Figure 5.2 (e) shows [152]. Being able to perceive thin optical fibers is also essential for workplace safety. The thin, sharp glass fibers are very fragile and can easily break off. The fiber shards are potentially harmful to human eyes and skin or may stick in users' clothing and be carried to other locations, posing potential hazards outside the laboratories [99].

Under these extreme conditions, many researchers work more than 20 hours per week in laboratories with laser safety goggles on.

5.3.5 Summary

In summary, besides missing an effective laser eye protection method, several additional challenges make the working conditions of optics laboratories quite hazardous and extreme. Firstly, there is a lack of high-quality visual aids and color vision in assisting the demanding, high-precision optics operation tasks, such as lens artifact inspections, optical system alignment, and optical fiber splicing. Secondly, there is a lack of effective HCI methods in the laboratories to retrieve digital information, when the user's hands are occupied and the need to interact with a computer arises.

5.4 Prototype Design

In this section, we describe the software and hardware design, development, and technical details of the stereoscopic VST-HMD prototype used in our user study.

5.4.1 System Hardware

To better meet the health and safety standards of high-power laser operation, we developed a stereoscopic VST-HMD prototype, which integrates a high-resolution stereoscopic camera into a standalone VR headset. The ZED Mini camera is a stereo camera that can

Figure 5.3: System overview for the VST-HMD prototype, illustrating the software stack and shareware components.

provide natural stereo vision similar to the human eyes and is widely used in the MR community for research and prototyping [247]. In addition, we choose the ZED Mini camera as its video capture wavelength range is extended to the infrared region, which can allow researchers to naturally perceive infrared laser radiation without having to fusion extra infrared sensors. As shown in Figure 5.3, while the ZED Mini camera is originally designed for the Oculus Rift VR headset, we integrate it with an Oculus Quest 2 VR headset [150], a standalone VR device that has a higher per-eye resolution and refresh rate, therefore can achieve better visual quality. The ZED Mini camera is mounted on the Oculus Quest 2 with the same camera mount provided for the Oculus Rift. Finally, our prototype enables an HD720 stereo see-through from the ZED Mini camera with a fixed framerate of 60 fps. The Oculus Quest 2 headset has a display resolution of 1832×1920 per eye with up to 120 Hz refresh rate.

To minimize the number of cable connections in our setup, we use a dedicated 5G streaming network for wireless VR content transmission, instead of connecting the VR headset to the computer via a link cable. The VR content is played through Steam VR and streamed via the Virtual Desktop Client [232]. The ZED Mini camera is connected to an Alienware m17 R2 laptop with an NVIDIA GeForce RTX 2080 graphics computation unit (GPU). The data is transmitted from the ZED Mini camera to the laptop via a high-speed type C USB3 cable. An Asu AX1800 WiFi6 5G router with a data transfer rate of 1200 Mbits per second is used for wireless streaming.

The average motion-to-photon latency measured by a laser pointer and a 240 fps slow-motion camera is $125.00ms \pm 16.67ms$. Although the end-to-end latency seems quite large, the ZED Unity SDK integrates an asynchronous time warp algorithm [236], which can significantly reduce the registration error of the VST-HMD through a robust re-projection technique. The time warp algorithm has proven to be highly effective in compensating for the perceived video latency and increasing the perceived frame rate for VST-HMDs [236, 73].

5.4.2 Software Implementation

On top of the hardware framework, we integrate several 3D software features to assist users' general operation workflows.

As summarized from the field studies, visual inspection of laser radiation and optical component contamination is very important in operating an optical system in a cleanroom environment. Therefore, we develop a display setting feature that serves as a visual aid that allows users to modify the contrast, brightness, saturation, sharpness, and hue of the see-through video via 3D sliders in real time. Being able to adjust these video display parameters dynamically can help users view the laboratory environment and the optical components with a more comfortable visual quality. For example, when observing a highly reflective mirror under bright room illumination, users can reduce the brightness and saturation level of the display to avoid excessive specular reflection, thus making mirror contamination easier to detect. By enhancing the contrast of laser radiation against the environmental background, users can also more clearly perceive laser radiation against its background. Figure 5.1 (e) presents a screenshot of the display settings user interface.

Secondly, we integrate a 3D CAD model of an optical beam walk system. Similar to the 3D visual aid that medical surgeons commonly use in performing fine surgery operations [189], overlaying a 3D CAD model of the optical system in the real world can greatly assist with researchers' workflows, especially for large-scale experiments that involve hundreds of optical components. In addition, multimedia information such as tutorial videos is provided. Figure 5.4 illustrates an example of a 3D CAD model display while a user watches a video tutorial. The main menu for selecting the different functionalities consistently follows the user, while other components such as the 3D CAD models and video tutorial components can be flexibly positioned and resized within the 3D environment using both VR controllers. As most users at laser laboratories have very little experience with VR, we only provide interaction modality with the 3D UIs via the Oculus touch controllers, which is the more stable and accurate interaction modality in comparison to hand tracking [82].

The 3D UIs are developed using the Unity game engine 2019.4.29f1, the ZED Unity Plugin 3.5.2, Oculus Unity integration SDK 33, and the Microsoft Mixed Reality Toolkit (MRTK) 2.7.2.

5.4.3 Camera View Calibration

The correct placement of the camera view is important in facilitating correct hand-eye coordination when using a VST-HMD [173]. To compensate for the translational placement offset of the stereoscopic camera, we utilize an iterative manual hand-eye calibration approach. Figure 5.4 (c) shows a screenshot of the calibration user interface. The Oculus Quest 2's inside-out tracking system provides accurate positions of the controller in the real world, which we use as a reference. Then, we compensate for the displacement of the ZED Mini camera by moving the ZED Mini's camera view so that the spatial positions of

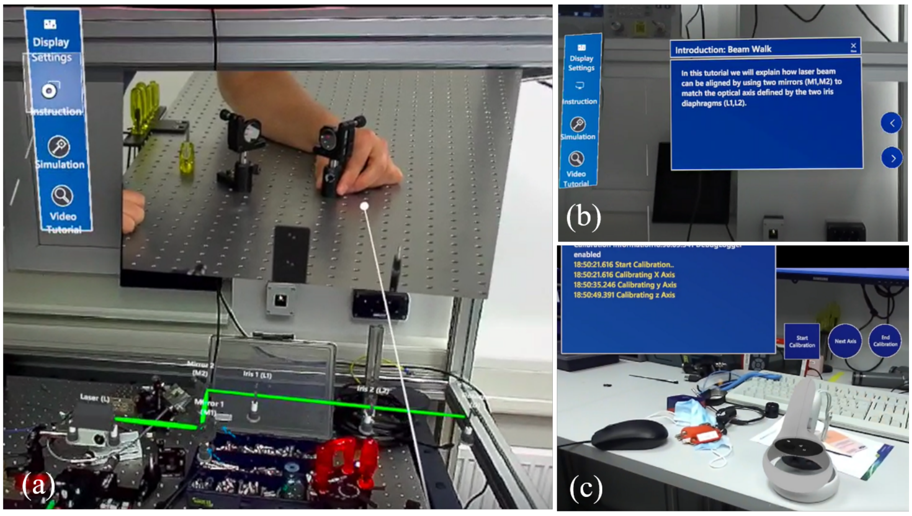


Figure 5.4: Screenshots from the VST-HMD prototype. (a) Display of a CAD model and an instruction video. (b) An instruction manual for the user study. (c) Screenshot of the hand-eye manual calibration user interface. The position of the virtual controller should exactly overlay on top of the actual controller for indication of the correct placement of the virtual camera planes.

the real-world Oculus Quest 2 controller displayed by the ZED Mini camera match the virtual controllers displayed by the Oculus Quest 2 tracking system.

5.5 User Study

The first part of the study includes a within-subject visibility test designed together with several laser safety experts from the laboratories. To evaluate if a VST-HMD could replace the conventional laser safety goggles, we must learn how well the VST-HMD can fulfill the visibility requirements at the advanced optics laboratories. As we have seen from the user observations and field studies, good visibility of the laboratory not only requires full VLT, but also a high-resolution vision for ensuring experiment and fiber safety. Therefore, the visibility test aims to evaluate participants' performance in color recognition, perception of small fiber glasses, and optical components contamination recognition. This will help future laser operators in deciding in what scenarios and tasks could the current VST-HMD prototype be an appropriate and acceptable alternative for laser laboratories.

The second part of the study is an exploratory operation study, where participants are asked to align a laser beam through two irises by performing a "beam walk" using two mirrors directing the beam to the irises. This "beam walk" task is a very common alignment procedure that could be seen in almost any optical system where precise alignment of laser beams is required. It involves tasks that researchers perform daily at an advanced optics laboratory, such as mounting optical components on an optics bench, adjusting the optical components to the correct height and angles to correctly redirect the laser beam, and performing fine adjustments to align the laser beam to the target location. The goal

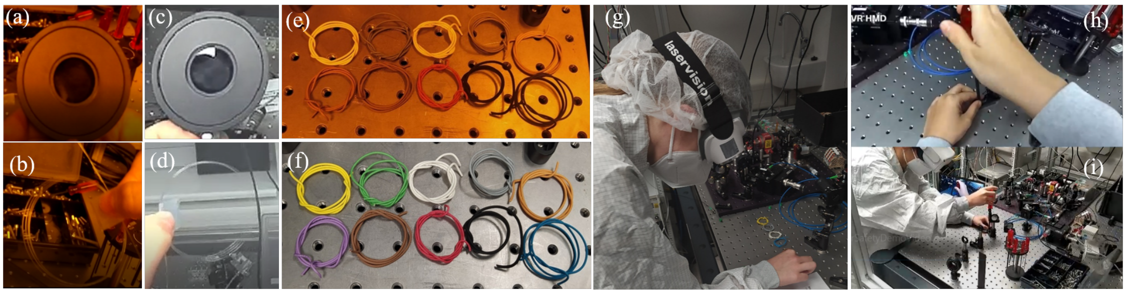


Figure 5.5: (a/c) Viewing a coated and contaminated mirror from the conventional laser safety goggle / the VST-HMD prototype. (b/d) Viewing an optical fiber from the conventional laser safety goggle / the VST-HMD prototype. (e / f) Viewing 10 electronic cables with different colors from the conventional laser safety goggles / the VST-HMD prototype. (g) Participants performing the visibility study task wearing conventional laser safety goggles. (h-i) Participants performing optical alignment tasks wearing the VST-HMD prototype.

of this study is to allow laser experts to explore our prototype to perform a typical optics operation task.

5.5.1 Participants

18 participants took part in the user study, among which 9 self-identified as men and 9 self-identified as women. Their age ranges from 22 to 47 (M = 32.6, SD = 6.49). 12 participants had never used a VR headset before, 4 participants had used it only a few times, and 2 participants have a lot of experience in VR. 10 participants worked more than 20 hours per week at advanced optics laboratories, 4 worked more than 10 hours per week, and 4 participants have never worked with an optical setup before. Since high-energy laser labs also employ student assistants, which are newly trained in simple laser operation, the 4 non-experts were included in our sample population to have a balanced representation of the target user group. All of them had a normal or corrected-to-normal vision, 8 wore glasses, and 2 wore contact lenses during the study. No other eye disorder, including color blindness, was reported.

5.5.2 Tasks

Visibility Tests In the visibility test, we used a conventional laser safety goggle with 30% VLT [122]. Although for the most extreme conditions, researchers need to use a laser safety goggle with only 1% of VLT, we did not choose a broadband laser safety goggle with 1% visibility [121], as such low visibility could lead to accidents for participants who are not familiar with working under such conditions.

The tasks for the visibility study are described below:

- Test 1 (**T1**): Participants were given 10 colored cables (yellow, green, white, gray, orange, purple, brown, red, black, and blue) in random order and were asked to arrange them in a specific sequential order, as shown in Figures 5.5 (e), (f), and (g).

- Test 2 (**T2**): Participants were asked to describe the conditions of four different mirrors and lenses, as shown in Figures 5.5 (a) and (c). The four optical components contained different contamination such as tiny dust, fingerprints, larger shining specks of dust, and scratches, which are the most common types of artifacts on optical components.

- Test 3 (**T3**): Without touching the fiber, participants were asked to identify how many circular revolutions a thin optical fiber has, as shown in Figures 5.5 (b) and (d).

Optics Operation Task The task of the operation study consisted of aligning a laser through two irises by "beam walk". The optical setup consisted of a 635 nm, 0.9 mW class 2 alignment laser [48], two mirrors, and two irises. Class 2 visible-light lasers are considered safe even with unintentional eye exposure and are used in laboratories without requiring wearing laser safety goggles. However, performing tasks with the actual class 2 laser on was not required and was only a voluntary option. Only expert optics lab users who had laser safety training before were allowed to perform this task with the laser on. Virtual video tutorials and CAD models were available from the VST-HMD to assist with the tasks.

5.5.3 Measures

Measurement of Time and Correctness For **T1** of the visibility study, we recorded the elapsed time between the moment when the participant was presented with a visual instruction and the moment the user indicated finishing the task. In addition, the number of correctly placed cables for each participant was also recorded. For **T2** of the visibility study, we recorded the kind of artifacts and contamination that each participant could identify. For **T3** of the visibility study, we recorded whether each participant was able to see the fiber as well as the correctness of the participant's response regarding the number of fiber revolutions.

Virtual Reality Motion Sickness Questionnaire Right after the participants finished the second part of the study, they were asked to fill out a standard virtual reality motion sickness questionnaire (VRSQ) [112] to evaluate the degree of motion sickness they experienced during the study. The questionnaire consists of 9 questions to be answered on a 4-point Likert scale. A copy of the VRSQ can be found in the Appendix A.1.1.

System Usability Scale Questionnaire After the participants had finished the second part of the study and the VRSQ, they were asked to fill out a system usability scale (SUS) [21] questionnaire. SUS is a standard questionnaire to reliably evaluate the usability of a software application. The SUS consists of 10 questions to be evaluated on an 11-point Likert scale (from 0 fully disagree to 10 fully agree). Since most SUS questions are aimed at first-time users and are generally not suitable for assessing non-digital physical devices, we did not ask participants to complete the same questionnaire for the conventional laser safety goggle condition. To compare the usability of the conventional laser safety goggles and the VST-HMD, we performed a qualitative analysis through semi-structured interviews. A copy of the SUS questionnaire can be found in the Appendix A.1.5.

Semi-structured Interviews After all the experiments, each participant was invited for an individual semi-structured interview. The discussion revolved around the following questions:

- **Q1** Would you prefer the VST-HMD as a laser safety goggle over the conventional laser safety goggle, and why?
- **Q2** Would you feel safe using the VST-HMD as a laser safety goggle working at an advanced optics laboratory, and why?

5.5.4 Procedures

The experiment started with each participant filling out a consent form and a demographic questionnaire that informed them about the tasks involved in the study, the potential risks of VR motion sickness, and data processing policies. The participant began with the visibility tests. Half of the participants were asked to first complete the tasks with the conventional laser safety goggles, and the other half started with the VST-HMD prototype. Which participant began with which laser safety goggles was randomly decided. Participants who had no previous experiences with VR were given a short, 3-5 minutes introduction to VR. When working with the VST-HMD prototype, participants followed the instructions given in the 3DUI and had the option to change the display settings (e.g., brightness, contrast, sharpness, hue, etc) of the video pass-through via the 3DUI, as shown in Figure 5.4 (b). When working with the conventional laser safety goggles, participants followed a paper form instruction. The first part of the study lasted an average of 15 minutes.

After completing the visibility tests, participants could take a 3-minute break. In the second part of the study, participants were asked to wear the VST-HMD prototype to perform a beam walk. They were given up to 5 minutes to explore the features of the VST-HMD before starting the task. Participants who had received laser safety training could voluntarily switch on the laser. All 14 experienced optics researchers voluntarily turned on the laser. The second part of the study lasted an average of 15 minutes. After completing all the tasks, participants were asked to fill out the SUS questionnaire and the VRSQ. The study concluded with a semi-structured interview. The entire study lasted on average 45 minutes.

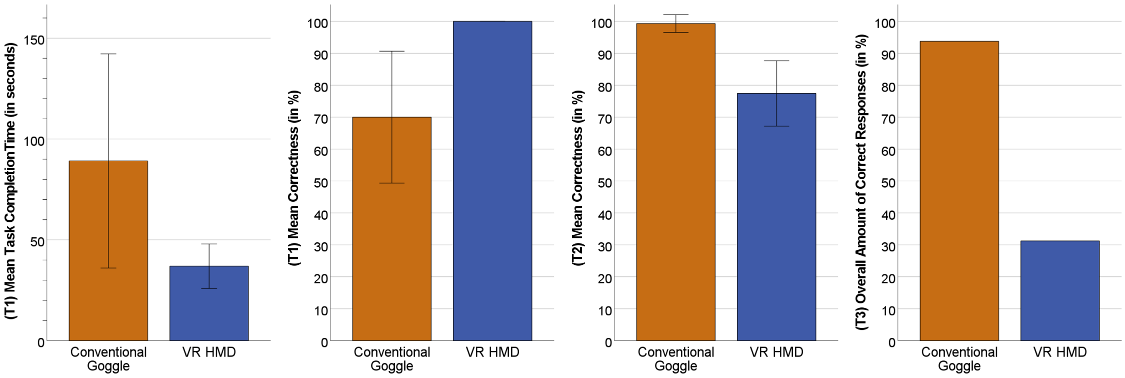


Figure 5.6: Pooled results of the visibility study, including (a) time measurements and (b) correctness in the color identification task, as well as (c) percentage of correctly identified lens contamination, and (d) relative amount of participants correctly recognizing the number of revolutions of a fiber. The vertical bars in (a)-(c) show the standard deviation.

5.6 Results

5.6.1 Visibility Study

For evaluating the results of the visibility tests, we computed differences between the measurements in the conventional goggles and VST-HMD conditions.

The distribution of the T1 task completion time difference was skewed (1.587), therefore violating the t-test's normality assumption. For this reason, we log 10 transformed the difference and confirmed the resulting normal distribution using the Shapiro-Wilk test and visual inspection of histograms as well as QQ-plots. Afterwards, we performed a one-sample t-test against 0 ($t(15) = 11.811$, $p < 0.001$, Cohen's $d = 2.953$), revealing a significant difference between conventional goggles ($M = 89.125$, $SD = 53.085$) and the VST-HMD ($M = 36.938$, $SD = 11.006$).

For the correctness measures of the tasks T1 and T2, the differences were approximately normally distributed according to the Shapiro-Wilk test and histograms as well as QQ plots. Paired t-tests showed a significant difference between conventional goggles ($M = 0.700$, $SD = 0.207$) and the VST-HMD ($M = 1.000$, $SD = 0.000$) for the color identification task T1 ($t(15) = -5.809$, $p < 0.001$, Cohen's $d = -1.452$). Differences between conventional goggles ($M = 0.993$, $SD = 0.028$) and the VST-HMD ($M = 0.774$, $SD = 0.102$) were also significant for the lens dirt identification task T2 ($t(15) = 8.904$, $p < 0.001$, Cohen's $d = 2.226$).

Finally, we performed an exact McNemar's test on the dichotomous correctness variable of task T3, the identification of circular revolutions of fiber. We found a significant difference between conventional goggles ($M = 0.938$, $SD = 0.250$) and VST-HMD ($M = 0.313$, $SD = 0.479$) with $p = 0.002$. A follow-up question to study participants revealed that 100% of correct responses were guessed for the VST-HMD condition, while only 13.3% of correct responses were guessed for the condition using conventional goggles. All results of the visibility study are illustrated in Figure 5.6.

5.6.2 Usability and VR Sickness

For comparing the usability of the VST-HMD system to industrial norms, responses to all ten SUS questions were added up to a total score with a maximum of 100. This yields an average SUS score of 77.906 (SD = 8.009), which converts to a percentile rank of roughly 80% (i.e., the perceived usability of the VST-HMD system is higher than for 80% of the products considered in a review of 500 studies by Sauro [190]). In addition, we found a large negative correlation between the individual SUS score and the participant's age (Pearson's $r = -0.57$).

For assessing the VR sickness of participants after using the VST-HMD prototype, we computed average scores for the oculomotor (M = 23.438, SD = 18.313) and disorientation (M = 12.917, SD = 17.464) components of the VRSQ as suggested by Kim et al. [112]. Averaging both components yields a total score of 18.177 (SD = 16.001). This indicates that users only suffered from minor to no motion sickness while using the prototype. It also demonstrates that the asynchronous time warp effect from the ZED Unity SDK has effectively compensated for the high end-to-end streaming latency.

5.6.3 User Preferences

In response to **Q1**, 6 participants indicated that they preferred the VST-HMD, 4 participants preferred the conventional laser safety goggle, and 8 indicated that they preferred having both options.

Preference for the VST-HMD The participants who preferred the VST-HMD reported that using the VST-HMD not only gives them the advantage of having full visibility of the environment but is also helpful for simplifying their workflows. In addition, The participants mentioned that their work could greatly benefit from having the overlay of the CAD models in the real world. Many spatial parameters such as distance, height, and angle of the optical components can be visualized and displayed in 3D. Among these participants, four were laser experts. These participants mostly reported experiencing no motion sickness and could imagine using the VST-HMD for their daily work.

Preference for both options The participants who preferred having both laser safety goggles also reported seeing the great potential of the solution, but they remained critical of the current VST-HMD prototype. The most widely criticized limitation is the cable connection, as it restricts their physical flexibility when working in a narrow and complex environment. The second main concern is the limited perceived headset resolution. Although the ZED Mini camera in HD720 mode can capture small details such as fingerprints and dust, the perceived resolution is less due to the VAC effect [116]. Many participants reported their perception sometimes felt "weird" when they tried to look at the optical components or the fiber glasses at a close distance. Therefore, many of the participants preferred to examine the components at a larger distance, leading to lower perceived resolution. In addition, the ZED Mini camera has a fixed focal length and only objects at distances from 28cm to infinity will be sharp. This could further reduce the actual perceived resolution at a close distance. Therefore, some researchers expressed the concerns that they would not be able to use the current prototype for tasks such as mounting optical components that require close examination of components contamination for cleanroom standards, or wearing the headset when performing optical fiber splicing. However, the participants remained very open to the VST-HMD option, as they believed that working with the VST-HMD with some limitations is still a more human-friendly solution than wearing conventional laser safety goggles with only 1% of VLT. Moreover, they mentioned that not all optics tasks they perform require examining component contamination and performing fiber splicing. For some experiment setups, contamination-free components are already mounted on the optical table, researchers only need to perform system alignment with the lasers by adjusting the height and angle of the optics mounts. Typically, the conventional broad-band goggles used for such system alignment tasks have the lowest VLT [121]. For such tasks, the current VST-HMD could already be an alternative.

Preference for the conventional goggles The participants who preferred only using the conventional laser safety goggles raised concerns about the long-term side effects on their health. They also experienced severe blurry vision when perceiving distant objects due to the fixed focal length of the ZED Mini camera and the convergence-accommodation conflict. In addition, two participants self-reported to be highly sensitive to VR motion sickness and would not like to use VR for a long time. Three participants are experts who were trained to work in the optics laboratories conventionally for decades. They mentioned that with no previous VR experiences, the current limitations of VST-HMD could reduce their productivity rather than improve it.

5.6.4 Perceived Safety

In response to **Q2**, the four participants who preferred the conventional safety goggles reported not feeling safe while working with the current VST-HMD. Their main concerns were the long-term side effects of using a VR headset rather than laser eye safety. One participant mentioned that if using the VST-HMD for too long, it might be hard for him to adjust to and comprehend the normal real-world vision again. Although the human visual system could adapt to unnatural displays even if the images are inverted, reversed, or displaced from their normal positions, studies show that motor and proprioceptive changes could persist even after the vision change restores [85, 86]. Therefore, future research needs to investigate the long-term side effects of using VST-HMD on human visual, motor, and proprioceptive systems. The remaining 14 participants reported that they feel safe using the current VST-HMD prototype, as there is no direct exposure of their eyes to lasers. As laser incidents are most likely to occur when the researchers are not aware of the position of the laser beam that is beyond visible wavelengths, several participants reported that they would feel safer wearing the VST-HMD where these laser radiations can be seen. However, all participants recognized that carrying a heavy laptop inside a complex environment could lead to additional hazards.

With some laser expert participants, we also discussed exceptional circumstances. On the occasion that unexpected headset failures occur in the laboratory, carrying a conventional laser safety goggle in the pocket of the lab coat as a fallback would be a viable solution. When a primary high-power laser beam hits a camera, the camera will likely be damaged as well. However, the costs of laser damaging laser safety goggles are much more than the costs of them damaging VST cameras, as our current prototype excluding the laptop is three to four times cheaper than a typical conventional broad-band laser safety goggles with special optical coating [120]. Moreover, the current prototype has the VR headset's built-in VST as a fallback. Even if the main cameras are damaged, users could still quickly access the built-in VST. As the built-in VST system on Oculus Quest 2 consists of four different cameras, and high-power lasers are operated in a laser shield box, it is very unlikely that radiation from the lasers will damage all four cameras at the same time. Nonetheless, future research could also investigate using a hybrid VST and OST headset [52], or applying similar optical density filters on the VST cameras and reconstructing the real-world illumination via image post-processing techniques.

The expert participants also mentioned that the future design of an advanced optics laboratory needs to consider protecting the expensive experiment equipment from accidental damages due to VR headset failures. An emergency switch-off-control of the laser facilities via an easy-to-access button needs to be available to protect the expensive laser system.

5.7 Discussion

5.7.1 VST-HMD Advantages and Current Limitations

On the one hand, VST-HMDs have some appealing benefits in providing laser safety for researchers at advanced optics laboratories. It could help researchers better perceive the environment when working under extreme conditions with multiple high-energy lasers without any reduction of view of the color. In addition, when using VST-HMD for operating lasers beyond visible wavelengths, researchers no longer need to hold an external infrared viewer or a laser detection card, as the laser light could be directly visible through the VST-HMD. Moreover, researchers could simplify their current optics operation workflows by directly accessing the lab manual and 3D drawings of the experiment setup through the VST-HMD.

On the other hand, there are several known limitations in our prototype, such as the cable connection, video streaming latency, and VAC. A high-end MR device such as the Varjo XR-3 headset [222], which has higher display resolution, VST resolution, and lower latency should be able to better meet the visibility demand of laser laboratories. However, when using such high-end devices, high-energy laser laboratory environments need to design larger free spaces for VR setup installation. In addition, using such a device will introduce higher hardware costs, with a single VR device being at least three times more expensive than a typical broad-band laser safety goggle, making it a less scalable and cost-effective solution. Thus, each laser laboratory needs to evaluate the performance and rendering quality needed when choosing the appropriate HMDs for their experiments.

5.7.2 Human Factors

Given the large end-to-end latency of the current prototype, it is unclear if the effects of cybersickness would increase for longer VR exposure. Nonetheless, many participants still favored the current prototype for static laser operation and control tasks that do not require inspection of optical contamination or optical fiber splicing. In addition, the optical design of existing VST-HMDs may lead to the undesirable convergence accommodation effect, which may result in eye strain when inspecting optical components at close distances. Weather using a VST-HMD for operation tasks causes more stress to the human eyes, especially compared to conventional laser safety goggles with only 1% visibility of the environment needs further investigation.

Our demographic questionnaire reveals that most participants, regardless of occupation, already spend a considerable amount of time in front of a computer screen daily. However, using a VST-HMD for daily laser operation involves blue light exposure at a larger field of view (FoV) than using a laptop. The long-term effects of such exposure need further investigation to estimate if a maximum daily VR exposure time needs to be recommended for getting the most out of the safety and health benefits of using the"*digital laser safety goggles*". Moreover, the user study also reveals that not all researchers will be willing to adapt to using a VST-HMD as an alternative laser safety goggle in the long term. As we found a large negative correlation between the individual SUS score and the participant's age, how to facilitate familiarization with new technology for older user groups will remain an active field of research.

5.7.3 Integrated Visual System and Diminished Reality

To accommodate specialized experimental conditions, innovation in the integrated visual system could further improve operator productivity. For example, experiments such as quantum imaging often require low-light conditions. In this context, a VST-HMD integrated with stereo night vision would be important. Additionally, for experiments utilizing ultrafast laser pulses, a trigger synchronization mechanism needs to be developed to align the camera trigger and laser trigger, such that the laser is perceived as continuous radiation rather than short pulses.

This work could also be further extended to the domain of diminished reality (DR), where the VST-HMD could diminish the user's perceived environment by concealing, eliminating, or replacing the real-world environment in real-time [156]. Some complex optical systems involve multiple lasers, each with different wavelengths and intensity. Such a system could be visually cluttered with various laser beam paths, making the setup confusing to work with. For example, a dull color, low-intensity laser beam path could be hard to perceive with a bright color, high-intensity laser beam nearby. A system with different laser beam paths overlapping with each other could make the visual search of a specific laser beam difficult. DR has the potential to cope with these challenges by giving users the option to toggle the display of each laser wavelength, making it possible to hide unwanted laser wavelengths when inspecting a primary laser beam.

5.7.4 Co-located MR Experiences

The design of conventional laser safety goggles, which cover the entire field of view of human vision, results in a large portion of the user's face being obscured, negatively impacting the ability to make eye contact and share social cues during group tasks. Recent advancements in integrated eye-tracking and face-tracking technology in commercial VR headsets present a potential solution to this challenge through the use of co-located multi-user MR experiences. A virtual avatar face could be displayed and merged with the real-world face and body of the co-located users. By wearing a VST-HMD, each individual's avatar could display the real-time facial expression of the user for better communication with each other.

5.7.5 Remote Telepresence and Collaboration

When conducting experiments with multi-spectral high-power lasers, it is beneficial to reduce the required personnel in order to reduce the risk of laser accidents. Additionally, in certain laboratory environments, the presence of multiple individuals can result in an increase in ambient temperature, which may negatively impact the temperature requirements of the experiment. The incorporation of an egocentric stereo 3D camera in the laser safety goggles offers the potential for remote telepresence and collaboration, providing an alternative means of reducing the number of personnel physically present in the laboratory. Furthermore, the use of such a camera can provide benefits for individuals operating complex optical systems, as it allows for the freeing of the user's hands from holding a separate camera to show remote experts a live feed of the experimental setup.

5.7.6 Additional Safety Concepts

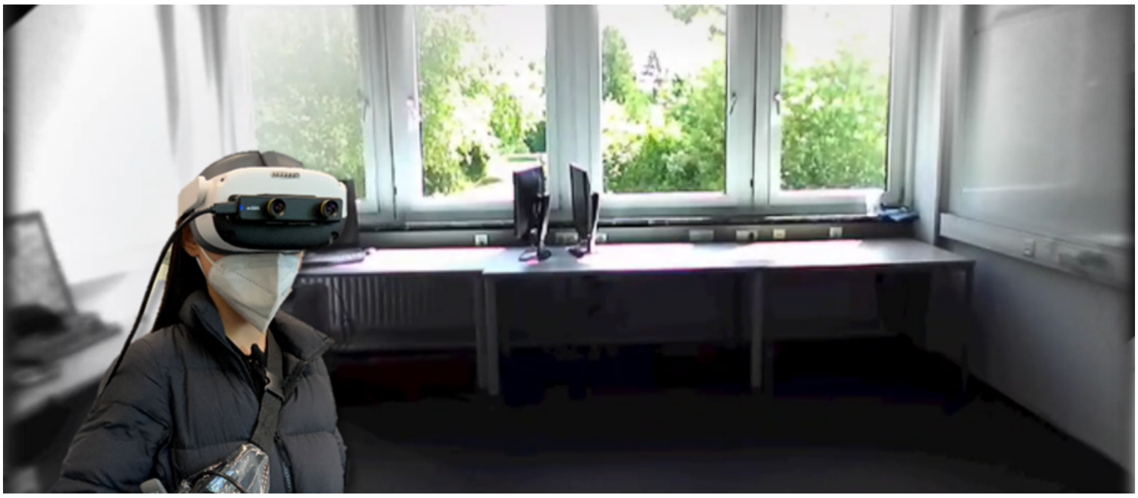
The current VST-HMD is developed using commercially available components and has a cost-effective scalability advantage, as the net cost, excluding the laptop, is less than that of advanced laser safety goggles, such as those with a thin dielectric coating. However, it is crucial to protect the VST-HMD from damage caused by high-power lasers to reduce the costs associated with a laser safety program. Future research in this area could investigate novel methods for protecting the VST-HMD, such as the application of optical filters to the stereo camera lenses and the use of state-of-the-art digital image filtering algorithms to reconstruct missing color information, either through the use of additional color sensors or through the utilization of rapidly advancing artificial neural network methods.

5.8 Conclusion

In this chapter, we identified an important application domain that is largely unknown and unexplored in the MR community: using stereoscopic VST-HMDs as laser safety goggles. Through field studies, user studies, and expert feedback, we revealed abundant evidence that for future photon science and accelerator science experiments, VST-HMD will likely be the only option due to the limits of the conventional safety goggle. Based on these results, we developed the first stereoscopic VST-HMD prototype for advanced optics laboratories. The exact requirements of using VST-HMD were thoroughly analyzed through an empirical evaluation which resulted in high user preferences for using VST-HMDs. We believe that this work will greatly motivate future work in the MR community for solving the challenging eye protection problems that optics researchers are facing.

6

Mixed Reality Tunneling Effects



In this chapter, we present MR tunneling, a novel sensor fusion technique to create a perceptual high-resolution stereoscopic untethered VST-HMD by exploiting the characteristics of the human vision system. MR tunneling investigates the following research question:

- **RQ2**: How to balance the trade-off between limited render performance and high visual quality of a stereoscopic untethered VST-HMD?

The content of this chapter is primarily based on the following publication:

- **Ke Li**, Susanne Schmidt, Reinhard Bacher, Wim Leemans, Frank Steinicke. (2022). Mixed Reality Tunneling Effects for Stereoscopic Untethered Video-See-Through Head-Mounted Displays. 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) [DOI: https://doi.org/10.1109/ISMAR55827.2022.00018]

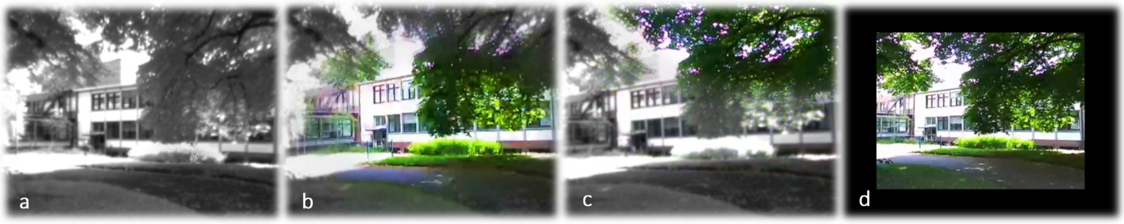


Figure 6.1: Screenshots of (a) the low-resolution grayscale VST from a standalone VR headset, (b) the MR Tunneling effect, which fuses the grayscale VST and an external high-resolution VST, and (c) the foveated MR tunneling effect, where users can control the center of the FoV restrictor through eye movements, and (d) the external high-resolution VST without MR tunneling.

6.1 Motivations

In the previous chapter, we demonstrated the significance of a VST-HMD for laser safety at R&D laboratories at particle accelerators. However, as revealed by the user study experiment and expert reviews, existing VST-HMDs still exhibit several limitations that mitigate their practicality for certain tasks in laser laboratories. As mentioned in Chapter 1, a high-resolution, wide FoV, colorful stereoscopic VST-HMD such as the Varjo-XR3 [230] still requires a tethered connection to an expensive and non-portable computer station. Such a VST-HMD limits users to a small space, making it a non-viable solution for real-world implementation as laser safety goggles, where users are required to equip a flexible VST-HMD that they can easily travel with when working in a complex and hazardous laboratory setting [133]. On the other hand, existing portable VST-HMDs such as standalone VR headsets often provide grayscale low-resolution VST. This view of the real world is referred to as the passthrough mode available, for instance, in the Oculus Quest 2 or VIVE Flow HMDs. However, the fish-eye lenses and multi-camera image stitching algorithms [71] lead to a distorted view of the real world [226]. An increase in the number and quality of VST camera sensors in VR headsets inevitably increases hardware costs, headset weight, and computational resource requirements.

