SCHOOL OF COMPUTATION, INFORMATION AND TECHNOLOGY - INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics: Games Engineering

A Virtual Reality Travel Application for Elderly People Focusing on Object Integration and Interaction with Augmented Virtuality

Denitsa Aleksandrova Asova

SCHOOL OF COMPUTATION, INFORMATION AND TECHNOLOGY - INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics: Games Engineering

A Virtual Reality Travel Application for Elderly People Focusing on Object Integration and Interaction with Augmented Virtuality

Eine Virtual-Reality-Reiseanwendung für Ältere Menschen mit Fokus auf Objektintegration und Interaktion mit Augmented Virtuality

Author:	Denitsa Aleksandrova Asova
Supervisor:	Prof. Gudrun Klinker, Ph.D.
Advisor:	Christian Eichhorn, M.Sc.
Submission Date:	15.11.2023

I confirm that this master's thesis in informatics: games engineering is my own work and I have documented all sources and material used.

Munich, 15.11.2023 Denitsa Aleksandrova Asova

Acknowledgments

I would like to thank my advisor, Christian Eichhorn, for his valuable advice and constant support during each stage of writing this thesis. A special thank you to Tobias Mesmer and Jonas Weigand for their assistance in the development of the AV smartphone and camera components. I would also like to thank my family for all the support and encouragement they gave me during the whole process, and for motivating me to outdo myself.

Abstract

In this project, the advantages of Virtual Reality (VR) and Augmented Virtuality (AV) are utilized to create an immersive travel experience for older adults. The existing Virtual Reality travel application GeoTravel was updated to incorporate VR design guidelines for seniors derived from current research. A realistic virtual avatar was integrated to represent users in the virtual environment, offering better social interaction cues. 360-degree videos with synchronized audio were utilized to achieve higher perceptual realism compared to static 360-degree images. To simulate real-world interactions that improve the feeling of presence, three AV objects — a water bottle, a camera, and a smartphone — are supported. Users can manipulate these objects in the real-world and observe their corresponding virtual counterparts performing the executed actions in the virtual environment. Due to an increase in COVID-19 and other viral cases, it was not possible to test the usability of the application with older adults. A user study with 17 students was conducted instead, asking the participants to evaluate the application from the perspective of a senior. Results showed "Good" (SUS) usability, however, an actual evaluation with the target group is still necessary.

Contents

Acknowledgments **iii**

Abstract **iv**

1 Introduction **1**

2 Related Works **3**
2.1 Terminology . . . 3
2.1.1 Immersion and Presence . . . 3
2.1.2 Virtual Reality . . . 6
2.1.3 Augmented Reality . . . 13
2.1.4 Augmented Virtuality . . . 14
2.2 Augmented Virtuality Applications . . . 16
2.2.1 AV for Human Connection and Wellness . . . 16
2.2.2 AV for Educational Experiences . . . 20
2.2.3 Summary . . . 25
2.3 Virtual Reality Applications for Older Adults . . . 26
2.3.1 VR for Promoting Overall Wellness . . . 26
2.3.2 VR for Health Assessment and Rehabilitation . . . 28
2.3.3 VR Exergames . . . 32
2.4 Analysis and Definition of Guidelines . . . 34

3 Background - *GeoTravel* **37**
3.1 Concept and Features . . . 37
3.2 Thesis Goals . . . 40

4 Implementation **41**
4.1 Hardware Setup . . . 41
4.2 Project Upgrade . . . 41
4.3 Virtual Avatar . . . 44
4.4 Virtual Environment . . . 46
4.5 Object Integration and Interaction with AV . . . 49
4.5.1 OpenCV Dynamic Link Library . . . 49
4.5.2 Camera Calibration . . . 51
4.5.3 Aruco Marker Tracking . . . 52
4.5.4 Vuforia Tracking . . . 57
4.5.5 Integrated Objects . . . 58

5 Evaluation **63**
5.1 Setup 63
5.2 Study Outcome 64

6 Discussion **66**

7 Future Work **68**

8 Conclusion **69**

List of Figures **70**

List of Tables **71**

Bibliography **72**

1 Introduction

10% of the global population in 2022 was older than 65, and this percentage has substantially increased since 2000, when it was around 7% [RR; Wai04]. Among this 10%, there are seniors with health issues that restrict their mobility and/or lead to some cognitive impairments. In such cases, the majority of older adults either do not leave their homes or stay in healthcare facilities. For them, traveling becomes nearly impossible, considering their health, safety, and the need for specialized assistance. Even for older adults who do not suffer from any serious health conditions, traveling might still be challenging due to the physical limitations that come with age, and factors such as money and accessibility. The confirmed positive effects of traveling on the elderly, such as the observed improvement of their physical and mental health and the corresponding sense of independence [Qia+22; AJ11; PS15], raise the question of how to make these benefits more accessible. The advancements of technology within the Mixed Reality (MR[2.3]) domain offer an innovative solution. Virtual Reality (VR[2.1.2]) allows individuals to experience highly interactive and immersive environments, without having to physically leave their homes. These virtual environments (VEs) can simulate completely unrealistic and imaginative worlds, as well as real-world locations that can be visited. Within them, people can have a first-person perspective of the simulated world, manipulate virtual objects, and even engage in social interactions with others. The social aspect of VR holds great potential benefits for the older generation. The impact of social isolation on the physical and mental well-being of older adults has been confirmed by researchers [CW09]. In VR, they could virtually meet and communicate with others even if they are physically separated by large distances. Despite the possible benefits that VR could offer to older adults, there is limited existing research on VR applications specifically designed for this target group [section 2.3]. Even fewer scientists have explored the impact of Augmented Virtuality (AV) on seniors. AV is part of Milgram's mixed reality continuum [Mil+95] and allows the incorporation of real-world objects into the virtual world [section 2.1.4]. Consequently, well-known everyday objects could be used to interact with the VE. For seniors, this would eliminate the need to use complex controllers that might not only be unfamiliar but also overwhelming for them. Furthermore, the familiarity with real-world objects and their interaction could enhance the overall VR experience, motivating older adults to engage more with the VR applications.

This work aims to address the limited research on MR applications designed for older adults by focusing on two essential components:

1. **Improving an existing VR travel application for older adults:** The existing application GeoTravel [3] was enhanced with high-quality 360-degree videos featuring synchronized audio, and a dynamic virtual avatar.

2. **Incorporating AV objects:** Three AV objects - a water bottle, a camera, and a smartphone

- were integrated into the application.

The research questions investigated in this work primarily focus on the AV objects, examining their integration, interaction, and potential to motivate older adults to use GeoTravel more frequently. Additionally, the study explores the quality and immersive nature of the virtual environment. The outlined research questions are as follows:

- **RQ1:** To what extent do users perceive the interaction with the AV objects as natural, and how many prefer this concept over a fully virtual implementation?
- **RQ2:** To what extent do the AV objects motivate users to engage with the application and enhance its overall immersive experience?
- **RQ3:** Are 360-degree videos with synchronized audio better perceived compared to static 360-degree images?

2 Related Works

2.1 Terminology

To ensure a better understanding of the concepts addressed within this thesis, this section offers precise definitions and explanations of the key terms involved.

2.1.1 Immersion and Presence

There has been an ongoing discussion about the difference between **immersion** and **presence**. In the context of computer games, most authors only define immersion, while presence is completely absent as a separate term in their papers, or only briefly mentioned without any exhaustive explanation. Brown and Cairns [BC04] define immersion as "the degree of involvement with a game" and establish three levels of involvement: "engagement, engrossment and total immersion". The user starts their playing experience at the engagement stage, during which they have to be willing to play a game. Afterward comes engrossment, the second stage, when the user connects with the game. The last stage, total immersion, occurs when the user completely loses their sense of self and merges with the game's world. According to Brown and Cairns [BC04], total immersion is synonymous with presence.

Ermi and Mäyrä [EM05] relate presence to its original definition in the context of teleoperations as a representation of transportation [Ste92] and, therefore, prefer only to use the term immersion when analyzing computer games. They see immersion as a multidimensional phenomenon and outline three distinct dimensions - sensory immersion, challenge-based immersion, and imaginative immersion. Furthermore, they claim that each dimension has a corresponding fundamental factor that improves it: audiovisual quality corresponds to sensory immersion, challenges and interaction to challenge-based immersion, and narrative to imaginative immersion.

- Audiovisual quality: There are three groups of factors that influence the audiovisual quality [AF17] and, therefore, sensory immersion. The **Human Influential Factors** are subjective and based on the user's unique perception of multimedia and include, for example, the user's age, mood, and cultural background. The **Technological Influential Factors** are technology-based, measurable attributes, such as audio bandwidth, resolution, sampling and frame rate. Lastly, the **Contextual Influential Factors** focus on how the temporal, spatial, and social contexts influence perception of media. According to Akhtar and Falk [AF17], the way users experience multimedia can vary based on where, when, and with whom it is experienced.

- Challenges and interaction: Challenge-based immersion focuses on the active participation of users and how it is encouraged with mental and physical challenges. The most important aspect, however, is the right balance between the challenges and the user's abilities. Nakamura and Csikszentmihalyi [Csi14] named it **flow**. They define flow as a mental state characterized by intense focus on enjoyable activities that leads to merging of action and perception, loss of self-consciousness, and a distorted notion of time. Moreover, the activities need to have clear goals and provide immediate feedback, and the user needs to feel in control of their actions.
- Narrative: Many computer games place significant importance on their story, world, and characters. Imaginative immersion is achieved when the narrative allows the users to fully engage in the game's fiction. Following Ermi and Mäyrä's [EM05] concept, Bastos et. al [Bas+17] assert that implicit story and symbology are crucial components of imaginative immersion, and that a narrative that lets the users draw their own conclusions about the game is more engrossing.

Bastos et. al [Bas+17] outline five more potentially immersive features. They hypothesize that audiovisual synchronization, impressive graphics, player's presentation, and mixed-reality technologies improve sensory immersion.

The audiovisual synchronization makes sure that the visual elements and sounds are synchronized with each other and with the actions of the players, which creates a more vivid and captivating game's world.

Impressive graphics, encompassing elements such as art style, 3D models, special effects, and animations, interest players. Independent studies conducted by Bastos et. al [Bas+17] and Haywood and Cairns [HC06] showed that the graphics have to be relevant to the game's narrative but they should not necessarily achieve a high level of realism.

The player's representation in the virtual world and their goals in the game are central to the gaming experience and have been studied extensively by Jaime Banks [Ban15; Ban+19]. He promotes the idea that, to achieve immersion, games should provide players with a set of characters and customization tools that let them express their identities and feel represented in the virtual world.

Mixed-reality technologies, such as motion capture and skeleton tracking, merge elements of the physical and virtual worlds and stimulate the sense of proprioception, or the sense of "one's own body in space" [HIT15]. According to Shinkle [Shi08], players feels more engaged when they feel their own body in the virtual world.

Although **immersion** and **presence** are frequently used interchangeably in the context of computer games, it is important to note that they are defined as distinct concepts within the field of virtual reality (VR). According to Mel Slater [Sla03], immersion refers to "what the technology delivers from an objective point of view" and presence "is a human reaction to immersion".

The level of immersion in a VR system is determined by the system's rendering software and display technology, more specifically by the type of sensory information they provide [BM07]. The level of immersion is objective and can be measured, which allows for a comparison where one system can be more immersive than another. In contrast, presence refers to the feeling of "being there" and is a context-specific, subjective user response. The level of presence within a given VR system can vary depending on the user and their current mindset.

With these definitions in mind, it is easier to identify the factors that enhance immersion since they are hardware- and software-specific and measurable. According to Bowman and Mcmahan [BM07], components like field of vision (FOV), display size and resolution, frame rate, and refresh rate affect the degree of visual immersion, or how closely the system's visual output reflects real-world visual signals. Presence, on the other hand, is a psychological phenomenon and its subjectivity, contextual dependency and multidimensionality make it more challenging to outline some influential factors. In his work "Immersion, engagement, and presence", McMahan [Mcm03] proposed a way to assess 3D games and discussed several presence-related components. The first element he examined is the quality of the social interaction that is offered within the virtual environment (VE). A social interaction is of high quality if it is perceived as "sociable, warm, sensitive, personal, and intimate". It includes interacting both with the environment and with other participants. The sense of "being with someone" and interacting with them in the VE was originally observed by Durlach and Slater [DS00]. They state that the feeling of presence is increased when environmental changes caused by one participant can be observed by the rest and when collaborative interactions also result in environmental changes. Biocca et al. [BHB03] define this aspect of presence as social presence in their work "Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria". According to them, social presence encompasses the feeling of "real" interaction and connectedness despite the absence of physical co-presence.

The next element McMahan[Mcm03] observed is social realism, or more specifically integrating avatars in the VE. He examined two points of view: an egocentric viewpoint, where the user perceives the environment "through the eyes of their avatar", in a similar way to how they perceive their own body in the real-world, and an exocentric viewpoint, where the user observes their avatar from an external position. According to Slater [SW97], the egocentric point of view is better for creating a sense of presence in the VE because it mimics how the real-world is perceived and closely links the user's actions and perceptions. Regarding the physical appearance of the avatars, McMahan [Mcm03] refers to two varying views expressed by Stone [Sto95] and Bolter [BG99], respectively. Stone believes that presence is achieved when the unique features of a person are transferred to the virtual world. On the other hand, Bolter sees presence as "the freedom to become someone (or something) else". Bolter's avatar can change its appearance, be used by many users at once, or even be invisible. Stone's avatar should closely resemble the real-world person behind it.

Perceptual realism is also important, according to McMahan [Mcm03], along with social realism. It depicts the degree to which the virtual world corresponds to the real-world visually and aurally. Prothero et al. [Pro+95] believe that the user's perception of space enhances presence. According to them, a wide field of view and separating foreground and background are essential. A wide FOV allows the user to observe a larger part of the virtual environment, which is similar to observing the real-world with peripheral vision. Separating foreground and background enhances the perception of depth. It helps the user to better understand the spatial relationships between the virtual objects and their own position in the virtual space. Similar to Prothero et al. [Pro+95], Clive Fencott [Fen99] believes that the virtual environment's content and its design are crucial for presence. The subject of his research is the perceptual realism required to generate a sense of presence. He defines three perceptual opportunities to allow users of the VE "to explore, strategize, and generally feel some sense of control over what they are doing" [Fen99]:

- **Sureties** are trivial details of the VE that are appealing due to their high predictability, such as benches, streetlights, traffic signs, paths, doorways, and background traffic sound.
- **Shocks** are design elements that disrupt the sense of reality in the VE, such as the end of the environment, incomplete buildings, and seeing through cracks. Fencott [Fen99] also considers latency and motion sickness, resulting from poor design or excessive hardware usage, as additional components within this perceptual opportunity.
- **Surprises** are unforeseeable elements that emerge from the VE. They can be objects that tempt the users to do a certain action, help them to navigate the VE, or encourage them to engage more with the VE.

Lastly, McMahan [Mcm03] examined the importance of a synthetic social actor and an intelligent environment in creating a sense of presence. Synthetic social actors are computer-generated characters designed to simulate human- or non-human-like behavior and interactions. In VEs, they are most commonly virtual guides and virtual pets that contribute to the feeling of presence by emotionally connecting with users. Interactions with them can evoke feelings of empathy and companionship. Moreover, they can offer assistance and guidance within the virtual environment.

Intelligent environments, according to McMahan [Mcm03], "follow basic social cues" and "are designed to make the machine seem more human". Artificial intelligence (AI) features, including natural language processing, context awareness, learning, and adaptation, are primarily used in the VE to achieve this objective. Intelligent virtual environments offer realistic simulations and interactive features that enhance the sense of presence.

This thesis follows Mel Slater's [Sla03] definitions of immersion and presence and, in chapter 4, addresses some of the influential factors related to both concepts that were used for the implementation of GeoTravel.

2.1.2 Virtual Reality

In 1989, Jaron Lanier first introduced the term virtual reality (VR) while working at a company specializing in virtual reality technology and tools [Kru91a]. He popularized the term through his work and study but Steuer argues that Lanier's "device-driven definition of virtual reality is unacceptable" in his article "Defining Virtual Reality: Dimensions Determining Telepresence" [Ste92]. Steuer identifies three issues with this technology-based definition. The first one is classifying a VR system based on the presence of a minimum number of certain hardware devices. The next problem is not providing a basic unit of measurement for VR since it is broadly defined as a hardware system. Lastly, Lanier did not specify any parameters to evaluate different VR systems and compare them with one another.

After examining Lanier's definition, Steuer [Ste92] explores three additional definitions of virtual reality that are widely acknowledged but still make reference to specific hardware. The first one is from Coates [Coa92], who defines VR as "electronic simulations of environments experienced via head-mounted eye goggle and wired clothing enabling the end user to interact in realistic three-dimensional situations". For Greenbaum [Gre92] "virtual reality is an alternate world filled with computer-generated images that respond to human movements. These simulated environments are usually visited with the aid of an expensive data suit which features stereophonic video goggles and fiber-optic gloves." And Krueger [Kru91b] states that VR " typically refers to three-dimensional realities implemented with stereo viewing goggles and reality gloves." According to Steuer [Ste92], all of these definitions share the same limitations as Lainer's definition.

To address these limitations, Steuer [Ste92] proposes a definition of VR that focuses on the sense of presence. He introduces two terms to differentiate between the user's perception of their physical environment, referred to as "presence," and the environment generated by a virtual system, known as "telepresence." By utilizing the term "telepresence", Steuer [Ste92] offers the following definition of virtual reality:

"A virtual reality is defined as a real or simulated environment in which a perceiver experiences telepresence."

His broad definition is explained by the novelty of the hardware components used for previous definitions at that time. The definition of VR as telepresence, however, can be applied to existing as well as emerging media technologies. It associates VR with the individual experience of the user rather than with specific hardware. Moreover, the unit of measurement is explicitly defined as the individual user, which in turn allows VR to target different sensory stimuli, to change the interaction with the virtual environment and the representation of the user. Steuer [Ste92] focuses mainly on two of these dimensions: **vividness**, or the ability of VR to create a sensory-rich experience, and **interactivity**, or the extent to which users can influence the content of the environment.

The vividness of a VR system is determined by its technical features. Steuer [Ste92] suggests two important contributing factors. The first one is the "sensory breadth", which describes how many sensory dimensions the VE targets at once. The second one is the "sensory depth", or the level of detail of each such dimension. According to J. J. Gibson [Gib14], there are five primary sensory dimensions:

- The **basic orienting system,** which helps individuals orient themselves by using environmental cues such as light, sound, and motion.
- The **auditory system**, which helps people interpret sound stimuli such as pitch, volume, and spatial location.

- The **haptic perception system,** which targets the sense of touch and allows people to feel temperature, pressure, and different textures.
- The **taste-smell system,** which is responsible for the perception of smells, and flavors.
- The **visual system,** which handles visual information such as color, shape, motion, and depth.

Traditional media, such as television, typically have limited breadth and focus only on the visual and auditory dimensions. Newer technologies, on the other hand, increase their breadth by integrating more of the perceptual systems. Most gaming consoles, for example, incorporate haptic feedback via game controllers [Bro+10]. VR systems take a step further by incorporating haptic gloves [PV18] and scent-emitting devices [FIO21], enhancing the experience.

The degree of "sensory depth" within each perceptual dimension is also of great significance. Steuer [Ste92] describes it as the quality of the perceived dimension. The depth of the image depends on factors, such as its brightness, contrast, focus, sharpness and llumination [Aba+14]. The depth of media systems, however, is limited by the amount of encoded and transmitted data, which is not the case for real-world perception. Engaging multiple perceptual systems simultaneously proves to be highly effective in creating a sense of presence, even when certain stimuli have minimal depth. Furthermore, Steuer [Ste92] believes that the breadth and depth mutually enhance each other and generate a sense of presence in a multiplicative manner.

Similar to vividness, a VR system's interactivity is also influenced by its technical specifications. Steuer [Ste92] outlines three main factors that influence interactivity: **speed of interaction**, or how fast the input can be processed by the system; **range**, or how many actions are presented by the system at a time; and **mapping**, or how well the system's controls represent human actions in the real world.

The speed of interaction, or response time, is the first variable Steuer [Ste92] examines. The highest value it can take is real-time, in which the user's input is immediately processed and affects the environment without a noticeable delay. VR systems aim to provide real-time interaction to evoke a sense of presence.

The range of interactivity is influenced by the number of attributes within the system that are modifiable and the extend to which they can be altered. Steuer [Ste92] discusses four range dimensions:

- **Temporal ordering** defines the chronological order in which events occur and are perceived [Man86]. Possible actions that affect the temporal ordering of a video-based system, for example, include turning it on or off, pausing it, and fast-forwarding or rewinding it. Considering these actions, a recorded video has a higher range of interactivity than a live broadcast because it supports all the actions mentioned above, while a live broadcast typically only allows the user to play and stop it.
- **Spatial organization** refers to the positioning of objects within the VE. It includes factors such as the objects' location, orientation, and proximity to one another. When

comparing a 2D video game to a 3D one, it can be seen that the 3D game has a higher range of interactivity due to the additional dimension available for object placement and rotation [Pis].

- **Intensity** encompasses various elements, including the volume of sounds, brightness of images, and the strength of smells. In today's gaming industry, it is a standard practice for video games to offer a settings menu where the user can adjust the volume and brightness [Bom]. Video players provide users with the same range of interactivity as video games in this dimension, offering similar options to adjust parameters such as the volume and brightness.

- **Frequency** is most commonly observed in the context of sound and colour. Steuer's [Ste92] explanation on this dimension is limited, as he associates frequency with timbre, in the context of sound, and with colour, likely referring to the frequency of light within the visible spectrum. Specialized tools are available that allow the modification of timbre [FH98], however, light frequency is a physical property and cannot be modified [Din]. Without providing additional explanation, it remains unclear which attributes Steuer [Ste92] believes can be altered to modify the color frequency.

Mapping actions from the real-world to actions in the VE can be seen as a spectrum, ranging from a completely random and unrelated mapping to a one-to-one natural mapping. According to Steuer [Ste92], mapping is influenced by the types of controllers used for interaction and the similarity between the performed actions in the real-world and the executed actions within the VE. In the case of a one-to-one natural mapping, such as using a steering wheel controller to move a virtual car, little to no adjustments are required between the performed controller action and its execution within the VE, which increases the feeling of presence and game enjoyment [Ska+11]. In other scenarios, an appropriate metaphorical representation can create a more cohesive interaction, as in a VR game where flicking the wrist creates fire. In the case of an arbitrary mapping, such as the WASD layout used for controlling a character in a game, the mapping first needs be learned. Therefore, it is imperative that the mapping remains consistent within the system, eliminating the need for users to learn different mappings multiple times.

Burdea and Coiffet [BC03] define VR in the following way:

"Virtual reality is a high-end user-computer interface that involves real-time simulation and interactions through multiple sensorial channels. These sensorial modalities are visual, auditory, tactile, smell, and taste."

This definition outlines two essential components of VR, which have been recognized in previous definitions as well: **interactivity** and **immersion**. However, Burdea and Coiffet [BC03] introduce an additional aspect of VR, previously unmentioned: **imagination**. It is derived from the capability of VR applications to solve real problems in different areas, such as engineering, and medicine, via simulations. The effectiveness of the VR simulations, according to the researchers, depends "very much on the human imagination" and "the mind's capacity to perceive nonexistent things." Combining these three features of VR results in the VR triangle, which can be seen in figure 2.1.

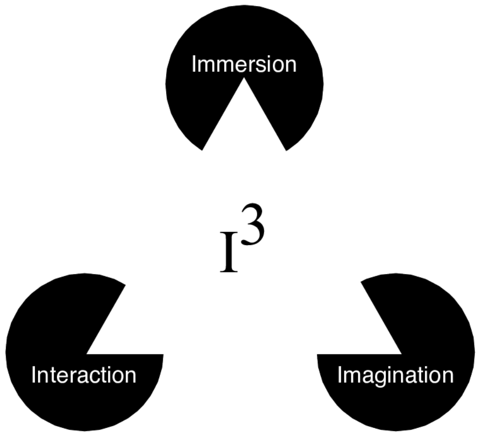


Figure 2.1: The three I's of VR, taken from [BC03]

Sherman and Craig [SC03] provide a similar definition to Burdea and Coiffe [BC03]. According to them,

"virtual reality is a medium composed of interactive computer simulations that sense the participant's position and actions and replace or augment the feedback to one or more senses, giving the feeling of being mentally immersed or present in the simulation (a virtual world)."

From this definition, four key elements of VR can be identified: a virtual world, immersion, sensory feedback, and interactivity.

The virtual world, as defined by Sherman and Craig [SC03], is a simulated space generated through a medium that describes a group of objects and the accompanying rules and relationships that control their behaviour and interactions. The terms "virtual environment" and "virtual world" are frequently used interchangeably, with virtual environment even being used as a synonym to VR. Although they do not argue with this interchangeable usage, Sherman and Craig [SC03] suggest that a virtual world can be understood as "an instance of a virtual world presented in an interactive medium such as virtual reality". Girvan [Gir18] supports this distinction between the two terms and believes that "a virtual world is a subset of virtual environment". In her perspective, virtual worlds aim to simulate the real-world and the social interactions it offers to individuals. A virtual environment, on the other hand, may not inherently promote social interaction.

Sherman and Craig [SC03] distinguish between two types of immersion: physical immersion and mental immersion. Their definition of mental immersion aligns with the concept of presence, as explored in section [2.1.1]. Physical immersion is explained as a "synthetic stimulus of the body's senses via the use of technology" [SC03]. In order to achieve physical immersion, the virtual world adapts to the user's position and orientation, and artificial stimuli are delivered to one or more of their senses in response to their actions. For example, when a user approaches an object, it seems larger. The extent to which real-world sensations are suppressed and the number of involved senses determine the degree of physical immersion.

The VR system provides direct sensory feedback, most commonly visual and/or haptic, to its users based on their physical position. In order to achieve that, the system must track the user's movement. Usually, their head and hands are tracked, but more advanced systems also support eyes [CKK19] and body joints tracking [Cas+19].

Sherman and Craig's [SC03] definition of interactivity does not differ from the one discussed in section 2.1.2, however, they address a crucial extension - collaborative environments. Collaborative environments allow multiple users to interact within the same virtual world. Each user is represented by an avatar that allows them to communicate with both the virtual world and other users' avatars [Ben+01]. Numerous collaborative environments have been created and evaluated concerning their ability to facilitate collaborative work [CS98; Hag96; Sno+96].