One common approach to enhance an existing integrated MR display system is sensor fusion, for example, by adding images from cameras with higher resolutions or different modalities [68, 169]. In this chapter, we extend the previous MR sensor fusion framework with MR tunneling, a novel method to enhance the VST functionality of a standalone VR headset. We merge the built-in grayscale camera inputs from the VR headset with the inputs from an external stereoscopic camera, as Figure 6.1b shows. MR tunneling has the potential to enhance VST-HMDs in several aspects. Firstly, MR tunneling can address the foveal-peripheral characteristics of human vision [194]. By displaying the external high-resolution VST in the central-foveal to para-peripheral region, users can perceive the world with high-level details without losing low-level environmental awareness in the peripheral region, where the low-resolution grayscale VST is shown [175]. Secondly, displaying the ultra-low latency VST in the peripheral region automatically creates a "tunneling effect" known for effective cybersickness reduction in VR [7], which might also reduce these undesired effects in MR.

To achieve the described MR tunneling effects, we developed a custom-built high-resolution RGB VST-HMD framework based on a small embedded system, a standalone VR headset, and an external stereoscopic camera. Our prototype is lightweight and flexible, which overcomes the portability issue of previous commercial and custom-built VST-HMDs [177]. To compensate for the different latencies of the two camera sensors, we integrated a "time warp" approach [236, 73] and developed two extended MR tunneling algorithms for latency compensation. The first one is the *head speed accommodation effect*, an algorithm that rapidly fades away the central high-latency video stream upon detection of rapid head movement. The second one is *foveated MR tunneling*, which facilitates users' visual search activities via eye-tracking, as illustrated in Figure 6.1 c.

To evaluate our system and the effects of the proposed methods on perceived usability, presence, cybersickness, and task load, we perform a user study involving 24 participants. In the study, users need to solve puzzle-assembling tasks that involve hand-eye coordination, untethered locomotion, and fine motor skills.

To summarize, the contributions of this chapter include:

- A custom-built, portable, high-resolution, and wide FoV VST-HMD framework improving the VST functionality of standalone VR headsets.
- Three novel MR tunneling effects for stereoscopic camera sensor fusion, FoV extension, and latency compensation.
- A thorough user study with puzzle assembling tasks to investigate the advantages and limitations of our system and methods.

The source code of the software system is made publicly available here: https://github.com/keli95566/MRTunnelingPico. A demo video of the system design and user study experiment process is available here: https://youtu.be/yIDXRc3FDJA.

6.2 Related Work

6.2.1 VST-HMDs and Sensor Fusion

One main barrier in the development of stereoscopic VST-HMDs is the trade-off between their performance and portability. To design a lightweight device, only a limited amount of computation units and sensors can be integrated into the wearable headset. This results in existing stereoscopic VST-HMDs either having untethered portability, but limited display and rendering performance (e.g narrower FoV, lower resolution, or frame rate) [164], or providing high display performance (e.g high resolution and frame rate with wide FoV) [230], but with restricted flexibility due to the tethered experience.

Sensor fusion is an important approach to balancing such trade-offs by merging essential external sensory information, which can not be retrieved or stored in the existing device. For example, by fusing geo-referenced 3D models of the urban environment, GPS, and gyroscope data, mobile devices can have robust 3D positional tracking relative to the world [251]. By combining the sensory inputs of a stereoscopic VST camera and a thermal camera, users can develop effective thermoception through thermal visual augmentations in MR [68, 169]. The panoramic image stitching algorithm, which merges the four wide-angle fish eye camera sensors on today's standalone VR headsets, is a form of sensor fusion itself [93]. Despite the rapid development of MR hardware and software, sensor fusion will most likely remain an important field of research for enhancing existing MR devices.

6.3 System Setup

Although high-end MR headsets such as the Varjo XR-3 [230] can already provide high-resolution, wide FoV, and low-latency color VST, the rendering and display quality can only be achieved with a tethered experience, thus reducing their applicability. The goal of our custom-built VST-HMD is to improve existing commercial standalone VR headsets. Even though the display quality and computational power of MR devices will continue to grow, it is reasonable to assume that the VST quality of standalone devices will remain inferior to high-end devices tethered to graphics workstations.

Our hardware system is configured to include a small embedded single-board computer, a high-resolution stereoscopic camera, and a commercial-grade standalone VR headset. Figure 6.2 illustrates a general overview of the system.

Single-board Edge AI Computer The single-board computer is a ZED Box [209], an edge AI device based on an Nvidia TX2 NX, a high-performance embedded system with an accelerated 256-core NVIDIA Pascal GPU. It is a system frequently used in nowadays' autonomous robotics or computer vision applications that require compact low-power edge AI processing [243]. The ZED Box of our system runs on Ubuntu 18.04.6 LTS, Jetpack 4.6, and ZED SDK version 3.7.3. As the initial ZED Box does not have an integrated Wifi module, we use a Linksys WUSB6300 AC1200 dual-band wireless USB 3.0 adapter for 5G wireless data streaming. The wifi adaptor is driven by an open-source Linux driver[157] which needs to be manually built and installed into the operating system. The ZED Box is connected to a TalentCell 72W 100WH power bank. Streaming commands to the ZED Box can be sent to the ZED Box remotely via the Secure Shell Protocol (SSH).

High-Resolution Stereo Camera A ZED Mini serves as an external stereoscopic camera, which provides a high fidelity VST functionality for the VR headset. The HD720 video capture mode of the original ZED Mini is recommended for MR VST applications [208]. The HD720 model has a vertical FoV of 54° and a horizontal FoV of 85°. The stereo camera is separated by a distance of 65mm, which is the assumed average interpupillary

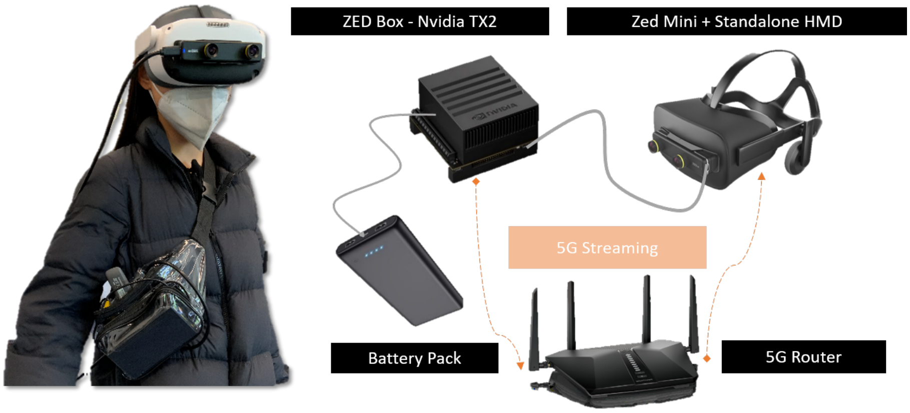


Figure 6.2: System overview of our custom-built portable stereoscopic VST-HMD, with the left image showing a user wearing our prototype. The ZED Mini camera is connected to the ZED Box via a type C USB cable. The ZED Box then encodes a pair of HD720 stereo images and streams it through a GPU-accelerated high-speed Gstreamer pipeline to the standalone HMD via a dedicated 5G router.

distance (IPD) of human eyes. The ZED Mini camera is connected to the single-board computer via a type C USB 3.0 cable.

Wireless Video Streaming The ZED Box TX2 NX hardware is compatible with the ZED Mini camera SDK, however, such an embedded device is not suitable for running VR applications. Therefore, the ZED Box is fully dedicated to high-speed video encoding and streaming. The VR application is then directly built into the standalone VR headset. We use an open-source multi-media streaming framework, Gstreamer version 1.14.5 [81], and the ZED Gstreamer plugin [210] for processing and encoding the video stream. The video stream is sent to the VR device through the user datagram protocol (UDP) via a TP-Link Archer MR600 AC1200 5G Router. We use the Nvidia accelerated Gstreamer Deepstream SDK [162], which drastically decreases the video encoding speed of a stereo image pair with a resolution of 2560×720 from 33ms per frame to only 6ms per frame.

VR headset Since we are interested in investigating the foveated MR tunneling effect as well, we choose Pico Neo 3 Pro Eye [228], a standalone VR headset with an integrated eye tracker. The headset has 1832×1920 per eye resolution, 98° horizontal FoV, and 90° vertical FoV. The ZED Mini camera is permanently mounted on the VR headset via extra-strong double-sided tapes. The movement tracking is turned off to enable users to freely travel without being bound to a small tracking space.

Software Integration The application that receives and displays video streams on the VR headset is developed using the Unity game engine version 2020.3.16f1, Pico Unity Plugin version 2.0.4, and Unity XR interaction toolkit version 2.0.1. However, as the ZED Unity SDK does not support standalone Android devices, we integrate a custom Gstreamer-Unity plugin for Android devices to receive the stereo video streams from the ZED Box. The plugin is based on the Android branch of an open-source Gstreamer-Unity package [188]. However, the package does not include build support for the Pico VR

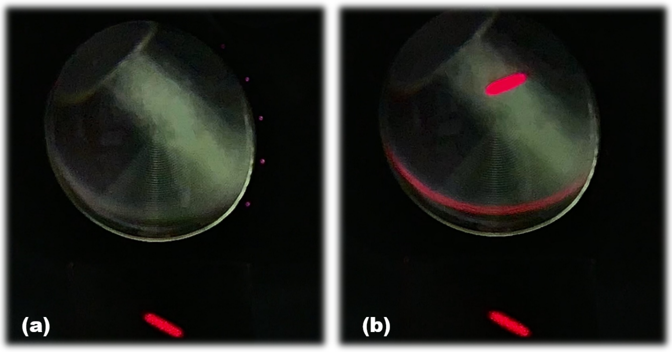


Figure 6.3: Motion to photon latency measurement using a laser pointer and a slow motion camera, with (a) showing the initial appearance of the laser pointer at frame f_0, and (b) showing the first appearance of the laser pointer seen by the VST-HMD at frame f_1. The number of elapsed frames C can be calculated via $C = f_1 - f_0$.

headset which has an arm64 architecture. We rebuild the plugin, targeting the correct architecture using the Android NDK tool. The plugin has a dependency on the Gstreamer Android package version 1.14.5. The custom Gstreamer-Unity plugin then allows dynamic access to Gstreamer-android application program interfaces (APIs) at Unity run time. The stereo camera video feed is rendered on two separate image planes with the separation of the assumed average human IPD of 65 mm.

Video Streaming Performance The stereo image pair is combined into a single 1280×1440 image streamed via a single Gstreamer UDP port, and rendered simultaneously in a single Unity update frame to ensure that the images for both eyes update at the same time. Our system can run on many commercial standalone VR headsets, including the Oculus Quest 2. However, as the Pico Neo 3 Pro Eye headset has 8GB RAM in comparison to the 6GB RAM on Quest2, the Pico headset yields higher performance in framerate and latency. The system with the Pico headset integration can run smoothly at 30 fps, with a peak framerate of 39 fps. The average estimated MTP latency is $153.47 \pm 33.33ms$.

Motion to Photon Latency Estimation The MTP latency in MR is defined as the delay of display update of the HMD when the user moves their head [168, 169]. For our system, the MTP latency is the end-to-end streaming latency from the moment the video is captured to the moment the video is rendered on the HMD. We measure the MTP latency with a 240 fps slow-motion camera on an iPad Pro. Both the slow-motion camera and the ZED Mini camera are focused on a piece of black paper. A laser pointer illuminates the black paper. As shown in Figure 6.3, due to the end-to-end latency, the occurrence of the laser pointer will appear several frames later in the HMD. We then can record the elapsed frames ($C = f_1 - f_0$), and compute the MTP latency $t_{lag} = \frac{1}{240fps} \times C \pm \frac{1}{30fps} ms$, where C is the difference measured in number of frames, and 30 fps is the fixed video streaming frame rate.

Manual Hand-Eye Calibration To display correct stereoscopic images for both human eyes, we performed an intrinsic and extrinsic calibration process [19]. The intrinsic calibration corrects for aberration caused by lens distortion by estimating the radial distortion coefficients, and the extrinsic calibration determines the global camera position in the world [61]. When using the ZED Gstreamer plugin, the pre-calculated intrinsic parameters are automatically included in the Gstreamer pipeline [210]. However, for an HMD to best facilitate hand-eye coordination tasks [173], it is also important to perform an extrinsic calibration that determines the transformation of the external stereo camera relative to the cameras on the VR headset. In a multi-view computer vision system, the extrinsic calibration is typically performed using checkerboard patterns where each camera needs to accurately locate the local position of the corners of the checkerboard. As the resolution of the cameras integrated into the VR headsets are not high enough to accurately detect the checkerboard patterns, we adopt an iterative manual hand-eye calibration process. Using the grayscale VST as background, we develop a program to incrementally move the image plane of the external ZED Mini display along the x, y, and z-axis. We iterate moving the axes until the color VST perceptually aligns with the background VST.

Economic Comparison Compared to existing commercial devices, our framework offers a scaleable intermediate VST-HMD solution in terms of costs and performance. While existing VR headsets such as the Oculus Quest 2 have low cost, the gray-scale video feeds offers limited capabilities. Compared to tethered high-end MR headsets, although our system could have a lower framerate and higher latency, users could be in favor of our system as it has lower hardware costs and offers untethered experiences. Moreover, the performance of our framework can be improved by investing in more powerful embedded devices, such as the Jetson Xavier by Nvidia [1].

6.3.1 MR Tunneling

Time Warp for Latency Compensation and Video Stabilization As the MTP latency of the ZED Mini camera is significantly larger than that of the VR headset, this can lead to unstable VST displays of the ZED Mini and can cause potential sensory conflicts [167]. Previous work in fusing multiple VST camera systems commonly uses the time warp technique to mitigate such registration errors [169]. In our system, we integrate a similar technique, which records the timestamps and the corresponding VR headset's transforms at the time of the video's capture and delays the update of the image plane transforms with the relative average MTP latency. As our system functions at a relatively low frame rate of 30 fps, we also apply a linear interpolation to estimate the transforms of the image planes if a corresponding timestamp could not be found.

Heterogeneous Image Fusion To merge the ZED Mini camera video feed with a narrower FoV into the grayscale VST with a wider FoV, we adjust the transparency of the ZED Mini video feed through a custom Unity shader. We divide the entire FoV in the VR headset into three regions according to the characteristics of the human visual

[1] https://www.nvidia.com/en-gb/autonomous-machines/embedded-systems/jetson-xavier-series/

system [22]. The color VST is displayed without any transparency reduction from the central fovea region to the para-peripheral region, corresponding to 60° horizontal FoV [37]. This then includes the para-fovea and the central-peripheral region of the human vision, which users rely on to perform visual searches in complex tasks [55]. From the mid-peripheral region at 60° horizontal FoV to the far peripheral region at 98° horizontal FoV, the color video gradually merges into the grayscale VST background with linearly reduced transparency. In the far peripheral region, only the grayscale low-resolution VST is displayed. As the merging of two VST video feeds results in applying an FoV restrictor on the high latency and high-resolution ZED Mini VST, it creates a ”tunneling effect", as Figure 6.4 (C3) shows. One problem when fusing the grayscale VST from the Pico headset remains in the image distortion effect that occurs at the upper left and right corners of the screen, as shown in all sub-figures 6.4 that include the grayscale VST. However, in our study users do not depend on their upper far peripheral vision to perform the tasks.

Head Movement Accommodated MR Tunneling Recent research established on the sensory conflict theory suggests that cybersickness when using HMDs is caused by differences in the user's virtual and physical head pose [171]. When using a high latency VST-HMD system, such head pose mismatch dramatically increases when users rapidly turn their heads. Therefore, we propose an extension to the MR tunneling effect which rapidly fades away the high latency VST, and fades back when the head movement speed reduces to below $45°/s$. This detection threshold is determined through a pilot test, where 4 users tested our prototype and selected their optimal value. The user's head movement speed is tracked using the physics tracker from "Unity Super Science", an open-source project that collects essential extensions to the Unity game engine [223]. The high latency VST image will fully re-appear to the user's central FoV after $150ms$, which is around its average MTP latency. Figure 6.4 (C4) demonstrates the effect when a user is turning their head, and the center fades away.

Foveated MR Tunneling Visual search is one of the most important activities when conducting complex real-world tasks. It is facilitated by a combination of head movement, which is responsible for visual scanning of a larger area, and eye movement, which scans within the user's FoV [235]. As the MR tunneling effect applies an FoV restrictor to the central high-resolution VST, it narrows the user's color FoV and could degrade users' task performance. The foveated MR tunneling effect can enable users to scan the environment within their FoV through eye movement. In our prototype, it is achieved by obtaining the users' eye gaze position from the VR headset's eye tracker and dynamically moving the center of the color VST to the users' eye fixation position. As abrupt, rapid eye gaze transition can cause viewing discomfort, we apply a linear interpolation between each detected gaze position to smoothen the effects of the high-frequency position change. The transition is interpolated over 100 ms, an optimal value determined by the four pilot test users. Figure 6.4 (C5) demonstrates the foveated MR tunneling effect where the center of the FoV restrictor is set to the fixation point of the user's tracked eye gaze.

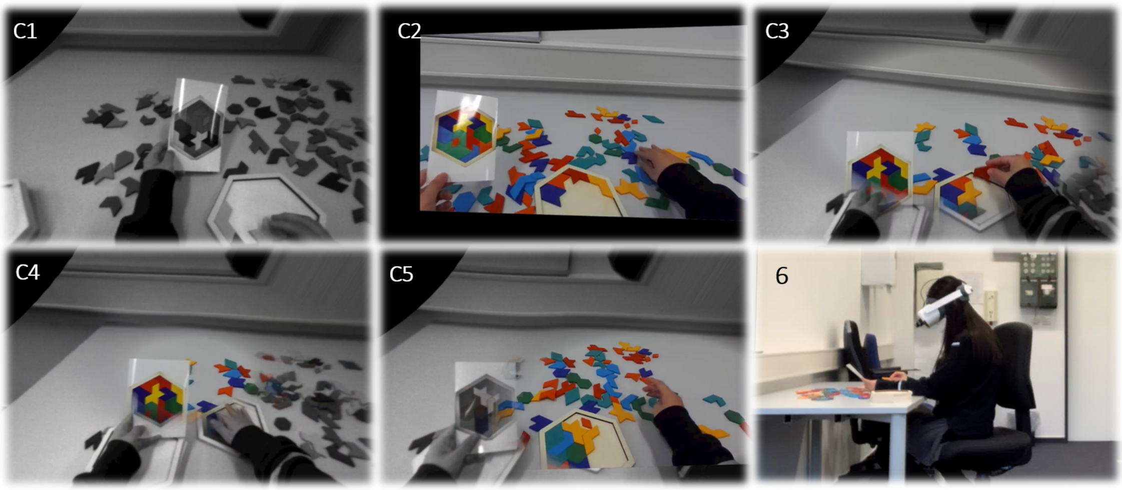


Figure 6.4: Screenshots of different VST effects for Condition C1-C5. (C1). only the grayscale VST from the VR headset; (C2). only the color VST from ZED Mini with a time warp effect for latency compensation; (C3). the basic MR tunneling effect merging C2 and C1; (C4). the high latency VST rapidly fades away upon detection of a sudden and fast head movement; (C5). the foveated MR tunneling effect by changing the center of the FoV restrictor to the user's gaze fixation point; and (6). a user performing the assembling tasks.

6.4 User Study

To evaluate the previously introduced MR tunneling methods, we conducted a user study with the setup as described in Section 7.4.

In the study, users need to solve puzzle assembling tasks that involve hand-eye coordination, untethered locomotion, and fine motor skills. We collect participants' objective task performance as well as their quantitative and qualitative subjective feedback under six conditions:

(C0) Natural vision

(C1) Grayscale VST from the VR headset

(C2) High-resolution color VST from the external stereo camera

(C3) MR tunneling effect (combining C1 and C2)

(C4) MR tunneling effect + fading during rapid head movements

(C5) Foveated MR tunneling effect + fading during rapid head movements

All conditions C2-C5 integrate a basic time warp effect described in Section 6.3.1 for MTP latency compensation.

6.4.1 Task

De Pace et al. [57] identified maintenance-assembly-repair as one of the five major application areas for AR in the industry domain, emphasizing the potential to augment the real-world environment with instructions in the form of audio tracks, and animated 3D models, or text labels. For testing our MR tunneling methods, we aimed to design an abstracted task that resembles such applications, thus requiring a combination of motor and cognitive skills. Tangram puzzles satisfy these conditions without requiring domain-specific knowledge or skills, which would complicate a study in a university context and reduce the generalizability of the results. Moreover, Tangram puzzles have been used for experimental research [239]. Tangrams consist of several polygon-shaped pieces that have to be arranged in such a way that they replicate a given pattern. While Tangram puzzles typically require some degree of creativity, we provided study participants with instructions showing the target pattern to resemble common assembly tasks more closely. To ensure a consistent level of difficulty between conditions, we selected a Tangram that requires users to place the same 16 puzzle pieces in an overall hexagonal shape, varying only the exact placement of the pieces within that shape. Similar to real assembly tasks, both the shape and the color of individual pieces can be used to solve the puzzle. This imposes a particular challenge in the black-and-white VST condition, since both the instructions and the puzzle pieces appear in grayscale, however, these conditions intentionally simulate realistic situations. Since we were mainly interested in the effects of the different VST techniques, we omitted all virtual content in this study to avoid the risk of introducing additional confounding factors such as the type, size, or positioning of the virtual elements. In addition, locomotion is integrated into the tasks to simulate the realistic and holistic usage of an untethered HMD, where moving in the environment is important at places such as industrial plants [57] and laser laboratories [133].

6.4.2 Measures

To obtain a comprehensive overview of the experience while using the techniques introduced in Section 7.4, we employed several subjective and objective measures.

Cybersickness As a subjective measure for cybersickness, we asked participants to rate the perceived severity of 16 symptoms that constitute the Simulator Sickness Questionnaire (SSQ) [106] on a 4-point Likert scale (ranging from *0 - none* to *3 - severe*). As suggested by Kennedy et al., results were grouped into nausea, oculomotor, and disorientation sub-scores as well as a total score. A copy of the SSQ can be found in the Appendix A.1.1

Task Performance For the task described in Section 6.4.1, we counted the number of correctly placed puzzle pieces, where a roughly correct positioning and orientation, rather than a perfect alignment, was considered correct. Each pattern consisted of a total of 16 pieces, and participants were instructed to tackle as many patterns as possible in the allotted time. Since participants had different amounts of time to complete the task depending on how quickly they walked from the starting point to the puzzle table, we divided the number of correctly placed puzzle pieces by the number of time participants sat at the table, yielding the objective measure of task performance "puzzle pieces per minute." In addition, we also compared the required walking time.

Perceived Workload The NASA TLX questionnaire [88] was used to measure various aspects of workload, including mental, physical, and temporal demand, as well as performance, effort, and frustration. The original scale, which ranges from 1 to 100 in increments of 5, was converted to an 11-point Likert scale with labels *0 - perfect / 100 - failure* for performance and *0 - very low / 100 - very high* for all other items. A copy of the NASA-TLX questionnaire can be found in Appendix A.1.2.

Presence As introduced in Section 3.4.2, sense of presence is typically measured in the context of experiencing virtual environments, with high scores indicating that users have the illusion of "being there" [199]. For our VST-HMD, we were interested in the reverse case, i.e., whether users feel physically present in the real environment despite their view being altered (in terms of field of view, color, latency, etc.). Therefore, we used an adapted version of the Spatial Presence sub-scale of the Presence Questionnaire (IPQ) [195], which consists of 5 items, each measured on a 7-point Likert scale. For all items, the word "virtual" was replaced with "real". Presence was measured only for the five conditions involving a VST display since a maximum score of 7 can be assumed for the natural condition. A copy of the original and the adapted IPQ can be found in Appendix A.1.3 and A.1.4.

Usability After each condition, participants were asked to answer the 10 questions of the System Usability Scale (SUS) [33], providing ratings between 1 (*strongly disagree*) and 5 (*strongly agree*), where higher values correspond to better usability. A copy of the SUS can be found in Appendix A.1.5.

Preference In a final questionnaire, participants were asked to describe their favorite and least favorite VST conditions.

6.4.3 Hypotheses

Concerning the previously described measures, we formulated the following hypotheses:

(H1) Natural vision is rated highest in all measures.

(H2) All MR tunneling effects (C3, C4, C5) reduce cybersickness compared to the color display without MR tunneling (C2), with greater reduction achieved when the colored center is faded during movements (C4 and C5) than when the colored center is continuously overlaid (C3).

(H3) The combination of a colored center and a grayscale periphery (C3, C4, C5) results in higher task performance and lower mental workload than only a colored center (C2), which in turn performs better than only a grayscale image (C1).

(H4) Techniques with peripheral vision yield higher presence ratings than those without peripheral vision.

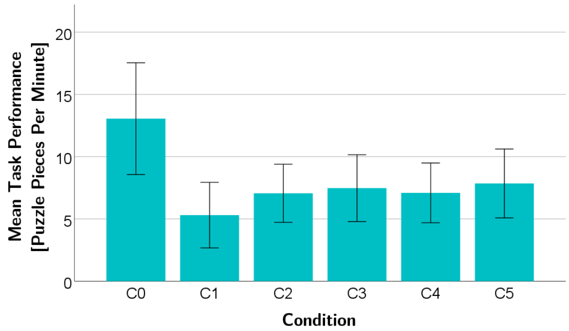
6.4.4 Procedure

Prior the completing the study, each participant gave informed consent and filled out a demographic questionnaire. Then, the experimenter demonstrated the puzzle task to the participants, with a Tangram different from the one used in the conditions. Participants were instructed to try to correctly place as many pieces as possible on the puzzle board, even if a pattern could not be fully completed in time. They were also reminded not to empty the puzzle board after completion, but to take a new one instead. Participants were then led through the lab space to show where the instruction cards were kept and where they would sit to assemble the puzzle. Each participant was randomly assigned to one of six different condition orders determined using a balanced Latin square. Puzzle patterns were assigned randomly but were not repeated.

In all conditions, including the natural vision condition C0, participants were required to wear the VST-HMD. For C0, the HMD was placed on the participants' foreheads to record their movements and provide auditory cues. This also allowed isolating the effects of the different visual presentations from other factors, such as the weight of the HMD. At the beginning of each condition, participants were guided to the entrance of the experiment room, without displaying an image on the VST-HMD. After confirmation by the experimenter, a see-through image was displayed in the headset (for C1-C5). Participants then had 10 seconds to familiarize themselves with the display and the environment. After 10 seconds, an audio cue indicated the start of the study, which was limited to 3 minutes per condition. During this time limit, participants first had to collect the puzzle instructions before sitting down at the puzzle table and assembling as many puzzles as possible. After 3 minutes, an audio cue indicated the end of the condition. Participants removed the headset and filled out a set of questionnaires as described in Section 6.4.2. Meanwhile, the experimenter took photos of the complete and incomplete puzzle boards and shuffled the puzzle pieces in preparation for the next condition. Participants were allowed to take a break (at least 1 minute after completing the questionnaires) and drink water before starting the next condition. Completing all 6 conditions required an average time of 50 minutes.

6.4.5 Participants

A prior G-Power analysis indicated a minimum sample size of $N = 23$ to detect medium effects ($\eta_p^2 = .006$) with 90% power using an ANOVA with an alpha level of 0.05. We invited 24 participants, 12 female and 12 male (aged between 19 and 45, $M = 26.9$). All of them were students or researchers of HCI or Computer Science, with HCI students receiving compensation in the form of course credit. 10 participants use VR systems regularly (at least once a month), and only 5 use them less frequently than once a year. The complete data set of two participants had to be excluded from the analysis because one did not understand the puzzle task correctly and for another, the eye-tracking was disturbed by the user's glasses. A screening question was asked for each participant prior



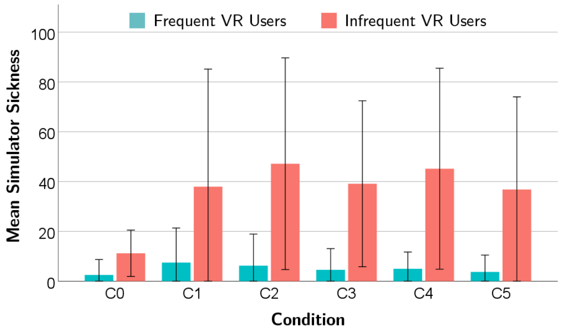


Figure 6.5: Leaft: mean task performance and Right: mean cybersickness per condition. Vertical bars represent the standard deviation.

	Task Performance		**Walking Time**		**Presence**	
	M	**SD**	**M**	**SD**	**M**	**SD**
(C0)	13.06	1.03	12.76	3.63	-	-
(C1)	5.31	0.60	**15.92**	3.54	3.88	1.43
(C2)	7.06	0.54	19.84	6.02	3.66	1.48
(C3)	7.48	0.62	18.60	3.77	3.81	1.27
(C4)	7.09	0.55	18.11	4.01	3.93	1.24
(C5)	**7.85**	0.64	16.84	3.58	**4.02**	1.40

Table 6.1: Means and standard deviations for spatial presence, task performance, and walking time with bold numbers indicate the best performance.

	Cybersickness		**Nausea**		**Oculomotor**		**Disorientation**	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
(C0)	7.65	9.12	5.64	10.05	7.92	9.20	5.69	9.27
(C1)	25.50	39.78	16.48	24.31	22.74	31.82	29.11	58.57
(C2)	30.43	38.97	16.91	21.62	27.22	35.30	38.60	53.99
(C3)	24.99	31.06	15.18	17.08	24.81	30.06	**25.31**	39.29
(C4)	28.73	36.83	17.78	24.33	28.25	31.57	29.11	46.84
(C5)	**23.29**	32.95	**13.44**	19.45	**21.36**	27.74	27.84	47.84

Table 6.2: Means and standard deviations for cybersickness and all su bscales of the SSQ, with bold numbers indicate the best performance.

to the study to ensure that all participants had a normal or corrected-to-normal vision.

6.5 Results

Cybersickness SSQ was measured not only after each condition but also before the first condition, which allowed for analysis of carryover effects. However, no increase in cybersickness was observed throughout the study, which can be attributed to the experimental protocol that required participants to take breaks between conditions until potential cybersickness symptoms subsided. We, therefore, considered the absolute values of the SSQ instead of the differences between each measurement and the previous one as the dependent variable. Since residuals were not normally distributed, we performed Friedman tests on all sub-scales of the SSQ. Cybersickness significantly differed between the six conditions ($\chi^2(5) = 21.874, p = .001$). To reveal between which conditions significant differences occurred, we performed multiple Wilcoxon signed-rank tests (the reported p values are adjusted with Holm-Bonferroni). C0 lead to a significantly lower Cybersickness than all other conditions, i.e., C1 ($Z = -2.893, p = .042$), C2 ($Z = -3.202, p = .018$), C3 ($Z = -3.212, p = .018$), C4 ($Z = -3.327, p = .013$), and C5 ($Z = -3.115, p = .022$). Cybersickness showed a strong, negative correlation to VR experience (Pearson's $r = -.611, p = .003$). Therefore, we grouped participants into frequent VR users (experiencing VR systems at least once a month) and infrequent VR users. The resulting mean cybersickness scores per condition are plotted in Figure 6.5.

Usability Usability was only measured for the five VST conditions (C1 to C5). A Friedman test did not reveal any significant differences ($\chi^2(4) = 2.162, p = .706$).

Spatial Presence As for usability, spatial presence was not measured for the natural vision condition C0, since a maximum value was assumed. A Friedman test for the remaining five conditions did not result in any significant differences ($\chi^2(4) = 1.337, p = .855$).

Workload We performed Friedman tests for all sub-scales of the NASA-TLX questionnaire, followed by Wilcoxon signed-rank tests with Holm-Bonferroni adjustment in case the former indicated a significant difference.

The viewing conditions had a significant effect on mental demand ($\chi^2(5) = 35.224, p < .001$). Wilcoxon pairwise comparisons showed that C0 yielded a significantly lower mental demand than C1 ($Z = -3.751, p = .003$), C4 ($Z = -3.456, p = .007$), and C5 ($Z = -3.559, p = .005$). Two Friedman tests also suggested significant effects on physical demand ($\chi^2(5) = 14.514, p = .013$) and performance ($\chi^2(5) = 12.640, p = .027$), however, none of the pairwise post-hoc tests were significant after controlling for multiple testing via Holm-Bonferroni. Temporal demand did not yield a significant difference. After observing a significant effect on effort ($\chi^2(5) = 17.816, p = .003$), Wilcoxon post-hoc tests revealed a significant difference between C0 and C1 ($Z = -3.112, p = .028$). Finally, the level of frustration was significantly different between conditions ($\chi^2(5) = 24.226, p < .001$). Post-hoc tests showed that C0 resulted in a significantly lower frustration than C1 ($Z = -3.517, p = .007$), C2 ($Z = -3.223, p = .018$), and C4 ($Z = -3.064, p = .028$).

Task Performance Task performance, computed as *puzzle pieces per minute*, was compared between all six conditions using a repeated-measures ANOVA. The data of 3 participants had to be excluded because the walking time was not measured for the natural condition. An ANOVA on the remaining 19 participants revealed a significant difference ($F(5, 90) = 26.851, p < .001, \eta_p^2 = .599$). Post-hoc tests with Holm-Bonferroni adjustment revealed that participants performed significantly better in C0 than in all other conditions, i.e., C1, C3, C4, and C5 ($p < .001$) as well as C2 ($p = .001$). In addition C1 yielded significantly worse results than C3 ($p = .025$), C4 ($p = .018$), and C5 ($p = .024$). In C1, two participants accidentally dropped all the instruction cards due to the low visibility of the environment. In C2, one participant accidentally bumped into the table due to the narrow vertical field of view.

Walking Time For the same 19 participants as in the previous section, we performed a repeated-measures ANOVA on walking time, resulting in a significant difference between conditions ($F(5, 90) = 11.758, p < .001, \eta_p^2 = .395$). According to post-hoc tests with Holm-Bonferroni adjustment, walking time was significantly lower in C0 than in all other conditions, i.e., C1 ($p = .016$), C2 and C3 ($p = .001$), C4 ($p = .002$), as well as C5 ($p = .003$). Furthermore, C1 resulted in significantly lower walking times than C2 ($p = .025$) and C3 ($p = .026$).

Preferences The least preferred conditions were C1 ($N = 9$) and C4 ($N = 5$). For C1, besides the obvious disadvantage of not having color to solve the puzzle task ($N = 6$), low image quality/blurriness was mentioned by five participants, and two participants reported the resulting difficulty in seeing the boundaries between the puzzle pieces. One participant reported a higher level of fatigue, dizziness, or eye strain due to the additional attention required. Surprisingly, two participants stated that they found the absence of colors beneficial, as they found it easier to focus only on the shape without the ambiguity caused by similarly colored pieces.

Negative evaluations of C4 were justified by the need to wait for the color to reappear, which interrupted the flow of participants' task execution ($N = 4$). Three participants suggested making the corresponding animation faster. The alternation between appearance and disappearance as well as the need to refocus on the colored center after each change was also described as straining to the eyes ($N = 3$), distracting ($N = 2$), and jarring ($N = 1$), making it difficult to concentrate ($N = 1$).