Following the definition of VR and its core components, Sherman and Craig [SC03] examine the fundamentals of VR technology and experience. They assert that the combination of user interface, virtual world, and real-world experience delivers a complete virtual reality experience. However, this thesis briefly focuses specifically on the user interface and its two levels, as depicted in figure 2.2.

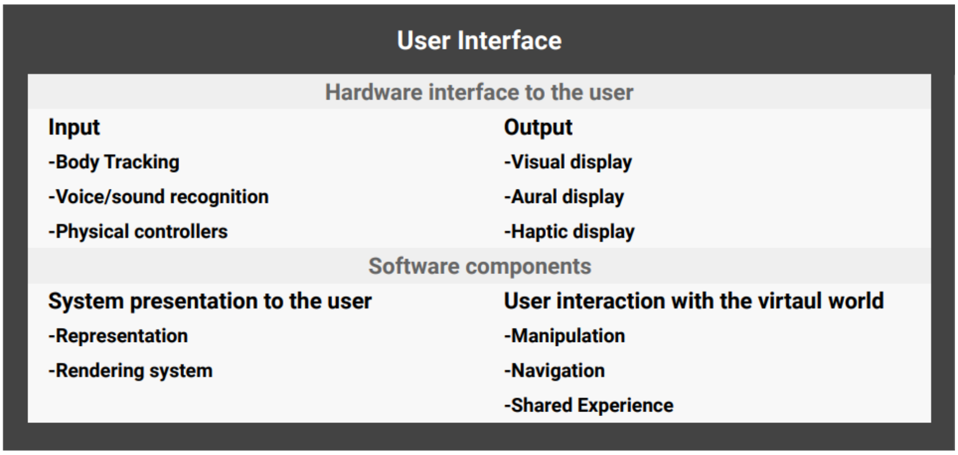


Figure 2.2: User Interface in VR, adapted from [SC03]

The interactivity and immersion of the VR system are directly influenced by the employed input devices. To facilitate user tracking and enable interaction within the virtual world, dedicated hardware is necessary. Sherman and Craig [SC03] discuss two types of input: **user monitoring** and **world monitoring**.

User monitoring includes the real-time tracking of the user's movement and actions. There are three important aspects that are considered:

- **Position tracking:** The VR system uses position tracking to determine the users' location and/or orientation within the VE. This is accomplished through the utilization of sensor devices that detect and track the position and/or orientation of the user, commonly used for head and hands tracking. Different types of position sensors exist - electromagnetic, mechanical, optical, videometric, ultrasonic, inertial, and neural [MAB92]. Each of them provides a different trade-off between accuracy, resistance to interference, and hardware limitations.

- **Body tracking**: It refers to the VR system's capability to detect and track positions and actions of various body parts of users, such as the head [Ama17], hands and fingers [Mar+20], and eyes [CKK19]. Tracking the user's head is necessary for almost all VR systems since its orientation and/or position determine how the VR world is displayed. Hand/finger tracking enables the integration of gestures that can be used for different actions within the virtual world, such as grabbing a virtual object or teleportation.
- **Physical controllers:** Physical controllers are devices that allow users to directly input actions into the VR system. They have various buttons, joysticks or touchpads that offer different functionality based on the use case. Some have built-in motion sensors to track the controller's movement, while others also provide haptic feedback [Son]. Furthermore, certain physical objects can be used as props and integrated into the virtual world. Their familiarity, ease of use and palpability allow for a more realistic and immersive experience [Hin+94].
- **Speech recognition:** Currently, speech recognition systems typically convert audio sounds into text strings [JR19]. In VR systems these text strings can subsequently be associated with a predefined response/action. Speech recognition enables a natural form of interaction, however, it encounters obstacles in real-time scenarios, such as ambient noise, varying accents, and the speed of speech [Pet16].

World monitoring examines the way the virtual experience adapts to a dynamic environment, creating a persistent virtual world. According to Sherman and Craig [SC03], a persistent virtual world evolves over time as users' actions have lasting effects that can lead to changes occurring even in the absence of a user. A non-persistent virtual world, on the other hand, always starts in the same state. In a persistent world, asynchronous communication is possible, as the changes made by one user are stored and their effects can be observed by another user at a later time. The client/server architecture is usually used to implement a persistent virtual world, with the server maintaining the virtual world and clients remotely accessing it. However, this architecture presents certain challenges, particularly in transferring the large amounts of data between the server and the clients, which can cause significant delays [Ali+20]. One strategy to address this challenge is to download the entire virtual world on the client side every time the VR application is started and then to only send incremental changes to the server. By applying this strategy, clients experience the delay only when launching the application.

The presentation of the virtual world to the user relies on the characteristics of the output devices. These devices typically stimulate the visual, aural, and haptic senses, and can be categorized into three primary types: stationary, head-based, and hand-based. Stationary displays are fixed screens or monitors that are positioned in front of the user [PSP93]. Head-based displays, or head-mounted displays (HMDs), are placed on the user's head and track its movements to adjust the displayed content accordingly [KS13]. Similarly, hand-based displays are fixed to or held by the user's hand [WS06].

Head-mounted displays are the most used type of visual displays. They have one dedicated display for each eye [KS13]. By tracking the orientation of the user's head, HMDs can adjust the displayed virtual environment in a way that aligns with the user's perspective in real-time. However, even slight delays in the tracking system can result in lagging between the user's movements and the corresponding changes in the virtual scene and cause motion sickness. Motion sickness includes symptoms such as eye fatigue, disorientation, and nausea [CKY20]. Another drawback of HMDs is its limited field of view (FOV) [SC03]. Expanding the FOV requires using a wider display to include the user's peripheral vision, which in turn could increase the overall size and weight of the HMD. However, HMDs have a complete field of regard that allows the user to see and interact with the whole virtual world from a first person point of view.

Headphones (head-mounted) and speakers (stationary) are the primary auditory displays in VR experiences. Headphones can be used by only one person at a time and are placed on top of their head. They deliver audio directly to the ears, which supports the immersive experience of 3D spatial sounds [Wyk]. Speakers allow multiple users to listen to the same audio simultaneously and are often more comfortable for longer use because they do not directly touch the ears. However, creating a 3D spatial sound is more challenging because multiple speakers are required and both ears hear the sound of each speaker.

Most haptic displays are designed for hands, with tactile displays being the most popular type. Tactile displays target the skin's ability to feel and interpret sensory stimuli such as pressure, temperature and texture [CMH08]. They generate these sensations by applying vibration and pressure. The two most commonly used methods to connect the user to the display are attaching actuators to the user's hand or having them hold a joystick.

2.1.3 Augmented Reality

The concept of augmented reality (AR) can be traced back to the 1960s and 1970s when Ivan Sutherland created the first head-mounted display [Sut68] and Myron Krueger [Kru83] explored the notion of "artificial reality", but the term was not officially defined until the early 1990s. Tom Caudell [CM92] introduced the term while developing a "heads-up, see-through, head-mounted display (HUDSET)", which was used to "augment the visual field of the user with information necessary in the performance of the current task".

According to Azuma's definition, augmented reality has the following three characteristics [Azu97]:

- **Combines Real and Virtual Environments:** Users perceive the real-world with seamlessly integrated virtual objects.

- **Interactive and Real-Time:** Users can interact with both the physical and virtual objects in real time.
- **Registered in 3D Space:** The virtual objects are aligned with the real-world physical environment in 3D space, giving the impression that both the virtual and real objects coexist within the same space.

These three characteristics also outline the technical specifications of an AR system. Specifically, it must have a display capable of combining real and virtual images, a computer system that can create interactive graphics that support real time interactions, and a tracking system that can accurately determine the user's viewpoint position, ensuring that the virtual objects are fixed within the real world [BCL15].

Based on the characteristics of AR and the details provided in section 2.1.2 about VR, the main differences between AR and VR can be summarized in the following way. AR users observe the real-world with integrated virtual objects, whereas in VR, they are completely immersed in the virtual environment, without having any view of the physical world. Furthermore, in AR the users can interact with both the physical and virtual environments, while in VR only the virtual environment can be manipulated. In AR, the physical and virtual objects share the same space and need to be spatially aligned, whereas VR does not require such alignment because the users only perceive the virtual world. Lastly, AR applications typically run on devices such as smartphones or smart glasses that overlay the virtual content onto the real-world. VR, on the other hand, utilizes more immersive devices such as VR headsets.

2.1.4 Augmented Virtuality

Milgram's virtuality continuum, shown in figure 2.3, depicts the connection between augmented reality (AR), virtual reality (VR) and augmented virtuality (AV). A real environment (RE) is any real-world scene without any virtual objects that a person observers directly or through a display. Opposed to it comes VR, which has been extensively discussed in section 2.1.2. A mixed reality (MR) environment contains both real-world and virtual objects. AR is a subset of mixed reality (MR) where virtual objects are integrated into the real-world. On the other hand, AV is another form of MR where real-world objects are integrated into the virtual world.

Despite the popularity of AR and VR, at the time of writing this thesis, there were only a limited number of scientific papers that explain the term AV as a part of broader discussions on mixed reality, and none were found that focus solely on exploring this specific subset of mixed reality and its characteristics. The similarity between AV and AR and Milgram's mixed reality continuum [2.3], however, can be utilized to derive some characteristics of AV. Taking into account the position of AV on the mixed reality continuum [2.3], it can be seen that AV's primary environment is virtual and it is augmented with real-world elements. Similarly to AR, AV is interactive, allowing users to engage with both the virtual and physical objects. Furthermore, interactions occur in real-time and have immediate effect on the environment.

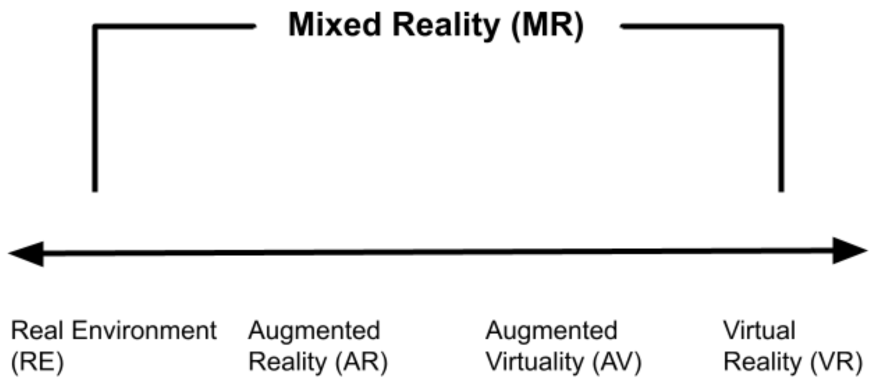


Figure 2.3: Milgram's mixed reality continuum, adapted from [Mil+95]

AV also requires spatial alignment between the virtual world and the physical objects for the user to perceive them as coexisting in the same space.

Although there is a limited number of papers concentrating on the theoretical aspects of AV, more research is available that describes its practical applications. Section 2.2 discusses some of these papers, exploring different use cases of AV.

2.2 Augmented Virtuality Applications

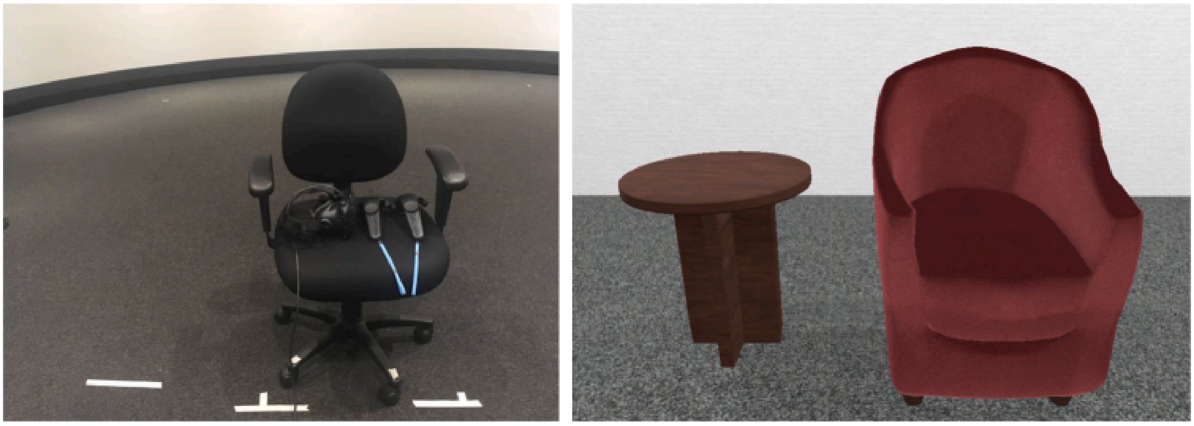
A handful of research studies have been conducted to explore the benefits of AV across various application areas, such as collaboration, teaching, training and tourism. This section explores the application areas of AV, examining its integration methods, as well as exploring the potential advantages and disadvantages of using AV in these specific contexts.

2.2.1 AV for Human Connection and Wellness

Three notable studies were conducted to examine how AV technology affects human connection and well-being. These researches aimed to understand how AV technologies could positively influence interpersonal relationships and contribute to the emotional and physical wellness of people. Gonzales et al. [ZRB21] investigated how AV technology could enhance presence and alleviate study-related stress in students. Korsgaard et al. [Kor+19] explored the effects of AV on remote social eating among solitary older adults, and Regenbrecht et al. [Reg+04] concentrated on the utilization of AV for remote collaboration.

Gonzales et al. [ZRB21] hypothesized that integrating real-world objects into the virtual environment would enhance the feeling of presence and lead to greater reductions in stress levels for students. Furthermore, they believed there would be correlation between the increased sense of presence experienced by the participants and the following decrease in their stress levels. To investigate these hypotheses, the researchers created a virtual counseling room. Two configurations were examined: one employing an AV setup and the other using a virtual setup. For the AV setup, a chair and a table were added to both the virtual and real-world environments. In the real-world, the students could sit on the chair and the table was on its right side. The virtual world had the same positioning of the two objects. They were also modeled to match the appearance and size of the physical objects. The virtual setup, on the other hand, used the same virtual environment, but in the real-world, the chair had a different appearance, and there was no table present. The two configurations can be observed in figure 2.4. Results of a user study revealed promising outcomes regarding the ability of AV to enhance the sense of presence and reduce stress levels; however, the correlation between presence and stress reduction was not confirmed. The AV configuration elicited a higher level of engagement with the visual aspects of the environment compared to the virtual setup. Since the chair and table were not interactive and could only be observed, Gonzales et al. [ZRB21] considered them as purely visual items and the main contributor to the feeling of presence. There was a decrease in the perceived stress (the subjective evaluation of the experienced stress [CKM83]) in both configurations. This observation contradicts the hypothesis about the correlation between presence and stress reduction since the decrease in stress levels was not solely attributed to the AV setup. However, in the AV setup there was also a reduction in the physical stress level, which is an objective measure and can be assessed using the heart-rate. This finding emphasizes the importance of incorporating real-world elements into the virtual experience. One aspect that has the potential to produce even more favorable outcomes but still requires further investigation is introducing interactivity to these elements.

(a) AV setup

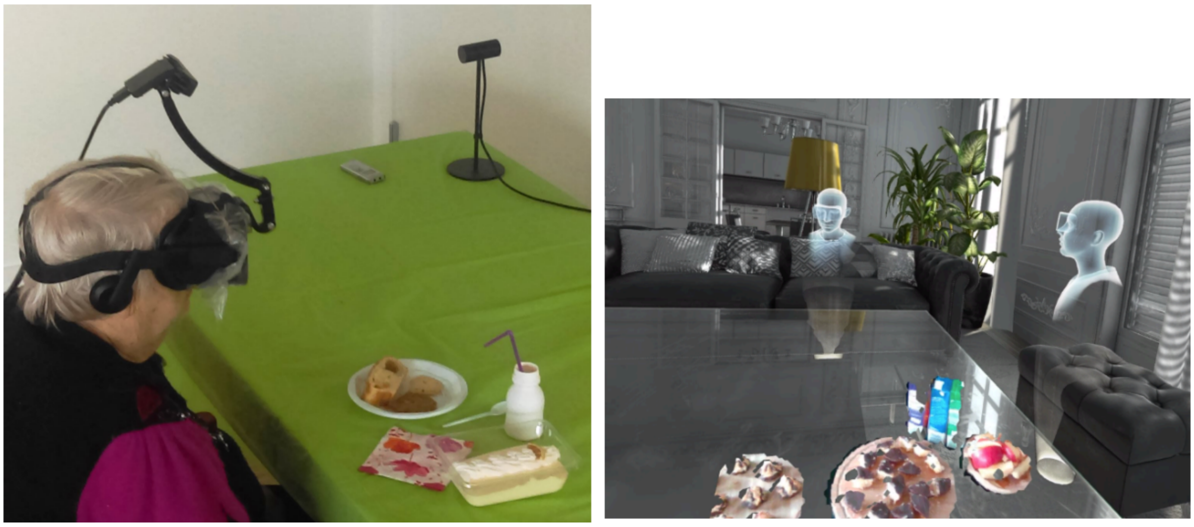


(b) Virtual setup

Figure 2.4: Left: Real life environment. Right: Virtual Reality Environment. Taken from [ZRB21]

Maintaining proper nutrition is of utmost importance for older adults, playing a critical role in preserving health, preventing illnesses, and ensuring a high quality of life [KØG04]. Moreover, malnutrition is often linked to feelings of social isolation and loneliness[San+20]. Building on earlier studies indicating that individuals tend to consume more food in social settings [De 90], and considering the demonstrated ability of VR technologies to evoke feelings of physical well-being and improved overall health perception among older adults [23], Korsgaard et al. [Kor+19] conducted research to address the undernourishment problems among older adults through the use of AV technology. They developed an AV-based system with a physical cake buffet and a virtual living room. The cake buffet was selected because it is often present at social gatherings and can be associated with positive perceptions [GTM12]. Older adults needed to wear an Oculus Rift HMD equipped with a depth sensor on the top. They could see and interact with the food, as the depth sensor captured and reconstructed it as textured geometry in the virtual world. A green tablecloth was used to set the table apart from the food placed on top of it. A custom rendering-plugin was then used with the data from the depth sensor to align the colour and depth images and create a point cloud. Subsequently, surface reconstruction was performed on the point cloud to obtain a surface mesh, which was then textured using the aligned colour image. The hands of the person were rendered in the same way. Figure 2.5a depicts the physical setup. Figure 2.5b displays the virtual environment of the system and the generated food meshes. As it can be seen, the virtual room had a realistic style, as it was preferred by older adults and suited the concept of the cake buffet [KBN19]. The position and orientation of the food meshes were computed by multiplying the matrix that described the HMD's position and orientation with the constant matrix transformation of the fixed angled depth sensor mounted on the HMD. The constant matrix was calculated based on the position and orientation differences between the mesh representations and a set of visual markers that were placed in the real-world. Three participants seated in separate rooms could use the system simultaneously. They were each represented by semi-transparent white genderless virtual avatars seating around a table. Each participant could only see the generated meshes of their own food and hands and not those of the others because the mounted depth sensor did not allow constant monitoring of the food and hands. The participants could also speak with one another in the virtual world. The feeling of co-presence was limited as the avatars lacked nonverbal communication cues. Nonetheless, the users of the system noted that they were aware of each other's presence. Overall, the system successfully created a pleasant atmosphere that facilitated communication, despite some flickering from the generated food meshes due to the lack of noise cancelling during rendering. Korsgaard et al. [KBN19] did not provide any information about the amount of food that was consumed by the older adults. One potential explanation for that could be that the novelty of the technology might have shifted the participants' focus from consuming the food to primarily observing and interacting with the virtual environment. Nonetheless, the researchers maintain a belief that the system could potentially yield health benefits similar to those associated with real social meals. Additionally, they acknowledge that a better virtual avatar could be advantageous in improving the communication among the participants and increasing the sense of co-presence. The used HMD was also found relatively heavy, consequently constraining the participants' comfort and movements. Furthermore, it restricted the ability to wear larger glasses, which are especially crucial for older adults. Additional research is required to evaluate the long-term benefits of the system and the impact of augmented virtuality on nutrition.

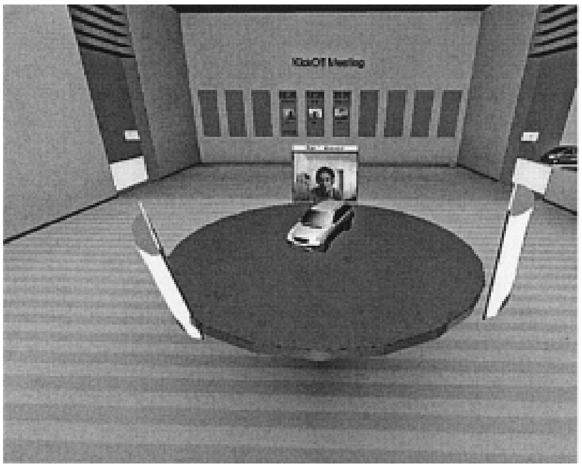
In the context of work culture, remote collaboration has emerged as an integral and foundational component. With the advancement of technology, people can work together, regardless of their geographical location. This approach to work fosters flexibility, and innovation, allowing individuals to connect, communicate, and collaborate from the comfort of their own spaces. VR's immersive and interactive features have been effectively employed to facilitate and enhance remote collaboration [Tru+21; LD97]. Regenbrecht et al. [Reg+04] explored the potential of AV to support remote collaboration. They designed an AV-based videoconferencing system capable of simulating face-to-face meetings over a network within an AV environment. The authors believed that using AV for remote collaboration could



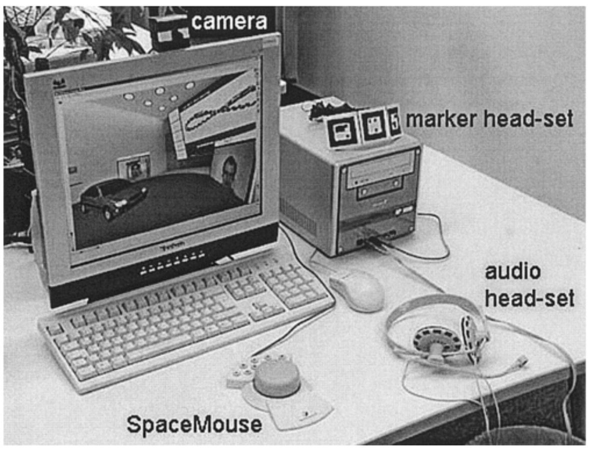
(a) Physical setup (b) Virtual environmen

Figure 2.5: AV-based prototype for social eating among solitary older adults, taken from [KBN19]

contribute to the feeling of natural communication, which was limited in existing videoconferencing systems. In the virtual world of the system, users could find themselves situated in a meeting room around a table, with a projection screen positioned in front of them. 3D models could be selected from boards on the wall and placed on the table to be interacted with. For user representation, live video was displayed on video planes that could follow the participants' movements in the real-world. The video streams were accompanied by synchronized audio. Additionally, participants had access to a status wall displaying the names and videos of all other participants, allowing them to assess their own visibility as well. The virtual room is depicted in figure 2.6a. Regenbrecht et al. [Reg+04] examined two hardware setups for their system. In the first setup, users were required to wear a HMD with headphones and a microphone. A camera positioned in front of them captured the video stream and monitored the user's head movements. The main benefit of this configuration was that users could observe the entire virtual environment simply by moving their head, however, they had to wear a HMD, which presented limitations, such as the field of view and resolution. In the second configuration, which can be seen in figure 2.6b, users were not required to wear any equipment. Instead, they had to sit in front of a monitor that was equipped with a camera, speakers and a microphone. Tracking the user's head was the main challenge of this setup, due to the absence of recognizable artificial features, such as the HMD. To overcome this challenge markers were placed on the user's head. Interaction with the virtual world focused on three key aspects - interacting with the 3D models, the projection screen, and other participants. This interaction was restricted to a seated position, which simplified the tracking process to only tracking the head position in the real-world, rather than the entire body. For the non-HMD hardware setup, the researches developed a technique called WorldWindow to recreate the free movement of the user's view into the virtual world that a HMD allowed. Utilizing computer vision techniques, the system detected the markers placed on the user's head and estimated their pose, enabling the video capture camera to also monitor head movements. The view into the virtual environment displayed on the monitor was then adjusted based on the detected head position. As an alternative to the WorldWindow approach, Regenbrecht et al. [Reg+04] integrated a space mouse that allowed rotation along two axes to change the field of view. The interaction with the 3D models and the projection screen used two pointing methods. The first one used the space mouse to send a virtual ray and the second one tracked a marker placed on a regular desktop mouse. In addition to the video streams of each participants and the synchronized audio as forms of interaction, the participants could point to each other with both mice and change the pose of each video plane. After conducting two user studies, Regenbrecht et al. [Reg+04] found that the AV-based videoconferencing system received positive feedback, with the majority of participants expressing a keen interest in employing it for their work-related tasks. The video quality was perceived as good, but the audio quality was a concern, with suggestions for improvement, such as better visual cues around the speaking avatar. However, the participants indicated a minimal sense of presence. Despite perceiving their partners as real, the virtual session failed to create a sensation of being in the same room with them. Some participants found the head-tracking less acceptable, potentially due to the system delays or the discomfort caused by the headbands with markers.



(a) Virtual conference room



(b) Marker-based hardware setup

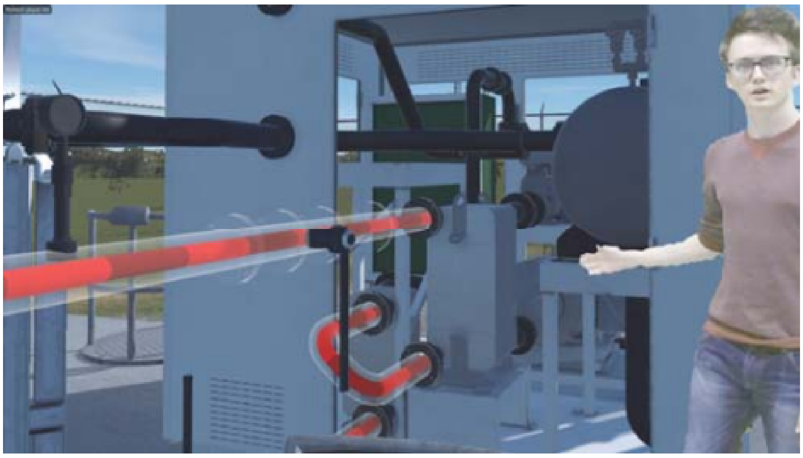
Figure 2.6: AV-based videoconferencing system, taken from [Reg+04]

2.2.2 AV for Educational Experiences

The intersection of education and new technology is crucial because it provides learners with dynamic experiences that inspire curiosity, and foster critical thinking. Examining the role of AV in education is important because it has the potential to further stimulate engagement and enhance learning by providing immersive and interactive experiences. Borst et al. [BLW18] and Vellingiri et al. [Vel+23] investigated how integrating photorealistic avatars into a virtual reality trip affects students. Ternier et al. [Ter+12] developed an open source tool for "creating learning games for augmented reality or augmented virtuality environments", and Albert et al. [Alb+14] decided to explore the potential of AV in improving hazard recognition skills of construction workers.