C5 was most often selected as the preferred VST condition. Participants acknowledged that they did not have to move their heads as much as in other conditions ($N = 2$) and always had the color information available when they needed it. The combination of large FoV and color vision was emphasized, with the eye-tracking condition being perceived as the MR tunneling method with the lowest latency or shift between center and periphery ($N = 2$). As can be seen in Figure 6.6, three participants selected C5 as the least preferred. Two participants elaborated that they disliked that the colored region did not exhaust the maximum possible area. As with C4, the (dis)appearance of the colored region was also mentioned negatively (N=5), with one participant stating that it was more difficult to control than in C4.

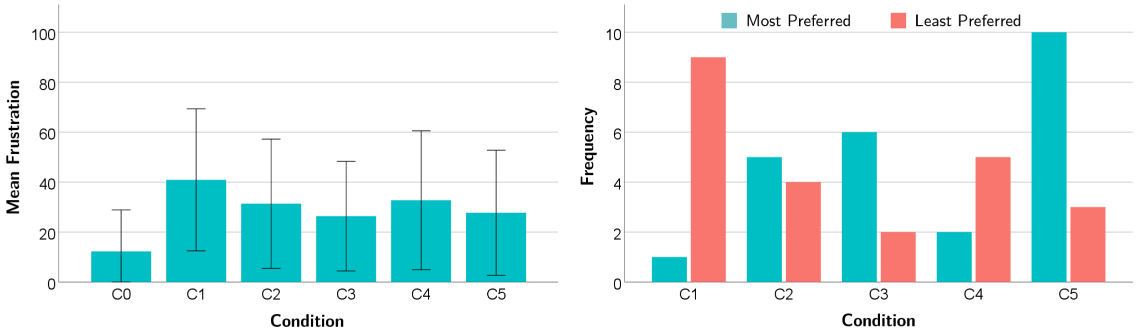


Figure 6.6: Left: Mean frustration score (vertical bars represent the standard deviation) and Right: most / least preferred condition.

	Mental Demand		Physical Demand		Temporal Demand	
	M	**SD**	**M**	**SD**	**M**	**SD**
C0)	40.45	26.09	12.27	17.16	46.82	26.62
C1)	63.64	26.82	**24.09**	27.02	50.00	28.79
C2)	52.73	25.30	28.18	27.54	49.55	25.54
C3)	**51.82**	22.81	25.91	26.31	**47.27**	24.72
C4)	55.91	24.82	24.55	25.95	49.09	29.42
C5)	55.91	25.01	25.00	24.25	47.73	27.24

Table 6.3: Means and standard deviations for Mental Demand, Physical Demand, and Temporal Demand, with bold numbers indicating the best performance.

	Performance		Effort		Frustration		Usability	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
C0)	25.45	29.56	45.45	28.41	12.27	16.60	-	-
C1)	45.00	26.86	70.45	22.57	40.91	28.44	62.16	18.90
C2)	41.82	24.42	58.18	24.62	31.36	25.87	66.82	20.49
C3)	37.73	19.98	60.91	21.36	**26.36**	21.94	**68.64**	21.11
C4)	**35.45**	23.65	**57.73**	24.68	32.73	27.81	66.48	16.01
C5)	39.55	20.58	58.64	21.22	27.73	25.06	67.05	19.89

Table 6.4: Means and standard deviations for Performance(NASA-TLX), Effort, Frustration, and usability, with bold numbers indicating the best performance.

6.6 Discussion

Superiority of natural vision (H1) As expected, natural vision outperformed all other conditions in task performance as well as cybersickness, and three of the other conditions in mental demand as well as frustration, confirming (H1). However, in the qualitative feedback, a learning effect and associated positive effects such as increasing comfort and decreasing sensitivity to lagged display over time were mentioned 9 times. Also, a subgroup analysis of frequent and infrequent VR users showed that the significant differences in natural vision versus VST conditions diminish for participants with regular VR use. This indicates that while VST still imposes limitations such as a restricted FoV, resolution, and color fidelity, users can become accustomed to it through regular use of such systems. Lastly, our study did not include overlaid virtual information as we aimed at focusing on the actual VST methods. Potential advantages that could arise for VST compared to natural vision, such as faster task completion due to superimposed instructions, were therefore not weighed in the evaluation.

Reduced cybersickness for MR tunneling (H2) No significant differences were found between the VST conditions, leading to rejection of (H2). For infrequent VR users, the differences between conditions C1 to C5 were larger but also not statistically significant. However, qualitative feedback suggests that some conditions elicited more symptoms of cybersickness than others. For example, for C2, three comments described that head movement was exhausting or even caused headaches. Two participants reported moving their heads more slowly than in the other conditions, and one participant even closed their eyes during head movements to avoid cybersickness. Whether these reported symptoms are individually different or reflect a general trend would need to be investigated in a future study with a larger sample of infrequent VR users ($N = 137$ for the same values as assumed in our study, but a small effect size).

Improved task performance for MR tunneling (H3) We hypothesized that performance in the puzzle task would be enhanced not only by the addition of color information but also by the fusion of a colored center (either in the geometric center of the visual field or in the region focused by the eyes) with a grayscale periphery. While the first part of the hypothesis could be confirmed, we could not find evidence for the latter part. This may be partly due to the assembly task, which can be approached differently. Diverse strategies were reported in the qualitative feedback, such as moving all puzzle pieces to the center of the table at the beginning of the task or placing the puzzle board further away so that it is completely covered by the video stream of the color camera. Moreover, contrary opinions were expressed about the advantages and disadvantages of the respective methods, which is reflected in Figure 6.6. In general, participants' preferences diverged, suggesting that for different users, customized methods including individually adapted values (e.g., fade times or the radius of the colored area) could be offered.

Higher spatial presence for peripheral vision (H4) Ratings for spatial presence were similar for all VST conditions, therefore not confirming (H4). Future studies would need to investigate whether these changes in tasks involve spatial orientation and/or multiple users, as these could be factors that make peripheral vision particularly useful. In our study, the walking phase was quite short (17.01 seconds on average), and the main part of the study was conducted in a seated position with limited interaction space in front of the participants. Three participants specifically noted that they would prefer VST without grayscale periphery for tasks that do not require a large FoV, and MR tunneling for other tasks. This suggests that the choice of the best VST technique depends not only on the individual user's preferences but also on the task and/or environment.

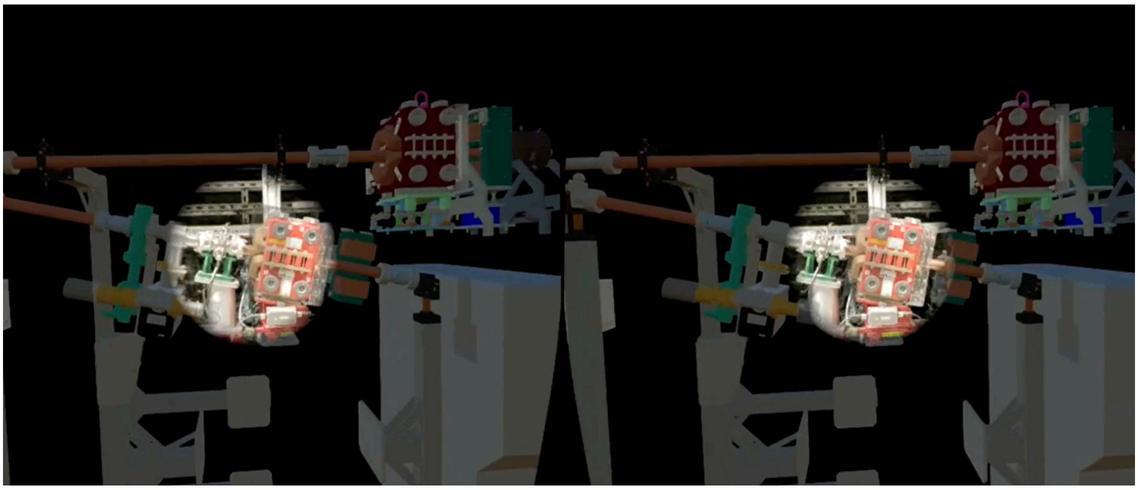
6.7 Conclusion

In this chapter, we presented MR tunneling and its two extensions for extending the VST functionality of a standalone VR headset through sensor fusion. The user study results demonstrate that overall, foveated MR tunneling was the most preferred VST technique. However, the non-foveated tunneling effect demonstrates the highest overall usability with an above-average SUS score. Based on participants' qualitative feedback, the foveated tunneling effect can be further optimized by making the head speed accommodation effect optional and providing an FoV customization for users who desires a larger central color region.

The techniques presented in this chapter are not limited to current standalone VR headsets. With improved headset capabilities, such as built-in high-resolution color VST, there will also be increasingly computation-intensive tasks such as real-time computer vision applications for VST-HMDs. With the portability-performance trade-off of VST-HMDs, future headsets will continue to demand wireless streaming of computationally heavy tasks to more powerful workstations in order to keep the untethered experiences for users. Therefore, the fusion of a low-latency video stream from the built-in cameras and delayed visual information from an external device will remain relevant for upcoming headset generations.

7

Magic NeRF Lens



In this chapter, we present *Magic NeRF Lens*, a VR framework that supports immersive photorealistic visualizations of complex industrial facilities leveraging the recent advancement of NeRF. This chapter aims to answer the following research questions:

- **RQ3**: *How to develop a user-friendly toolkit for visualizing and interacting with real-world NeRF models in immersive MR, particularly for virtual inspection of complex physics facilities?*
- **RQ4**: *How well can such a neural rendering system perform for real-world immersive MR applications? What are users' perceptions of the overall usability, spatial presence, and their task performance during virtual facility inspection tasks?*

The content of this chapter is primarily based on the following publication:

- **Ke Li**, Susanne Schmidt, Tim Rolff, Reinhard Bacher, Wim Leemans, Frank Steinicke. (2024). Magic NeRF Lens: Interactive Fusion of Neural Radiance Fields for Virtual Facility Inspection. Frontiers in Virtual Reality. [DOI: https://doi.org/10.3389/frvir.2024.1377245]

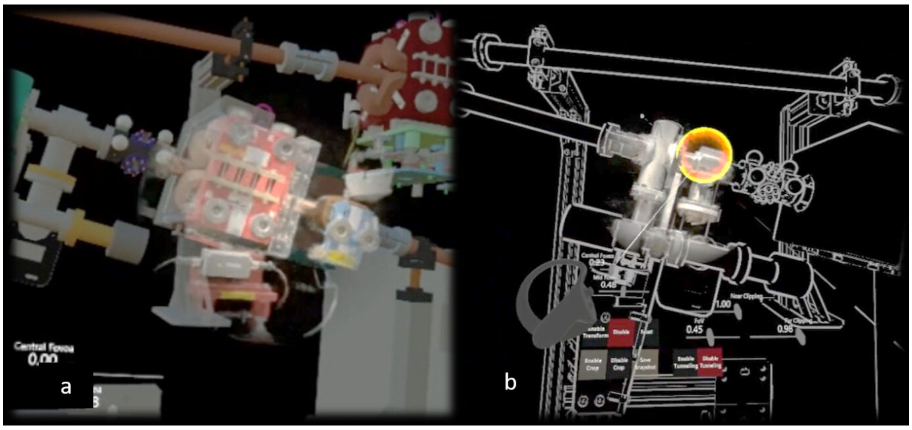


Figure 7.1: Screenshots from the Magic NeRF Lens framework illustrating (a) the data fusion of a high-resolution NeRF rendering and the polygonal representation of its CAD model, and (b) the 3D NeRF drawing interaction using the polygonal CAD representation as context.

7.1 Motivation

As introduced in Chapter 3, the possibility for users to spatially interact with industrial designs and thoroughly inspect these facilities in a telepresence environment can significantly improve productivity, task engagement, and workflows [39]. Moreover, at critical infrastructures such as particle accelerators and nuclear power plants, where human onsite visits to the facilities are limited due to safety hazards and operation constraints [42, 58], VR visualization systems that accurately represent the complex facility conditions and fully immersive users in the inaccessible remote environments is crucial for online facility inspection and maintenance planning. As mentioned in Chapter 1, existing industrial VR systems typically visualize the virtual facilities using a polygonal representation of their CAD models [182] or meshes reconstructed from photogrammetry [192], RGBD cameras [252], or LiDAR measurements [79]. These 3D representations often provide only limited realism in modeling complex geometries of real-world conditions, introducing incompleteness or inaccuracy in their visual appearances and restricting their potential to support real-world tasks such as detailed facility inspection, quality control, and maintenance planning [54].

As introduced in Chapter 4, NeRF [154] has recently emerged as a novel volumetric 3D representation that can replicate and store the intricate details of our complex realities through training a Fourier feature neural network [219] with a set of sparse 2D images and their camera poses as input. As NeRF can generate highly realistic 3D reconstructions with a relatively small amount of input data, it offers a new approach to establishing visualization frameworks for virtual facility inspection. However, developing interactive VR experiences for industrial NeRF models faces several challenges. Unlike traditional render pipelines based on the rasterization of geometric primitives, a NeRF model needs to be rendered through an expensive volumetric ray casting process which includes a large number of feed-forward queries to the neural network. As a result, the render performance scales largely with resolution and the size of the rendered volume. Due to the substantial volume of real-world industrial facilities, the temporal complexity of NeRF rendering hinders its implementation in real-time, stereoscopic, high-resolution, high-frame-rate industrial VR applications. Furthermore, although there are vast amount of recent work on NeRF [158, 25], prior efforts have mainly concentrated on demonstrating proof-of-concept experiment results, rather than delivering a user-friendly toolkit for visualizing and interacting with real-world NeRF models in immersive VR environments.

As a first step towards adapting VR NeRF to real-world industrial processes and applications, we present *Magic NeRF Lens*, a visualization toolkit for virtual facility inspection leveraging NeRF as photorealistic representations of complex real-world environments. Our framework provides a native render plugin that enables interoperability between a low-level network inference engine [158] and a high-level game engine [130], making customized VR NeRF experience creation possible through typical VR application development workflows using the Unity game engine. To reduce the temporal complexity of VR NeRF rendering, we introduce a data fusion technique that combines the complementary strengths of volumetric rendering and geometric rasterization. As Figure 8.1(a) illustrates, the photorealistic rendering of a NeRF model is merged with the polygonal representation of its corresponding CAD models, creating a 3D-magic-lens-style visualization [231] and achieving render volume reduction without sacrificing users' immersion and presence in the VR environments. As Figure 8.1(b) illustrates, such data fusion also can be realized through dynamic editing of the NeRF model. Users can flexibly select target render volume through an intuitive 3D drawing interaction by dynamically revealing or concealing a portion of the NeRF model using the polygonal representation of the industrial CAD model as context.

We systematically benchmarked our framework using real-world industrial data, highlighting the advantages of the proposed magic lens interactions in optimizing NeRF rendering within an integrated VR application. Furthermore, we explored users' perceptions of the magic-lens-style visualizations for visual search inspection tasks, thoroughly analyzing their impact on system usability, task performance, and spatial presence to derive design insights. Additionally, confirmatory expert reviews were conducted with five facility control and management specialists at an industrial facility, demonstrating the framework's benefits in supporting real-world virtual facility inspection processes, including facility redesign and maintenance planning.

In summary, the contribution of this chapter includes:

1. The development and open-source implementation of an immersive visualization toolkit to support general virtual facility inspection tasks leveraging photorealistic NeRF rendering.

2. The design and implementation of two magic-lens-style interaction and visualization methods for optimizing VR NeRF rendering through data fusion.

3. Systematic evaluation of the performance of our framework and visualization design through technical benchmarking, user study, and expert reviews.

For a better understanding of our methods, interaction designs, and user evaluation, we strongly encourage readers to refer to our supplementary video: https://youtu.be/2U4X-EaSds0. The source code of the framework is made publicly available here: https://github.com/uhhhci/immersive-ngp.

7.2 Related Work

7.2.1 Magic Lens Techniques

Magic lens techniques were first developed by [29], allowing users to change the visual appearance of a user-defined area of the UI by overlaying a transparent lens over the render target. Interactive magic lenses are widely used in modern visualization systems [229], where context-aware rendering of large information spaces is needed to save computational costs [234, 231]. In 3D computer graphics, several 3D magic lens effects have been developed to allocate computational resources to more resolution-important features for the visualization of volumetric medical scans [234, 231] or context-aware AR applications [35, 23]. Magic lens style visualization is also used in immersive VR HMD rendering to perform sensor fusion with different resolutions, frame rates, and latency [132]. Our framework adapts 3D magic lens-style interactions for photorealistic NeRF rendering in immersive VR while introducing a novel magic-lens-style visualization design for industrial facilities.

7.3 System Design

7.3.1 Design Goals

Our visualization system is designed as a step towards tackling the facility inspection and maintenance planning challenges at industrial facilities such as particle accelerator tunnels. As introduced in Chapter 2, the accelerator facilities must operate continuously for more than 5,000 hours per year, during which time on-site human access is not possible, and any unexpected interruptions to operation result in high energy and setup costs. While VR can immerse users in inaccessible remote environments to support various inspection tasks, the design of such visualization systems needs to thoroughly consider various human factors to avoid negative effects of VR usage such as motion sickness while delivering a system with high usability. Moreover, the design of our framework targets creating a more interactive visualization experience that allows users to modify and customize their virtual experiences to enhance user task engagements and provide a wider range of functionalities [16]. With considerations of interactivity, task requirements, and human factors, we formalize the following design goals (**DG**):

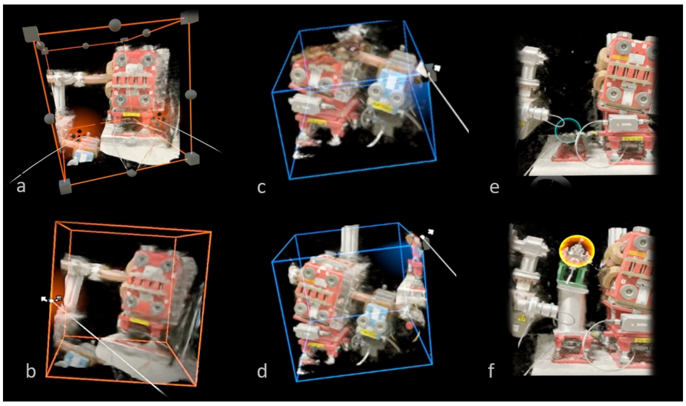


Figure 7.2: Illustration of our system extension to instant-ngp. (a-b): NeRF model manipulation, (c-d): NeRF model crop box manipulation, (e-f): Volume editing via 3D drawing.

DG1: The visualization system can render high-quality photorealistic representations of complex facility environments without introducing undesirable effects such as motion sickness or reducing users' immersion in the virtual environment.

DG2: If requested, the immersive visualization of the virtual facility can be the one-to-one real-world size of the real facility to provide a realistic scale of the industrial environments.

DG3: The visualization toolkit allows users to naturally and intuitively perform facility inspection tasks through both exploratory and interactive VR experiences.

7.3.2 Interaction Techniques

7.3.2.1 Basic Interactive Virtual Inspection

According to [16], the design of a VR application can be categorized into three different levels: passive, exploratory, and interactive. Passive VR experiences such as those with 360° videos enable observations of the remote environment but without giving users control of what they perceive. Exploratory VR experiences allow users to freely change their positions in the virtual environment, however, there are limited actions available to control and modify the virtual content. While passive and exploratory VR is useful in certain facility inspection tasks, the design of our framework employs both exploratory and interactive VR designs to improve user task engagements and provide a broader range of virtual inspection functionalities.

As Figure 7.2 (a,b) illustrates, with consideration of **DG3**, our framework integrates natural and intuitive 3D interactions with the NeRF model, allowing users to perform spatial transformation by translating, rotating, and scaling the NeRF model using VR controllers, making it possible to dynamically explore and manipulate the virtual facilities. As Figure 7.2 (c,d) illustrates, we also provide the possibility to select target render volume using a volumetric crop box. Users can manipulate the crop box by rotation, translation, and scaling using VR controllers, making it possible to precisely define the region of interest for rendering and ensuring optimal focus in the virtual environment. Figure (e-f) demonstrates an interactive volumetric editing design, where users can remove and erase a portion of the NeRF model through a responsive 3D drawing effect. Such interaction can facilitate detailed customization and fine-tuning of the virtual environments to meet specific design and simulation requirements.

7.3.2.2 Magic NeRF Lens with FoV Restrictor

While designs described in Section 7.3.2.1 enable dynamic manipulation and editing of NeRF models, several considerations from **DG1** and **DG2** are not met. For example, when inspecting facility equipment that covers a substantial area of volume, the increased latency in VR NeRF rendering on devices with limited computational resources will lead to the negative effects of motion sickness. The magic NeRF lens with FoV restrictor is designed to visualize facilities with larger volumes with considerations of various human factors (**DG1**) and inspection task requirements (**DG2**). Applying a FoV restrictor to the NeRF model creates a "VR tunneling" effect [125], which is a typical motion sickness reduction technique in VR gaming to restrict the optical flow and sensory conflicts of the peripheral region of the human eye [196]. However, a FoV restrictor could significantly reduce the user's sense of presence and immersion. To maintain rendering performance without sacrificing users' immersion in the VR environment, we propose using the "Mixed Reality (MR) tunneling" effect [132] by merging the NeRF model with the polygonal representation of the CAD model in the peripheral regions of the user's vision field. Such visual data fusion techniques can improve users' overall perception (**DG1**), while the polygonal CAD model can be used as the reference for the NeRF model to achieve a one-to-one real-world size of the physical facility (**DG2**).

Figure 7.3 (a) shows the schematic relationship between the main VR camera, the NeRF camera, and the active NeRF rendering volume for the magic NeRF lens effect with a FoV restrictor. Figure 7.3 (b) shows the volumetric crop box of the NeRF model, which is dynamically aligned with the user's viewing direction and the near clipping plane of the VR camera, so that the volume crop box acts as an interactive lens that automatically selects the target rendering volume according to the size of the image plane W and the distance of the far clipping plane L. The parameters W and L are both user-defined values to adjust the total number of sampling points for ray casting and could be adjusted during application runtime to avoid frame rate jitter. For the static MR tunneling effect, the high-resolution NeRF rendering is displayed in the central region of the HMD. However, an eye tracker could be integrated to achieve foveated MR tunneling and reduce the need for frequent head movements while inspecting the facilities [132].

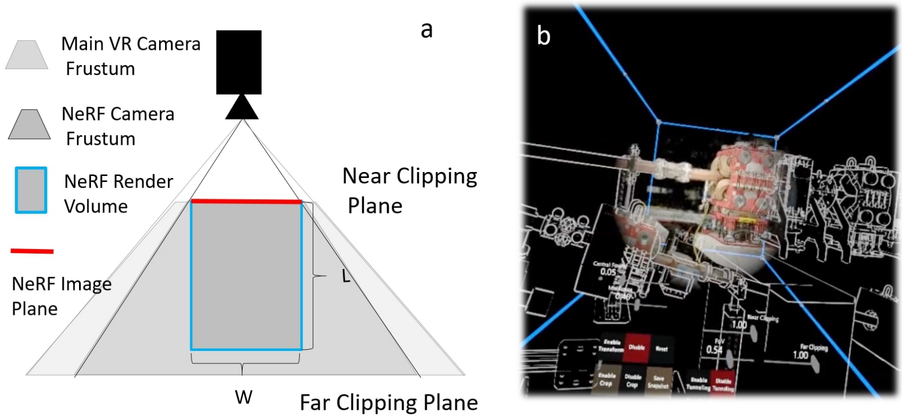


Figure 7.3: (a): Schematic sketch of the design of our interactive lens effect where the FoV of the NeRF camera is reduced and the actual NeRF render frustum is defined as a box rather than a pyramid to reduce NeRF render load. (b): Screenshot of the magic lens effect, where the blue box visualizes the NeRF crop box that is dynamically following the user's head movement.

7.3.2.3 Magic NeRF Lens with Context-aware 3D Drawing

The second magic NeRF lens effect is designed to enable more interactive visualization of the facility (**DG3**) while further optimizing VR NeRF rendering performance (**DG1**). For many facility inspection tasks, users typically do not need to see the entire NeRF model. For example, components such as walls, simple electrical boxes, floors, and ceilings are static elements that normally do not need to be inspected and modified and could be visualized simply by the polygonal representations of their CAD models. The magic NeRF lens effect with 3D drawing interaction enables users to dynamically select the target render region using the polygonal presentation of the CAD model as context. As shown in Figure 8.1 (b), an edging rendering of the polygonal CAD model visualizes the overall facility environment. Users can adjust the radius of a 3D sphere attached to the VR controller and point the 3D sphere at a spatial location where the NeRF render volume should be revealed on demand by using the edge polygonal CAD rendering as a guideline. With the empty space skipping render techniques that are implemented in most NeRF render pipelines, the network query at the location where the volume density value is zero, indicating that this space is essentially empty, could be automatically ignored for that sample location. As a result, the magic NeRF lens effect with the 3D drawing technique could potentially improve the overall rendering speed (**DG1**) while providing users with more dynamic and responsive visualization experiences (**DG3**).

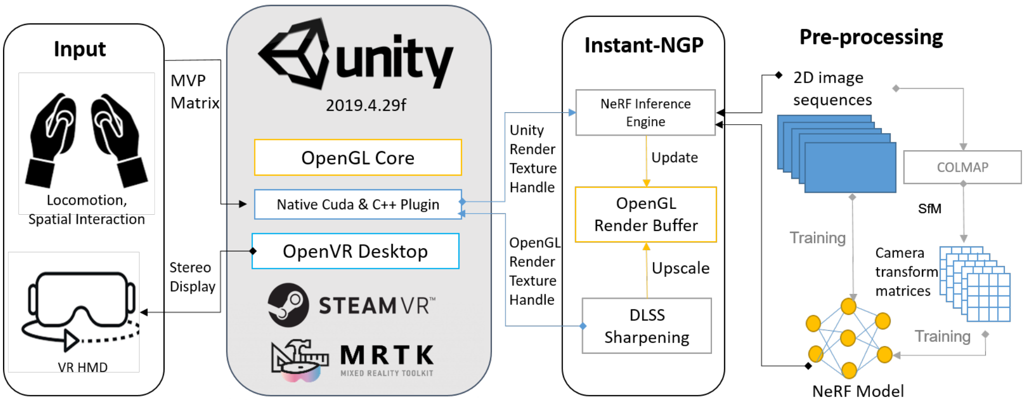


Figure 7.4: The system architecture based on the framework proposed in our arXiv preprint [130], depicting the individual processes from left to right: Starting with the input, a view model projection (VMP) matrix is computed from VR input devices and the HMD. The VMP is applied in the native Unity plugin, which provides a communication layer to the instant-ngp backend as well as the final rendering. The instant-ngp [158] backend performs the volume rendering through NeRF by updating the provided texture. The pre-processing refers to NeRF model training.

7.4 System Implementation

7.4.1 VR NeRF Rendering and Interaction Implementation

7.4.1.1 Render Plugin

A key technical challenge in implementing interactive VR experiences using neural rendering lies in bridging the interoperability gap between the low-level inference implementations and high-level game engines that are typically used for the VR application development process. Our system implementation tackles this challenge by introducing a native render plugin that enables data sharing between instant-ngp NeRF inference implementation [158] and the Unity VR application runtime. As illustrated in Figure 7.4, a pre-trained NeRF model and its' associated 2D image sequences are loaded into the instant-ngp application and the render device's GPU memory. OpenGL render buffers are created and upscaled by Nvidia's Deep Learning Super Sampling (DLSS) technique to improve image quality and increase the perceived rendered resolution in real-time [1]. A customized CUDA and C++ native plug running in Unity can access the OpenGL render buffer and render textures of instant-ngp through pre-compiled dynamic link libraries (DLL). We enable efficient data exchange between Unity and instant-ngp by sharing only the texture handle pointers for render event updates. In immersive VR rendering on a HMD, two render textures are created and placed as screen space overlays in front of the user's eyes. Synchronization problems are avoided by updating both render textures simultaneously in a single render frame. In terms of software, our framework implementation uses Unity Editor version 2019.4.29f1 with the OpenVR desktop and SteamVR

[1] https://www.nvidia.com/de-de/geforce/technologies/dlss/

runtime.

7.4.1.2 Manipulating a NeRF Model as an Object

To support spatial manipulation of a NeRF model in VR, we create a model space for the NeRF model whose spatial properties such as position, rotation, and scale are defined by a volume bounding box. As Figure 7.2 (a-b) shows, the bounding box is represented as a transparent cube whose translation, rotation, and scale matrices are combined into one transformation matrix that is applied to the view matrix of the instant-ngp camera to render the correct view of the NeRF model. The model-view-projection (VPM) matrix is applied to instant-ngp renderer through the native CUDA and C++ plugin. The object manipulation interaction from MRTK [2] is attached to the cube, allowing users to intuitively rotate, scale, and translate the box and its associated NeRF model using one- or two-handed control.

7.4.1.3 Crop Box Editing

To enable spatial transformation of a volumetric crop box, represented as the axis-aligned bounding box (AABB) in instant-ngp, we attach a second 3D cube game object with an object manipulator where the spatial transformations of the Unity AABB bounding box, such as translation and rotation, are applied to the AABB crop box defined in instant-ngp's coordinate system. As shown in Figure 7.2 (c-d), users can extend, reduce, or rotate the AABB render volume using the object manipulation interaction provided by MRTK.

7.4.1.4 Volumetric Editing

We adapt instant-ngp's implementation of a volumetric editing feature in our framework. Similar to instant-ngp's preliminary implementation, a 3D sphere object is attached to the controller at a certain distance to indicate the intended drawing region. As Figure 7.2 (e) shows, users can interactively select the region where voxels within the sphere should be made transparent by moving the sphere to the target region in 3D space while pressing a button on the VR controller to confirm the drawing action. The same interaction and action could be performed to reveal the volume at the region set to transparent, as shown in Figure 7.2 (f). In an instant-ngp NeRF model, the volume density grid stores the transparency of the scene learned by the network. A binary bitfield can be generated based on a transparency threshold and can be used to quickly mask the transparent region from being sampled during raytracing [158].

In summary, our system implementation provides many core components for building interactive VR NeRF applications using the Unity game engine to streamline the VR NeRF development process. A full demo video of the current system can be found in the supplementary video.

[2] https://github.com/microsoft/MixedRealityToolkit-Unity

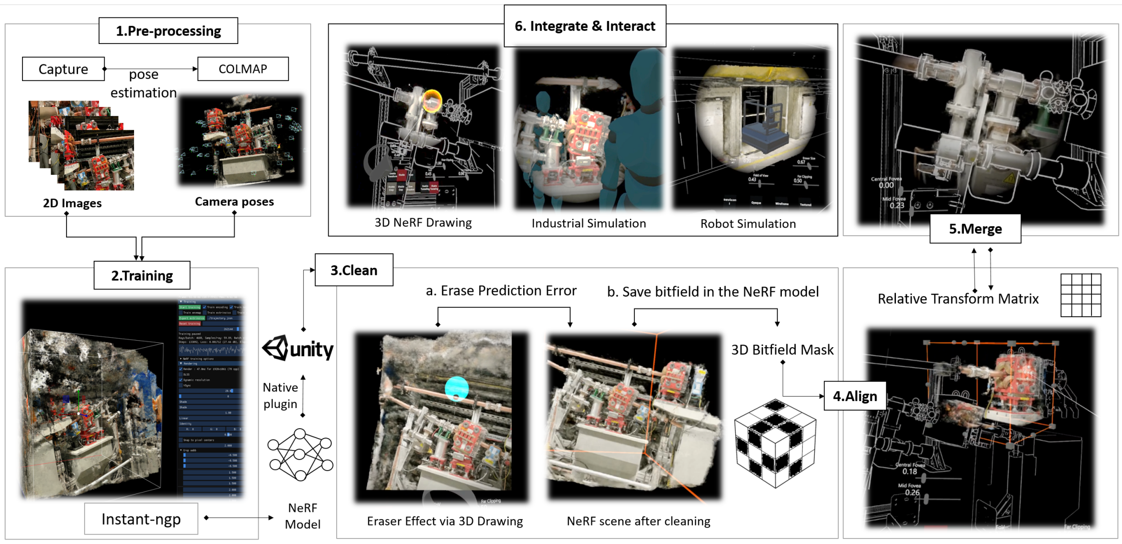


Figure 7.5: Overview of the data fusion pipeline to merge a NeRF model with the polygonal representation of its corresponding CAD model, illustrating sub-processes from image pre-processing, NeRF model training, scene cleaning, scene alignment, scene merging, and examples of final integration and interaction using different features of our framework.

7.4.2 Data Fusion Pipeline

In this section, we focus on developing a data fusion pipeline to merge a NeRF model with the polygonal representation of its corresponding CAD model to visualize the facility at a one-to-one real-world size and provide a realistic scale of the industrial environments. We choose CAD models as CAD models of industrial facilities are often created in the early stages of facility design and have complementary features to their corresponding NeRF models. Moreover, CAD models of industrial facilities typically lack realistic textures because the detailed environments are usually too complex to be modeled accurately [79]. Semantic information typically embedded in CAD models is also not represented in NeRF models without additional expensive network training and voxel-wise segmentation. Other benefits of fusing a polygonal representation of a CAD model with a NeRF model include the possibility of creating more realistic MR experiences that include depth occlusion and physical interactions using the mesh representation of the CAD model without the need to perform mesh reconstruction of the NeRF model to achieve comparable effects.

Figure 7.5 provides an overview of the data fusion pipeline, which is divided into the following six steps.

1. **Preprocessing.** Since most 2D images do not contain their camera poses, the 2D images must be preprocessed using the conventional SfM algorithm to estimate the camera poses using software such as COLMAP [192]. However, this step can be skipped for cameras that can track their own poses.

2. **Training.** The processed data is trained in the instant-ngp framework to generate

an initial estimate of the scene function and an initial occupancy grid whose size is defined by a preset AABB bounding box.

3. **Scene cleaning.** The quality of NeRF rendering can be affected by artifacts of real-world 2D images, such as motion blur, lens distortion, or insufficient images around the viewing angle. This often results in false clouds in the 3D NeRF scene, which can degrade the user's viewing experience and needs to be cleaned up before the NeRF model can be applied to practical applications. Therefore, our VR NeRF framework provides the ability to modify the pre-trained instant-ngp NeRF model. The initial density grid estimation is inspected and cleaned using the interactive eraser function described in section 7.4.1. Users can manually remove regions with cloudy prediction errors or low render quality. The edited density grid and the binary bitmask of the density grid are stored in the NeRF model for reuse. When reloading the cleaned NeRF model, our framework automatically checks the saved bitmask, so that the erroneous network prediction that has already been removed will no longer be sampled and rendered.

4. **Scene Alignment.** The user could manually align the cleaned NeRF model with the CAD model using the NeRF object manipulation and crop box editing functions described in section 7.4.1. If the NeRF model is adjusted to the one-to-one scale of the real facility, alignment using the entire NeRF model will result in a rendering volume too large to be efficiently manipulated in VR. Therefore, it is recommended to perform the alignment process by focusing on a small part of the scene for accurate object manipulation.

5. **Scene Merging**. In the merge step, the user could perceptually validate the alignment by reducing the FoV of the NeRF camera and adjusting the transparency of the shader that renders the NeRF images so that both the CAD drawing and the NeRF rendering are simultaneously visible. The user can iteratively move back and forth between the alignment and fusion steps until the two 3D representations are spatially aligned. The relative transformation matrix between the NeRF model and the CAD model can be saved for reuse.

6. **Integrate & Interact**. The merged visualization system can be used for various customized virtual inspection applications, such as facility upgrades and redesign, maintenance planning, and robot simulation.