Developing photorealistic avatars is one of the more frequently researched application areas of AV. It holds significance for various reasons. First, a photorealistic avatar offers social interaction cues that conventional avatar representations often lack, such as gestures and body language [YAB18]. Furthermore, it serves as a frame of reference for the scale of objects in the virtual world, which is often misjudged [Jon+08]. Third, photorealistic avatars have demonstrated the potential to enhance learning [Shi17] and popularize AV-based educational tours.

The photorealistic avatars of teachers designed by Borst et al. [BLW18] were created using depth-based camera images. The researchers focused on two approaches for integrating the avatar. In the first approach, the avatar embodied a live teacher, captured via live-streamed depth camera video, guiding networked student groups. The second method was a non-networked variant, in which students could experience the field trip independently, and the avatar was created using depth camera recordings of the same teacher. The students were equipped with a HMD and a controller. Within the virtual world, the field trip took place at a solar energy plant, offering students the opportunity to engage with various educational stations and explore embedded educational components using ray-based pointing interactions. They could move between the stations via teleportation by selecting the desired target with the ray. The teacher was positioned in front of a TV that displayed a virtual mirrored view of the environment with an exaggerated FOV, as seen in figure 2.7b. With this view, the teacher could see themselves and point at virtual objects behind them while facing the students. Borst et al. [BLW18] acknowledged the drawbacks of such a view, namely the limited immersion of the teacher, however they focused on improving the students' experience instead. The teacher also had access to images of the students' view of the virtual environment to ensure that they were examining the correct stations. Furthermore, live webcam videos of the students were integrated into the teachers' virtual environment. They were positioned considering the teacher's head tracking, ensuring that the teacher's gaze was accurately directed towards a student. The avatar of the teacher, depicted on figure 2.7, was created as a 3D video-like mesh, generated by the RGBD information captured by a Kinect sensor. Moreover, the teacher's voice was also synchronized. The virtual energy plant was designed based on a real-world facility. Additionally, it was expanded with educational content to stimulate student interaction and facilitate the learning process. In the networked version of the trip, all students were locally situated at the same virtual location, with representations of other students appearing slightly offset to the left and right. The representations of the students were kept minimal, featuring generic head avatars. The voices of the students were broadcast to everyone, alongside the teacher's voice. The teacher asked various questions and responded to the students' answers and follow-up questions. According to the researchers, such two-way communication



(a) Teacher's avatar

(b) Teacher's view

Figure 2.7: VR field trip, taken from [BLW18]

enhanced the perceived level of presence of students and their motivation to learn, and better represented a real life trip. The non-networked version with prerecorded teacher did not support any communication with the teacher. The student was given multiple-choice quizzes instead and in the case of an incorrect answer, a brief explanation was provided.

Similar to Borst et al. [BLW18], Vellingiri et al. [Vel+23] designed an educational AV tour for students to learn how to recognise trees. However, they created photorealistic avatars of the students. The researchers utilized two distinct hardware configurations. In the initial setup, a single rear Kinect sensor was positioned behind the user to capture and display the user's photorealistic avatar moving forward through the tour. Additionally, it identified a Walking-in-Place (WIP) [NSN16] gesture as well as hand raising for questions. The second configuration involved the deployment of two Kinect sensors: one positioned in front of the user and another behind. The rear sensor was tasked with avatar creation and WIP gesture detection, while the front sensor was responsible for recognizing hand gestures. Previous studies suggested that employing multiple sensors resulted in enhanced accuracy for skeleton tracking, and frontal perspectives were generally more precise than rear perspectives [Mor+14; Wei+15]. However, this approach raised deployment costs and requirements for educational applications. Consequently, the primary aim of Vellingiri et al. [Vel+23] was to compare a single rear sensor against a front-and-rear sensor setup and their effects on educational systems. In addition to the Kinect sensors, both configurations of the AV nature tour system used an Oculus Rift HMD to place the user in the virtual world. The virtual world contained seven virtual tree exhibits that were positioned on either side of a virtual path. Each exhibit was augmented with audio, text and images to provide information about the trees. Following each exhibit, students were given three questions, each having three possible answers. The answers were presented in the form of images, and students could indicate their choice by using a hand-raising gesture. The system then displayed the right answer and gave more feedback if it was not correctly identified. The students could move between the exhibits with the WIP gesture.

User studies from both Borst et al. [BLW18] and Vellingiri et al. [Vel+23] indicated improvements in the knowledge acquisition of students. The networked approach with a live teacher of Borst et al. [BLW18] reported notably higher levels of social presence but the non-networked approach also produced positive ratings. Similarly, the students reported very high levels of attraction to the live teacher and the prerecorded version was well-received. However, the competence of the prerecorded teacher was evaluated notably lower compared to the live teacher. The perceived sense of presence, on the other hand, did not receive very high scores. The students expressed high motivation for both approaches, with a significantly higher engagement with the live teacher. Vellingiri et al. [Vel+23] did not observe a significant impact of the sensor placement on the acquisition of knowledge. However, the front-and-rear setup exhibited faster response times than the rear-only. The researchers believed, it resulted from the better tracking and gesture recognition facilitated by the front Kinect sensor. Despite that, the participants in the front-and-rear configuration took longer to complete the AV nature walk than participants in the rear-only condition. Vellingiri et al. [Vel+23] speculated that this behavior originated from frustration due to the notably longer response times associated with the rear-only setup's inferior tracking and gesture detection capabilities. They reached the conclusion that despite the limitations in tracking and gesture detection inherent to a single sensor, rear-only systems could offer cost-effective educational AV tour solutions that could be advantageous for students. Borst et al. [BLW18] confirmed the effectiveness of a live-guided educational VR tour. They believed this stemmed from the teacher's ability to capture the students' interest and effectively address misconceptions and inquiries. Nonetheless, the educational gains achieved through the networked version were not compared with the gains obtained from in-person live lectures. Conducting such a comparison would determine the degree to which the networked VR tour could function as an effective substitute or enhancement to traditional teaching techniques.

AV technology offers a unique way to engage learners with historical and cultural artifacts and support cultural heritage learning. By integrating real-world objects into the virtual environment, AV allows users to explore a virtual representation of historical landmarks and interact with physical props of heritage objects and artifacts. Furthermore, interactive AV simulations can enhance learners' understanding of cultural heritage and their motivation to learn more about it. Ternier et al. [Ter+12] created ARLearn to facilitate the development of serious games that can be positioned at various points along Milgram's mixed reality continuum [2.3] without requiring technological expertise. While this work does not explain the implementation specifics of ARLearn, it explores one of its application scenarios: how it is utilized to enhance motivation for cultural heritage learning. For this scenario, an adventure game integrating AV was created. The game followed a police story in Amsterdam based on real historic events. The player assumed the role of a police officer conducting an investigation into Amsterdam's drug scene. As part of the investigation, the officer was required to explore various famous locations within the city and learn their historical backgrounds. The researches believed that such locations would be interesting for student to explore and learn more about. The locations were created in the virtual world by using real-world data from StreetLearn, which is "a Google StreetView based user interface, that offers intuitive navigation and visualisation facilities" [Ter+12]. The game's concept received positive feedback from the majority of its users, and the user interface was generally considered attractive. However, it was noted that the interface, while appealing, was not particularly user-friendly and posed some obstruction to the overall experience. The learning content related to Amsterdam was generally regarded as engaging and effectively presented by most participants. However, some individuals found it overly complex. Additionally, the effectiveness of the game story in promoting cultural heritage learning was not unanimously supported, as half of the participants found it confusing, distracting, and overly simplistic. Even though it was based on historical events, the connection between the game's narrative and the actual locations was not clearly expressed. The evaluation of the Amsterdam case study provided Ternier et al. [Ter+12] with useful information on how to improve the game. Firstly, the content and narrative should be less complex and their connection should be better illustrated. Secondly, an alternative to StreetView should be chosen as StreetView challenges the learning process due to its inability to incorporate individual buildings, leading to additional distractions for students. Additionally, images of high quality should be integrated to support the selection of locations based on historical significance rather than image quality.

Hazard recognition is "the ability of individuals and groups of workers to identify and communicate hazards" [AHK13]. Hazards take various forms and are potential sources of harm and danger to individuals. Recent research by Carter and Smith [CS06] revealed that people struggle to identify hazards in dynamic and unpredictable environments and between 10% and 33.5% of hazards remain undetected. Traditional approaches to hazard identification rely on experienced instructors for knowledge transfer. However, hazard recognition skills are mainly tacit and cannot be easily taught with passive training methods [CCZ03]. Given the acknowledged advantages of using VR and AR applications to enhance safety education, Albert et al. [Alb+14] developed an interactive AV training environment known as SAVES (System for Augmented Virtuality Environment Safety), utilizing the benefits of AV to improve hazard recognition skills. SAVES is a hyper-realistic 3D environment where workers encounter a range of representative work scenarios. The system evaluates the workers' capability to identify hazards, offers feedback, and teaches them how to detect hazard stimuli in a safe virtual environment through retrieval mnemonics. The virtual world of SAVES simulates a petro-chemical facility, allowing users to engage with various construction scenarios. Numerous real-world construction images are integrated in the system, with a subset of them depicting different hazard types and severity levels. SAVES features an interactive interface that offers guidance for more precise hazard localization within the virtual environment. Additionally, it provides users with real-time feedback as they engage in hazard recognition activities. Once a user detects a hazard stimulus, an interface form emerges, prompting the user to identify the hazard sources. Using the user's input, the system provides the respective feedback by using a points system. Albert et al. [Alb+14] utilized energy-based retrieval mnemonics and serious gaming theory to facilitate the training process. A retrieval mnemonic is a memory technique that helps individuals recall information or specific details by associating them with easily remembered cues or triggers, such as patterns [Scr+10]. To provide construction workers with reliable cognitive cues and mnemonics, Albert et al. [Alb+14] categorized the hazard sources simulated in SAVES based on their energy type, for example mechanical, electrical, chemical, biological, etc. The use of the mnemonic was then demonstrated to the workers in SAVES. Serious games have a purpose other than pure entertainment [DAJ11] and, in the context of training and education, utilize different learning techniques in combination with gameplay elements. Charsky [Cha10] believes that serious games, which have clear objectives, well-defined rules to achieve these objectives, decision-making elements, and a feedback system for performance evaluation, are effective in stimulating and engaging learners. Furthermore, as highlighted by Garris et al. [GAD02], the integration of instructional content within serious games can foster self-motivation and self-directed learning among learners. Following these guidelines, Albert et al. [Alb+14] specified a clear goal within SAVES, namely the accurate detection of hazards in the virtual world. The system was also designed to provide instant feedback on the worker's performance. The instructional content was developed using cognitive mnemonics and representative construction scenarios. A user study of SAVES reported a substantial improvement of 27% in hazard recognition skills, demonstrating the advantages of incorporating retrieval mnemonics, serious gaming concepts and augmented virtual environments into safety training. AV offers workers the opportunity to engage in realistic construction scenarios without exposing them to actual risks. This allows them to practice identifying and responding to potential hazards in a controlled environment as many times as needed. A missing aspect within SAVES that could potentially yield even more substantial improvements is integrating representative audio.

2.2.3 Summary

Table 2.1 provides a concise summary of the previously discussed application areas of AV. It shows both the broader application domain and the specific use cases, the adopted AV integration methods, and some noteworthy insights gained from the research papers examined.

Authors	Application Domain	Use Case
Gonzales et al.[ZRB21]	Human Connection and Wellness	Presence and Stress Reduction
Korsgaard et al.[Kor+19]	Human Connection and Wellness	Remote Social Eating for Solitary Older Adults
Regenbrecht et al.[Reg+04]	Human Connection and Wellness	Remote Collaboration
Borst et al.[BLW18]	Educational Experiences	Photorealistic Avatars and Educational Tours
Vellingiri et al.[Vel+23]	Educational Experiences	Photorealistic Avatars and Educational Tours
Ternier et al.[Ter+12]	Educational Experiences	Cultural Heritage Learning
Albert et al.[Alb+14]	Educational Experiences	Hazard Recognition

AV Integration Method	Remarks
Real-world objects into the VE	Interactive objects needed
Dynamically generated mesh representations of food	Better virtual avatars needed
Live video of users	Better audio quality needed
3D video-like avatar meshes	Higher engagement with live teacher
3D video-like avatar meshes	Inferior tracking and gesture detection of rear-only systems
Real-world images of locations	High quality images needed
Real-world images of hazards	Audio needed

Table 2.1: AV Applications Summary

2.3 Virtual Reality Applications for Older Adults

Virtual reality has significantly influenced a wide range of fields, such as healthcare, education, training, and entertainment. According to Lee et al. [LKH19], existing studies of VR technologies for older adults are limited because researchers primarily focus on improving their design and usability for younger users, neglecting the specific needs and preferences of older adult users. However, certain researchers recognize the potential of VR to enhance the quality of life of older adults [Hug+17a]. Several studies have been conducted to evaluate the potential of VR in detecting and managing health conditions that affect not only older adults but are often more closely related to aging [Lee+03; CSD08; Kar+21]. The following section explores the application areas of VR for older adults, along with its challenges and design issues. A brief summary of the main reviewed papers can be seen in table 2.2.