7.5 Performance Benchmarking

While instant-ngp represents a significant advancement in achieving real-time interactive NeRF rendering, previous benchmarking efforts have primarily focused on demonstrating proof-of-concept render results for monoscopic image generation at a fixed FoV [158]. Notably, the system performance of instant-ngp implementations for stereoscopic VR rendering within an integrated VR application for real-world usecases is unknown. In this section, we present a performance benchmark experiment and its results using our magic NeRF lens visualization design. In particular, our experiment aims to provide insights into how parameters such as FoV and VR HMD's pixel density affect VR NeRF rendering performance. This enables us to evaluate the overall performance and efficiency of our system implementation while deriving configuration recommendations for using the magic NeRF lens in rendering volume reduction.

7.5.1 Experiment Design

We examine the performance trend when varying the NeRF rendering FoV from 10°, which covers the foveal region of human vision, to 60°, which covers the average central visual field for most people [170]. We also vary the PPD value to match the resolution requirements of VR displays of different quality, including medium display resolution (PPD=15, e.g. Oculus Quest 2), medium to high-end displays (PPD=20, e.g. Meta Quest Pro), or high-end displays (PPD=25, e.g. Varjo XR-3). The final NeRF render resolution per eye ($R \times R$) for each configuration is calculated using the following equation:

$$R = FoV \times PPD \times 2 \tag{7.1}$$

where we multiply the pixel density per degree by 2 to account for the upsampling required for aliasing reduction and edge smoothing via supersampling [9].

7.5.2 Materials

We evaluate the performance of our system on a self-generated real-world dataset consisting of 60 2D images at 3689×2983 resolution reconstructing a section of a complex particle accelerator facility. The images were captured by particle accelerator operation specialists during the facility's maintenance shutdown period. Its VR NeRF reconstruction is shown in Figure 7.7. Since the CAD model is designed to be at an exact scale of the real facility, we aligned the NeRF model with its' polygonal representation of CAD model using the data fusion pipeline described in section 7.4.2 and were able to measure that the NeRF rendered volume of this NeRF model corresponds to approximately 2.2 m $\times$ 1.47 m $\times$ 2.08 m in real-world scale, which would require a full FoV VR rendering to examine closely. All of our benchmarks are run on an Nvidia Geforce RTX 3090 GPU and an Intel(R) Core(TM) i7-11700K CPU with 32GB of RAM. As demonstrated in previous NeRF rendering implementations [158, 130], DLSS has a significant influence on VR NeRF rendering performance. As illustrated in Figure 7.4, DLSS is a crucial part of the NeRF render pipeline in upscaling the render buffer such that images rendered at low resolutions can be perceived at a higher resolution, thereby improving performance without sacrificing visual quality [3]. As a result, DLSS is enabled for all benchmarking efforts.

[3]For evaluating the impact of DLSS on VR NeRF render performance, we refer readers to our arXiv preprint on immersive-ngp [130]

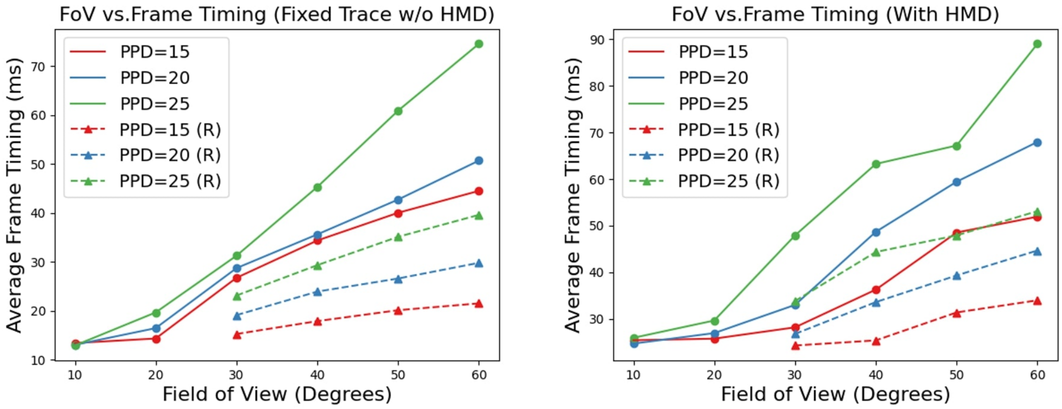


Figure 7.6: Systematic benchmark results showing (left): the relationship between rendering FoV and average frame time following a predefined fixed path without a VR HMD attached, and (right): the trend between rendering FoV and average frame timing for a test user following only approximately the same 3D trace. (R) indicates reduced rendering via 3D NeRF drawing.

7.5.3 Systematic Benchmark (w/o HMD)

To simulate the system performance of how a user would use our framework to inspect different parts of the NeRF model, we first asked a test user to closely inspect three fixed locations within the NeRF model. A custom script recorded the user's exact 3D trajectory with the camera transformation information at each frame. The relative position of the three target components concerning the polygonal representation of the CAD model was visualized as a schematic sketch, as shown in Figure 7.7-(Instruction Card). A video recording of the 3D trace is also available in the supplementary material.

First, we performed a systematic benchmark where we collected the average frame timing ($\overline{f_t}$) for stereoscopic rendering of different display FoV and PPD configurations by replaying the same pre-recorded 3D trace. Figure 7.6 (left) plots the average frame timing result for each rendering configuration. As expected, the rendering latency scales linearly with increasing FoV for the basic NeRF lens effect. Additionally, it is shown that our system could theoretically achieve real-time VR rendering of 30 frames per second ($\overline{f_t} < 33.33ms$) at less than 50° FoV.

To evaluate whether additional contextual rendering could lead to real performance gains, we create an edited NeRF model by 3D drawing interaction, where we define the 3D density grid so that only the target components are visible. Since FoV settings below 20° could already achieve good real-time performance without reducing the rendering volume, we only examine FoV and IPD settings above 30° in this section. Figure 7.6 (left) also plots the average frame timing for each rendering configuration for rendering the selected regions. As expected, exposing only the target render volume reduces the overall render load due to empty space skipping and early ray termination implementation. This confirms the performance optimization insight that removing unimportant scene details using the 3D NeRF drawing effect could be an option to gain additional performance.

7.5.4 Empirical User Benchmark Results (With HMD)

In the second benchmarking experiment, we evaluate the same set of rendering configurations. However, instead of replaying a pre-recorded 3D trace, a test user was asked to repeat each configuration, following roughly the same path to inspect the three target components to evaluate the overall performance of the actual VR system. For high-resolution configurations where real-time performance could not be achieved, the test user was able to use keyboard control instead, while keeping the VR HMD connected to the entire rendering pipeline. Figure 7.6 (right) shows the general trend between FoV and $\overline{f_t}$. Compared to the systematic performance results, the actual total frame timing is about $5 - 10$ ms higher because of the additional computational resources needed to drive the HMD. However, the general performance trend in an actual system with VR HMD is similar to the systematic benchmark result. Additionally, we could derive that for an actual VR NeRF system, the optimal render resolution is 1200×1200 at 30° FoV (PPD=20) or at 40° FoV (PPD=15) using an RTX 3090 GPU.

7.6 User Study

Understanding human factors is an important step in the development of new VR systems and interaction techniques. In this section, we present a user study experiment that aims to:

1. Systematically evaluate the overall performance and capabilities of the proposed visualization system in supporting actual virtual facility inspection processes through a visual search task,
2. Quantify the user experiences through perceptual metrics such as system usability, perceived motion sickness, and perceived task load to assess the effectiveness of the system implementation,
3. Understand the impact of different magic NeRF lens visualization and interaction styles on the overall system usability, task performance, and spatial presence.

In addition, we conducted confirmatory expert reviews at an industrial facility with five control and management specialists to validate the benefits of our framework in supporting real-world virtual facility inspection tasks.

7.6.1 Study Design

7.6.1.1 Conditions

We investigate the two magic lens designs with the two most common types of CAD model visualizations in large-scale industrial facilities. As shown in Figure 7.7, (C1) is the baseline condition with only FoV restriction but no data fusion for comparison, (C2, C4) implement the magic NeRF lens effect with FoV restriction, and (C3, C5) implement the magic NeRF lens effect with 3D drawing. (C2, C3) use a polygonal representation of a CAD model with colored abstract texture, while (C4, C5) use a polygonal representation of the CAD model with only edging rendering.

7.6.1.2 Materials

The user study was conducted on the same graphics workstation and software configuration used for the system benchmark described in section 7.5. For all conditions, the FoV of the NeRF rendering camera was set to 30° with a per-eye render resolution of 1200×1200 pixel, which is one of the optimal rendering configurations determined in our benchmark experiments. We used an Oculus Quest Pro VR HMD, which has a PDD value of 22. However, our VR NeRF framework is compatible with other VR headsets that support the SteamVR and OpenVR desktop runtime.

7.6.1.3 Tasks

To simulate how a user would perform virtual inspection tasks, we designed a visual search activity where the user was asked to locate three detailed components in the NeRF model based on a schematic sketch. As shown in Figure 7.7-(Instruction Card), the schematic sketch consisted of an overview of the CAD model, with each search target highlighted in a dialog box. Arrows are provided to indicate an approximate location where the detailed component might be found. The instructions for each search target were generated from the real-world images and therefore represent updated real-world conditions compared to the abstract CAD model. All search targets are detailed elements that could not be effectively updated in the CAD model due to maintenance activities or the difference between the actual installation and the original design. A pilot test with 3 users confirmed that all target components in each scene could be found in 3-5 minutes for all proposed conditions. To avoid learning effects in the course of the study, 5 different scenes were prepared, each showing a different part of an actual large-scale industrial facility.

7.6.1.4 Participants

We invited 15 participants, 4 female, 10 male, and 1 who preferred not to disclose their gender. 5 participants were between 18 and 24 years old, and 10 participants were between 25 and 34 years old. All were students or researchers in HCI, Computer Science, or Physics, with HCI students receiving compensation in the form of course credits. 8 participants use VR systems regularly (at least once a month), and only 2 use them less than once a year, 5 never used VR before.

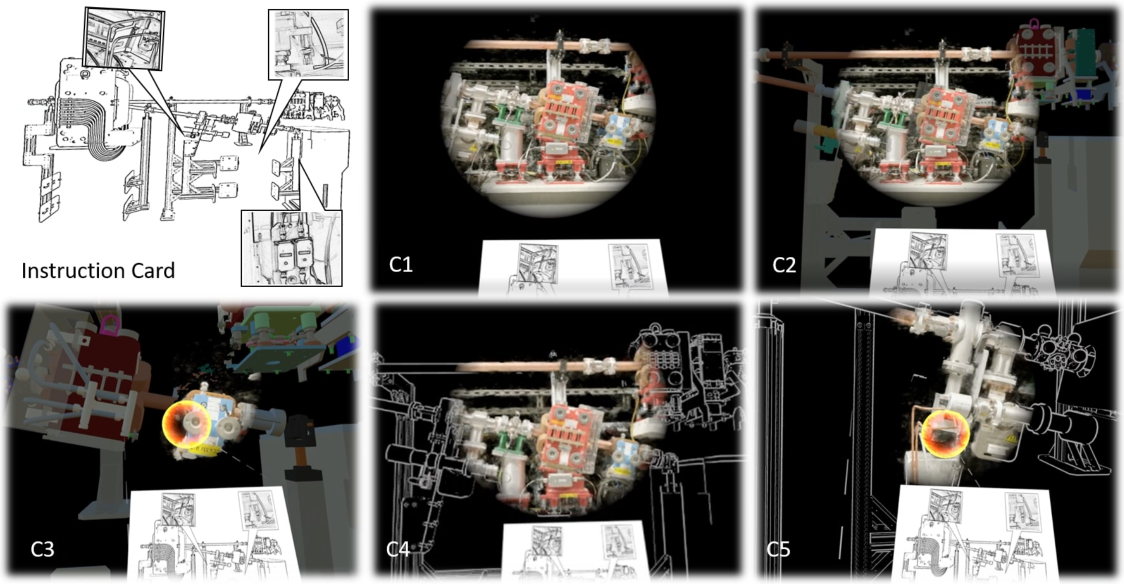


Figure 7.7: Screenshots for different visualization and interaction conditions of Magic NeRF Lens. (C1): Baseline condition with only the NeRF model and the FoV restrictor. (C2): Magic NeRF Lens with a textured CAD model as context. (C3): 3D NeRF drawing with a textured CAD model as context. (C4): Magic NeRF Lens using only the wireframe representation of the CAD model. (C5): 3D NeRF drawing with wireframe representation of the CAD model.

7.6.1.5 Procedures

Each participant first signed a consent form and completed a demographic questionnaire. The experimenter then presented a test virtual environment in which participants could practice locomotion and 3D VR drawing controls for a maximum of 3 minutes. Each participant was then presented with the 5 conditions in order. To counterbalance the order and assignment of scenes to conditions, we used a replicated Latin square design with one treatment factor (condition) and three blocking factors (participant, trial number, and scene). As a result, each condition/scene combination was experienced by 3 participants. During each condition, participants could press a "start" button to begin the study and begin time recording to assess task performance. The experimenter monitored the participants' VR interactions via a secondary screen that mirrored the VR displays. After the participant informed the experimenter when they had found a component, the experimenter confirmed its correctness, after which the participant could begin searching for the next search target. Once all three components were found, the participant could press an "end" button, which marked the end of time recording for the search tasks. The total study time was approximately 60 minutes.

7.6.1.6 Hypotheses

We hypothesized that displaying the polygonal representation of the facility's CAD model in the user's periphery would provide context to the user and thus improve spatial orientation within the scene. Since the use of NeRF drawing tools introduces additional complexity, we further hypothesized that using 3D NeRF drawing will have less usability and higher cognitive load on users than Magic NeRF lens effect through only MR tunneling.

This led to the following hypotheses:

- **H1**: Magic NeRF lens effects (C2-C5) reduce perceived mental demand and effort, and lead to higher performance, both subjectively and objectively, as measured by time spent on the visual search task.
- **H2**: Magic NeRF lens effects with MR tunneling display (C2 & C4) yield higher usability scores than those based on the 3D drawing techniques (C3 & C5).

Based on the results of [132], we did not expect significant differences between the conditions in terms of presence and cybersickness. Nevertheless, we collected subjective ratings on both metrics to quantify how well users perceive these effects in general.

7.6.2 Results

Table 7.1: Means and standard deviations for the subscales of the NASA TLX (incl. Mental / Physical / Temporal Demand). Best values are marked in **bold**.

	Search Time		Physical D.		Temporal D.	
	M	SD	M	SD	M	SD
(C1)	36.00	18.05	12.00	13.99	18.67	16.95
(C2)	22.00	14.86	**8.33**	5.88	**14.33**	9.61
(C3)	36.67	25.75	11.33	16.09	15.33	12,46
(C4)	**19.00**	15.49	9.00	13.65	15.33	15.06
(C5)	39.33	19.35	13.67	12.46	19.67	15.52

Table 7.2: Means and standard deviations for the subscales of the NASA TLX (incl. Performance / Effort / Frustration). Best values are marked in **bold**.

	Performance		Effort		Frustration	
	M	SD	M	SD	M	SD
(C1)	19.67	19.04	36.33	19.77	16.67	17.49
(C2)	17.67	13.61	23.33	17.29	**12.00**	10.14
(C3)	24.00	20.63	40.33	29.24	30.67	23.21
(C4)	**12.00**	11.62	**22.00**	18.30	15.67	16.68
(C5)	24.67	17.67	40.00	21.30	20.00	13.76

Table 7.3: Means and standard deviations for the task performance (i.e., search time), usability, and cybersickness. Best values are marked in **bold**.

	Search Time		Usability		Cybersickness	
	M	**SD**	**M**	**SD**	**M**	**SD**
(C1)	194.78	129.11	76.00	14.60	9.23	14.34
(C2)	157.78	102.24	**81.83**	10.07	7.73	8.05
(C3)	291.66	154.22	67.17	19.95	12.47	16.46
(C4)	**128.09**	63.18	78.00	12.18	**6.98**	10.94
(C5)	286.14	161.00	72.67	15.94	8.98	9.67

Table 7.4: Means and standard deviations for each subscale of the SSQ. Best values are marked in **bold**.

	Nausea		Oculomotor		Disorientation	
	M	**SD**	**M**	**SD**	**M**	**SD**
(C1)	3.18	5.89	10.11	14.79	11.14	21.18
(C2)	4.45	6.11	**7.07**	7.29	9.28	17.18
(C3)	9.54	13.00	11.62	15.12	11.14	19.83
(C4)	**2.54**	7.62	7.58	10.72	**8.35**	15.61
(C5)	5.09	7.09	9.10	8.20	9.28	16.36

Table 7.5: Means and standard deviations for all subscales of the IPQ (i.e., Sense of being there, Spatial Presence, Involvement, Experienced Realism). Best values are marked in **bold**.

	IPQ-PRES		IPQ-SP		IPQ-INV		IPQ-REAL	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
(C1)	1.00	1.07	3.84	.84	3.28	1.20	2.15	.96
(C2)	.73	1.58	**3.92**	.90	3.08	1.38	**2.45**	.88
(C3)	.87	1.36	3.85	1.09	3.15	1.27	2.27	.91
(C4)	.47	2.13	3.60	.91	3.18	1.47	2.33	.85
(C5)	**1.33**	.98	3.76	.86	**3.35**	1.09	**2.45**	.92

We collected multiple objective and subjective measures to assess the user experience as study participants interacted with our magic NeRF lens framework. The mean and standard deviation of the six subscales of the NASA-TLX index are recorded in Table 7.1 and Table 7.2. The overall ratings of usability, cybersickness, and overall task performance are recorded in Table 7.3. Each subscale of the SSQ is recorded in Table 7.5.

7.6.2.1 Usability

Usability was measured using the SUS with 10 questions providing ratings between 1 (strongly disagree) and 5 (strongly agree), where higher values correspond to better usability [21]. The overall rating of usability converts to a value range [0,100]. A repeated measures ANOVA with Greenhouse-Geisser correction revealed a significant effect of the visualization method on usability, $F(2.514, 35.202) = 4.269, p = .015, \eta_p^2 = .234$. Post-hoc tests with Bonferroni correction showed a significant difference between C3 and C4 ($p = 0.048$). A copy of the SUS questionnaire can be found in Appendix A.1.5.

7.6.2.2 Perceived Workload

We measured six aspects of perceived workload using the NASA-TLX questionnaire [88]. Because each aspect was represented by a single Likert scale ranging from 0 to 100 (with 21 levels), we used nonparametric Friedman tests to analyze the responses. There was a statistically significant difference in mental demand depending on the condition, $\chi^2(4) = 17.857, p = .001$. Post-hoc analysis with Wilcoxon signed-rank tests and Bonferroni-Holm p-value adjustment revealed a significant difference between C2 and C5 ($Z = -3.257, p = .011$). Friedman tests also showed significant effects of the visualization method on effort ($\chi^2(4) = 15.366, p = .004$) and frustration ($\chi^2(4) = 10.183, p = .037$), although none of the post-hoc tests were significant after p-value adjustment. No significant differences were found for physical demand, temporal demand, and performance. A copy of the NASA-TLX questionnaire can be found in Appenxid A.1.2.

7.6.2.3 Task Performance

Task performance was measured as the time it took participants to find the three locations indicated on the instruction card for each scene. A Shapiro-Wilk test indicated that the residuals of search time were not normally distributed, which was confirmed by visual inspection of the QQ plots. Therefore, a Friedman test was performed, which revealed a significant effect of the visualization method on search time ($\chi^2(4) = 23.307, p < .001$). Post-hoc Wilcoxon signed-rank tests with Bonferroni-Holm adjustment of p-values showed significant differences between C2 and C3 ($Z = -3.010, p = .018$), C2 and C5 ($Z = -3.124, p = .014$), C4 and C3 ($Z = -3.237, p = .011$), and C4 and C5 ($Z = -3.351, p = .008$).

7.6.2.4 Presence

We measured the sense of presence using the IPQ [195] with all three subscales *Spatial Presence*, *Involvement*, and *Experienced Realism*, as well as a single item assessing the overall "sense of being there". Three ANOVAs for the subscales and a Friedman test for the single item revealed no significant differences. A copy of the IPQ can be found in Appendix A.1.3.

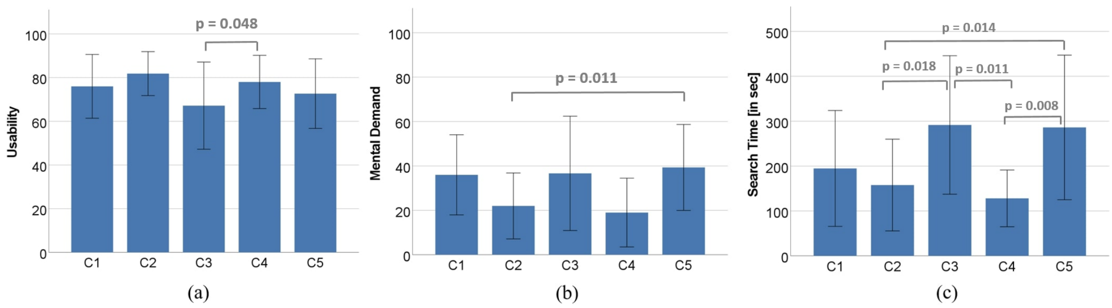


Figure 7.8: Mean (a) usability, (b) mental demand as measured by the NASA-TLX, and (c) search time to finish the task in seconds. Vertical bars indicate the standard deviation. Any significant differences were labeled with their corresponding p values between conditions.

7.6.2.5 Cybersickness

As a subjective measure of cybersickness, participants rated the perceived severity of 16 symptoms covered by the SSQ [106]. From the ratings, we calculated subscores for nausea, oculomotor, and disorientation, as well as a total cybersickness score, as suggested by [106]. No increase in cybersickness across trials was observed, so we analyzed absolute values rather than relative differences between trials. Since the residuals were not normally distributed, we performed a Friedman test for each of the four (sub)scores. For nausea, a significant difference between conditions was found, $\chi^2(4) = 10.330, p = .035$, but this could not be confirmed by post-hoc pairwise Wilcoxon signed-rank tests with Bonferroni-Holm correction. For oculomotor, disorientation, and total cybersickness scores, no significant effects were found using Friedman tests. A copy of the SSQ can be found in Appendix A.1.1.

7.6.3 Discussion

7.6.3.1 Spatial Orientation

The magic lens effects with FoV restriction (C2, C4) achieved better ratings in terms of mental effort and higher task performance than C1, though not significant, therefore H1 was rejected. Nonetheless, qualitative user feedback indicated that C2 and C4 "*make(s) it more confident /easier to navigate around the machines*" (N=2). Without the contextual guidance of CAD models, "*it was difficult to get a feeling of scale or to identify different places correctly.*" Detailed comments mentioning spatial orientation confirmed that these effects further "*helped with general orientation*" and "*provided a sense of [the user's] position*". In contrast, the 3D NeRF drawing interaction (C3, C5) performed significantly worse than the 3D NeRF drawing effect (C2, C4), confirming H2. The main problem reported with these conditions was "*difficulty to find the correct depth*" (N=4). This could be because most participants were unfamiliar with the complex environments, and using only a 2D schematic sketch as visual instruction does not provide users with sufficient spatial orientation to navigate efficiently to the search target. As potential future use of our framework and techniques needs to accommodate users' diverse skill sets and backgrounds, future application of the magic NeRF lens effect with 3D drawing could provide specific 3D points or markers to help orient users and provide reference points to support users who need additional visual aids and guidance for them to effectively navigate within the intricate industrial systems.

7.6.3.2 Task Load, Performance, Usability, and Cybersickness

Across all conditions, there were no significant differences in perceived cybersickness, with participants experiencing little to no motion sickness. In general, users found the magic lens effects with FoV restrictor (C2, C4) highly usable (SUS_{C2} = 81.83, and SUS_{C4} = 78), with better objective task performance and lower perceived mental effort. However, as assumed by H2, the magic NeRF lens effects with 3D drawing received lower usability scores, with a significant difference between (C3) and (C4). As Figure 7.8 shows, usability scores were moderately correlated with cognitive load ($r = -.458$) and task performance ($r = -.543$). Qualitative feedback indicates that participants often repeatedly unfolded and erased the NeRF models because they were unsure where to find the components, leading to frustration and even framerate jitters when more areas than necessary were revealed. In addition, participants mentioned that the CAD is often outdated and incomplete compared to the NeRF model. This discrepancy makes it even more difficult for users without a facility management background to complete the tasks. Nevertheless, two participants mentioned the 3D NeRF drawing effect as their preferred condition because "*it was more fun* " and has "*a nice property of reducing complexity and putting the focus on the spots you want to investigate*". Therefore, we recommend future use of 3D NeRF drawing effect mainly for facility management experts but still make it an option for non-expert users to accommodate different preferences.

7.6.3.3 Context Rendering Style

Concerning the rendering style of the polygonal representation of the CAD model (colored solid vs. edge rendering), the conditions yielded similar results in both subjective ratings and task performance. Since users reported different preferences in the open-ended feedback, a customization option could be offered in a practical application.

7.6.4 Expert Feedback

To explore how our framework could be used by practitioners in a real industrial setting, we validated our system through expert reviews at DESY. Five facility managers and control system specialists participated in the reviews. All participants have a leading position in the design, coordination, or control of particle accelerators at DESY, and two of them are also experienced VR experts who have already developed VR systems for facility inspection.

The expert reviews were conducted using an exploratory application that illustrates a section of the NeRF model of a large-scale industrial facility. It provides many flexibility and customization options, allowing the user to freely adjust the NeRF camera's field of view, change the size of the NeRF editing sphere, vary the translucency of the merging effects, as well as the manipulation interactions mentioned in section 7.4. The application ran on an Alienware m17 R2 laptop with 16GB of RAM and an RTX 2080 GPU and was displayed through an Oculus Quest Pro HMD. We reduced the resolution of the application to 800×800 pixels to achieve real-time performance.

Overall, participants felt "*very confident to use the system*" and "*it is something (they) could work with*", even though the application was running on a laptop with moderate performance and moderate resolution. All expert participants confirmed that using NeRF for virtual facility inspection could benefit their workflow. One facility management expert commented that NeRF is a compelling, low-cost alternative for 3D facility documentation: "*I think the system has a good advantage. It is quite nice to project the NeRF model on the CAD model, as it is a lot more effort to take laser scans of the facility*". In addition, most expert participants preferred to inspect the facilities in VR because the system "*helps them to see if (they) could reach anything*" or "*if (they) could fit any equipment through the existing environment*". In addition, two experts mentioned that having a one-to-one real-world scale NeRF model aligned with the polygonal CAD model in immersive VR also gives them a better spatial awareness of complex machines than working with a 2D desktop application. For example, they mentioned that the VR NeRF environment could help them assess in advance if "*an operator's hand would fit through a narrow gap to handle components*" or if special equipment would need to be prepared in advance.

The expert who leads the design and upgrade of their facilities mentioned the benefits of data fusion visualization: "*With this system, I see the possibility to test something in theory before you build it in practice. For example, when you have a machine, and you want to test if you have enough space for installing it, it is quite nice you could test everything in the virtual area before you do it in reality*". He also mentioned that their CAD models usually only show the initial design of the facility. Once the facility is operational, these CAD models can become incomplete and outdated. He found the contextual 3D NeRF drawing effect particularly helpful in comparing the difference between the original design and the actual implementation, which could even help operators update the original CAD designs accordingly.

For future development of the facility inspection system, one expert suggested the interesting idea of integrating contextual QR code scanning to further support information retrieval from their large inventory database. Sometimes, facility inspection tasks require scanning labels containing QR codes with manufacturing and maintenance information that are attached to all cables and equipment. The ability to retrieve such labels directly from the NeRF model would further streamline facility inspection processes. Concerning safety-critical processes such as immersive robot teleoperation, participants mentioned that although NeRF could provide photometrically accurate results, its geometric accuracy also needs to be verified and compared with conventional 3D sensors. Nevertheless, as our proposed NeRF magic lens effects and data fusion pipeline could be applied to other types of conventional 3D models, we encourage further investigation of experimenting with data fusion with other types of 3D data of large-scale facilities.

7.7 Limitations and Future Work

Our framework still has several limitations that could be addressed in future work. First, future research could investigate an automatic CAD-NeRF alignment approach to skip the manual hand-eye calibration process. Detailed investigation of an interactive point-matching algorithm could be a promising approach, taking into account the real-world mismatch between the NeRF model and the CAD model [28]. In addition, optimization techniques such as empty space skipping and early ray termination could also lead to frame rate jitter as the number of network queries becomes view-dependent. Although the use of a FoV restrictor could reduce such effects, such framerate jitter will be more noticeable on medium and low-end graphics hardware. Therefore, we encourage future work to investigate the integration of further optimization techniques, such as foveated rendering, to enable a comfortable VR NeRF system even on low-end graphics devices.

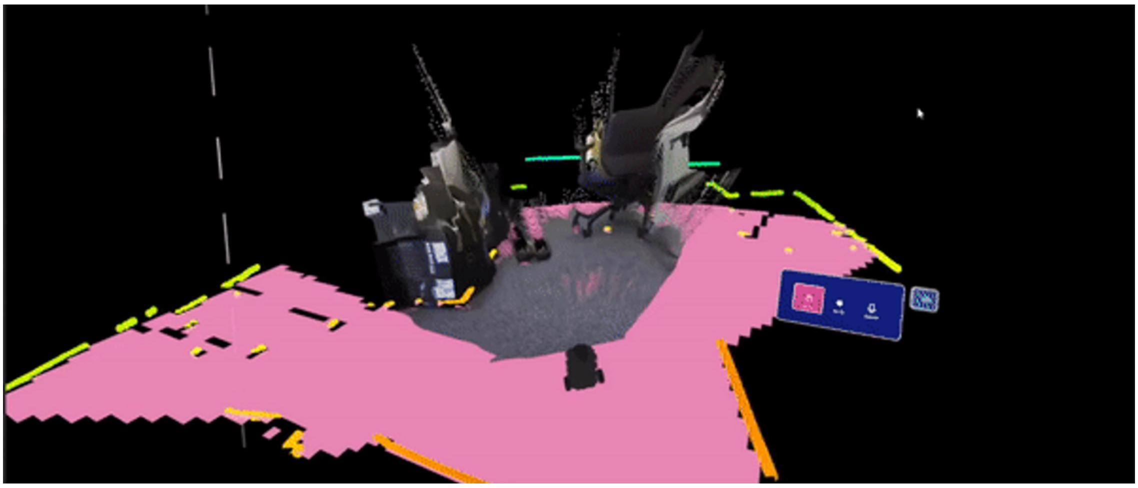
It is also important to note that the field of photorealistic view synthesis is evolving rapidly. Relevant work on representing real-world scenes with 3D Gaussian functions as an alternative to multi-layer perceptron has emerged in the last two months, paving the way for more efficient rendering of radiance fields via rasterization rather than ray casting [107]. We encourage future work to further extend our system to support 3D Gaussian splatting in immersive VR [107], and to compare the different trade-offs between the two types of representation and rendering for user interaction and system performance.

7.8 Conclusion

We presented *Magic NeRF Lens*, an interactive immersive visualization toolkit to support virtual facility inspection using photorealistic NeRF rendering. To support the rendering of industrial facilities with substantial volume, we proposed a multimodal data fusion pipeline to visualize the facility through magic-lens-style interactions by merging a NeRF model with the polygonal representation of its' CAD models. We designed two magic NeRF lens visualization and interaction techniques and evaluated these techniques through systematic performance benchmark experiments, user studies, and expert reviews. We derived system configuration recommendations for using the magic NeRF lens effects, showing that the optimal configuration for visualizing industrial facilities at one-to-one real-world size is 20 PPD at 30° FoV, or 15 PPD at 40° FoV within an integrated VR application. Through a visual search user study, we demonstrate that our MR tunneling magic NeRF lens design achieves high usability and task performance, while the 3D NeRF drawing effect is more interactive but requires future integration of more visual guidance to support users who are not familiar with the complex facility environments. Follow-up system reviews with 5 experts confirmed the usability and applicability of our framework in support of real-world industrial virtual facility inspection tasks such as facility maintenance planning and redesign. Finally, we believe that the interdisciplinary and open-source nature of this work could benefit both industrial practitioners and the VR community at large.

8

Reality Fusion



In this chapter, we introduce *Reality Fusion*, a novel robot teleoperation system that localizes, streams, projects, and merges a typical onboard depth sensor with a photorealistic, high resolution, high framerate, and wide FoV rendering of the complex remote environment represented as 3DGS. This chapter aims to investigate the following research question:

- **RQ5**: *How to provide low-latency and high-quality volumetric visual feedback to the operators during robot teleoperation tasks at particle accelerator tunnels while ensuring that operators have a high level of situation awareness of the complex remote environments?*

This chapter is primarily based on the following paper:

- **Ke Li**, Reinhard Bacher, Susanne Schmidt, Wim Leemans, Frank Steinicke. (2024). Reality Fusion: Robust Real-time Immersive Mobile Robot Teleoperation with Volumetric Visual Data Fusion. 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (accepted, to appear)

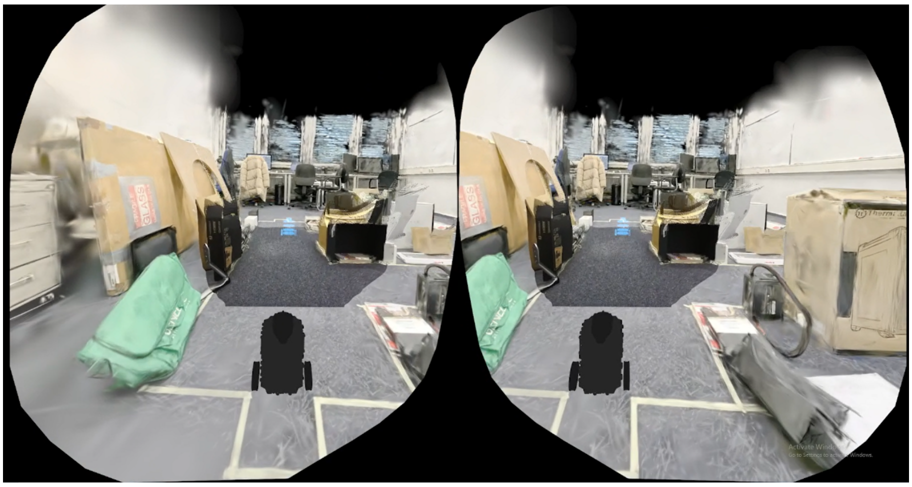


Figure 8.1: Screenshot from our robot teleoperation system with Reality Fusion rendering in immersive VR. Real-time point points from the robot's onboard RGBD sensor are localized, streamed, and projected onto a photorealistic scene of the remote environment represented as 3D Gaussian splats.

8.1 Motivation

As introduced in Section 2.3, the possibility to display 3D spatial cues about the robot's environment to the operators through an immersive HMD enables a remote "telepresence" experience that has the potential to improve the operator's task performance significantly. However, building a robust real-time immersive robot teleoperation system is still faced with many technical challenges. On one hand, real-time 3D data capturing and streaming to the HMD is crucial in providing operators with accurate situational awareness of the robot's environment. On the other hand, spatial data streaming and processing presents a trade-off between visual quality and processing latency. While telepresence systems using omnidirectional camera [249] or multi-camera setups [89] can provide operators with a high level of immersion and detailed information of the robot's environment, the latency in streaming, processing, and rendering these 3D data introduce the undesirable effect of VR motion sickness [249] and a delay in robot control and intervention [217]. Due to the limitation of robots' payload and the limited computational resources of an embedded system, many real-world mobile platforms can only provide operators with real-time visual feedback through low-cost sensor setups such as a single 2D video camera or a single stereo depth camera. These sensors often capture visual information with a restricted FoV, which hinders users from observing the remote environment from obscured angles and therefore restricts them from establishing a concrete mental model of the robot's surroundings.

In this chapter, we present *Reality Fusion*, a novel immersive robot teleoperation framework that implements a multi-modal data fusion method to address the issue of spatial data streaming trade-offs in immersive robot teleoperation systems. As Figure 8.1 illustrates, our framework first introduces a system that can render high resolution, high framerate, wide FoV, and photorealistic 3D scenes of complex environments that are represented as 3DGS [107] in VR. Using this type of photorealistic scene representation as the digital replica of the robot's remote environment, the VR system enables offline visualization of complex scenarios that are typically difficult to accurately model with conventional 3D meshes or point clouds. To enable operators to view the real-time status of the remote environment, we introduce a data fusion method, where real-time point cloud from the onboard RGBD sensor is localized, streamed, projected, and merged with the 3DGS environment, with the 3DGS rendering of the remote environment effectively extending the FoV of the RGBD point cloud. Such a teleoperation system design allows us to implement a lightweight, flexible, and cost-efficient immersive telepresence mobile robot with only a single RGBD sensor and a small embedded system.

We systematically benchmarked our framework through a controlled user study experiment, in which 24 participants performed a mobile robot navigation task through a real-world maze. Furthermore, we compare the efficiency of egocentric and exocentric teleoperation, from which we discuss the advantages and trade-offs of these two teleoperation modes to derive further design insights for immersive telepresence systems using *reality fusion*.

In summary, the contributions of this chapter include:

- A VR teleoperation framework that allows robust, intuitive, and efficient remote control of mobile robots through a novel visual data fusion technique.
- The implementation of the framework in terms of hardware and software including a flexible telepresence robot, a high-performance VR 3DGS renderer, and an immersive robot control package.
- A controlled user study experiment with robot teleoperation tasks that systematically evaluate various human factors of the framework while deriving design insights.

The source code of the framework will be made publicly available here: https://github.com/uhhhci/RealityFusion. A demo video of the system is available here: https://youtu.be/qrnzmbWyXRA.

8.2 Related Work

8.2.1 Immersive Robot Telepresence and Teleoperation

A key challenge in building an immersive robot teleoperation system is to provide operators with high-quality and low-latency volumetric visual feedback of the robot's environment. Ferland et al [72]'s teleoperation system can perform a stereoscopic projection of a binocular camera into the user's world space, allowing for 6 DoF exocentric immersive teleoperation in VR. However, operators' situation awareness can be largely restricted by the narrow FoV of the onboard camera. Although several previous works investigated using a multi-camera setup [216] or streaming videos from an omnidirectional camera [249], these solutions introduce a significant increase in data streaming latency, leading to undesirable effect of VR motion sickness [249] and delay in robot control and intervention [216]. Another approach suggests displaying real-time mesh reconstruction results to the operator using dynamic SLAM algorithms [211]. However, 3D reconstruction with dynamic SLAM can introduce temporal delays of several seconds, making it unsuitable for applications requiring immediate intervention by the operator. Tefera et al. [225] propose reducing live point cloud streaming bandwidth through a foveated point cloud segmentation and streaming framework, but this could introduce visual degradation and aliasing effects during third-person teleoperation, possibly reducing user flexibility and control options. Various prior robot teleoperation systems use a cockpit-like design, integrating multiple sensor data sources—including 2D videos and 3D point clouds—into immersive user interfaces within a simulated control-room environment [104, 233, 13]. These systems, similar to our approach, enable monitoring of a robot's movements from both egocentric and exocentric perspectives [13]. However, cockpit-like teleoperation frameworks typically use a large, immersive 3D space to display a wide range of visual information, which can vary in latency and spatial dimensions. This can lead to confusing interface designs and cognitive overload for users. Our work extends previous teleoperation systems by combining limited online volumetric data with a high-quality offline photorealistic 3D representation of the environment. The offline representation serves as a contextual guideline to enhance operators' immersion, while the online data provides primary visual feedback for monitoring the robot's surroundings. This creates a more robust data streaming solution and a more coherent visualization style for real-time remote control.

8.2.2 3D Representations for Robot Teleoperation

As introduced in Chapter 4, a key challenge in immersive robot teleoperation is creating robust 3D representations to visualize the remote environment. Currently, most immersive robot teleoperation systems use conventional explicit representations such as point clouds and meshes to visualize the remote environment. However, 3D meshes for robot teleoperation are typically created with dynamic SLAM algorithms for real-time feedback [211], which can yield inaccurate results with complex geometries, making them unsuitable to represent complex real-world conditions such as an industrial facility. Although point cloud is also a popular 3D representation in immersive robot teleoperation, such discrete representations introduce holes and occlusions, degrading the visual quality of the 3D representation and restricting users' understanding of the modeled environments [238, 36]. As introduced in Chapter 4, the latest breakthrough in photorealistic scene rendering proposes a radiance field representation [154]. However, due to the early stage of research of 3DGS and NeRF for robotics, existing 3DGS-SLAM or NeRF-SLAM methods result in degraded visual quality or an increase in rendering frame timing, making them unsuitable for real-time immersive VR applications. Therefore, our framework focuses on developing an efficient integration of the current 3DGS method to an immersive robot teleoperation system for rendering 3DGS models offline rather than reconstructing dynamic SLAM 3DGS visual feedback online. Nonetheless, this is, to our best knowledge, the first usable robot teleoperation system based on immersive photorealistic rendering in VR.

8.3 Reality Fusion

We define the term "*Reality Fusion*" as the merging of two photorealistic 3D representations of a real-world environment with the purpose of data augmentation by combining the complementary features of different types of spatial data and naturally integrating them into a coherent spatial UI. In this work, we develop a *reality fusion* method that can combine real-time 3D projection of a stereo camera and the 3DGS of the real world to create a fully immersive telepresence experience for robot operators. This section presents the design goal and theoretical background related to such a *reality fusion* method.

8.3.1 Design Goals

Our framework specifically targets a typical application scenario of a robot teleoperation system for remote industrial facility inspection, where human access to the facility is limited due to various hazards or operation constraints, however, the environment of the facility is static and is unlikely to undergo immediate large structural change. A typical example of such an environment is a particle accelerator tunnel, as introduced in Chapter 2. Building an effective and robust immersive teleoperation system for mobile robot navigation for such an application needs to consider the following design goals:

- **G1** The mobile robot needs to be flexible and lightweight, capable of visiting areas such as narrow gaps between facility components.
- **G2** The operator needs to have high situation awareness and spatial orientation of the robot's environment for task planning and navigation control.
- **G3** The operator needs to receive real-time visual feedback on the robot's current status with minimum latency to perform timely intervention and robot control.
- **G4** The implementation of the VR application needs to enable natural and intuitive control of the robot without introducing undesirable effects such as VR motion sickness.

8.3.2 Stereoscopic 3D Projection in World Space

To fulfill the requirement of **G2**, our framework adapts 3DGS models for high performance photorealistic scene rendering in immersive VR [107]. Rendering a 3DGS model can only provide offline passive visual feedback of the environment. To fulfill the requirement of

G3, we integrate stereoscopic 3D projection of a depth camera to provide operators with active real-time spatial feedback. Given a stereo camera, depth information (d) can be estimated based on the correspondence established from the two camera views. Given a homogeneous pixel vector with depth estimation $P_d = [x_d, y_d, d, 1]$, it's 3D coordinate $P_C = [x_c, y_c, z_c, w]$ can be calculated through stereoscopic 3D projection transformation [72]:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ w \end{bmatrix} = \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & f \\ 0 & 0 & -\frac{1}{b} & 0 \end{bmatrix} \begin{bmatrix} x_d \\ y_d \\ d \\ 1 \end{bmatrix} \tag{8.1}$$

Here, b is the stereo camera baseline, a is the camera aspect ratio, and f is the camera focal length, all of which could be obtained through camera calibration. Then, P_C can be converted to homogeneous 3D coordinate P_h through perspective division.

To calculate the point cloud's 3D coordinate in world space $P_W = [x, y, z, 1]$, we transform P_h through a transformation matrix obtained using the view projection matrix of the virtual camera M_{VP} and transformation matrix of the stereo camera M_{camera} which describes the tracked camera's position and orientation in global world space:

$$P_W = M_{VP} \times T_{offset} \times M_{camera} \times P_h. \tag{8.2}$$

As the robot's tracked position has a different center of origin than the actual position of the camera, an additional translation matrix T_{offset} obtained through manual calibration is applied to compensate for the translation offset.

As *reality fusion* only depends on a single depth camera for receiving real-time visual feedback, it reduces the overall required payload on the robot (**G1**). In addition, the fusion of both a 3DGS model and real-time point clouds creates a coherent and natural visual appearance with low streaming and rendering latency (**G4**). Moreover, both types of volumetric data enable 6DoF changes of perspective and robust exocentric robot control, making it easy for users to adjust their viewpoints and increasing the flexibility of robot planning and navigation control tasks (**G4**).

8.4 Framework Implementation

As shown in Figure 8.2, our framework consists of three main components: (i) the robot operator equipped with a VR HMD and VR controllers for interaction with the spatial UI; (ii) a Unity application that handles the robot control logic, communicates with remote ROS endpoint, and perform high-performance graphics rendering through native CUDA and C++ plugins; and (iii) a telepresence robot with an onboard visual sensor providing real-time feedback to the remote operator. This section presents relevant details of our framework and implementations.

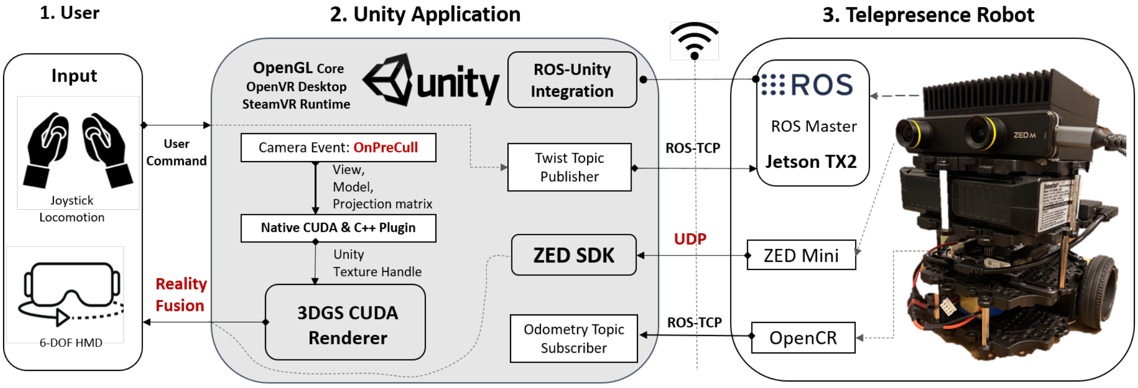


Figure 8.2: System overview of our immersive teleoperation framework which includes: 1. the operator equipped with a VR-HMD and sends command to the remote robot via VR controller inputs; 2. a Unity application which manages robot control logic, communicates with remote ROS endpoint, and perform data fusion and graphics rendering; 3. a custom-built telepresence mobile robot equipped with a SBC and a RGBD sensor.

8.4.1 The Telepresence Robot

Overview We designed and developed a compact, lightweight, and modulized telepresence mobile robot that can be easily reassembled and replicated from commercially available hardware. As shown in Figure 8.2, the robot is modified from the open-source Turtlebot 3 burger robot platform. Our custom-made robot has three core hardware components: an OpenCR board for low-level control of the robot's motion, a single ZED Mini stereo camera as the spatial vision sensor, and an SBC as the robot's central computing unit. As with the original Turtlebot burger, the robot has two differential wheel drives with a maximum linear speed of $0.22m/s$ and a maximum angular speed of $2.84rad/s$.

The Single-board Computer The SBC is an edge AI device based on an Nvidia TX2 NX, a high-performance embedded system with an accelerated 256-core NVIDIA Pascal GPU with CUDA version 10.2. The ZED Box runs on Ubuntu 18.04.6 LTS, Jetpack 4.6, ZED SDK version 3.7.3 [1], and ROS Melodic. The SBC is powered by a TalentCell 72W 100WH power bank.

High-Resolution Stereo Camera and Video Streaming A ZED Mini stereoscopic camera is mounted facing the forward direction of the robot. Stereoscopic videos are streamed at HD720 resolution with a vertical FoV of 54° and a horizontal FoV of 85°. The video stream is wirelessly sent via a local 5G network to the VR device through a user datagram protocol (UDP) with the ZED SDK, where streaming latency is minimized through highly optimized GPU video encoding and decoding processes. From a similar video streaming configuration, we can estimate the motion-to-photon latency of such a setup to be $153.47 \pm 33.33ms$ at 30 frames per second (fps) [132] within a local 5G network.

[1] https://github.com/stereolabs/zed-unity

8.4.2 Unity 3DGS VR Renderer

Overview Efficient 3DGS rendering relies on a sorting process that can rapidly re-order each Gaussian primitive based on the update of the camera poses and their clipping planes [107]. Therefore, we developed a custom Unity VR renderer through Unity's native render plugin to utilize Kerbl's original CUDA kernels for parallel sorting and tiled-based rendering. As a result, compared to the currently available 3DGS Unity integration [179], our renderer can provide performance equivalent to the original CUDA implementation and is better optimized for immersive VR rendering.

Native Renderer Architect As our custom renderer does not include the 3D Gaussian points as built-in Unity game assets, we directly displayed the final rendered images as screen quad objects in the HMD. Our 3DGS Unity renderer takes the user's tracked head pose, converts it into the camera view-projection matrix, combines it with other user-defined values such as resolution, FoV, geometric transformation, and updates these parameters in the native CUDA/C++ renderer at each frame. An effective VR plugin should also correctly synchronize the user's head movement and the rendered images to avoid undesirable temporal aliasing effects such as scene jittering, which occurs in a recent attempt for native CUDA Unity-3DGS integration [46]. Our rendering pipeline prevents this problem by triggering native render events inside the *onPreCull* Unity camera event to ensure that all native rendering jobs are completed before displaying the final render texture to users. Finally, the rendered images are directly copied from CUDA memory to Unity textures on the GPU through a CUDA kernel.

Spatial Registration To correctly project 3D Gaussians in Unity, we converted the coordinate systems of 3DGS from COLMAP [192] to Unity. For easy registration of the 3D Gaussians with the Unity world space, we define a reference object in the real world with known scale, rotation, and position whose one-to-one digital copy is available in Unity. Through the reference object, we performed manual calibration to obtain a relative transformation matrix to register the 3DGS model. In addition, we also assume the reference object as the initial position of the robot for using the OpenCR's odometry sensor for markless tracking and pose estimation.

8.4.3 Unity Turtlebot Control Module

Overview We developed a control module in Unity for handling user inputs, managing robot motion, communicating with the remote ROS master, and visualizing real-time robot states. Users can control the robot's movement using the joysticks of the VR controllers with the robot's linear and angular speed linearly mapped to the joystick inputs. The communication module for sending and receiving ROS messages is based on the Unity-TCP-Connector package [2]. Messages with low bandwidth consumption such as odometry and IMU data are sent via the ROS-TCP connector. However, multi-media data such as

[2] https://github.com/Unity-Technologies/ROS-TCP-Connector

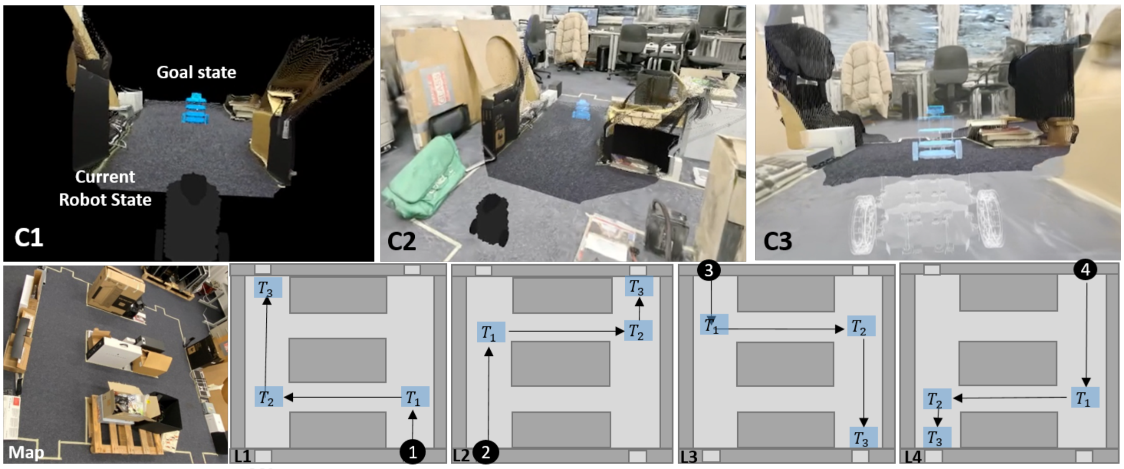


Figure 8.3: Illustration of our user study experiment design, with C1 showing the exocentric stereo projection condition, C2 showing the exocentric reality fusion condition, C3 showing the egocentric reality fusion condition. The second row illustrates the real-world maze and the four planned teleoperation trajectories.

videos and point clouds are transmitted via UDP for faster processing. To visualize the real-time poses of the robot, a digital twin of the robot is rendered whose transform is updated at every frame based on the pose estimation obtained from OpenCR odometry.

Exocentric Control As Figure 8.3 C2 demonstrates, our framework allows robust control of the robot from third-person (exocentric) perspective. In the exocentric control mode, an operator can observe the robot's current state in the world from any desired perspective. Users can use the joystick inputs of the VR controllers to translate their positions in the virtual world. While pressing the trigger button of the VR controllers, users can switch the joystick inputs to control the robot's movement instead.

Egocentric Control As Figure 8.3 C3 demonstrates, our framework also provides an egocentric robot control mode, where the operator sees the environment from the robot's perspective. In egocentric control mode, the user's head position in the virtual environment automatically follows the robot's tracked movement. In addition, we place the user's virtual head position right behind the virtual robot state indicator rather than at the position of the stereo camera so that operator is aware of the robot's position relative to its environment. Notice that in egocentric control mode, we only render half of the digital twin of the robot state indicator, such that the rendering does not obscure the operator's fovea vision of the real-time stereo projection.

Other Materials and Software Our framework was developed on Unity version 2019.4.29f1 based on OpenGL graphics API and uses the OpenVR desktop runtime and steamVR runtime. In addition, we use the MRTK version 2.8 to develop spatial UIs and to manage user inputs in VR.

8.4.4 Overall System Performance

Our framework can achieve real-time performance on both high-end and moderate workstations. On moderate hardware such as the Alienware m17 R2 laptop with an RTX 2080 graphics card, the overall performance of the entire system, including video decoding, stereo correspondence estimation, stereo projection, point cloud rendering, robot control, and 3DGS rendering runs at $40-45fps$ for an industrial facility environment with $622,335$ Gaussians at 900×960 per eye resolution (50% resolution of a Meta Quest Pro headset with 106° horizontal FoV, and 96° vertical FoV). On a high-end device such as the Nvidia RTX 3090 GPU, the overall performance is $30-35$ fps for a room-scale environment with $727,019$ Gaussians at 1536×1440 per eye resolution (80% of the maximum resolution of a Meat Quest Pro headset with 106° horizontal FoV, and 96° vertical FoV).

8.5 User Study Experiment

In this section, we present a user study experiment that benchmarks user performance and evaluates the effectiveness of our framework. The following research questions guided our user study and experiment design:

- **RQ1**: What is the effect of reality fusion on the operator's cognitive load, task performance, and situation awareness?
- **RQ2**: For reality fusion, how does the teleoperation perspective change users' performance and perception?

8.5.1 Conditions

We used a within-subject experiment design through which participants need to complete a mobile robot navigation task through the following three types of immersive robot teleoperation UIs in VR. Figure 8.3 C1-C3 presents application screenshots of the three different conditions.

C1 Exocentric Stereo Projection Only [72]: users see only the real-time stereo projection while navigating the robot and their position in the virtual world in exocentric control mode.

C2 Exocentric Reality Fusion: users see both the real-time stereo projection and 3DGS rendering while navigating the robot and their position in the virtual world in exocentric control mode.

C3 Egocentric Reality Fusion: users see both the real-time stereo projection and the 3DGS rendering while navigating the robot in egocentric control mode.

C1 is a reference VR system as proposed by Freland et al. as our comparison baseline [72]. As abundant previous research already revealed the superiority of robot teleoperation systems with VR HMD compared to conventional 2D displays and videos [72, 211], this study focuses on comparison across different immersive VR robot teleoperation designs only. The ordering of the three conditions for each participant is counter-balanced using a balanced Latin Square method to compensate for carry-over effects.

8.5.2 Participants

An a priori power analysis indicated a sample size of approximately $N = 27$ to detect medium effects ($\eta_p^2 = .006$) with 80% power using an ANOVA with an alpha level of .05. We invited 24 participants (10 female and 14 male) to make sure that each condition and route can be equally balanced. Two participants were between 18-24 years old, 17 were between 25-34 years old, 3 were between 35-44, 1 was between 45-54, and 1 was 65 years old or higher. All were students, researchers, or scientists in HCI, computer science, physics, or robotics. 10 participants use VR systems regularly (at least once a month), and 8 rarely use them (once or less than once a year). 10 participants never operated a robot before, 11 rarely operated a robot before, and 3 participants worked with a robot regularly. All participants had normal or correct to normal vision.

8.5.3 Tasks

As illustrated in Figure 8.3, we designed a $2.2m \times 2.2m$ maze with four symmetrical different entrance points. Inside the maze, there are three $0.6m \times 0.15m$ obstacle areas which form two $0.6m \times 0.875m$ pathways. We designed four different trajectories through which users can navigate the robot from one entrance of the maze to the target exit. Each trajectory consists of three subpaths and the level of difficulty for navigation of these subpaths is the same according to the steering law which can predict the amount of time (T) users need to navigate through a 2D tunnel given the width of the tunnel W and the length of the tunnel A: $T = a + b\frac{A}{W}$ [4]. As Figure 8.3 C1-C3 presents, a goal state indicator (rendered as blue) is presented to inform the participants of the target robot position they need to navigate the robot to. Participants need to sequentially navigate the robot to reach all three goal states ($T1, T2, T3$) in the designated trajectory for each condition. For each different condition, participants are assigned a different navigation trajectory. The maximum speed of the robot was adjusted to $0.05m/s$ linearly and $0.5rad/s$ angularly to ensure teleoperation safety.

8.5.4 Materials

The experiment was performed on a Meta Quest Pro headset and a Windows PC with a Nvidia 3090 GPU. The 3DGS model of the real-world maze was reconstructed from 69 images with 3990×2985 resolution. We generated the 3DGS model following the original model training pipeline developed by Kerbl et al. [107]. The training output was post-processed using the Unity 3DGS 2D editing toolkit developed by Pranckevičius

[179], where we removed outliners and erroneous results to improve the overall visual appearance. The post-processed 3DGS model has 727,019 Gaussians representing around $3.7m \times 3.7m \times 1.5m$ real-world volume which covers the entire maze and its surroundings. The robot teleoperation application runs at $30 - 35fps$ at 1536×1440 per eye resolution as recorded in 8.4.4.

8.5.5 Measures

Task Performance To objectively compare users' performance in different conditions, we record the total elapsed time for users to complete each condition, starting from the moment when the first goal state indicator was displayed in the HMD until the robot successfully reaches the last goal.

Perceived Workload For evaluating users' subjective perceived task loads, we use the standard NASA-TLX questionnaire, which measures various aspects of workload, including mental, physical, and temporal demand, as well as performance, effort, and frustration during the teleoperation tasks [88]. The original scale, which ranges from 1 to 100 in increments of 5, was converted to an 11-point Likert scale with labels *0 - perfect / 100 - failure* for performance and *0 - very low / 100 - very high* for all other items. A copy of the NASA-TLX questionnaire can be found in appendix A.1.2.

Situation Awareness For evaluating users' situation awareness of the remote environment, we used the Situation Awareness Rating Technique (SART) questionnaire [221]. The SART questionnaire consists of 9 questions that evaluate the supply (S), demand (D), and user understanding (U) on a 7-point Likert scale [221]. As proposed by Taylor et al., the final SA can be obtained from the following formula: $SA = U - (D - S)$. A copy of the SART can be found in Appendix A.1.6.

Cybersickness To measure the amount of induced cybersickness, we use a standard SSQ. The questionnaire was completed before the user study and immediately after each VR exposure [106]. As suggested by Kennedy et al., the questionnaires provide users' self-report ratings on common cybersickness symptoms on a 4-point Likert scale (ranging from *0 - none* to *3 - severe*) with the results grouped into nausea, oculomotor, and disorientation sub-scores as well as a total score. The questionnaire was completed before the user study and immediately after each VR exposure [106]. A copy of the SSQ can be found in the Appendix A.1.1.

Qualitative Feedbacks In a final questionnaire, participants were asked to indicate and describe their preferred conditions for teleoperating the robot as qualitative feedback.

8.5.6 Hypothesis

We formulate the following hypothesis concerning previously described measures and conditions:

Table 8.1: Means and standard deviations for Mental Demand, Physical Demand, and Temporal Demand from NASA-TLX with bold numbers indicate the best performance.

	Mental Demand		Physical Demand		Temporal Demand	
	M	**SD**	**M**	**SD**	**M**	**SD**
(C1)	6.13	3.89	3.71	3.69	3.08	2.86
(C2)	4.75	4.29	2.92	2.81	3.04	3.11
(C3)	**3.63**	3.67	**2.04**	1.74	**3.0**	3.91

Table 8.2: Means and standard deviations for Frustration, (Perceived) Performance, Efforts from NASA-TLX with bold numbers indicate the best performance.

	Peformance		Frustration		Efforts	
	M	**SD**	**M**	**SD**	**M**	**SD**
(C1)	4.79	4.08	3.29	2.41	4.16	2.65
(C2)	**3.08**	2.70	2.92	2.63	**2.50**	2.65
(C3)	3.21	3.80	**1.75**	1.13	2.96	4.24

(H1) Reality Fusion (C2, C3) leads to less perceived cognitive workload while improving the operator's situation awareness and overall performance.

(H2) Egocentric teleoperation (C3) results in better user performance and lower task load than exocentric teleoperation (C2). However, due to non-self-induced motion, C3 leads to higher motion sickness than C2.

8.6 Results

In this section, we present a summary of statistically significant results and discuss their design implications. The mean and standard deviation of the six subscales of the NASA-TLX index are recorded in Table 8.1 and Table 8.2. The overall ratings of usability, cybersickness, and overall task performance are recorded in Table 8.3. Each subscale of the SSQ is recorded in Table 8.3.

Task Performance Task performance, computed as total elapsed time (measured in seconds), was not normally distributed according to Shapiro-Wilk tests. Therefore, we performed a Friedman test which shows that there is a significant effect across conditions ($\chi^2(2) = 13.083, p = .001$). Wilcoxon signed-rank tests with Bonferroni-Holm adjustment show that participants perform significantly better in C2 than C1 ($Z = -2.543, p = .033, r = -.519$) and perform significantly better in C3 than C1 ($Z = -2.971, p = .009, r = -.606$).

Perceived Workload As the answers of NASA-TLX were not normally distributed for each condition, we performed Friedman tests for all sub-scales of the questionnaires. Friedman

Table 8.3: Means and standard deviations for each subscale of the SSQ. Best values are marked in **bold**.

	Nausea		Oculomotor		Disorientation	
	M	SD	M	SD	M	SD
(C1)	**71.67**	13.60	**65.92**	16.77	**108.00**	31.55
(C2)	72.42	13.17	66.29	18.00	111.5	36.02
(C3)	79.67	21.04	67.83	16.51	122.58	38.69

Table 8.4: Means and standard deviations for the task performance (i.e., task completion time, measured in seconds), situation awareness, and cybersickness. Best values are marked in **bold**.

	Task Performance		Situation Awareness		Cybersickness	
	M	SD	M	SD	M	SD
(C1)	161.75	69.93	12.46	4.20	**91.21**	22.72
(C2)	130.38	30.38	**17.29**	6.00	92.00	24.12
(C3)	**119.42**	24.49	16.71	5.11	98.46	26.03

tests show that there is a significant effect on mental demand ($\chi^2(2) = 8.000, p = .018$), performance ($\chi^2(2) = 6.194, p = .045$), and frustration ($\chi^2(2) = 8.758, p = .013$). Wilcoxon signed-rank tests with Bonferroni-Holm adjustment show that C1 has significantly higher mental demand ($Z = -2.584, p = .029, r = -.527$) and a higher level of frustration ($Z = -3.162, p = .005, r = -.645$) than C3. Post-hoc tests with adjustment did not reveal further significant differences in perceived performance.

Situation Awareness As the answers of SART follow a normal distribution according to Shapiro-Wilk tests, we compared all three conditions using a repeated-measures ANOVA test. The ANOVA test reveals a significant difference ($F(2, 46) = 7.691, p = 0.0013$). Post-hoc tests with Bonferroni adjustment show that participants obtained significantly higher situation awareness in C3 than in C1 ($p = 0.015$) as well as higher situation awareness in C2 than in C1 ($p = 0.003$).

Cybersickness SSQ was measured not only after each condition but also before the first condition, which allowed for analysis of carryover effects. However, no increase in cybersickness was observed throughout the study, which can be attributed to the experimental protocol that required participants to take breaks between conditions until potential cybersickness symptoms subsided. We, therefore, considered the absolute values of the SSQ instead of the differences between each measurement and the previous one as the dependent variable. Since residuals were not normally distributed, we performed Friedman tests on all sub-scales of the SSQ. Cybersickness significantly differed between the three conditions ($\chi^2(2) = 7.148, p = .028$). Multiple Wilcoxon signed-rank tests (the

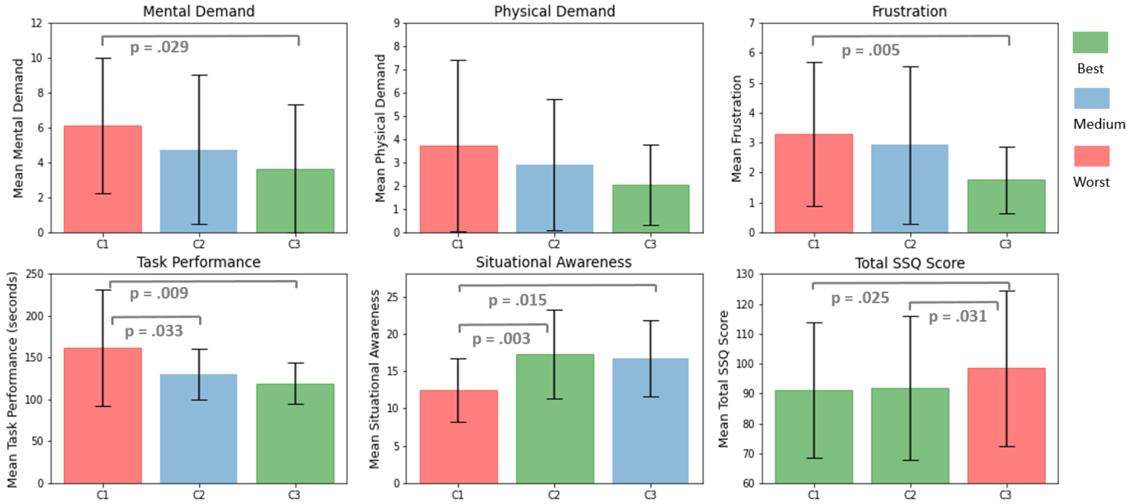


Figure 8.4: Mean mental demand, physical demand, frustration, task performance, situational awareness, and SSQ score per condition. Vertical bars represent the standard deviation. Any significant differences were labeled with their corresponding p values between conditions.

reported p values are adjusted with Holm-Bonferroni) reveal that C3 leads to significantly higher nausea ($Z = -2.539, p = .033, r = -.518$) and overall motion sickness symptoms ($Z = -2.633, p = .025, r = -.537$) than C1. Moreover, C3 also leads to significantly higher overall motion sickness than C2 ($Z = -2.565, p = .031, r = -.524$).

Preferences In the post-study questionnaire, 12 participants indicated clear preferences for C2, 8 participants indicated clear preferences for C3, 2 indicated equal preferences for C2 and C3, and 2 did not indicate any clear preferences.

8.7 Discussions

8.7.1 Improved Performance with Reality Fusion (H1)

As plotted in Figure 8.4, statistical analysis confirmed that reality fusion (C2, C3) results in significantly higher situation awareness and better task performance, partially confirming H1. According to participants' qualitative feedback, reality fusion enables "*a better overview/understanding of the whole environment*" (N=7), makes it easier to "*determine the robot's global position*" (N=3), and therefore makes it "*easier to plan routes*" (N=4). Without reality fusion, participants were "*not sure if there was a direct route between the robot and the target*" and therefore tend to "*scan around the environment much more*", resulting in worse task performance.

However, although C2 and C3 lead to lower overall mental demand, physical demand, and frustration, a comparison between C1 and C2 alone does not reveal a significant difference in perceived task load. Therefore, this part of H1 can not be confirmed. This indicates that displaying a global 3D map may introduce extra cognitive demands to users while these extra efforts help them achieve better task performance and gain more situational awareness.

In terms of SSQ, participants experienced only none to minor motion sickness in both C1 and C2, despite that users were exposed to more complex virtual environments in C2. This is attributed to our technical implementation, which is highly optimized for high-framerate and high-resolution rendering with low video streaming latency. Therefore, it is also safe to conclude that by using reality fusion, users can teleoperate robots in VR in real-time without discomfort.

8.7.2 Exocentric and Egocentric Comparison (H2)

In answering RQ2, we compare the results of different measures between C2 and C3. Although C3 leads to lower overall mental demand, physical demand, frustration, and better task performance than C2, the differences were not significant. Therefore, the first part of H2 can not be confirmed. Nonetheless, participants mentioned in the qualitative feedback that it is more demanding to have to control both the robot's movement and their movements in the virtual environment (N=4). In addition, they found the egocentric teleoperation mode (C3) more immersive and it helps them "*pay more attention to the real-time point cloud*" and become "*more aware of the (robot's) environment*".

The post-study qualitative questionnaire also revealed participants' split preferences for the two teleoperation modes, with those who preferred C3 believing that egocentric teleoperation is more "*natural*" and "*intuitive*" (N=2), while others preferring moving freely (N=3) and looking for the best perspective (e.g. a top-down view) on their own.

In terms of SSQ, as expected, participants reported stronger motion sickness symptoms in the egocentric mode due to continuous non-self-induced motion, with a significant difference between C3 and C2 in total SSQ score confirming the second part of H2. This indicates that while the egocentric teleoperation mode presents certain advantages, it might not be suitable for long-duration teleoperation tasks and could be offered as an option the user could switch to, rather than the main teleoperation mode.