2.3.1 VR for Promoting Overall Wellness

Lee et al. [LKH19] examined the potential of VR to promote wellbeing in older adults. They analyzed existing literature to outline the primary factors that impact the physical, emotional, and psychological wellbeing of older adults in VR.

The physical wellbeing of older adults encompasses their ability to engage in physical activities and fulfill social roles effectively. The main factors that influence it are:

- **Motivation:** encouraging older adults to develop an interest for physical activities. As a novel approach, VR could transform the typically disliked by seniors physical exercises into enjoyable activities. Given VR's capability to construct immersive and detailed virtual environments, Lee et al. [LKH19] suggest that these environments could greatly interest older adults and increase their willingness to interact with them. Therefore, participating in physical activities within these virtual worlds could have the potential to reduce the negative perceptions often associated with exercising and increase the motivation and interest of seniors. Hughes et al.'s [Hug+17b] research aligns with this viewpoint, confirming the potential of VR when used with existing gaming systems to engage older adults in healthcare facilities and within their residences. Baragash et al. [BAG22] also have a similar opinion.

- **Training:** training older adults to perform physical activities they might be hesitant about because of the associated risks, with the goal of enhancing their overall physical strength. VR enables the creation of risk-free simulations of various environments that could be utilized for both cognitive and physical training of elders [Has+21; COM20], eliminating the need to expose them to potentially hazardous situations.

- **Reminder:** displaying useful virtual information about the health of older adults in real-time. VR has been used to assist people with different physical and cognitive impairments with their daily activities, such as managing medications [Kur+07], preparing meals [Fol+18], and retaining memory [Opt+10].

- **Accessibility:** ensuring good accessibility for older adults to utilize the VR technology and minimizing any physical challenges they could experience while using the system. For seniors to safely use VR systems, it is necessary to familiarize them with the equipment, provide assistance for using it and avoid over-exertion [Ija+22].

Designing VR applications taking into consideration the physical wellbeing of older adults is challenging. According to Porter [Por05], older adults often worry about being seen as physically weak or ill. Consequently, the use of external devices, such as a HMD, could potentially amplify these concerns. Motion sickness and dizziness caused by VR should also be taken into consideration because older adults are more susceptible to them [Gol06]. Such discomfort could negatively impact their overall experience and willingness to engage with VR technology. Another challenge to address is the reduced attentional capacity among older adults. Designers should aim to minimize the cognitive load by limiting the number of tasks or objects that older adults need to focus on simultaneously [Eis+18].

The social wellbeing examines the feelings of belonging and reduced loneliness experienced by seniors when they interact with technologies. The four main factors that need to be considered according to Lee et al. [LKH19] are:

- **Remote participation:** Allowing older adults to interact and participate in social events without being physically present at the same location.

- **Virtual interaction:** Creating a virtual environment that supports interaction with other participants.

- **Emotional relationship:** Ensuring older adults experience a similar emotional connection as in the real- world while socializing within the virtual environment.

- **Community services:** Providing assistance and guidance to accommodate the wide range of activities provided by VR systems.

Remote participation is particularly important for older adults due to their reduced mobility, travel limitations, and health concerns. With technological advancements facilitating remote participation, the integration of such capabilities into VR does not present significant challenges. VR allows seniors to explore locations they cannot visit in the real-world and to engage with others. However, achieving high levels of immersion and co-presence is challenging when designing VR applications for older adults for several reasons. First, many older adults are less familiar with emerging technologies, such as VR, and need more time to learn how to use them. Therefore, minimising the use of controllers and including natural mapping of controls for interaction is essential [YC21; PT12]. Second, the concept of virtual avatars is also new to seniors. To create a higher sense of immersion and allow older adults to identify with their virtual representation, the avatars need to be highly realistic and reflect the individual's appearance [Bak+19].

Lastly, the psychological wellbeing of older adults describes their emotional and subjective wellness, that is how they manage their emotions and their overall sense of well-being and life satisfaction [Die00]. It is influenced by:

- **Positive emotions:** Eliciting positive emotions while using the VR system.
- **Virtual content :** Creating enticing virtual content that can continuously engage the user.
- **Independence:** Supporting older adults' independent lifestyle.
- **Environmental mastery:** Supporting the ability of seniors to effectively manage and navigate through the virtual environment.

According to Lee et al. [LKH19], the occurrence of system errors, challenges related to hardware setup and software updates, and issues in user interaction could evoke negative emotions among older adults, potentially impacting their sense of independence and control. Consequently, simplifying the setup process and implementing measures to prevent errors becomes imperative. In the event that errors do occur, it is crucial that they can be easily resolved. Enriching the virtual content with interactive games, music, and creative activities, such as drawing, could provide a stimulating and engaging environment for older adults. However, this process should not hinder the environmental mastery. Starting with less complex VR environments and gradually increasing their complexity could help older adults to adapt. Furthermore, providing clear and intuitive instructions and interfaces is essential for navigating through the virtual environment.

2.3.2 VR for Health Assessment and Rehabilitation

Traumatic brain injuries (TBI) and Alzheimer's disease (AD) impact a significant number of older adults. Around 32% of hospitalizations and 28% of deaths resulting from traumatic brain injuries (TBI) in the United States occur among individuals aged 75 years and older [DP]. Approximately one in every nine seniors older than 65 is affected by Alzheimer's disease, and one in every three seniors passes away with Alzheimer's [Ass]. TBI can result in lasting cognitive impairments and consequent behavioral challenges, affecting the patients' ability to perform activities of daily living (ADLs). Older adults with AD often have difficulties navigating and become disoriented and lost while walking or driving, which poses risks to both their health and public safety, and undermines their personal autonomy and quality of life [SFT06]. It has been observed that many traditional methods of assessing brain-injured individuals can be insensitive to cognitive impairments [LO06] and that methods for assessing the navigational abilities of individuals in the early stages of AD (EAD) are scarce. VR with its ability to create immersive and controlled virtual environments that are both realistic and safe for users offers a promising solution for overcoming these limitations and supporting rehabilitation.

In their paper "A Virtual Reality System for the Assessment and Rehabilitation of the Activities of Daily Living" Lee et al. [Lee+03] introduced a virtual supermarket to assess and train the cognitive ability of brain injured patients in performing daily life activities. The supermarket contains different display stands and refrigerators. To interact with objects in the VE, the patient needs to be close to them and then press a joystick button to select them. After an item is selected, it is placed inside a virtual shopping handcart. The patient is assigned two tasks: an exercise task and a main task. In the exercise task, the participant is initially introduced to the virtual supermarket, instructed on navigation techniques, and guided on how to interact with objects. The specific navigation method, however, was not explained in the paper. In the main task, the participant has to find and collect a list of specified items and place them in the handcart. Lee et al. [Lee+03] conducted an experiment with five patients undergoing rehabilitation treatment for traumatic brain injuries or stroke. Each patient first performed the exercise task. Subsequently, they completed the main task five times within a span of five days. The system recorded various metrics including elapsed time, distance covered, collisions with walls, selected goods count, opened refrigerator doors count, joystick button presses count, and error rate. When excluding the data from the first day, there was a tendency for the time, distance, and collisions to decrease. Conversely, the number of selected goods and button presses increased, while the error rate decreased. On the initial day, the patients struggled to execute two tasks simultaneously. By the third day, they were more successful in navigating the supermarket aisle and retrieving items from the shelves. Hence, the effectiveness of repeated training in VR was demonstrated. One of the participants was a 65 years old man who performed poorly in the experiment, indicating the need of the system's design to be updated for older adults. The experiment was supervised by a physiatrist, who noted that the patients faced challenges with the handheld joystick due to its weight and instability. Moreover, they encountered challenges while navigating the virtual supermarket, primarily due to physical impairments resulting from paralysis. Nevertheless, the physiatrist believed in the effectiveness of VR as an ADL training method, provided that interfaces could be designed to address the physical challenges faced by the patients.

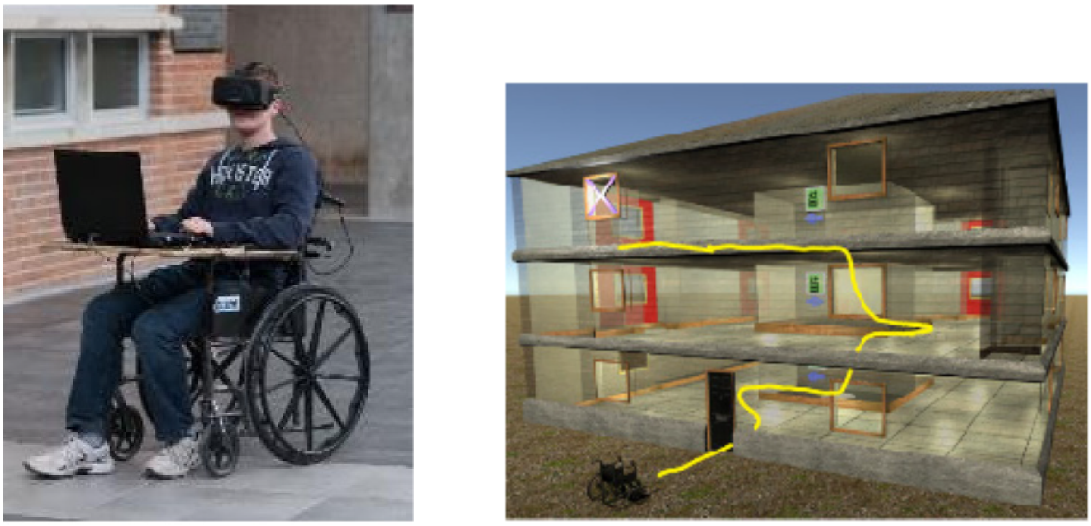
Cushman et al. [CSD08] hypothesized that a VR test environment would help the early detection of navigational impairment in AD and Mild Cognitive Impairment (MCI). They created a highly realistic virtual environment based on a real-world hospital from detailed floor plans and photographs of the hospital lobby. The real-world hospital and the created VR hospital are illustrated on figure 2.8. The researchers did not use a HMD but a desktop PC to display a 3D view of the lobby from wheelchair height since the system was to be used from a seated position to avoid physical discomfort. Different tasks could be executed in the VE to assess the navigational skills of patients. For example, they were tasked with memorizing a route they had navigated within the real-world hospital. Subsequently, they had to retrace this route within the VE while recalling and pointing at various objects along the way. The results of a conducted user study confirmed the researchers' hypothesis that virtual environment testing could be used to identify impaired navigation abilities, however, the observed scores were slightly lower in this type of testing compared to real-world assessment methods. Despite that, VE assessment systems have certain advantages compared to traditional real-world testing methods. They could simulate high-risk or dangerous situations in a controlled environment; the VE could easily be adjusted to the specific patient's situation; data, such as response time and accuracy could be collected to assist analyzing the patient's behaviour.

(a) Real world hospital

(b) VR hospital

Figure 2.8: VE for early detection of navigational impairment in AD and MCI, taken from [CSD08]

White and Moussavi [WM16] further researched the ability of VR to improve the cognitive capacity of people with AD. They believed that performing navigational exercises within a VE could be used as a treatment approach at the early stages of AD. Since standard interaction devices, such as joysticks, could be confusing to older people, the researchers decided to integrate a novel input system called “VRNChair”, which was designed by Moussavi et al. [MM14]. “VRNChair” uses a standard wheelchair that is equipped with a motion capture unit. It can estimate the magnitude and the direction of movement of the wheelchair and translate it to the VE. In such a way, the user can sit and move the wheelchair to navigate within the VE and use a HMD to freely observe their surroundings. Figure 2.9a shows the “VRNChair”. The VE consists of a symmetric three-story building. In this VE, the user is initially presented with an external view of the building, which has a randomly selected window highlighted with an "X". Subsequently, the user enters the building and is tasked with locating the designated window from the inside. In figure 2.9b, an illustrative path is depicted, in which the user has to make turns and navigate to upper floors. One person at an early stage of AD tested the system. The researchers noted significant progress in his navigating within the VR Building. Moreover, the participant’s spouse noticed improvements in his everyday life, notably in his driving skills and overall emotional state. The VR system of White and Moussavi [WM16] demonstrated the capability of VR to transfer learned skills to the real world. Furthermore, over a short period of 4 weeks of using the system, the participant significantly improved his navigational skills. The “VRNChair” also allowed for a natural and intuitive interaction and Moussavi et al. [MM14] observed a notably reduced motion sickness.