8.7.3 Empirical Evaluation

As shown in Figure 8.5, to further examine the effectiveness of our framework in real-world accelerator facilities, we performed an exploratory empirical evaluation where the two operators of the MARWIN robot [58] tested the MR system at the EuXFEL LINAC tunnel. One of the operators found it easier to control the robot through the egocentric control mode, mentioning that " *it is nice if one can move the camera position close to the ground to see the robot's view, it is easier to operate the robot and see the environment from a perspective you can't normally see.* " Moreover, the operator mentioned that learning the input mapping on joysticks for controlling both the motion controls of one's movement in the virtual environment and the robot's motions can be "*confusing*". Future work could investigate utilizing other types of user inputs for separating users' self-motion from the

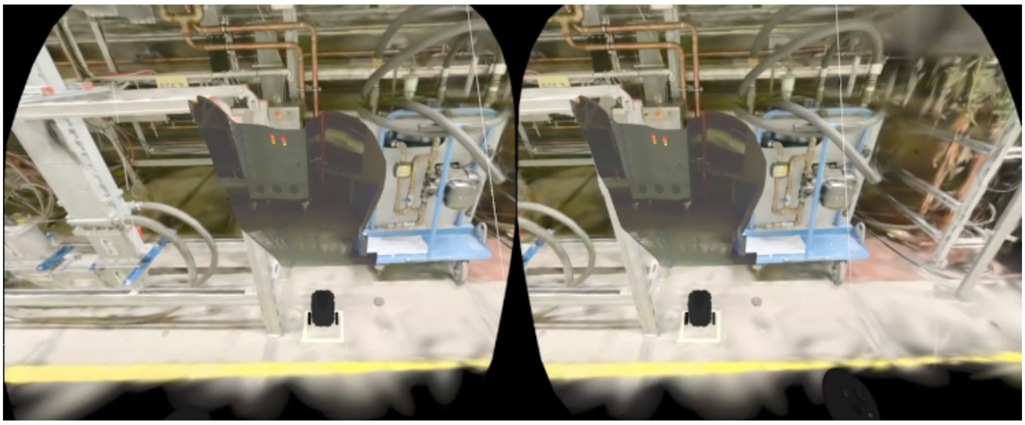


Figure 8.5: Screenshot of Empirical Onsite Testing of Reality Fusion at the EuXFEL LINAC tunnel.

robot's motion for the exocentric control mode, for example, by using a foot-controlled teleoperation design [95].

One major limitation of the current system implementation is that tracking the robot's poses in the environment relies on the OpenCR board's odometry, which obtains data from sensors like encoders to derive an estimated position relative to a starting point. In a non-robot-friendly environment such as particle accelerator tunnels, various factors could lead to tracking error accumulation, such as uneven floor layout and small obstacles on the path. Such motion drifts could result in errors in visual registration of the volumetric data fusion process, potentially leading to confusion and teleoperation errors. Removing such visual registration errors is the key to real-world application of our teleoperation system. The robot operator who designed and developed the MARWIN robot recommended adapting the current framework to the vison-based tracking method implemented for the MARWIN robot, which is based on 3D LiDAR SLAM combined with vision marker-based tracking. Such an existing tracking system already set up in the tunnel could achieve center-meter accuracy and mitigate the tracking errors caused by motion drifts and hardware failures.

8.8 Conclusion

In this chapter, we presented a novel immersive robot teleoperation framework that allows natural, intuitive, and robust remote control of mobile robots in complex semi-structured environments through the reality fusion technique. Our open-source implementation includes a high-performance Unity application for high resolution, photorealistic 3DGS VR rendering, low-latency point cloud streaming, and intuitive mobile robot motion control, as well as a telepresence mobile robot system design that can be easily replicated. We thoroughly evaluated various human aspects of our framework with 24 participants and demonstrated the significant improvement of reality fusion in objective task performance as well as perceived situation awareness.

In future work, we encourage researchers to improve the reality fusion methods by integrating dynamic SLAM capturing techniques to update the 3DGS global environments to extend our technique in accommodating more application domains where dynamic changes in the environments are involved.

9

Exploratory Systems and Designs

In this chapter, we present a collection of exploratory systems and designs that are related to the main research question of this dissertation but further investigate the technical and interaction design of the proposed application domains beyond the specific questions (**RQ1** - **RQ5**) introduced in Chapter 1. In Section 9.1, we develop a multi-modal HRI framework that allows users to robustly operate a mobile robot using intuitive gesture controls. This system demonstrates the potential of using diverse input modalities to further simplify the interactions between humans and robots at particle accelerators. Section 9.2 presents a novel conceptual design and the initial implementation of a CR version control framework based on NeRF to support the complex optical system development workflow at advanced optics and laser laboratories. Despite presenting only conceptual designs and initial implementations, the work in this chapter has been published as extended abstracts and short papers at international workshops and conferences. These publications provide a foundation for future research and development of MR interfaces at particle accelerators.

9.1 Multi-modal Robot Teleoperation Interfaces

9.1.1 Motivation

As introduced in Chapter 2, developing effective and easy-to-use operation and manipulation methods for mobile robots is an important research topic in the field of HRI. The robot teleoperation system we introduce in Chapter 8 primarily focuses on providing high-fidelity visual feedback to the operators through volumetric data fusion. In this section, we extend previous research in MR robot teleoperation by experimenting with multi-modal HRIs in immersive MR. To streamline the implementation and testing of new interaction designs, we establish the VAMCozmo framework, an open-source project for the development of novel VR, AR, or MR HRI methods using the Anki Cozmo education robot [1].In addition, we present several example interaction designs and implementations for common teleoperation tasks such as robot path planning through gesture controls or controller inputs.

In the following, we first provide an overview of the system architecture. Then, we describe the implementation and design of a path-following algorithm. Moreover, we present two interaction techniques for exocentric robot path planning in MR using different input modalities. Finally, we discuss the current limitations of the system and suggest future work. This section is primarily based on the following publication:

- **Ke Li**, Reinhard Bacher, Wim Leemans, Frank Steinicke. (2022). Towards Robust Exocentric Mobile Robot Tele-Operation in Mixed Reality. ACM Human-Robot Interaction (HRI) Workshop on Virtual, Augmented and Mixed Reality for Human-Robot Interaction (VAM-HRI). [https://openreview.net/pdf?id=HYIes841hJc]

The source code of the framework is made publicly available here: https://github.com/keli95566/VAMCozmo. A demo video is available here: https://youtu.be/WhreJMzNcKs.

9.1.2 System Setup

Figure 9.1 illustrates the basic components and structures of our MR mobile robot interaction system. The client side of the system is an application built with the popular game engine Unity, which can be run on any MR device that supports the OpenXR backend [110]. The MR client can be an OST HMD such as the Microsoft Hololens2 [2], or a VST HMD such as the Oculus Quest2 VR headset with passthrough API [164]. Using OpenXR backend allows our system to be compatible with a wide range of HMDs. Additionally, we use the MRTK, a popular MR development framework for creating immersive UIs.

Our system targets Anki Cozmo, an education mobile robot that is more accessible to a wider population than industrial-grade mobile robots. The Cozmo robot has four differential wheels and a movable lift. The Cozmo Driver [166] is a Python package that connects the Cozmo SDK with ROS to send commands and receive sensory feedback from the robot. We establish communication between the MR HMD client and the robot via TCP connection over the local network using the ROS-TCP connection package provided by Unity [224].

[1] https://www.digitaldreamlabs.com/pages/cozmo
[2] https://www.microsoft.com/en-us/hololens/

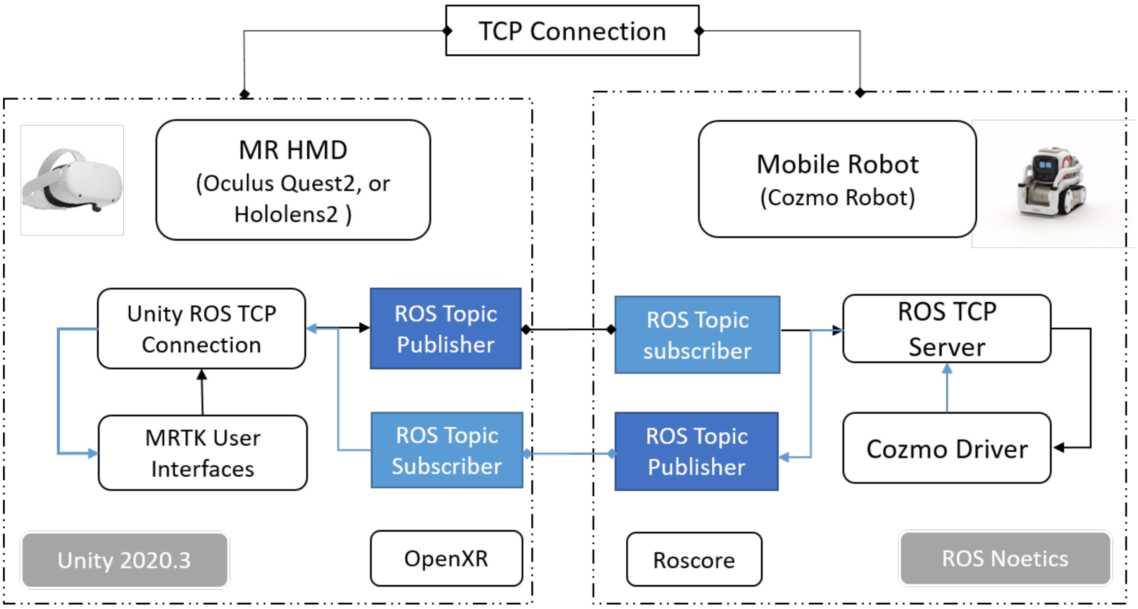


Figure 9.1: Illustration of the system overview: The system consists of an MR HMD client running on Open XR backend, and a Cozmo robot running on ROS. The robot and client are connected via TCP connection over a local network.

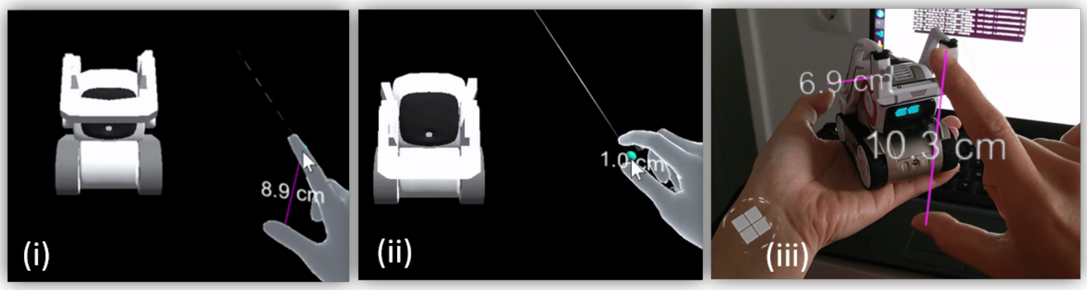


Figure 9.2: An example interaction with the Cozmo robot in MR with our system. Sub-figure (i) and (ii) illustrate our interaction designs in simulation. Sub-figure (iii) illustrates live control of the robot lift using our system with Microsoft Hololens 2.

The proposed system creates a simple and easy-to-scale design space for developing HRI techniques in MR and could be used for both VST HMD and OST HMD. Figure 9.2 illustrates an interaction example for the Cozmo robot and Hololens 2 using the proposed system. In the example, users could control the movable lift of the robot via hand tracking in real time. A line is rendered between the index and thumb fingers to measure and control the height of the movable lift.

9.1.3 Differential Drive Robot Kinematics

Anki Cozmo is a simple differential wheel drive robot. The movement and direction change of the robot depends on the relative rate of rotation of wheels on either side of the robot body and does not require additional steering. To correctly simulate a mobile

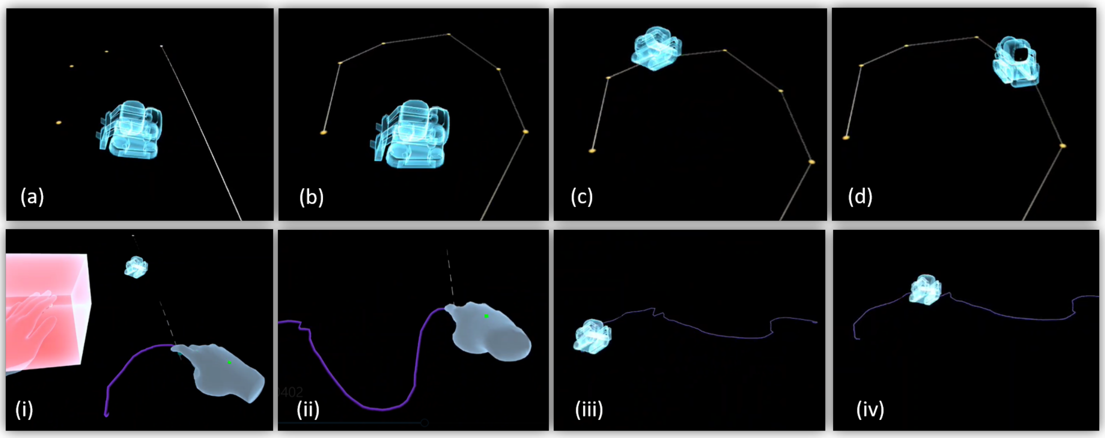


Figure 9.3: Illustration of two different MR mobile robot path planning interaction designs: Sub-figure (a-d) illustrate a raycast pointing method, where users define and select waypoints via hand or controller pointing. Sub-figure (i-iv) illustrates a free-hand drawing method, where the drawn trajectory is projected onto a 2D plane as the robot's final motion trajectory.

robot's locomotion behavior in MR, we implement an articulation wheel controller based on the classical kinematics model for differential drive robots [113]. In contrast to a typical two-differential-wheel drive robot, the Cozmo robot is a four-wheel differential drive robot. Therefore, to apply the classical kinematics model, we make the following assumptions:

1. The robot is moving on a 2D plane with constant friction. Therefore, the robot's locomotion state could be expressed by a state vector $R = (x, y, \theta)$, where x and y are the 2D positions of the robot, and θ is the angle that describes the direction the robot is facing.

2. The two wheels on the same sides drive at the same linear speed. Therefore, we could use the same kinematic model for a robot with two differential wheels for the four differential wheel Cozmo robot.

Given a target linear speed $v(m/s)$ and angular speed $\omega(rad/s)$, the expected linear speed of the right wheels (v_r) and the speed of the left wheels (v_l) of the robot could be calculated using the following equation:

$$v_r = \frac{l \cdot \omega}{2} + v \tag{9.1}$$

$$v_l = -\frac{l \cdot \omega}{2} + v \tag{9.2}$$

where l is the distance between the left wheel and the right wheel.

With the linear and angular velocity, we could calculate the joint speed of individual differential wheel $\omega_i = \frac{v_i}{R_i}$, where v_i is the target linear speed of the wheel, and R_i is the radius of the wheel.

9.1.4 Path-following Algorithm

Based on the differential drive kinematics model, we develop a simple path-following algorithm that allows robots to follow a pre-planned path defined through a list of way-points. As described in Algorithm 1, the path following algorithm utilizes the simplest forward kinematics commands such as turn in place and drive straight, and does not depend on an inverse kinematics calculation [187] which requires more constraint consideration.

Algorithm 1 A Simple Path Following Algorithm for Differential Drive Robot

Input: A list of way-points P that define the path for the robot to follow, and the robot's initial state $R_0 = (x_0, y_0, \theta_0)$.
Output: Robot trajectory

for each point $p(x, y)$ in P **do**
  while Robot not pointing towards the target point $p(x, y)$ **do**
    Turn robot in place with target angular velocity ω.
    Update robot state vector (x_i, y_i, θ_i).
  end while
  Calculate distance d between robot's current position (x_i, y_i) and the target point $p = (x, y)$.
  while Robot not traveling sufficient distance d **do**
    Drive robot forwards in the target direction with the target linear velocity v.
    Update robot state vector (x_i, y_i, θ_i).
  end while
end for

9.1.5 Multi-modal Interaction Designs for Path Following

As demonstrated in Chapter 8, one major advantage of an exocentric MR environment for mobile robot operation is the possibility of accurately planning the robot's future trajectory inside the entire virtual or MR environment. Figure 9.3 illustrates two interaction designs for trajectory planning of the mobile robot.

In the first design, the user can define the robot's future path via controller or hand pointing. An arrow-like pointer originates from the user's hand or controller and intersects with a point on the 2D plenary surface, thus providing accurate visual feedback of the pointed position. After selecting multiple target points in the 3D environment (Figure 9.3 a), the user could confirm the path (Figure 9.3 b), and the robot will sequentially move to different way-points selected by the users (Figure 9.3 c-d).

In the second design, the robot follows a trajectory that the user defines via free-hand drawing. The drawing begins with the user touching a trigger button (Figure 9.3 i). A trajectory is created by tracking the 3D position of the index finger of the user's hand (Figure 9.3 ii), and the drawing terminates with the user's other hand leaving the trigger button. Further, the initial list of the tracked 3D points is filtered, sampled, and optimized, before being projected onto the 2D floor. The robot then runs the path following algorithm

1 (Figure 9.3 iii-iv) by reaching different waypoints on the drawn trajectory.

Past research has shown that controller or hand pointing is the most effective selection method for selection task that requires accurate pointing, such as VR text selection and typing [203]. Due to a lack of visual feedback, freehand drawings could be less accurate and more confusing. However, it could offer a quicker and more intuitive way for the operator to draft and illustrate the general robot trajectory [186]. Although for accurate positioning of the robot, the controller and hand pointing method could be more practical, the freehand drawing method could be useful when it comes to path illustration or robot path drafting on a world in miniature (WIM) representation or a 2D map of the environment. Future user studies could evaluate and validate the system usability [21] and workload index [88] for each of the proposed interaction designs.

9.1.6 Discussion and Conclusion

In this section, we propose a system and several interaction designs for multi-modal robot operation in MR, utilizing gesture control and VR controllers as input methods. We implemented a basic articulation wheel controller and a path-following algorithm for an educational robot as a case study, presenting two MR interaction designs for robot trajectory planning. A significant future enhancement for the framework involves aligning the robot's simulated kinematic behavior in MR with real-world conditions. The current simulated kinematic model makes assumptions that overlook important real-world factors such as friction, skid, and wheel slip, which can cause deviations between the robot's actual trajectory and the simulated path. Implementing and testing an advanced kinematics model will enhance the accuracy and performance of our system and interaction designs.

9.2 RealityGit: Cross Reality Version Control

Multi-user collaboration at R&D workbenches in advanced optics laboratories involves updating and documenting the status of hundreds of components. Although there are abundant MR systems to support the collaboration of spatially co-located users, limited methods are available to facilitate cooperation in the spatial-temporal domain. Inspired by the version control workflow used in software development, we propose RealityGit, a novel CR system design that leverages the recent advancements of NeRF. We illustrate how the NeRF model can be used to channel users with different XR experiences by providing an accurate immersive visual documentation of the historical states of a complex R&D workbench. We demonstrate the feasibility of such a system through an implementation where users could contribute to the version control workflows by integrating historical NeRF scans into their MR devices or providing status reviews by annotating or editing a NeRF model in VR.

This section is primarily based on the following publication:

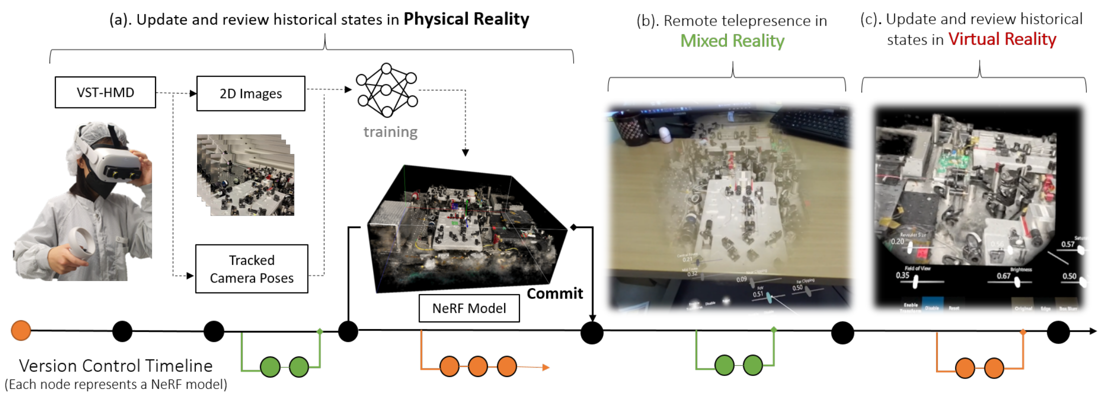


Figure 9.4: Illustration of the version control timeline of RealityGit, starting from (a), where a user wears a VST-HMD while physically present at the workbench. The VST-HMD saves the 3D states of the workbench by training a NeRF model and supplies these NeRF models to remote users or virtual users. (b) illustrates how remote users can contribute to the workflows by merging the current state of the NeRF model into their own workbench in MR. (c) illustrates how virtual users can contribute by performing assembly reviews and update tasks in immersive VR.

- **Ke Li**, Tim Rolff, Reinhard Bacher, Frank Steinicke. (2023). RealityGit: Cross Reality Version Control of R&D Optical Workbench. 2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). [DOI: https://doi.org/10.1109/ISMAR-Adjunct60411.2023.00178] [3]

A demo video of the design and concept is available here: https://youtu.be/JtIMI3W0naw

9.2.1 Motivation

As demonstrated in Chapter 5, optical R&D workbenches present a highly complex workspace involving many assembly components. This introduces a demanding spatial memory re-call challenge for researchers to keep track of each historical state of the system development process, especially for large projects where multiple users must collaborate on a workbench at different times. Although different visualizations of the setups such as a 3D CAD model could be brought to immersive MR to assist users with their workflows, the exact details of the workspace are often too complex to be accurately modeled in the traditional rasterization pipelines. Moreover, at R&D laboratories where iterative trial-and-error approaches are taken for rapid system prototyping, it is unrealistic to manually model every single step through conventional 3D modeling software.

To enable more efficient workflows when building setups at R&D workbench, we propose RealityGit, a CR version control system that could bring together users at different spatial and temporal locations with one unified version control timeline. As Figure 9.4 illustrates, the timeline records the accurate states of reality by converting the 2D image

[3]This work received the honorable mentioned award at the student design competition at ISMAR 2023

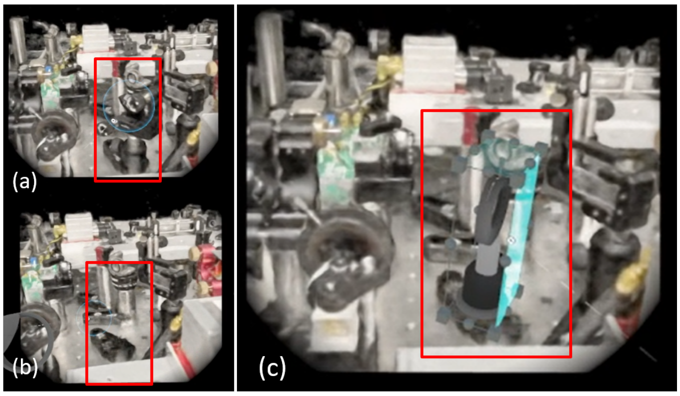


Figure 9.5: Application screenshots showing how a NeRF model can be asynchronously modified and updated in VR, with (a) showing the NeRF model before modification, and (b) showing the NeRF model after an optical component was erased using 3D NeRF drawing effect proposed in [131], and (c) showing how a virtual optical component is integrated onto the NeRF model to illustrate new design suggestions.

sequences captured by a VST-HMD to a NeRF [158]. Each NeRF model can be accessed by users at different points in time through different devices at different locations. For example, the NeRF model can be blended into a remote user's FoV at a real-world scale through an immersive MR device to provide a telepresence experience, where the remote user could also contribute to the physical setups. Additionally, users in VR can be engaged as well by reviewing the historical state of the development process while making suggestions for design updates and revisions. Finally, we present two system demonstrations showing the feasibility of our designs where both real-time MR and VR NeRF experiences are possible using instant-ngp [158] as training and rendering engine, and the Magic NeRF Lens framework as rendering and interaction framework [131].

9.2.2 The CR Version Control Workflows

Figure 9.4 presents an overview of the version control timeline of RealityGit with each node containing a 3D representation of the optical workbench. We distinguish 3 different node types, with a black node containing a NeRF model which records the actual physical state of the workbench, a green node containing a NeRF model edited or annotated by a remote user in real-time, and the orange nodes containing 3D representations asynchronously designed in VR.

An R&D project at advanced optics laboratories typically begins with a virtual 3D representation, such as a CAD model illustrating the initial design of the setup. Then, users start the initial prototyping processes at the physical optical workbench where users are equipped with a VST-HMD all the time to meet the laser health and safety standards. The VST-HMD can document each historical milestone by converting the 2D images captured by the see-through camera to a NeRF model upon the user's request. When a remote user can not be physically present at the laboratories, an MR branch can be opened, through which the user can annotate and edit the workbench through the recorded NeRF model to communicate with the onsite user synchronously. The onsite user can pull the change suggestions from the remote users and reconfigure the physical setup accordingly. As shown in Figure 9.4 (b), if the remote user also wears a stereoscopic VST-HMD, the NeRF model can display the remote scenario by seamlessly merging the NeRF model with the workbench of the remote user, potentially achieving realistic telepresence effects with the remote user having high situational awareness.

At different points in time, a VR branch can be opened asynchronously for virtual users to review different historical states of the setup. In the VR branch, users can edit and modify the physical design and create new designs based on the current or historical state of the physical setup rather than starting a new CAD modeling process from scratch. Figure 9.5 presents application screenshots of how a NeRF model could be asynchronously modified in the VR branch. To modify a NeRF model, the user can use the 3D NeRF drawing technique implemented by Li et.al in the Magic NeRF Lens framework [131] to dynamically and accurately erase a portion of the NeRF model. A virtual model can then be integrated and blended into the original NeRF model to demonstrate suggestions for future physical implementations, as shown in Figure 9.5 (c).

Similar to typical version control workflows, users working on the main branch (the actual physical workbench) can check out the designs of different users from the MR and VR branches, as well as the historical states of the physical setup in the main branch. Advanced version control features such as a comparison of the changes between the current physical state and a historical state can be potentially achieved by merging a historical NeRF model back to reality through the VST-HMD. Changes can be revealed spatially by comparing the images captured by the VST-HMD with the rendered NeRF images using techniques such as structural similarity index (SSIM).

9.2.3 System Demonstration

To demonstrate the feasibility of our design, we present two immersive XR experiences. Both experiences use the magic NeRF lens framework introduced in Chapter 7, where users can manipulate, edit, and update the NeRF model in immersive VR. The framework can be run on the Unity game engine version 2019.4 and uses instant-nap [158] as the backbone for NeRF training and rendering. To allow using the magic NeRF lens framework for users in immersive MR, we extend the previous system with a ZED Mini stereoscopic camera which can be mounted on an Oculus Quest 2 VR headset to provide high-resolution VST experiences. Both VR and MR NeRF experiences can be run in real-time at high resolution on a high-end graphics device such as an RTX 3090 GPU.

9.2.4 Discussion and Conclusion

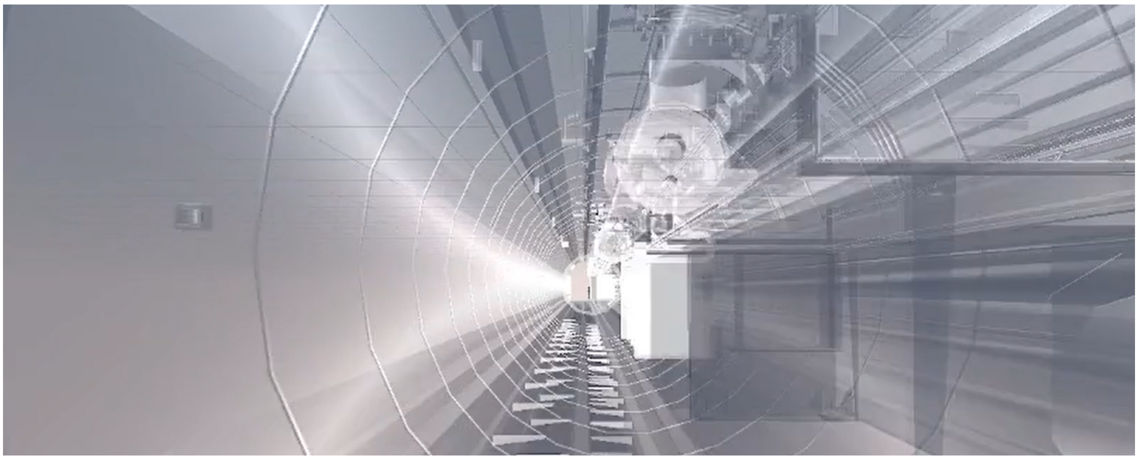
The possibility to collaborate across different spatial and temporal realities has the potential to significantly improve the efficiency of R&D workflows in empirical settings. For example, at high-power laser laboratories, onsite users' view of the workbench can be restricted, as users need to avoid moving their heads at the same height level of the optical workbench to prevent direct exposure to primary laser radiation. However, remote users or virtual users can freely navigate within the workbench through the NeRF model to perform close inspection and design tasks that are not possible in physical reality. Although this design concept mainly targets users at advanced optics laboratories, our CR design can potentially benefit users in other XR domains, such as electronic engineering, industrial manufacturing, and architectural planning.

Part IV

Conclusion

10

Summary



This dissertation investigated the core research question: *"How can MR technology improve HCI for users working in extreme, complex, and hazardous conditions such as high-energy laser facilities and particle accelerator tunnels?"* In Chapter 1, we motivated this research question and introduced three target application domains: **i)** using MR HMD as laser safety goggles, **ii)** immersive photorealistic visualizations for virtual facility inspection, and **iii)** teleoperation of robots at particle accelerator tunnels.

In Part II of the dissertation, we reviewed the general background of particle accelerators and introduced the key challenges of their operations. In addition, we reviewed various fundamental concepts of MR, ranging from visual displays, human vision systems, perception, as well as photorealistic 3D scene representations, eliciting their relevance to MR technology development at particle accelerators.

In Part III of the dissertation, we designed, implemented, and evaluated several novel MR interfaces through an HCD approach towards the "*particle accelerator metaverse*". In Chapter 5, we implemented a stereoscopic VST-HMD prototype with MR UIs for laser eye safety at advanced optics laboratories. We analyzed user requirements through field studies and user observations, and empirically evaluated the prototype with 14 laser experts at DESY for the first time. The empirical evaluation confirmed that the complex and hazardous working conditions at high-energy laser laboratories could be significantly improved with MR technology, not only in terms of laser eye safety but also in enhancing various aspects of the optical system development workflows. Based on the empirical evaluation and feedback, we thoroughly discussed and analyzed the advantages and limitations of VST-HMDs for laser safety, and introduced a wide range of future use cases and research directions towards the widespread utilization of VST-HMDs as "*digital laser safety goggles*".

Chapter 5 revealed that the initial VST-HMDs prototype still exhibits several limitations that mitigate their practicality for certain tasks such as fine-motor assembly work or those requiring locomotion within the laboratory environment. In Chapter 6, we presented MR tunneling, a novel method for stereoscopic untethered VST-HMD design through sensor fusion to balance the trade-off between limited rendering performance and high visual quality. We designed and implemented three MR tunneling effects and evaluated them in a within-subject study with 24 participants. The user study demonstrates the potential of our prototype based on the example of a puzzle assembly task that requires hand-eye coordination, spatial movement, and fine motor skills. The results confirm that the MR tunneling effects lead to higher overall usability, less perceived motion sickness, and a higher sense of presence with the foveated MR tunneling method being the most preferred design by users. Such a sensor fusion method could be potentially applicable in future laser safety VST-HMD solutions to improve the perceived resolution and framerate or to efficiently integrate external sensor views onto the user's vision field.

In Chapter 7, we presented *Magic NeRF Lens*, an open-source VR framework that aims to support virtual facility inspection through immersive photorealistic visualizations and interactions of NeRF. We introduced a data fusion technique to merge a NeRF model with the polygonal representation of its corresponding CAD model, which optimizes VR NeRF rendering through magic-lens-style interactions while introducing a novel industrial visualization design that can support practical tasks such as facility maintenance planning and redesign. Our data fusion method achieves a SUS score of > 80, indicating the system's high user acceptance and usability. Moreover, expert reviews further revealed the advantages of our framework compared to conventional virtual inspection software and confirmed the usability and applicability of the *Magic NeRF Lens* framework for real-world accelerator inspection tasks.

In Chapter 8, we introduced *Reality Fusion*, a novel robot teleoperation system that localizes, streams, projects, and merges a typical onboard depth sensor with a photorealistic, high resolution, high framerate, and wide FoV rendering of the complex accelerator environment represented as 3DGS. We systematically evaluated the *reality fusion* method through a user study where 24 participants performed robot steering tasks within a real-world maze [4]. The user study revealed that *Reality Fusion* enables robust egocentric and exocentric mobile robot teleoperating, leading to significantly better user performance, situation awareness, and user preferences.

Finally, in Chapter 9, we presented a collection of exploratory systems and designs, including an open-source framework aimed at streamlining the implementation and testing of multi-modal HRI designs, as well as a conceptual design and initial implementation of a CR version control workflows based on NeRF. These initial designs and implementations provide further technical foundation and design guidelines for future R&D of the "*particle accelerator metaverse*".

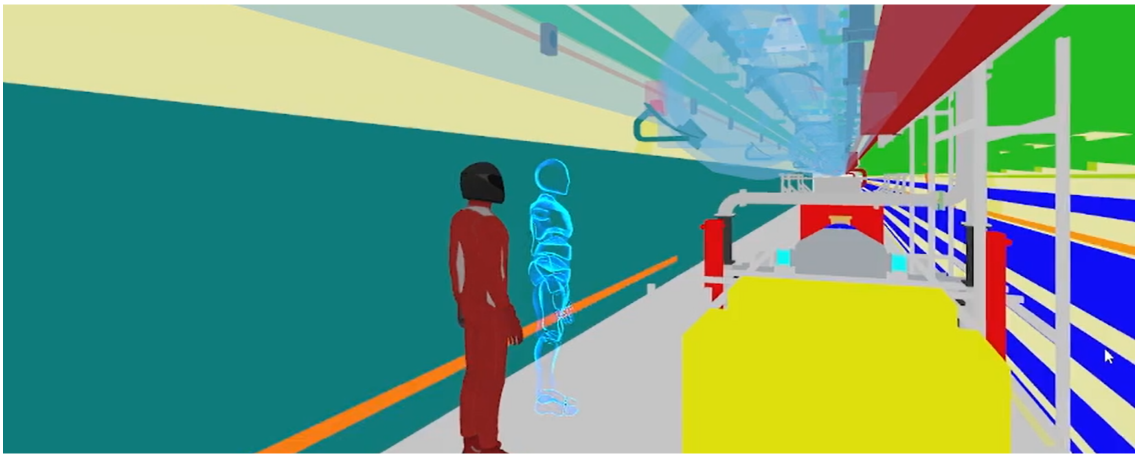
In summary, this dissertation presents the first step towards realizing a "*particle accelerator metaverse*", a concept where physical reality, virtual reality, artificial intelligence, and human intelligence converge within a shared MR space to streamline the operation and maintenance of large-scale physics facilities. Through a HCD process, we identified numerous areas in particle accelerator operations that could significantly benefit from MR technology. While certain applications such as robot teleoperation and virtual facility inspection are not unique to particle accelerators, the research results could be extended and benefit many other industrial settings as well. Other applications such as using VST-HMDs for laser safety are novel, diversifying the research and application domains of MR, and are poised to enhance the safety, productivity, and workflows of particle accelerator operation.

Despite demonstrating the feasibility and potential of MR in the identified application domains, the current systems have limitations that necessitate further research and development. For example, our mobile robot teleoperation system could be extended to a larger framework that supports robust teleoperation of more complex manipulators [217]. Although advancements in VST-HMD technology bring resolution and framerate closer to human visual capabilities, commercial-grade HMDs often overlook the specific human factors required in hazardous and complex optics laboratories. Therefore, specialized HMDs tailored for laser safety need ongoing research to enhance usability. The rendering of NeRF needs to be further accelerated to fully represent the large-scale accelerator facility at high resolution and framerate in immersive MR. The *Reality Fusion* and the *Magic NeRF Lens* framework could greatly benefit from research on a more robust and accurate 3D registration method for sensor fusion and data fusion. The development of a real-time dynamic SLAM algorithm for comparing and updating a “baked” 3DGS scene could further enhance the *Reality Fusion* teleoperation framework and present a feasible implementation for the *Reality Git* CR version control concept.