(a) “VRNChair” (b) Target path

Figure 2.9: VR Building spatial navigation exercise, taken from [WM16]

2.3.3 VR Exergames

Exergames, categorized as serious games that combine cognitive stimulation with physical activities, have demonstrated their potential to effectively motivate people with dementia to engage in physical exercises [Van+18]. Padala et al. [Pad+17] studied the effects of a supervised Wii-Fit program on balance in older adults with dementia. The results of their experiment indicated enhanced balance and decreased fear of falling among the patients. According to Unbehaun et al. [Unb+18], exergames could additionally improve the social interaction between dementia patients and their family members. Dementia involves various cognitive impairments, often accompanied by behavioral and psychological symptoms, such as aggressive motor behavior, depression, or hallucinations [Löv+08]. The World Health Organization reports that over 55 million individuals worldwide are affected by dementia, with approximately 10 million new cases emerging each year [Org]. Furthermore, dementia mainly develops in older people and eventually most patients require assistance from others with their daily life activities. Despite ongoing efforts, finding a cure for dementia remains a challenge [Löv+08]. However, current research indicates that engaging in cognitive and physical exercises can mitigate the risk and slow the progression of this condition [HAO04]. These methods present certain challenges, such as lack of motivation in patients, distractions from external stimuli, and the absence of proper supervision [SG04]. Because the disease's impact varies from person to person [Org], the need for personalized treatment approaches arises. Technological advancements have outlined VR and serious games as promising options. As discussed in section 2.1.2, VR creates immersive environments through multiple sensorial channels. According to Karaosmanoglu et al. [Kar+21] these characteristics of VR have the potential to reduce distractions from external stimuli and help dementia patients focus on the tasks presented. VR environments could also be customized for the specific needs of the user. Tabbaa et al. [Tab+19] explored the effect of VR on hospitalized dementia patients. Their research demonstrated the ability of virtual environments to invoke past memories of patients with dementia. However, VR technology comes with specific drawbacks, including motion sickness and limited range of motion, which must be carefully considered while developing a VR system for older adults with dementia. The high level of immersion in VR could also potentially cause undesirable effects, such as feelings of distress.

Karaosmanoglu et al. [Kar+21] believe that VR-based exergames facilitate the development of customizable and immersive virtual environments, enabling safe and supervised participation in physical activities. To address the limited research regarding the best design practices for such systems for dementia patients, they introduced a human-centered design approach for developing VR exergames for dementia patients and designed their own VR exregame called "Memory Journalist VR". Human-center design approaches focus on understanding and incorporating the user's requirements into the development process to ensure the usability of the system [MAG01]. This approach outlined four themes regarding the user requirements of dementia patients: engaging in immersive environments simulating natural settings, such as beaches and forests, and in sports activities; involving spectators and caregivers in social gaming activities; providing clear game instructions, comprehensive hardware explanations, intuitive controls, task repetitions and different difficulty levels; and utilizing stationary VR systems to prevent potential injuries and having constant presence of caregivers to maintain connection with the real-world and offer support. Taking the four themes into consideration, Karaosmanoglu et al. [Kar+21] developed "Memory Journalist VR" to stimulate the cognitive and physical abilities of people with dementia and to invoke some of their past memories. The patients could experience a 360-degree recording of famous locations in their capital and city of residence. For an intuitive interaction, the researchers created a 3D printed camera and replicated it in the virtual world. A Vive tracker [Viv] and a button were placed on the printed camera to track its pose and simulate real-world image capture interaction. The game could be played from a seated position to reduce the risk of falling and motion sickness. Moreover, caregivers could control it via a browser application, Remote App, and the VR experience was displayed on a TV screen to share it with spectators. The patients needed to wear a HMD and its tracking base station was fixed to the wall to ensure the patients' safety [Fig.2.10 a)]. The patient played as a reporter, whose job was to take photos of famous landmarks. The role and tasks were presented with audio-visual interactive tools within the VE. An introduction to the camera controller was also given, with enough time for the patients to become familiar with it. Remote App allowed the caregivers to start and end the experience, to preview and switch to another location, and to choose the difficulty of the task. Two difficulty levels were supported: easy and hard. The easy task involved detecting a landmark by only moving the head. The hard task, on the other hand, required arm movements as the patient needed to use the camera to zoom or rotate it to change the orientation of the photo. The patient was guided by a virtual parrot that was explaining the tasks, as depicted in figure 2.11a. A short explanation of the task was also given on a sticky note attached to the camera and through text displayed above the camera's view frame which can be seen on figure 2.10 b). As an additional reminder for the patients, the target locations were highlighted with sparkles. If the patient encountered any difficulties while performing the task, the caregiver could also send a hint through the parrot. If the patient pressed the camera button while pointing at the right location, a newspaper with the captured image appeared to indicate the completion of the task [Fig.2.11b]. If a wrong landmark was captured, only the taken picture appeared without a newspaper. Furthermore, the caregivers could activate bonus scenes, in which no tasks were given and free exploration was possible. After finishing the VR exergame session, both the patient and caregivers could review the captured photographs on the Remote App, fostering a shared sense of accomplishment.

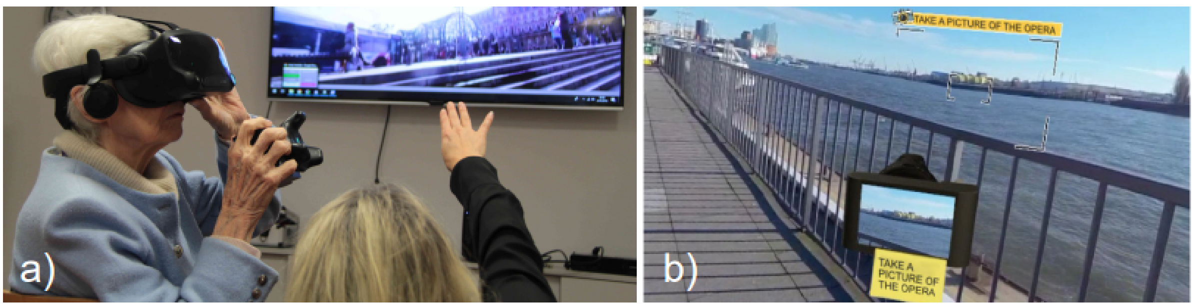
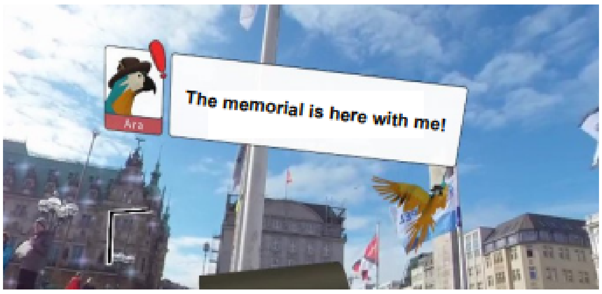
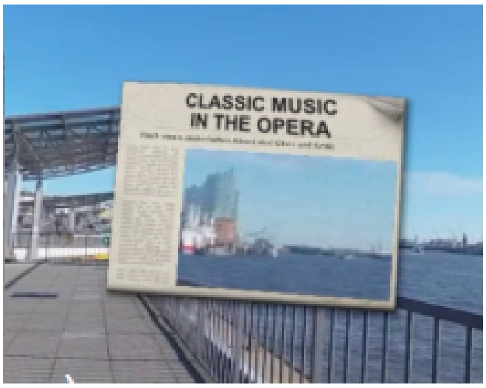


Figure 2.10: The hardware setup (a), the VE and virtual camera (b), taken from [Kar+21]



(a) The parrot



(b) The newspaper

Figure 2.11: Memory Journalist VR, taken from [Kar+21]

The researchers confirmed their hypothesis that VR exergames for people with dementia could be effectively integrated into their daily life. "Memory Journalist VR" was able to trigger positive emotions and social interactions in dementia patients and even the recollection of some of their memories. The VEs that were included in the game were interesting and immersive and cognitive stimulation of patients was observed. The 3D camera device proved to be an intuitive controller without requiring significant cognitive effort and the system was not physically demanding except for one participant that experienced some discomfort. However, there are certain limitations associated with using a Vive tracker. Depending on the object to which they are attached, Vive trackers could add additional volume and weight that could be uncomfortable for users. They are also relative expensive and are not compatible with all VR systems.

Authors	Application Domain	Use case
Lee et al.[LKH19]	Promoting Overall Wellness	Defining Key Impact Factors
Lee et al.[Lee+03]	Health Assessment & Rehabilitation	VR Supermarket for TBI Patients
Cushman et al.[CSD08]	Health Assessment & Rehabilitation	Navigational VE for AD and MCI Patients
White and Moussavi[WM16]	Health Assessment & Rehabilitation	Navigational VE for AD Patients
Karaosmanoglu et al.[Kar+21]	Exergames	Stimulating Dementia Patients

Table 2.2: VR Research Summary

2.4 Analysis and Definition of Guidelines

The reviewed papers in this chapter outlined some important guidelines for the development of an immersive VR application for older adults with integrated AV objects. The following guidelines were derived and need to be considered:

1. **High audiovisual quality:** To achieve better perceptual realism, the audio and video used in the application need to be of high quality and synchronized. Therefore, real-

world images and videos with accompanying audio could be incorporated, which could also stimulate the memory of seniors, create a sense of familiarity for them, evoke emotional responses and make the VR experience more accessible and enjoyable.

2. **Realistic avatars:** The player's representation in the virtual environment is important for establishing a high level of presence. Realistic virtual avatars are needed that offer social interaction cues. Moreover, an egocentric view is advantageous because it mimics how the real-world is perceived and how people interact with their physical surrounding.

3. **Social interaction:** Facilitating social interaction is an essential aspect of VR applications designed for older adults. It fosters a sense of social presence within the VE, reduces social isolation, and promotes overall wellbeing.

4. **Synthetic social actors:** Incorporating virtual guides and virtual pets enhances the sense of presence and allows for the establishment of emotional bonds with the users. Furthermore, it elicits feelings of empathy and companionship while providing valuable assistance and guidance within the virtual world.

5. **Rich sensory experience:** Sensory-rich VR environments are essential for seniors because they not only compensate for sensory limitations but also enhance engagement, stimulate cognitive functions, and promote the emotional wellbeing and the overall quality of the VR experience.

6. **Natural mapping of actions:** It ensures that the controls within the VE mimic real-world actions, which is of utmost importance for older adults. It allows for intuitive interaction that minimizes the cognitive load. Controllers designed with natural mapping are user-friendly and have a minimal to no learning curve. Additionally, real-world physical objects could be used as such controllers due to their familiarity to users.

7. **Limited number of objects and tasks:** Virtual environments should not overwhelm older adults with numerous objects and tasks that need to be observed and executed at the same time. This guideline prevents seniors from experiencing physical and cognitive overload, reduces their levels of stress and anxiety, and allows them to concentrate more effectively.

8. **Remote participation:** It addresses various challenges faced by older adults, including mobility limitations, social isolation, and health concerns. Moreover, it promotes social engagement and connectivity with others, and stimulates the cognitive functions.

9. **Stationary VR systems:** To reduce the risk of accidents while moving around, stationary VR systems are preferred for older adults. These systems allow seniors to remain seated during the VR session, which minimizes physical discomfort. Furthermore, since they require less movement, they decrease the likelihood of motion sickness.

10. **UI:** The UI for older adults should prioritize simplicity and clarity. It should feature large text and a clean, uncluttered design while maintaining consistency throughout

the application. Clear and concise instructions are essential, and the interface should provide feedback, such as visual and auditory cues, to indicate user interactions.

11. **Virtual environments:** The preferred VEs for seniors simulate real-world settings, such as beaches, forests, and cities. Engaging in social and physical activities within these virtual environments further enhances their appeal.

12. **Interactive AV objects:** Integrating real-world objects into the virtual world that are interactive is essential because they create more immersive and engaging VR experience. They contribute to the realism of VEs by simulating real-world interactions, which is useful for educational and healthcare applications.

13. **Universal tracking methods:** The tracking methods used to support AV should not be limited to a particular VR system. Moreover, the size and weight of the tracking devices needs to be considered to prevent any physical discomfort among older adults.

3 Background - *GeoTravel*

GeoTravel is an ongoing project that focuses on the development of a VR travel application for older adults. It was first created by George Nassif for his master's thesis [Nas20]. The main goals he outlined for the application targeted the solitary and monotonous life of seniors. GeoTravel was designed to enrich the daily lives of older adults by allowing them to virtually travel to different locations, and to improve their mental health, motivation and fulfillment while ensuring a user-friendly technology experience. Alexander Williams continued to work on the project in the scope of his bachelor's thesis [Wil21] and integrated AV features to support the interaction of seniors with the VE. Julia Schwan then further expanded the project for her bachelor's thesis [Sch21] by incorporating a virtual dog for guidance and companionship. This chapter examines the main concepts and features that were included in GeoTravel prior to the start of this thesis and outlines the objectives established for it.

3.1 Concept and Features

Initially, GeoTravel was designed for the Oculus Quest HMD, as it offered the best trade-off between cost and performance at the time. Julia Schwan [Sch21] then updated the application for the Oculus Quest 2 HMD to utilize its superior resolution, storage, and refresh rates. The right touch controller of the Oculus Quest 2 is used to interact with the VE and only one of its buttons is pressable. Within the virtual environment, it is represented as a walking stick to align with the concept of natural mapping, as its shape closely resembles that of the Oculus controller. Furthermore, the walking stick is a familiar object to seniors and was selected to prevent confusion and fear of use. The walking stick is augmented with a virtual button positioned at the top, approximately at the same location as the physical button of the controller, to assist seniors in recalling which button to press. The instruction word "wählen" is constantly displayed next to the button to indicate its purpose. Moreover, the virtual button changes its colour from yellow to green when the right physical button is touched or pressed as a visual feedback. A ray is displayed starting from the handle of the stick to assist the interaction. When the ray intersects an interactive UI element it turns red along with the instruction word. Figure 3.1 shows the walking stick.