To conclude, the complex and hazardous operation conditions of particle accelerators require the development of specialized MR systems that are often beyond the capability of commercially available hardware and software systems. Therefore, we look forward to future research to further bring the “particle accelerator metaverse” into practice.

11

Outlook



As mentioned in Chapter 1, this dissertation only covers a small portion of what the future "*particle accelerator metaverse*" could look like. Given the complexity of designing, developing, maintaining, and upgrading modern accelerator facilities, many other aspects of particle accelerators could greatly benefit from immersive MR UIs. Future work could potentially investigate the following application domains: **i)** immersive and situated analytics for scientific visualization, **ii)** cross-modal MR interfaces for dexterous manipulators, and **iii)** remote collaboration and telepresence.

i) Immersive and Situated Analytics As introduced in Chapter 2, the operation of a modern particle accelerator generates a tremendous amount of data, ranging from machine operational data for facility inspection and monitoring to essential scientific data produced by beamline users. Immersive analytics presents enormous potential for big data visualizations by utilizing the entire 3D space available for interacting with large-scale, multi-dimensional data [67]. For instance, rather than examining facility operational data on a 2D screen, this data can be accessed through immersive situated visualization [32]. By displaying the data in situ within a digital twin of the accelerators, the data are positioned at the relevant location where they are generated, making them easier to retrieve and understand. Moreover, data analysis in an immersive environment could enable more effective data exploration [50]. Hundreds of thousands of imaging snapshots are created at free electron laser beamlines to generate 3D reconstructions of different samples, ranging from protein structures to nano-material surfaces. Observing these 3D structures in detail in immersive MR could enhance users' understanding of a wide range of complex data that is inherently spatial.

ii) Cross-modal MR for Dexterous Manipulators As introduced in Chapter 2, with the growing complexity of accelerator facilities, the development of safe robot intervention is crucial for more efficient accelerator operations. In Chapter 8, we developed an immersive robot teleoperation visualization framework for a custom-made mobile robot. However, real-world implementations of robot interventions require more complex robot platforms which include dexterous manipulators and robot arms whose motions are more difficult to program and accurately control [42]. To develop an intuitive interface for more advanced robot platforms, utilizing cross-modal MR interfaces presents enormous potential. For example, in Chapter 9, we presented how users could naturally interact with a simple robot manipulator using hand gestures, saving the efforts in learning the complex control input mapping of using a controller or keyboard. Combining similar types of natural user inputs such as eye gaze and speech with spatial computing and MR could enable more robust robot programming and more accurate communications of users' intents to the robot during teleoperation.

iii) Remote Collaboration and Telepresence As introduced in Chapter 3, an important goal of the "*metaverse*" is to connect users with different physical locations to a shared MR experience. While remote collaboration and immersive telepresence are well-established topics for MR research [70], their real-world implementation in extreme and hazardous environments like particle accelerator beamlines remains sparse. During the scheduled beam time, the accelerator facility needs to be continuously operated for 24 hours per day, seven days per week [3]. Onsite visits of an on-call expert could be rather expensive, particularly when the requests are made outside regular working hours. As illustrated in the conceptual design of the Microsoft Mesh application [100], an immersive MR telepresence system could significantly enhance remote collaboration between beamline scientists and users. This system would allow scientists to provide detailed instructions through virtual spatial presence instead of being physically onsite, thereby maximizing the efficient use of human resources and reducing travel costs. The *Reality Git* framework introduced in Chapter 11 presented initial designs of immersive telepresence and remote collaboration systems for complex physics facilities using NeRF. Future exploration of this application domain could improve our *Reality Git* framework by studying how to efficiently store, compare, and update a large number of historical states of 3D volumes represented as NeRF or 3DGS.

In summary, the complexity and scale of modern large-scale physics facilities present unique challenges and research opportunities for the HCI, MR, and robotics communities. Future work in these areas will not only advance MR technology but also pave the way for innovative solutions to address the complex HCI challenges in particle accelerator operations. This dissertation presents foundational concepts and systems that future research can build upon to fully realize the potential of a "*particle accelerator metaverse*".

Bibliography

[1] *1st Joint Workshop on Cross Reality*. Accessed: 2024-06-21. URL: https://cross-realities.org/ (cit. on p. 24).

[2] G. Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: 2012. URL: https://api.semanticscholar.org/CorpusID:16912492 (cit. on p. 14).

[3] R. Abela et al. "XFEL: The European X-Ray Free-Electron Laser - Technical Design Report". In: 2006. URL: https://api.semanticscholar.org/CorpusID:109047600 (cit. on pp. 15, 17, 19, 141).

[4] J. Accot and S. Zhai. "Performance evaluation of input devices in trajectory-based tasks: an application of the steering law". In: *International Conference on Human Factors in Computing Systems*. 1999. URL: https://api.semanticscholar.org/CorpusID:207247723 (cit. on pp. 17, 118, 138).

[5] G. Adamides et al. "Usability Guidelines for the Design of Robot Teleoperation: A Taxonomy". In: *IEEE Transactions on Human-Machine Systems* 45 (2015), pp. 256–262. URL: https://api.semanticscholar.org/CorpusID:39748728 (cit. on p. 7).

[6] B. Adelstein, T. Lee, and S. Ellis. "Head Tracking Latency in Virtual Environments: Psychophysics and a Model". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 47 (2003-10). DOI: 10.1177/154193120304702001 (cit. on p. 29).

[7] I. B. Adhanom et al. "The Effect of a Foveated Field-of-view Restrictor on VR Sickness". In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2020), pp. 645–652 (cit. on pp. 33, 68).

[8] D. Aiteanu, B. Hillers, and A. Gräser. "A step forward in manual welding: demonstration of augmented reality helmet". In: *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.* (2003), pp. 309–310 (cit. on p. 47).

[9] K. Akeley. "RealityEngine graphics". In: 2000 (cit. on pp. 40, 95).

[10] A. Akhmetzyanov et al. "Exploration of Underinvestigated Indoor Environment Based on Mobile Robot and Mixed Reality". In: *IHIET*. 2019 (cit. on p. 21).

[11] M. B. Alatise and G. P. Hancke. "A Review on Challenges of Autonomous Mobile Robot and Sensor Fusion Methods". In: *IEEE Access* 8 (2020), pp. 39830–39846 (cit. on p. 20).

[12] R. A. Albert et al. "Latency Requirements for Foveated Rendering in Virtual Reality". In: *ACM Transactions on Applied Perception (TAP)* 14 (2017), pp. 1–13 (cit. on p. 30).

[13] J. Allspaw, L. Heinold, and H. A. Yanco. "Design of Virtual Reality for Humanoid Robots with Inspiration from Video Games". In: *Interacción*. 2019. URL: https://api.semanticscholar.org/CorpusID:195877417 (cit. on p. 111).

[14] *Apple Vision Pro-Apple*. Accessed: 2024-05-23. URL: https://www.apple.com/apple-vision-pro/ (cit. on pp. 25, 29).

[15] C. M. Aragon et al. "Developing a Research Agenda for Human-Centered Data Science". In: *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion* (2016). URL: https://api.semanticscholar.org/CorpusID:22849993 (cit. on p. 3).

[16] S. Aukstakalnis and D. Blatner. "Silicon Mirage: The Art and Science of Virtual Reality". In: 1992. URL: https://api.semanticscholar.org/CorpusID:109556966 (cit. on pp. 87, 88).

[17] R. Bacher. "Web2cHMI: A Multi-Modal Native User Interface Implementation for Accelerator Operations and Maintenance Applications". In: 2017. URL: https://api.semanticscholar.org/CorpusID:218529522 (cit. on pp. 17–19).

[18] H. Bae, M. Golparvar-Fard, and J. White. "High-precision vision-based mobile augmented reality system for context-aware architectural, engineering, construction and facility management (AEC/FM) applications". In: *Visualization in Engineering* 1 (2013), pp. 1–13 (cit. on p. 4).

[19] M. Bajura. "Camera Calibration for Video See-Through Head-Mounted Display". In: 1993 (cit. on p. 71).

[20] M. Bajura and U. Neumann. "Dynamic registration correction in video-based augmented reality systems". In: *IEEE Computer Graphics and Applications* 15 (1995), pp. 52–60 (cit. on p. 29).

[21] A. Bangor, P. T. Kortum, and J. T. Miller. "An Empirical Evaluation of the System Usability Scale". In: *International Journal of Human–Computer Interaction* 24 (2008), pp. 574–594 (cit. on pp. 47, 57, 102, 131).

[22] M. S. Banks et al. "Stereoscopy and the human visual system". In: *SMPTE motion imaging journal* 121.4 (2012), pp. 24–43 (cit. on p. 73).

[23] D. Barčević et al. "A hand-held AR magic lens with user-perspective rendering". In: *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2012), pp. 197–206 (cit. on p. 87).

[24] H. Bargman. "Laser classification systems." In: *The Journal of clinical and aesthetic dermatology* 3 10 (2010), pp. 19–20 (cit. on p. 46).

[25] J. T. Barron et al. "Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields". In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (2021), pp. 5835–5844. URL: https://api.semanticscholar.org/CorpusID:232352655 (cit. on pp. 6, 41, 42, 86).

[26] P. Bartkiewicz and P. Duval. "TINE as an accelerator control system at DESY". In: *Measurement Science and Technology* 18 (2007), pp. 2379–2386. URL: https://api.semanticscholar.org/CorpusID:37034982 (cit. on p. 17).

[27] V. I. BeforeAfter. *3D Gaussian Splatting Editing on Smartphones, for the First Time*. 2024. URL: https://beforesandafters.com/2024/01/09/3d-gaussian-splatting-editing-on-smartphones-for-the-first-time/ (cit. on p. 43).

[28] P. J. Besl and N. D. McKay. "Method for registration of 3-D shapes". In: *Sensor fusion IV: control paradigms and data structures*. Vol. 1611. Spie. 1992, pp. 586–606 (cit. on p. 106).

[29] E. A. Bier et al. "Toolglass and magic lenses: the see-through interface". In: *Proceedings of the 20th annual conference on Computer graphics and interactive techniques* (1993) (cit. on p. 87).

[30] D. A. Bowman and R. P. McMahan. "Virtual Reality: How Much Immersion Is Enough?" In: *Computer* 40 (2007). URL: https://api.semanticscholar.org/CorpusID:18730123 (cit. on pp. 33, 34).

[31] D. A. Bowman et al. "3D User Interfaces: Theory and Practice". In: 2004. URL: https://api.semanticscholar.org/CorpusID:60454016 (cit. on pp. 18, 31, 32).

[32] N. Bressa et al. "What's the Situation with Situated Visualization? A Survey and Perspectives on Situatedness". In: *IEEE Transactions on Visualization and Computer Graphics* PP (2021), pp. 1–1. URL: https://api.semanticscholar.org/CorpusID:237235842 (cit. on p. 140).

[33] J. Brooke. "SUS: a "quick and dirty" usability scale". In: *Usability evaluation in industry* 189.3 (1996) (cit. on p. 76).

[34] S. V. Broucke and N. Deligiannis. "Visualization of Real-Time Heterogeneous Smart City Data Using Virtual Reality". In: *2019 IEEE International Smart Cities Conference (ISC2)* (2019), pp. 685–690. URL: https://api.semanticscholar.org/CorpusID:216042284 (cit. on p. 17).

[35] L. D. Brown and H. Hua. "Magic Lenses for augmented virtual environments". In: *IEEE Computer Graphics and Applications* 26 (2006), pp. 64–73 (cit. on p. 87).

[36] G. Bruder, F. Steinicke, and A. Nüchter. "Immersive Point Cloud Virtual Environments". In: 2014. URL: https://api.semanticscholar.org/CorpusID:46987885 (cit. on p. 111).

[37] G. Bruder et al. "Tuning self-motion perception in virtual reality with visual illusions". In: *IEEE Transactions on Visualization and Computer Graphics* 18.7 (2011), pp. 1068–1078 (cit. on p. 73).

[38] J. Butime et al. "3D RECONSTRUCTION METHODS, A SURVEY". In: 2018. URL: https://api.semanticscholar.org/CorpusID:4871280 (cit. on p. 37).

[39] S. Büttner et al. "The Design Space of Augmented and Virtual Reality Applications for Assistive Environments in Manufacturing: A Visual Approach". In: *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments* (2017). URL: https://api.semanticscholar.org/CorpusID:7271974 (cit. on pp. 3, 6, 85).

[40] C. P. Cain et al. "Retinal damage and laser-induced breakdown produced by ultrashort-pulse lasers". In: *Graefe's Archive for Clinical and Experimental Ophthalmology* 234 (2006), S28–S37 (cit. on p. 48).

[41] P. Caserman et al. "Cybersickness in current-generation virtual reality head-mounted displays: systematic review and outlook". In: *Virtual Reality* 25 (2021), pp. 1153–1170. URL: https://api.semanticscholar.org/CorpusID:234876584 (cit. on pp. 32, 33).

[42] M. di Castro, M. Ferre, and A. Masi. "CERNTAURO: A Modular Architecture for Robotic Inspection and Telemanipulation in Harsh and Semi-Structured Environments". In: *IEEE Access* 6 (2018), pp. 37506–37522 (cit. on pp. 3, 6, 19, 20, 85, 141).

[43] CERN. *Short history of particle accelerators*. [Accessed 09-04-2024]. URL: https://cas.web.cern.ch/sites/default/files/lectures/zakopane-2006/tazzari-history.pdf (cit. on p. 13).

[44] A. Chen et al. "TensoRF: Tensorial Radiance Fields". In: *ArXiv* abs/2203.09517 (2022). URL: https://api.semanticscholar.org/CorpusID:247519170 (cit. on p. 41).

[45] Z. Chen et al. "MobileNeRF: Exploiting the Polygon Rasterization Pipeline for Efficient Neural Field Rendering on Mobile Architectures". In: *The Conference on Computer Vision and Pattern Recognition (CVPR)*. 2023 (cit. on pp. 41, 42).

[46] CLARTE-LAB. *Gaussian Splatting VR Viewer*. 2023. URL: https://github.com/clarte53/GaussianSplattingVRViewerUnity (cit. on p. 115).

[47] *Coated filters Manufacture Page*. URL: https://www.uvex-laservision.de/en/laser-safety-eyewear/laser-safety-filter/coated-filters/ (cit. on pp. 48, 49).

[48] *Compact Laser Modules with USB Connector*. Accessed: 2022-08-05. URL: https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=12994&pn=PL202&YVI=1 (cit. on p. 57).

[49] R. Cools, A. Esteves, and A. L. Simeone. "Blending Spaces: Cross-Reality Interaction Techniques for Object Transitions Between Distinct Virtual and Augmented Realities". In: *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2022), pp. 528–537. URL: https://api.semanticscholar.org/CorpusID:255188689 (cit. on p. 26).

[50] M. Cordeil et al. "IATK: An Immersive Analytics Toolkit". In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2019), pp. 200–209. URL: https://api.semanticscholar.org/CorpusID:201067338 (cit. on p. 141).

[51] F. Cutolo, P. D. Parchi, and V. Ferrari. "Video see through AR head-mounted display for medical procedures". In: *ISMAR*. 2014 (cit. on p. 26).

[52] F. Cutolo et al. "Hybrid Video/Optical See-Through HMD". In: *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)* (2017), pp. 52–57 (cit. on p. 62).

[53] M. Czernuszenko et al. "The ImmersaDesk and Infinity Wall projection-based virtual reality displays". In: *COMG*. 1997. URL: https://api.semanticscholar.org/CorpusID:16370783 (cit. on p. 26).

[54] F. Dai et al. "Virtual Reality for Industrial Applications". In: *Computer Graphics: Systems and Applications*. 1997. URL: https://api.semanticscholar.org/CorpusID:6386523 (cit. on pp. 3, 26, 85).

[55] E. J. David, J. Beitner, and M. L.-H. Võ. "The importance of peripheral vision when searching 3D real-world scenes: A gaze-contingent study in virtual reality". In: *Journal of Vision* 21 (2021) (cit. on p. 73).

[56] J. Davis, Y.-H. Hsieh, and H.-C. Lee. "Humans perceive flicker artifacts at 500Hz". In: *Scientific Reports* 5 (2015). URL: https://api.semanticscholar.org/CorpusID:16585986 (cit. on p. 29).

[57] F. De Pace, F. Manuri, and A. Sanna. "Augmented reality in industry 4.0". In: *Am. J. Comput. Sci. Inf. Technol* 6.01 (2018), pp. 1–7 (cit. on pp. 74, 75).

[58] A. Dehne et al. "MARWIN: A Mobile Autonomous Robot for Maintenance and Inspection". In: *Proc. 16th Int. Conf. on Accelerator and Large Experimental Physics Control Systems (ICALEPCS'17)*. 2017, pp. 76–80 (cit. on pp. 6, 7, 16, 19–21, 85, 123).

[59] N. Deng et al. "FoV-NeRF: Foveated Neural Radiance Fields for Virtual Reality". In: *IEEE Transactions on Visualization and Computer Graphics* 28 (2021), pp. 3854–3864 (cit. on p. 41).

[60] C. Dewen et al. “Design and manufacture AR head-mounted displays: A review and outlook”. In: *Light: Advanced Manufacturing* (2021). URL: https://api.semanticscholar.org/CorpusID:239725976 (cit. on pp. 27, 28).

[61] A. Distante and C. Distante. “Camera Calibration and 3D Reconstruction”. In: 2020 (cit. on p. 72).

[62] A. X. Dong et al. “Mitigating bit flips or single event upsets in epilepsy neurostimulators”. In: *Epilepsy & Behavior Case Reports* 5 (2016), pp. 72–74. URL: https://api.semanticscholar.org/CorpusID:16345961 (cit. on p. 21).

[63] A. Douplik et al. “3 - The response of tissue to laser light”. In: *Lasers for Medical Applications*. Ed. by H. Jelínková. Woodhead Publishing Series in Electronic and Optical Materials. Woodhead Publishing, 2013, pp. 47–109. ISBN: 978-0-85709-237-3. DOI: https://doi.org/10.1533/9780857097545.1.47. URL: https://www.sciencedirect.com/science/article/pii/B9780857092373500035 (cit. on p. 47).

[64] P. Duval et al. “Automating Operation Statistics at PETRA-3”. In: 2017. URL: https://api.semanticscholar.org/CorpusID:65181901 (cit. on p. 18).

[65] D. Eggert. *What is foveated rendering?* Accessed: 2024-05-23. URL: https://www.tobii.com/blog/what-is-foveated-rendering (cit. on p. 30).

[66] B. Ens, J. D. Hincapié-Ramos, and P. Irani. “Ethereal planes: a design framework for 2D information space in 3D mixed reality environments”. In: *Proceedings of the 2nd ACM symposium on Spatial user interaction* (2014). URL: https://api.semanticscholar.org/CorpusID:12579845 (cit. on p. 26).

[67] B. Ens et al. “Grand Challenges in Immersive Analytics”. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021). URL: https://api.semanticscholar.org/CorpusID:231836916 (cit. on p. 140).

[68] A. Erickson et al. “Is It Cold in Here or Is It Just Me? Analysis of Augmented Reality Temperature Visualization for Computer-Mediated Thermoception”. In: *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2019), pp. 202–211 (cit. on pp. 67, 69).

[69] W. J. Ertle, K. M. Donnelly, and C. E. Moss. “Laser accidents: A fifty-year review of medical and non-medical exposure C101”. In: *International Laser Safety Conference* 2011.1 (2011), pp. 280–283. eprint: https://lia.scitation.org/doi/pdf/10.2351/1.5056762. URL: https://lia.scitation.org/doi/abs/10.2351/1.5056762 (cit. on p. 47).

[70] A. J. Fairchild et al. “A Mixed Reality Telepresence System for Collaborative Space Operation”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 27 (2017), pp. 814–827. URL: https://api.semanticscholar.org/CorpusID:29646584 (cit. on p. 141).

[71] X. Fan et al. “Stereoscopic Image Stitching via Disparity-Constrained Warping and Blending”. In: *IEEE Transactions on Multimedia* 22 (2020), pp. 655–665 (cit. on p. 67).

[72] F. Ferland et al. “Egocentric and exocentric teleoperation interface using real-time, 3D video projection”. In: *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (2009), pp. 37–44. URL: https://api.semanticscholar.org/CorpusID:2810636 (cit. on pp. 7, 110, 113, 117, 118).

[73] J. P. Freiwald, N. Katzakis, and F. Steinicke. “Camera Time Warp: Compensating Latency in Video See-Through Head-Mounted-Displays for Reduced Cybersickness Effects”. In: *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (2018), pp. 49–50 (cit. on pp. 32, 33, 53, 68).

[74] J. Frey et al. “Assessing the zone of comfort in stereoscopic displays using EEG”. In: *CHI ’14 Extended Abstracts on Human Factors in Computing Systems* (2014). URL: https://api.semanticscholar.org/CorpusID:6402323 (cit. on p. 31).

[75] W.-k. Fung et al. “A case study of 3D stereoscopic vs. 2D monoscopic tele-reality in real-time dexterous teleoperation”. In: *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems* (2005), pp. 181–186 (cit. on p. 49).

[76] C. G. R. Geddes et al. “High-quality electron beams from a laser wakefield accelerator using plasma-channel guiding”. In: *Nature* 431 (2004), pp. 538–541. URL: https://api.semanticscholar.org/CorpusID:4342539 (cit. on p. 5).

[77] *Gestalt Robotics - Intelligente Automatiserung*. Accessed: 2024-05-25. URL: https://www.gestalt-robotics.com/ (cit. on p. 25).

[78] Gigahertz-Optik. *Basic radiometric quantities*. Accessed: 2024-05-25. URL: https://light-measurement.com/basic-radiometric-quantities/ (cit. on p. 39).

[79] L. Gong et al. “Development of virtual reality support to factory layout planning”. In: *International Journal on Interactive Design and Manufacturing (IJIDeM)* (2019), pp. 1–11 (cit. on pp. 85, 93).

[80] U. Gruenefeld et al. “VRception: Rapid Prototyping of Cross-Reality Systems in Virtual Reality”. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022). URL: https://api.semanticscholar.org/CorpusID:248419869 (cit. on p. 24).

[81] Gstreamer. *Gstreamer: open source multimedia framework*. 2022. URL: https://gstreamer.freedesktop.org/ (visited on 2022-05-30) (cit. on p. 70).

[82] E. Gusai et al. “Interaction in an Immersive Collaborative Virtual Reality Environment: A Comparison Between Leap Motion and HTC Controllers”. In: *ICIAP Workshops*. 2017 (cit. on p. 54).

[83] X.-F. Han, H. Laga, and Bennamoun. “Image-Based 3D Object Reconstruction: State-of-the-Art and Trends in the Deep Learning Era”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43 (2019), pp. 1578–1604. URL: https://doi.org/10.1109/TPAMI.2019.2954885 (cit. on pp. 37, 38).

[84] A. Haque et al. “Instruct-NeRF2NeRF: Editing 3D Scenes with Instructions”. In: (2023) (cit. on p. 41).

[85] C. S. Harris. “Adaptation to Displaced Vision: Visual, Motor, or Proprioceptive Change?” In: *Science* 140 (1963), pp. 812–813 (cit. on p. 61).

[86] C. S. Harris. “Perceptual adaptation to inverted, reversed, and displaced vision.” In: *Psychological review* 72 6 (1965), pp. 419–44 (cit. on p. 61).

[87] S. G. Hart. “Nasa-Task Load Index (NASA-TLX); 20 Years Later”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50 (2006), pp. 904–908. URL: https://api.semanticscholar.org/CorpusID:6292200 (cit. on pp. 34, 35).

[88] S. G. Hart and L. E. Staveland. “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research”. In: *Advances in psychology* 52 (1988), pp. 139–183 (cit. on pp. 35, 76, 102, 119, 131).

[89] M. Hofbauer et al. “Adaptive Multi-View Live Video Streaming for Teledriving Using a Single Hardware Encoder”. In: *2020 IEEE International Symposium on Multimedia (ISM)* (2020), pp. 9–16. URL: https://api.semanticscholar.org/CorpusID:231715981 (cit. on p. 109).

[90] D. M. Hoffman et al. “Vergence-accommodation conflicts hinder visual performance and cause visual fatigue.” In: *Journal of vision* 8 3 (2008), pp. 33.1–30. URL: https://api.semanticscholar.org/CorpusID:23021920 (cit. on p. 31).

[91] *Horizon Workrooms: Virtual Office and Meetings | Meta for Work*. Accessed: 2024-05-25. URL: https://forwork.meta.com/de/en/horizon-workrooms/ (cit. on p. 25).

[92] I. P. Howard and B. Rogers. “Binocular Vision and Stereopsis”. In: 1996. URL: https://api.semanticscholar.org/CorpusID:54161880 (cit. on p. 31).

[93] C.-Y. Hsu et al. “Fish-Eye Lenses-Based Camera Calibration and Panoramic Image Stitching”. In: *2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)* (2018), pp. 1–2 (cit. on p. 69).

[94] H. Hua and B. Javidi. “A 3D integral imaging optical see-through head-mounted display.” In: *Optics express* 22 11 (2014), pp. 13484–91. URL: https://api.semanticscholar.org/CorpusID:30367704 (cit. on p. 27).

[95] Y. Huang et al. “Design and Evaluation of a Foot-Controlled Robotic System for Endoscopic Surgery”. In: *IEEE Robotics and Automation Letters* 6 (2021), pp. 2469–2476. URL: https://api.semanticscholar.org/CorpusID:232316474 (cit. on p. 124).

[96] K. A. Hussey, S. E. Hadyniak, and R. J. Johnston. "Patterning and Development of Photoreceptors in the Human Retina". In: *Frontiers in Cell and Developmental Biology* 10 (2022). URL: https://api.semanticscholar.org/CorpusID:248153484 (cit. on pp. 29, 30).

[97] J. S. Hvass et al. "Visual realism and presence in a virtual reality game". In: *2017 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)* (2017), pp. 1–4. URL: https://api.semanticscholar.org/CorpusID:46804926 (cit. on p. 34).

[98] M. J. Ibahim et al. "An Evaluation of Dose Equivalence between Synchrotron Microbeam Radiation Therapy and Conventional Broadbeam Radiation Using Clonogenic and Cell Impedance Assays". In: *PLoS ONE* 9 (2014). URL: 10.1371/journal.pone.0100547 (cit. on p. 14).

[99] T. F. O. A. Inc. *Safety In Working With Optical Fiber*. 2018. URL: https://www.thefoa.org/tech/ref/safety/safe.html (visited on 2022-08-11) (cit. on p. 52).

[100] *Introducing Microsoft Mesh | Connect Like Never Before*. Accessed: 2024-05-25. URL: https://www.microsoft.com/en-us/microsoft-teams/microsoft-mesh (cit. on pp. 25, 141).

[101] C. Jambon et al. "NeRFshop: Interactive Editing of Neural Radiance Fields"". In: *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 6.1 (2023-05). URL: https://repo-sam.inria.fr/fungraph/nerfshop/ (cit. on p. 41).

[102] L. Jensen and F. Konradsen. "A review of the use of virtual reality head-mounted displays in education and training". In: *Education and Information Technologies* 23 (2017), pp. 1515–1529 (cit. on p. 34).

[103] J. Jerald. "The VR Book: Human-Centered Design for Virtual Reality". In: 2015. URL: https://api.semanticscholar.org/CorpusID:108749274 (cit. on pp. 4, 5, 17).

[104] S. J. Jorgensen et al. "Cockpit Interface for Locomotion and Manipulation Control of the NASA Valkyrie Humanoid in Virtual Reality (VR)". In: *NASA New Technology Report (NTR): MSC-27278-1* (2022). URL: https://ntrs.nasa.gov/citations/20220007587 (cit. on p. 111).

[105] H. Kato and M. Billinghurst. "Marker tracking and HMD calibration for a video-based augmented reality conferencing system". In: *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)* (1999), pp. 85–94 (cit. on p. 27).

[106] R. S. Kennedy et al. "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness". In: *The international journal of aviation psychology* 3.3 (1993), pp. 203–220 (cit. on pp. 33, 75, 103, 119).

[107] B. Kerbl et al. "3D Gaussian Splatting for Real-Time Radiance Field Rendering". In: *ACM Transactions on Graphics (TOG)* 42 (2023), pp. 1–14. URL: https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/ (cit. on pp. 36, 38, 42, 106, 110, 112, 115, 118).

[108] H. Kharoub, M. Lataifeh, and N. Ahmed. "3D User Interface Design and Usability for Immersive VR". In: *Applied Sciences* (2019). URL: https://api.semanticscholar.org/CorpusID:209796781 (cit. on p. 34).

[109] M. Khorasaninejad and F. Capasso. "Metalenses: Versatile multifunctional photonic components". In: *Science* 358 (2017). URL: https://api.semanticscholar.org/CorpusID:3794855 (cit. on p. 27).

[110] Khronos. *OpenXR*. 2021. URL: https://www.khronos.org/openxr/ (cit. on p. 127).

[111] H.-S. Kim et al. "Improvement of Realism of 4D Objects Using Augmented Reality Objects and Actual Images of a Construction Site". In: *KSCE Journal of Civil Engineering* 22 (2018), pp. 2735–2746 (cit. on p. 19).

[112] H. K. Kim et al. "Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment." In: *Applied ergonomics* 69 (2018), pp. 66–73 (cit. on pp. 57, 60).

[113] G. Klanar, D. Matko, and S. Blai. "Mobile Robot Control on a Reference Path". In: *Proceedings of the 2005 IEEE International Symposium on, Mediterrean Conference on Control and Automation Intelligent Control, 2005.* (2005), pp. 1343–1348 (cit. on p. 129).

[114] K. Kobayashi, S. Ishigame, and H. Kato. "Skill training system of manual arc welding". In: *IWEC*. 2002 (cit. on p. 3).

[115] P. Kourtesis et al. "Cybersickness, Cognition, & Motor Skills: The Effects of Music, Gender, and Gaming Experience". In: *IEEE Transactions on Visualization and Computer Graphics* 29 (2023), pp. 2326–2336. URL: https://api.semanticscholar.org/CorpusID:257127571 (cit. on p. 33).

[116] G. Kramida. "Resolving the Vergence-Accommodation Conflict in Head-Mounted Displays". In: *IEEE Transactions on Visualization and Computer Graphics* 22 (2016), pp. 1912–1931 (cit. on pp. 31, 60).

[117] S. G. Kratz et al. "Evaluating Stereoscopic Video with Head Tracking for Immersive Teleoperation of Mobile Telepresence Robots". In: *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts* (2015). URL: https://api.semanticscholar.org/CorpusID:263975 (cit. on p. 7).

[118] C. Kyrlitsias and D. Michael-Grigoriou. "Social Interaction With Agents and Avatars in Immersive Virtual Environments: A Survey". In: *Frontiers in Virtual Reality*. 2022. URL: https://api.semanticscholar.org/CorpusID:245859496 (cit. on p. 25).

[119] E. Langbehn et al. "Bending the Curve: Sensitivity to Bending of Curved Paths and Application in Room-Scale VR". In: *IEEE Transactions on Visualization and Computer Graphics* 23 (2017), pp. 1389–1398. URL: https://api.semanticscholar.org/CorpusID:7112698 (cit. on p. 34).

[120] *Laser Safety Filter T1C02 Manufacture Page*. URL: https://www.uvex-laservision.de/en/laser-safety-eyewear/laser-safety-eyewear-by-frame/r14/laser-safety-goggle-r14t1b09l/ (cit. on pp. 48, 62).

[121] *Laser Safety Spectacle R14T1B09L Manufacture Page*. URL: https://www.uvex-laservision.de/en/laser-safety-eyewear/laser-safety-eyewear-by-frame/r14/laser-safety-goggle-r14t1b09l/ (cit. on pp. 46, 48, 56, 61).

[122] *Laser Safety Spectacle R14T1L05L Manufacture Page*. URL: https://www.uvex-laservision.de/en/laser-safety-eyewear/laser-safety-eyewear-by-frame/r14/laser-afety-spectacle-r14t1l05l/ (cit. on p. 56).

[123] *Laser Safety Spectacle R14T1P04A Manufacture Page*. URL: https://www.uvex-laservision.de/en/laser-safety-eyewear/laser-safety-eyewear-by-frame/r14/laser-safety-spectacle-r14t1p04a/ (cit. on p. 46).

[124] V. Lazova et al. "Control-NeRF: Editable Feature Volumes for Scene Rendering and Manipulation". In: *arXiv preprint arXiv:2204.10850* (2022) (cit. on p. 41).

[125] J.-Y. Lee et al. "Estimating the simulator sickness in immersive virtual reality with optical flow analysis". In: *SIGGRAPH Asia 2017 Posters* (2017) (cit. on pp. 33, 89).

[126] Y.-h. Lee, T. Zhan, and S.-T. Wu. "Prospects and challenges in augmented reality displays". In: *Virtual Real. Intell. Hardw.* 1 (2019), pp. 10–20. URL: https://api.semanticscholar.org/CorpusID:67896618 (cit. on p. 27).

[127] D. Li et al. "12 - Understanding the impact of building thermal environments on occupants' comfort and mental workload demand through human physiological sensing". In: *Start-Up Creation (Second Edition)*. Ed. by F. Pacheco-Torgal et al. Second Edition. Woodhead Publishing Series in Civil and Structural Engineering. Woodhead Publishing, 2020, pp. 291–341. ISBN: 978-0-12-819946-6. DOI: https://doi.org/10.1016/B978-0-12-819946-6.00012-6. URL: https://www.sciencedirect.com/science/article/pii/B9780128199466000126 (cit. on p. 35).

[128] K. Li et al. "Single Shot Phase Shift 3D Scanning with Convolutional Neural Network and Synthetic Fractals". In: 2022 (cit. on p. 37).

[129] K. Li et al. "Bringing Instant Neural Graphics Primitives to Immersive Virtual Reality". In: *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (2023), pp. 739–740. URL: https://api.semanticscholar.org/CorpusID:258445556 (cit. on p. 36).

[130] K. Li et al. "Immersive Neural Graphics Primitives". In: *ArXiv* abs/2211.13494 (2022) (cit. on pp. 86, 91, 95).

[131] K. Li et al. "Magic NeRF Lens: Interactive Fusion of Neural Radiance Fields for Virtual Facility Inspection". In: *ArXiv* abs/2307.09860 (2023). URL: https://api.semanticscholar.org/CorpusID:259982525 (cit. on pp. 41, 133, 134).

[132] K. Li et al. "Mixed Reality Tunneling Effects for Stereoscopic Untethered Video-See-Through Head-Mounted Displays". In: *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2022), pp. 44–53. URL: https://api.semanticscholar.org/CorpusID:255188395 (cit. on pp. 23, 87, 89, 100, 114).

[133] K. Li et al. "Taming Cyclops: Mixed Reality Head-Mounted Displays as Laser Safety Goggles for Advanced Optics Laboratories". In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (2022), pp. 544–545 (cit. on pp. 6, 26, 47, 67, 75).

[134] R. Li, M. Tancik, and A. Kanazawa. "NerfAcc: A General NeRF Acceleration Toolbox". In: *ArXiv* abs/2210.04847 (2022) (cit. on p. 40).

[135] S. Li et al. "SteerNeRF: Accelerating NeRF Rendering via Smooth Viewpoint Trajectory". In: *ArXiv* abs/2212.08476 (2022) (cit. on pp. 40, 41).

[136] C. Liao et al. "Long-term effect of YAG laser iridotomy on corneal endothelium in primary angle closure suspects: a 72-month randomised controlled study". In: *British Journal of Ophthalmology* 105 (2020), pp. 348–353 (cit. on p. 48).

[137] S. Liu, D. Cheng, and H. Hua. "An optical see-through head mounted display with addressable focal planes". In: *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality* (2008), pp. 33–42 (cit. on pp. 27, 28).

[138] S. Liu et al. "Learning to Infer Implicit Surfaces without 3D Supervision". In: *Neural Information Processing Systems*. 2019. URL: https://api.semanticscholar.org/CorpusID:202785348 (cit. on p. 38).

[139] J. M. Lourenço. *The NOVAthesis LaTeX Template User's Manual*. NOVA University Lisbon. 2021. URL: https://github.com/joaomlourenco/novathesis/raw/main/template.pdf (cit. on p. iv).

[140] F. Lukačević et al. "Spatial Perception of 3D CAD Model Dimensions and Affordances in Virtual Environments". In: *IEEE Access* 8 (2020), pp. 174587–174604. DOI: 10.1109/ACCESS.2020.3025634 (cit. on p. 19).

[141] D. J. Lund and E. S. Beatrice. "Near infrared laser ocular bioeffects." In: *Health physics* 56 5 (1989), pp. 631–6 (cit. on p. 47).

[142] G. Lunghi et al. "Multimodal Human-Robot Interface for Accessible Remote Robotic Interventions in Hazardous Environments". In: *IEEE Access* 7 (2019), pp. 127290–127319 (cit. on p. 21).

[143] Y. Luo et al. "Monoscopic vs. Stereoscopic Views and Display Types in the Teleoperation of Unmanned Ground Vehicles for Object Avoidance". In: *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)* (2021), pp. 418–425 (cit. on p. 31).

[144] *Magic Leap 2*. Accessed: 2024-05-14. URL: https://www.magicleap.com/magic-leap-2 (cit. on p. 27).

[145] *Magic Leap 2 Product Specification Version 5.8*. Accessed: 2024-05-14. URL: https://cdn.sanity.io/files/af85qi1m/production/75fe4d0dbeb396bfe282789a644aecf52cbe7b28.pdf?dl= (cit. on p. 27).

[146] M. E. McCauley and T. J. Sharkey. "Cybersickness: Perception of Self-Motion in Virtual Environments". In: *Presence: Teleoperators & Virtual Environments* 1 (1992), pp. 311–318 (cit. on pp. 29, 32).

[147] W. R. Mccluney. "Introduction to Radiometry and Photometry". In: 1994. URL: https://api.semanticscholar.org/CorpusID:111310035 (cit. on p. 39).

[148] D. Meier. "Structured Design Fundamentals Of A Discipline Of Computer Program And Systems Design". In: 2016. URL: https://api.semanticscholar.org/CorpusID:63000047 (cit. on p. 17).

[149] L. M. Mescheder et al. "Occupancy Networks: Learning 3D Reconstruction in Function Space". In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2018), pp. 4455–4465. URL: https://api.semanticscholar.org/CorpusID:54465161 (cit. on pp. 37, 38).

[150] *Meta Quest 2: Our Most Advanced New All-in-One VR Headset*. Accessed: 2022-08-05. URL: https://store.facebook.com/quest/products/quest-2/ (cit. on pp. 26, 29, 53).

[151] *Meta Quest Pro: Premium Mixed Reality*. Accessed: 2024-05-14. URL: https://www.meta.com/de/en/quest/quest-pro/ (cit. on pp. 25, 26, 29).

[152] M. Michalska and J. Swiderski. "Low-loss fiber fusion splicing: a main requirement for the developing of all-fiber laser system technology". In: *Laser Technology 2012: Progress in Lasers*. Ed. by W. L. Wolinski, Z. Jankiewicz, and R. S. Romaniuk. Vol. 8702. International Society for Optics and Photonics. SPIE, 2013, pp. 15–20. DOI: 10.1117/12.2013434. URL: https://doi.org/10.1117/12.2013434 (cit. on p. 52).

[153] *Microsoft Hololens | Mixed Reality Technology for Busniess*. Accessed: 2024-05-25. URL: https://www.microsoft.com/en-us/hololens/ (cit. on p. 27).

[154] B. Mildenhall et al. "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis". In: *Commun. ACM* 65 (2020), pp. 99–106. URL: https://doi.org/10.48550/arXiv.2003.08934 (cit. on pp. 6, 19, 36, 38–40, 85, 111).

[155] P. Milgram and F. Kishino. "A Taxonomy of Mixed Reality Visual Displays". In: *IEICE Transactions on Information and Systems* 77 (1994), pp. 1321–1329. URL: https://api.semanticscholar.org/CorpusID:17783728 (cit. on pp. 4, 23, 24, 26, 27).

[156] S. Mori, S. Ikeda, and H. Saito. "A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects". In: *IPSJ Transactions on Computer Vision and Applications* 9 (2017), pp. 1–14 (cit. on p. 63).

[157] morrownr. *88x2bu-20210702*. 2022. URL: https://github.com/morrownr/88x2bu-20210702 (cit. on p. 69).

[158] T. Müller et al. "Instant neural graphics primitives with a multiresolution hash encoding". In: *ACM Transactions on Graphics (TOG)* 41 (2022), pp. 1–15. URL: https://api.semanticscholar.org/CorpusID:246016186 (cit. on pp. 6, 40–42, 86, 91, 92, 94, 95, 133, 134).

[159] K. Nagatani et al. "Emergency response to the nuclear accident at the Fukushima Daiichi Nuclear Power Plants using mobile rescue robots". In: *Journal of Field Robotics* 30 (2013). URL: https://api.semanticscholar.org/CorpusID:1502938 (cit. on p. 19).

[160] A. K. T. Ng, L. K. Y. Chan, and H. Y. K. Lau. "A Study of Cybersickness and Sensory Conflict Theory Using a Motion-Coupled Virtual Reality System". In: *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2018), pp. 643–644 (cit. on p. 32).

[161] T.-N. Nguyen, H.-H. Huynh, and J. Meunier. "3D Reconstruction With Time-of-Flight Depth Camera and Multiple Mirrors". In: *IEEE Access* 6 (2018), pp. 38106–38114. URL: https://api.semanticscholar.org/CorpusID:51877693 (cit. on p. 37).

[162] Nvidia. *DeepStream SDK*. 2022. URL: https://developer.nvidia.com/deepstream-sdk (visited on 2022-05-30) (cit. on p. 70).

[163] Oculus. *Asynchronous Timewarp Examined*. 2022. URL: https://developer.oculus.com/blog/asynchronous-timewarp-examined/ (visited on 2022-05-30) (cit. on p. 33).

[164] Oculus. *Mixed Reality with Passthrough*. 2022. URL: https://developer.oculus.com/blog/mixed-reality-with-passthrough/ (visited on 2022-05-30) (cit. on pp. 68, 127).

[165] *Oculus Rift S: PC-powered VR gaming headset*. Accessed: 2024-05-25. URL: https://www.oculus.com/rift-s/ (cit. on p. 28).

[166] T. Ogura. *Cozmo Driver for ROS*. 2017. URL: https://github.com/OTL/cozmo_driver (cit. on p. 127).

[167] C. M. Oman. “Motion sickness: a synthesis and evaluation of the sensory conflict theory.” In: *Canadian journal of physiology and pharmacology* 68 2 (1990), pp. 294–303 (cit. on pp. 32, 72).

[168] Optofidelity. *Measuring Head-Mounted Display's (HMD) Motion-To-Photon (MTP) Latency*. 2022. URL: https://www.optofidelity.com/blog/measuring-head-mounted-displays-hmd-motion-to-photon-mtp-latency (visited on 2022-05-30) (cit. on p. 71).

[169] J. Orlosky et al. “VisMerge: Light Adaptive Vision Augmentation via Spectral and Temporal Fusion of Non-visible Light”. In: *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2017), pp. 22–31 (cit. on pp. 67, 69, 71, 72).

[170] G. Osterberg. “Topography of the Layer of Rods and Cones in the Human Retina”. In: *Journal of the American Medical Association* 108.3 (1937-01), pp. 232–232 (cit. on p. 95).

[171] S. A. Palmisano, R. S. Allison, and J. Kim. “Cybersickness in Head-Mounted Displays Is Caused by Differences in the User's Virtual and Physical Head Pose”. In: *Frontiers in Virtual Reality*. 2020 (cit. on p. 73).

[172] M. Park et al. “Design and evaluation of an augmented reality welding helmet”. In: *Human Factors and Ergonomics in Manufacturing & Service Industries* 17 (2007), pp. 317–330 (cit. on p. 47).

[173] M. Park et al. “Hand-Eye Coordination Using a Video See-Through Augmented Reality System”. In: *The Ergonomics Open Journal* 1 (2008), pp. 46–53 (cit. on pp. 54, 72).

[174] S.-M. Park and Y.-G. Kim. “A Metaverse: Taxonomy, Components, Applications, and Open Challenges”. In: *IEEE Access* 10 (2022), pp. 4209–4251. URL: https://api.semanticscholar.org/CorpusID:245723446 (cit. on pp. 3, 24).

[175] A. Patney et al. “Perceptually-based foveated virtual reality”. In: *ACM SIGGRAPH 2016 Emerging Technologies* (2016) (cit. on pp. 29, 67).

[176] A. Patney et al. “Towards foveated rendering for gaze-tracked virtual reality”. In: *ACM Transactions on Graphics (TOG)* 35 (2016), pp. 1–12 (cit. on pp. 29, 30).

[177] K. P. Pfeil et al. “Distance Perception with a Video See-Through Head-Mounted Display”. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021) (cit. on p. 68).

[178] M. Pharr, W. Jakob, and G. Humphreys. *Physically Based Rendering: From Theory to Implementation*. 3rd. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2016. ISBN: 0128006455 (cit. on p. 39).

[179] A. Pranckevičius. *Gaussian Splatting playground in Unity*. 2023. URL: https://github.com/aras-p/UnityGaussianSplatting (cit. on pp. 42, 115, 119).

[180] F. Quercioli. "Augmented reality in laser laboratories". In: *Optics & Laser Technology* 101 (2018), pp. 25–29. ISSN: 0030-3992. URL: https://www.sciencedirect.com/science/Article/pii/S0030399217311167 (cit. on pp. 3, 49).

[181] F. Quercioli. "Beyond laser safety glasses: augmented reality in optics laboratories". In: *Applied Optics* 56 (2017), pp. 1148–1150 (cit. on pp. 3, 5, 47, 49).

[182] A. B. Raposo et al. "Towards the use of cad models in VR applications". In: *VRCIA '06*. 2006 (cit. on p. 85).

[183] F. Remondino. "Heritage Recording and 3D Modeling with Photogrammetry and 3D Scanning". In: *Remote. Sens.* 3 (2011), pp. 1104–1138 (cit. on pp. 4, 37, 38, 40).

[184] S. E. Reutebuch, H.-E. Andersen, and R. J. McGaughey. "Light detection and ranging (LIDAR): an emerging tool for multiple resource inventory." In: *Journal of Forestry* 103 (2005), pp. 286–292 (cit. on pp. 19, 37).

[185] J. P. Rolland, R. L. Holloway, and H. Fuchs. "Comparison of optical and video see-through, head-mounted displays". In: *Other Conferences*. 1995. URL: https://api.semanticscholar.org/CorpusID:29976274 (cit. on p. 28).

[186] E. Rosales, J. Rodriguez, and A. Sheffer. "SurfaceBrush: From Virtual Reality Drawings to Manifold Surfaces". In: *ACM Trans. Graph.* 38 (2019), 96:1–96:15 (cit. on p. 131).

[187] K. Saitoh et al. "A 2D-3D integrated interface for mobile robot control using omnidirectional images and 3D geometric models". In: *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality* (2006), pp. 173–176 (cit. on p. 130).

[188] Y. Saraiji. *GStreamer Integration with Unity using a Native plugin*. 2020. URL: https://github.com/mrayy/mrayGStreamerUnity (cit. on p. 70).

[189] I. M. Sauer et al. "Mixed Reality in Visceral Surgery: Development of a Suitable Workflow and Evaluation of Intraoperative Use-cases". In: *Annals of Surgery* 266 (2017), pp. 706–712 (cit. on p. 54).

[190] J. Sauro. *A practical guide to the system usability scale: Background, benchmarks & best practices*. Measuring Usability LLC, 2011 (cit. on p. 60).

[191] B. Schmitz. "Testing and assessment of laser safety eyewear - The European regulatory framework". In: 2013 (cit. on pp. 46, 48, 49).

[192] J. L. Schönberger and J.-M. Frahm. "Structure-from-Motion Revisited". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 4104–4113. URL: https://api.semanticscholar.org/CorpusID:1728538 (cit. on pp. 19, 37, 85, 93, 115).

[193] J. L. Schönberger et al. "Pixelwise View Selection for Unstructured Multi-View Stereo". In: *European Conference on Computer Vision*. 2016. URL: https://api.semanticscholar.org/CorpusID:977535 (cit. on p. 37).

[194] W. Schroyens et al. "Eye Movement Control during Reading: Foveal Load and Parafoveal Processing". In: *Quarterly Journal of Experimental Psychology* 52 (1999), pp. 1021–1046 (cit. on p. 67).

[195] T. W. Schubert. "The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness." In: *Z. für Medienpsychologie* 15.2 (2003), pp. 69–71 (cit. on pp. 34, 76, 102).

[196] A. F. Seay et al. "Simulator sickness and presence in a high FOV virtual environment". In: *Proceedings IEEE Virtual Reality 2001* (2001), pp. 299–300 (cit. on pp. 33, 89).

[197] V. A. Semenyuk. "Prediction of temperature and damage in an irradiated human eye during retinal photocoagulation". In: *International Journal of Heat and Mass Transfer* (2018) (cit. on p. 48).

[198] R. Skarbez, M. Smith, and M. C. Whitton. "Revisiting Milgram and Kishino's Reality-Virtuality Continuum". In: *Frontiers in Virtual Reality*. 2021. URL: https://api.semanticscholar.org/CorpusID:232327990 (cit. on p. 24).

[199] M. Slater. "Immersion and the illusion of presence in virtual reality". In: *British Journal of Psychology* 109.3 (2018), pp. 431–433 (cit. on pp. 33, 76).

[200] M. Slater and S. Wilbur. "A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments". In: *Presence: Teleoperators & Virtual Environments* 6 (1997), pp. 603–616. URL: https://api.semanticscholar.org/CorpusID:9437981 (cit. on p. 33).

[201] D. H. Sliney. "Laser and LED Eye Hazards: Safety Standards". In: *Opt. Photon. News* 8.9 (1997-09), p. 31. DOI: 10.1364/OPN.8.9.000031. URL: http://www.optica-opn.org/abstract.cfm?URI=opn-8-9-31 (cit. on p. 48).

[202] A. A. Snigirev et al. "On the possibilities of x-ray phase contrast microimaging by coherent high-energy synchrotron radiation". In: *Review of Scientific Instruments* 66 (1995), pp. 5486–5492. URL: https://api.semanticscholar.org/CorpusID:54086950 (cit. on p. 14).

[203] M. Speicher et al. "Selection-based Text Entry in Virtual Reality". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018) (cit. on p. 131).

[204] M. Speicher, B. D. Hall, and M. Nebeling. "What is Mixed Reality?" In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019). URL: https://api.semanticscholar.org/CorpusID:140449351 (cit. on p. 4).

[205] K. M. Stanney, R. S. Kennedy, and J. M. Drexler. “Cybersickness is Not Simulator Sickness”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 41 (1997), pp. 1138–1142. URL: https://api.semanticscholar.org/CorpusID:70690770 (cit. on p. 32).

[206] S. Starke et al. “Evolutionary multi-objective inverse kinematics on highly articulated and humanoid robots”. In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2017), pp. 6959–6966. URL: https://api.semanticscholar.org/CorpusID:30687068 (cit. on p. 21).

[207] J.-P. Stauffert, F. Niebling, and M. E. Latoschik. “Latency and Cybersickness: Impact, Causes, and Measures. A Review”. In: *Frontiers in Virtual Reality*. 2020. URL: https://api.semanticscholar.org/CorpusID:227164260 (cit. on p. 29).

[208] Stereolabs. *Getting Started with Unity and ZED*. 2022. URL: https://www.stereolabs.com/docs/unity/ (visited on 2022-05-30) (cit. on pp. 33, 69).

[209] Stereolabs. *ZED Box: Edge AI Gateway*. 2022. URL: https://www.stereolabs.com/zed-box/ (visited on 2022-05-30) (cit. on p. 69).

[210] stereolabs. *ZED GStreamer plugins*. 2022. URL: https://github.com/stereolabs/zed-gstreamer (cit. on pp. 70, 72).

[211] P. Stotko et al. “A VR System for Immersive Teleoperation and Live Exploration with a Mobile Robot”. In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2019), pp. 3630–3637. URL: https://api.semanticscholar.org/CorpusID:199501666 (cit. on pp. 111, 118).

[212] H. Strasburger, I. Rentschler, and M. Jüttner. “Peripheral vision and pattern recognition: a review.” In: *Journal of vision* 11 5 (2011), p. 13. URL: https://api.semanticscholar.org/CorpusID:15181693 (cit. on pp. 29, 30).

[213] H. Sun et al. “OptoBridge: assisting skill acquisition in the remote experimental collaboration”. In: *Proceedings of the 28th Australian Conference on Computer-Human Interaction* (2016) (cit. on p. 25).

[214] I. E. Sutherland. “A head-mounted three dimensional display”. In: *Proceedings of the December 9-11, 1968, fall joint computer conference, part I* (1968). URL: https://api.semanticscholar.org/CorpusID:4561103 (cit. on pp. 22, 26).

[215] K. A. Szczurek et al. “From 2D to 3D Mixed Reality Human-Robot Interface in Hazardous Robotic Interventions with the Use of Redundant Mobile Manipulator”. In: *ICINCO*. 2021 (cit. on pp. 7, 20).

[216] K. A. Szczurek et al. “Mixed Reality Human–Robot Interface With Adaptive Communications Congestion Control for the Teleoperation of Mobile Redundant Manipulators in Hazardous Environments”. In: *IEEE Access* 10 (2022), pp. 87182–87216. URL: https://api.semanticscholar.org/CorpusID:251605024 (cit. on pp. 7, 25, 26, 111).

[217] K. A. Szczurek et al. “Multimodal Multi-User Mixed Reality Human–Robot Interface for Remote Operations in Hazardous Environments”. In: *IEEE Access* 11 (2023), pp. 17305–17333. URL: https://api.semanticscholar.org/CorpusID:256969028 (cit. on pp. 7, 20, 109, 139).

[218] M. Tancik et al. “Block-NeRF: Scalable Large Scene Neural View Synthesis”. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), pp. 8238–8248 (cit. on p. 40).

[219] M. Tancik et al. “Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains”. In: *ArXiv* abs/2006.10739 (2020). URL: https://api.semanticscholar.org/CorpusID:219791950 (cit. on pp. 38, 85).

[220] M. Tancik et al. “Nerfstudio: A Modular Framework for Neural Radiance Field Development”. In: *arXiv preprint arXiv:2302.04264* (2023) (cit. on p. 41).

[221] R. M. Taylor. “Situational Awareness Rating Technique (Sart): The Development of a Tool for Aircrew Systems Design”. In: 2017. URL: https://api.semanticscholar.org/CorpusID:108133178 (cit. on p. 119).

[222] *Technical Specifications of Varjo XR-3*. Accessed: 2022-08-05. URL: https://varjo.com/products/xr-3/ (cit. on p. 62).

[223] U. Technologies. *Science! Super Science!* 2022. URL: https://github.com/Unity-Technologies/SuperScience (cit. on p. 73).

[224] U. Technologies. *Unity Robotics Hub*. https://github.com/Unity-Technologies/Unity-Robotics-Hub. 2021 (cit. on p. 127).

[225] Y. T. Tefera et al. “Towards Gaze-contingent Visualization of Real-time 3D Reconstructed Remote Scenes in Mixed Reality”. In: *Proceedings of the 2023 6th International Conference on Advances in Robotics* (2023). URL: https://api.semanticscholar.org/CorpusID:264974195 (cit. on p. 111).

[226] K. Terzic and M. E. Hansard. “Causes of discomfort in stereoscopic content: a review”. In: *ArXiv* abs/1703.04574 (2017) (cit. on p. 67).

[227] *The PHASER platform: innovative designs to bring X-ray-based FLASH to the clinic*. [Accessed 09-04-2024]. URL: https://physicsworld.com/a/the-phaser-platform-innovative-designs-to-bring-x-ray-based-flash-to-the-clinic/ (cit. on p. 13).

[228] Tobii. *Pico Neo 3 Pro Eye: with native Tobii eye tracking*. 2022. URL: https://vr.tobii.com/integrations/pico-neo-3-pro-eye/ (visited on 2022-05-30) (cit. on p. 70).

[229] C. Tominski et al. “A Survey on Interactive Lenses in Visualization”. In: *Eurographics Conference on Visualization*. 2014 (cit. on p. 87).

[230] Varjo. *Varjo-Ready Computers for XR-3 and VR-3*. 2022. URL: https://varjo.com/varjo-ready-computers/ (visited on 2022-05-30) (cit. on pp. 6, 67–69).

[231] J. Viega et al. “3D magic lenses”. In: *ACM Symposium on User Interface Software and Technology*. 1996 (cit. on pp. 86, 87).

[232] *Virtual Desktop: Your PC in VR*. Accessed: 2022-08-05. URL: https://www.vrdesktop.net/ (cit. on p. 53).

[233] M. E. Walker et al. “The Cyber-Physical Control Room: A Mixed Reality Interface for Mobile Robot Teleoperation and Human-Robot Teaming”. In: *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (2024). URL: https://api.semanticscholar.org/CorpusID:268344480 (cit. on p. 111).

[234] L. Wang et al. “The magic volume lens: an interactive focus+context technique for volume rendering”. In: *VIS 05. IEEE Visualization, 2005.* (2005), pp. 367–374 (cit. on p. 87).

[235] C. Ware. *Visual thinking for information design*. Morgan Kaufmann Elsevier, 2022 (cit. on p. 73).

[236] J. M. P. van Waveren. “The asynchronous time warp for virtual reality on consumer hardware”. In: *VRST*. 2016 (cit. on pp. 32, 53, 68).

[237] S. Weech, S. Kenny, and M. Barnett-Cowan. “Presence and Cybersickness in Virtual Reality Are Negatively Related: A Review”. In: *Frontiers in Psychology* 10 (2019). URL: https://api.semanticscholar.org/CorpusID:59553682 (cit. on p. 33).

[238] D. Wei, B. Huang, and Q. Li. “Multi-View Merging for Robot Teleoperation With Virtual Reality”. In: *IEEE Robotics and Automation Letters* 6 (2021), pp. 8537–8544. URL: https://api.semanticscholar.org/CorpusID:237598023 (cit. on pp. 7, 111).

[239] H. Wei et al. “Vision-Guided Hand–Eye Coordination for Robotic Grasping and Its Application in Tangram Puzzles”. In: *AI* 2.2 (2021), pp. 209–228 (cit. on p. 75).

[240] T. Wilksen et al. “A Bunch-Synchronized Data Acquisition System for the European XFEL Accelerator”. In: 2017. URL: https://api.semanticscholar.org/CorpusID:56019192 (cit. on p. 15).

[241] L. Winkelmann et al. “The European XFEL Photocathode Laser”. In: 2019. URL: https://api.semanticscholar.org/CorpusID:212667140 (cit. on p. 5).

[242] E. XFEL. *European XFEL Overview*. [Accessed 09-04-2024]. URL: https://www.xfel.eu/facility/overview/index_eng.html (cit. on p. 15).

[243] D. T. W. Xun, Y. L. Lim, and S. Srigrarom. “Drone detection using YOLOv3 with transfer learning on NVIDIA Jetson TX2”. In: *2021 Second International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP)* (2021), pp. 1–6 (cit. on p. 69).

[244] D. A. Yarotski. “Laser Incident Lessons Learned and Action List”. In: (2016-02). DOI: 10.2172/1239898. URL: https://www.osti.gov/biblio/1239898 (cit. on pp. 47, 48).

[245] A. Yeo et al. “Entering the Next Dimension: A Review of 3D User Interfaces for Virtual Reality”. In: *Electronics* (2024). URL: https://api.semanticscholar.org/CorpusID:267489972 (cit. on p. 34).

[246] K. Yin et al. “Advanced liquid crystal devices for augmented reality and virtual reality displays: principles and applications”. In: *Light, Science & Applications* 11 (2022). URL: https://api.semanticscholar.org/CorpusID:249184717 (cit. on p. 28).

[247] *ZED Mini: Mixed-Reality Camera*. Accessed: 2022-08-05. URL: https://www.stereolabs.com/zed-mini/ (cit. on pp. 28, 53).

[248] J. Zhang et al. “A 360° Video-based Robot Platform for Telepresent Redirected Walking”. In: *Proceedings of ACM SIGCHI Conference on Human-Robot Interaction (HRI) workshop on Virtual, Augmented and Mixed Reality for Human-Robot Interaction*. ACM, 2018. URL: http://basilic.informatik.uni-hamburg.de/Publications/2018/ZLKKS18a (cit. on p. 3).

[249] J. Zhang et al. “Detection Thresholds for Rotation and Translation Gains in 360° Video-Based Telepresence Systems”. In: *IEEE Transactions on Visualization and Computer Graphics* 24 (2018), pp. 1671–1680 (cit. on pp. 7, 109, 111).

[250] S. Zhang. “High-speed 3D shape measurement with structured light methods: A review”. In: *Optics and Lasers in Engineering* (2018) (cit. on p. 40).

[251] S. Z. Zhou et al. “Robust Pose Estimation for Outdoor Mixed Reality with Sensor Fusion”. In: *HCI*. 2009 (cit. on p. 69).

[252] M. Zollhöfer et al. “State of the Art on 3D Reconstruction with RGB-D Cameras”. In: *Computer Graphics Forum* 37 (2018) (cit. on pp. 4, 40, 85).

A
Appendix

A.1 Questionnaires

A.1.1 Simulator Sickness Questionnaire

Please rate how strongly you are affected by these symptoms NOW:

	None	Slight	Moderate	Severe
General discomfort	○	○	○	○
Fatigue	○	○	○	○
Headache	○	○	○	○
Eyestrain	○	○	○	○
Difficulty focusing	○	○	○	○
Increased salivation	○	○	○	○
Sweating	○	○	○	○
Nausea	○	○	○	○
Difficulty concentrating	○	○	○	○
Fullness of head	○	○	○	○
Blurred vision	○	○	○	○
Dizzy (eyes open)	○	○	○	○
Dizzy (eyes closed)	○	○	○	○
Vertigo	○	○	○	○
Stomach awareness	○	○	○	○
Burping	○	○	○	○

A.1.2 NASA-TLX

How mentally demanding was the task? (Mental Demand)

Very Low Very High

○ ○

How physically demanding was the task? (Physical Demand)

Very Low Very High

○ ○

How hurried or rushed was the pace of the task? (Temporal Demand)

Very Low Very High

○ ○

How insecure, discouraged, irritated, stressed, and annoyed were you? (Frustration)

Very Low Very High

○ ○

How successful were you in accomplishing what you were asked to do? (Performance)

Perfect Failure

○ ○

How hard did you have to work to accomplish your level of performance? (Efforts)

Very Low Very High

○ ○

A.1.3 Igroup Presence Questionnaire

In the computer-generated world I had a sense of "being there"

	-3	-2	-1	0	1	2	3	
not at all	○	○	○	○	○	○	○	very much

Somehow I felt that the virtual world surrounded me.

	-3	-2	-1	0	1	2	3	
fully disagree	○	○	○	○	○	○	○	fully agree

I felt like I was just perceiving pictures.

	-3	-2	-1	0	1	2	3	
fully disagree	○	○	○	○	○	○	○	fully agree

I did not feel present in the virtual space.

	-3	-2	-1	0	1	2	3	
did not feel	○	○	○	○	○	○	○	felt present

I had a sense of acting in the virtual space, rather than operating something from outside.

	-3	-2	-1	0	1	2	3	
fully disagree	○	○	○	○	○	○	○	fully agree

I felt present in the virtual space.

	-3	-2	-1	0	1	2	3	
fully disagree	○	○	○	○	○	○	○	fully agree

How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)?

	-3	-2	-1	0	1	2	3	
extremely aware	○	○	○	○	○	○	○	not aware at all

I was not aware of my real environment.

	-3	-2	-1	0	1	2	3	
fully disagree	○	○	○	○	○	○	○	fully agree

I still paid attention to the real environment.

-3	-2	-1	0	1	2	3

fully disagree	◯	◯	◯	◯	◯	◯	◯	fully agree

I was completely captivated by the virtual world.

	-3	-2	-1	0	1	2	3	
fully disagree	◯	◯	◯	◯	◯	◯	◯	fully agree

How real did the virtual world seem to you?

	-3	-2	-1	0	1	2	3	
completely real	◯	◯	◯	◯	◯	◯	◯	not real at all

How much did your experience in the virtual environment seem consistent with your real world experience ?

	-3	-2	-1	0	1	2	3	
not consistent	◯	◯	◯	◯	◯	◯	◯	very consistent

How real did the virtual world seem to you?

	-3	-2	-1	0	1	2	3	
about as real as an imagined world	◯	◯	◯	◯	◯	◯	◯	indistinguishable from the real world

The virtual world seemed more realistic than the real world.

	-3	-2	-1	0	1	2	3	
fully disagree	◯	◯	◯	◯	◯	◯	◯	fully agree

A.1.4 Adapted Igroup Presence Questionnaire for VST-HMD

Somehow I felt that the real world surrounded me.

	-3	-2	-1	0	1	2	3	
fully disagree	◯	◯	◯	◯	◯	◯	◯	fully agree

I felt like I was just perceiving pictures.

	-3	-2	-1	0	1	2	3	
fully disagree	○	○	○	○	○	○	○	fully agree

I did not feel present in the real environment

	-3	-2	-1	0	1	2	3	
did not feel	○	○	○	○	○	○	○	felt present

I had a sense of acting in the real environment, rather than operating something from outside.

	-3	-2	-1	0	1	2	3	
fully disagree	○	○	○	○	○	○	○	fully agree

I felt present in the real environment.

	-3	-2	-1	0	1	2	3	
fully disagree	○	○	○	○	○	○	○	fully agree

A.1.5 System Usability Questionnaire

I think that I would like to use this system frequently.

	1	2	3	4	5	
strongly disagree	○	○	○	○	○	strongly agree

I found the system unnecessarily complex.

	1	2	3	4	5	
strongly disagree	○	○	○	○	○	strongly agree

I thought the system was easy to use.

	1	2	3	4	5	
strongly disagree	○	○	○	○	○	strongly agree

I think that I would need the support of a technical person to be able to use this system.

	1	2	3	4	5

strongly disagree	○	○	○	○	○	strongly agree

I found the various functions in this system were well integrated.

	1	2	3	4	5	
strongly disagree	○	○	○	○	○	strongly agree

I thought there was too much inconsistency in this system.

	1	2	3	4	5	
strongly disagree	○	○	○	○	○	strongly agree

I would imagine that most people would learn to use this system very quickly.

	1	2	3	4	5	
strongly disagree	○	○	○	○	○	strongly agree

I found the system very cumbersome to use.

	1	2	3	4	5	
strongly disagree	○	○	○	○	○	strongly agree

I felt very confident using the system.

	1	2	3	4	5	
strongly disagree	○	○	○	○	○	strongly agree

I needed to learn a lot of things before I could get going with this system .

	1	2	3	4	5	
strongly disagree	○	○	○	○	○	strongly agree

A.1.6 Situation Awareness Questionnaire

How changeable is the situation? Is the situation highly unstable and likely to change suddenly (High) or is it very stable and straightforward (Low)?

1 2 3 4 5 6 7

○ ○ ○ ○ ○ ○ ○

How complicated is the situation? Is it complex with many interrelated components (High) or is it simple and straightforward (Low)?

1 2 3 4 5 6 7

○ ○ ○ ○ ○ ○ ○

How many variables are changing within the situation? Are there a large number of factors varying (High) or are there very few variables changing (Low)?

1 2 3 4 5 6 7

○ ○ ○ ○ ○ ○ ○

How aroused are you in the situation? Are you alert and ready for activity (High) or do you have a low degree of alertness (Low)?

1 2 3 4 5 6 7

○ ○ ○ ○ ○ ○ ○

How much are you concentrating on the situation? Are you concentrating on many aspects of the situation (High) or focussed on only one (Low)?

1 2 3 4 5 6 7

○ ○ ○ ○ ○ ○ ○

How much is your attention divided in the situation? Are you concentrating on many aspects of the situation (High) or focussed on only one (Low)?

1 2 3 4 5 6 7

○ ○ ○ ○ ○ ○ ○

How much mental capacity do you have to spare in the situation? Do you have sufficient to attend to many variables (High) or nothing to spare at all (Low)?

1 2 3 4 5 6 7

○ ○ ○ ○ ○ ○ ○

How much information have you gained about the situation? Have you received and understood a great deal of knowledge (High) or very little (Low)?

1	2	3	4	5	6	7
○	○	○	○	○	○	○

How familiar are you with the situation? Do you have a great deal of relevant experience (High) or is it a new situation (Low)?

1	2	3	4	5	6	7
○	○	○	○	○	○	○

A.2 Sample Prompts for ChatGPT

As stated in the declaration, OpenAI's ChatGPT was utilized to aid in the writing process of this dissertation. However, the original research and analysis were conducted independently by me. Specifically, the GPT-3.5 and GPT-4o models, available through the free version of ChatGPT, were used for assistance.

A list of sample prompts for enhancing writing clarity and grammar is documented below.

- "Please help me improve the clarity and rigor of the following sentences. "
- "Please rewrite this sentence to make it more concise and clear."
- "Help me rephrase the following sentence to enhance its clarity and coherence."

  Sample Input: The MARWIN mobile robot was developed for inspection and maintenance at the EuXFEL accelerator tunnel, helping operators to detect radiation hot spots originating from malfunctioning cavities while the cavities are pushed to their limits or to measure the radiation level to detect any residual radiation after facility shut-down to ensure secure human entrance to the facility.

  Sample Output: The MARWIN mobile robot was developed for inspecting and maintaining the EuXFEL accelerator tunnel. It assists operators in detecting radiation hot spots from malfunctioning cavities when they are pushed to their limits. Additionally, it measures residual radiation levels after the facility is shut down to ensure the safety of human entry.

A list of sample usages where ChatGPT was employed as a search aid and as inspiration for writing the introduction of various concepts is documented below.

- **Definition and Explanation**: "What is cross reality? Give me the most important academic paper as references."
- **Historical Context**: "What are some famous quotes about mixed reality from Paul Milgram? Give me academic papers as references."
- **Literature Research**: "What is immersive analytics, what are its benefits in big data analysis, give me academic papers as references."
- **Literature Research**: "How can varifocal HMD design address vergence accommodation conflict, give me academic paper as references. "
- **Technical Background**: "Why are varifocal devices not prevalent in commercial HMD? Give me academic papers as references."
- **Technical Summary**: "Summarize the main findings of the *Representing Scenes as Neural Radiance Fields for View Synthesis* paper."

ChatGPT was also used in assisting the coding process during research. Typical usage and sample prompts for this purpose are listed below.

- "Please explain the following code to me step by step."
- "Explain the concept of smart pointers in C++ and provide an example."
- "How do I properly allocate and deallocate memory using new and delete in C++?"
- "Explain each step of the Unity event loop."
- " Explain each step of the Unity Built-In Render Pipeline."

Additionally, ChatGPT was utilized to translate the abstract and declaration sections of this dissertation from English to German. It is important to note that ChatGPT was not publicly available until November 2022. Consequently, publications and work completed before this date were written without its assistance. However, other writing tools, such as DeepL Write[1] and Grammarly [2], were used to rephrase certain sentences and enhance the grammar and clarity of the writing.

[1]https://www.deepl.com/en/write
[2]https://www.grammarly.com/grammar-check