When the application starts, the player is introduced to a mountain scene. Positioned in front of them is an interactive panel displaying several locations that could be explored. Initially, only Munich and Hamburg could be visited, but later Egypt was also added. After a place is selected, the user is showed a 360-degree image of it. Three locations, each featuring three significant landmarks, were included in the application. Navigating between these landmarks is supported in three ways: via a menu that follows the view of the player; with floating images in the VE; and through a 3D interaction system. The menu consists of two sections. The "Orte" section displays all locations that are integrated, and the "Sehenswürdigkeiten" section - all the different landmarks of the respective location. For example, "München" as location offers the landmarks "Mariensäule", "Fischbrunnen", and "Neues Rathaus". After selecting "München" from the "Orte" section, "Mariensäule" could be explored as a first landmark. The VE floating images show the two remaining landmarks that could be visited. And the 3D interaction system includes mini 3D models of each landmark within a given location placed on a shelf next to the user. The menu and the floating images can be selected with the assistance of the walking stick's ray, and the 3D interaction system requires the stick to touch one of the models before pressing the button. Figure 3.2 illustrates the three navigation methods. To access additional locations and landmarks not found in the "Orte" and "Sehenswürdigkeiten" sections, a web-based search feature was implemented. Users have the option to input the name of the place they wish to explore, and various images of that place are presented for selection. Such places, however, are not permanently stored in the application and do not support any of the navigation methods.

(a) ...when physical button is not pressed (b) ...when physical button is not pressed

Figure 3.1: Virtual button on walking stick, taken from [Sch21]

To assist the navigation within the VE and provide companionship to the user, a virtual dog (fig.3.3a) was integrated into GeoTravel. Initially, it is positioned beside the player. Following a brief greeting period, it walks towards the floating images of the landmarks within the virtual environment, trying to direct the user's attention toward them. When all the landmarks within a specific location have been visited or if the web-based search is used, the virtual dog lies down next to the player. The incorporation of a virtual dog aligns with the principles for developing a VR application tailored to the needs of older adults, particularly in the context of synthetic social actors that offer guidance and companionship (guideline 4). While Julia Schwan [Sch21] did not introduce any additional communication mechanisms for the virtual dog as Karaosmanoglu et al. [Kar+21] did for their parrot, the virtual pet still provides some limited interaction by welcoming the player and engaging their attention through tail wagging.

The AV functionality serves two main purposes: firstly, to prevent panic attacks within the virtual environment triggered by a feeling of disconnection from the real-world, and secondly,

(a) "Orte" section

(b) "Sehenswürdigkeiten" section

(c) VE floating images

(d) 3D interaction system

Figure 3.2: Navigation methods

to remind the user to stay hydrated. To accomplish this, an external camera is used to capture a live video feed of the user's surroundings and to track a physical cup of water. The video is displayed on a virtual table (fig.3.3b) in front of the user, and a 3D model resembling the physical cup is incorporated into the virtual environment, mirroring the position of the actual cup. The video feed displays only the region surrounding a detected marker. The physical cup must be green to distinguish it from the background and enable tracking. However, this approach does not facilitate the estimation of a 3D pose.

The last important feature of GeoTravel is its multiplayer option. Two players can simultaneously use the application and explore it together. Their HMDs need to be connected to the same LAN and one of the players takes the role of a host to synchronize the VE. Players have the option to upload a photo of themselves before launching the application. This allows their face to be extracted and used either alone or together with a static generic avatar. However, as the multiplayer feature was not working when this thesis started, a more detailed explanation of how it was fixed and improved is provided in chapter 4.

(a) Virtual dog (b) Live video feed

Figure 3.3: Images taken from [Sch21]

3.2 Thesis Goals

The thesis goals outlined in this section are derived from the state of GeoTravel as it was presented to the author before the start of the study, and the guidelines discussed in section 2.4. The objectives are more precisely examined in chapter 4.

- **Project upgrade:** GeoTravel was running on the Unity Game Engine [Teca], version 2019.3.5f1, and used version 1.2.0 of the Oculus XR Plugin [Man], which were both already outdated at the time. Furthermore, the multiplayer option and other features were either nonfunctional or working only partially. Therefore, the whole project has to be upgraded and the code base needs to be revisited.

- **Better virtual avatar:** The user's virtual avatar remained static in the VE. To create a more immersive and interactive experience, a responsive and lifelike avatar is essential, capable of mimicking the user's movements.

- **More immersive virtual environments:** The VE consisted of static 360-degree images of famous landmarks, offering a certain degree of immersion. However, a notable enhancement could be achieved by replacing these static images with dynamic 360-degree videos, complemented by synchronized audio. This modification would introduce motion and auditory elements, thereby fostering a more immersive and interactive environment.

- **Improved object integration and interaction with AV:** The AV marker detection and cup tracking were implemented using a C# version of OpenCV, integrated as a Unity plugin [Too]. However, the plugin was no longer updated or supported and no marker pose estimation was available. Consequently, the marker detection and tracking need to be updated. Furthermore, to enhance the realism of the VE, it is essential to incorporate real-world objects that are better suited to the environment that is captured in the 360-degree videos. It is also equally important for older adults to feel at ease when using these objects.

4 Implementation

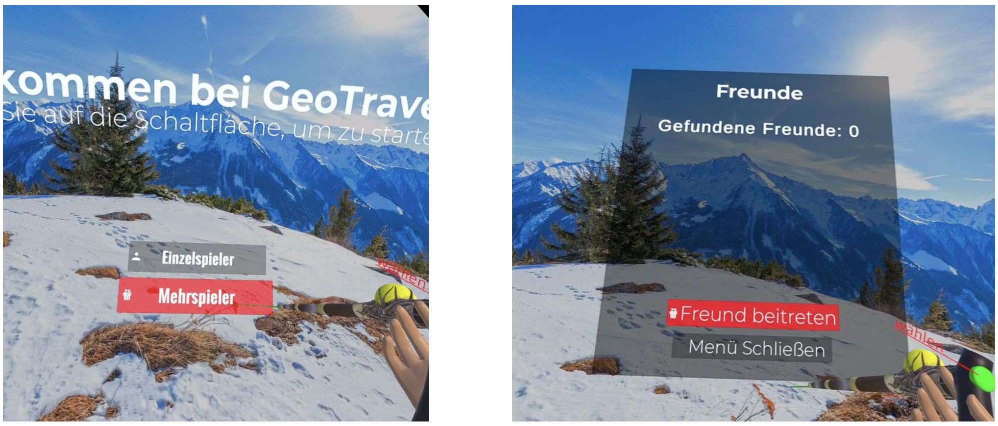
In this chapter, a detailed examination of the thesis goals listed in section 3.2 is conducted, with a specific focus on understanding the rationale behind them and the methodologies employed to achieve them.

4.1 Hardware Setup

GeoTravel runs on the Oculus Quest 2 HMD, an all-in-one VR headset that operates independently without the need for a PC, and has two touch controllers [Met]. It enables Six Degrees of Freedom inside-out tracking, accurately capturing both the user's head and body movements without the need for external sensors. The headset has four cameras on the front, however, no access to their content is provided to developers. Consequently, a small external camera attached to the top of the headset is used to enable the AV features. During the development process, the Logitech C270 HD Webcam was employed [Log]. Since the Oculus Quest 2 HMD does not support any external cameras, the headset and the Logitech camera had to be connected to a PC. Using the Quest Link connection, the HMD could run applications from the Unity Editor and the external camera could be accessed. A PC with an Intel Core i7-4790K Processor running at 4.00GHz and a NVIDIA GeForce GTX 1070 Graphics card was used for the development.

4.2 Project Upgrade

The project was upgraded from Unity version 2019.3.5f1 and Oculus XR Plugin version 1.2.0 to the Long Term Support(LTS) Unity version 2021.3.15f1 and Oculus XR Plugin version 3.2.3. The upgraded game engine and plugin introduced new features that conflicted with the functionality of the older versions. Therefore, the project was modularly reintegrated into the new environment. During the reintegration, a range of problems surfaced, with one of the most significant ones being the nonfunctional multiplayer feature, intended to meet the criteria outlined in guideline 8. As a networking library, Mirror was utilized [Mir]. Mirror has dedicated components for managing the networking aspects of a game and detecting servers on the LAN. Connecting to servers on the Internet/WAN is also possible, but for GeoTravel, the LAN coverage suffices. This decision aligns with the application's initial purpose, which was to help seniors suffering from dementia and Alzheimer's disease [Sch21], the majority of whom reside together in healthcare facilities. Mirror operates as a server authoritative system, however, one of the participants can function as both a client and a server simultaneously, acting as a host and eliminating the need for a dedicated server. GeoTravel is designed to support two players at the same time, a host and a client. Initially, each player directly started the application as a host, loading an online scene that would be synchronized across the network. To connect with a friend, the player had to access the "Optionen" menu and click the "Freunde finden" button. Subsequently, the game would search the LAN for an active server, and if one was detected, an attempt would be made to join it. However, this process presented an issue. When attempting to join the detected server, all scripts were reloaded and the player again operated as a host. To resolve this problem, an offline scene was created, allowing users to choose between playing alone or with a friend. The scene can be seen on figure 4.1a. If the player prefers the single-player mode, they assume the role of the host and load the online scene. Being a host is a prerequisite to allow friends to join the server at any time. If the player opts for the multiplayer mode, the application displays the IPs of all active servers (fig.4.1b), allowing the player to select the server they wish to join as a client. Upon successfully joining the server, the client's environment is synchronized with that of the host. If, while hosting in single-player mode, the user decides to join a friend, they can navigate to the "Optionen" menu and select "Freunde beitreten." This action again loads the offline scene, where the player can press the "Mehrspieler" button and open the multiplayer panel. To synchronize the online scene across the network, the existing code managing the scene elements had to be rewritten. A challenge that was faced here was synchronizing landmarks that the host downloaded via the web-based search feature. Given that these landmarks are still represented by 360-degree images, the image sizes proved too large to be transmitted directly over the network. Therefore, only the URL of the image is sent and the client internally downloads the image on their side from the web-based search feature. In the multiplayer mode, both the host and the client can choose which landmarks to explore together. All UI elements, the 3D navigation system, the virtual dog, and the AV objects are only locally displayed to each client. However, the virtual avatars of the players and their movements are synchronized over the network.



(a) The two modes (b) The multiplayer panel

Figure 4.1: The offline scene

The next problem identified during the reconstruction of GeoTravel was the missing functionality of the "Zuletzt besuchte Orte" button (fig.4.2a). In her thesis, Julia Schwan [Sch21] described the purpose of the button. It was created to allow players to revisit previously explored landmarks from the web-based search within a VR session without the need for new searches or image downloads. Such functionality is needed because the navigation methods described in chapter 3 are not applicable for web-based visited locations. In addition to implementing this functionality, the author extended it by displaying the most recent three landmarks that were explored by the user. The UI panel displayed to the user after pressing the button is illustrated on figure 4.2b and it is similar to the first navigation method (fig.3.2b) to ensure consistency.

(a) ... button

(b) ... panel

Figure 4.2: Recently visited places

Additional minor issues were addressed, including fixing a nonfunctional button designed to reset the VR session, which is responsible for restoring the game to its initial state once the online scene is loaded. If the player is connected to a friend's server, resetting the experience does not disconnect them. Instead, both the host and the client reload the online scene and can continue their shared experience. Some functions were invoking obsolete methods, and as a result, they were updated. Certain 3D models of the landmarks were missing in the 3D interaction system, thus necessitating the integration of new models. The orientation of the helper ray for the walking stick was also modified following a discussion with two seniors regarding their preferred way of pointing with a cane. Instead of pointing at an angle from the handle of the stick, the ray now extends directly along its stem to ensure a natural mapping of the pointing action (guideline 6). When the "Optionen" menu was activated, it caused an overlap with the navigation menus, resulting in a cluttered interface. To resolve this issue, when a particular menu is opened, all other menus are automatically disabled (guideline 10).

4.3 Virtual Avatar

The importance of the player's representation in the VE was thoroughly examined in Chapter 2 and resulted in the formulation of guideline 2. A virtual avatar had already been incorporated into the GeoTravel application. The avatar had a standard human body appearance, but it did not track or mimic the user's movements, remaining static within the environment. Instead of featuring a 3D model of a head, the avatar displayed a 2D image of the user's head and shoulders. Users had the option to upload a headshot-style photos of themselves, from which their head and shoulders could be extracted. A default photo was used in cases where the users did not provide their own photo. The static virtual avatar lacked social interaction cues that are vital for creating an immersive and interactive experience. Furthermore, considering the novelty of VR technology for older adults and the concept of virtual avatars, a virtual avatar capable of mirroring real-world movements not only aligns with guideline 6 but could also prove advantageous in enhancing engagement and user satisfaction. As a result, a new virtual avatar was incorporated, sourced from the VR application Sebastian Walchshäusl developed for his master's thesis [Wal22]. He designed a generic 3D avatar of an elderly male, which featured rigging and the ability to mimic the user's head and arm movements, as detected by the Oculus Quest 2. Tracking was limited to the head and arms since Sebastian Walchshäusl's [Wal22] application used teleportation as a navigation method within the VE, which was activated through hand gestures. In GeoTravel, on the other hand, the users observe the VE from a stationary standpoint, and no teleportation within the VE itself is supported. This decision was made to minimize the potential occurrence of motion sickness among seniors and to align with guideline 9. In a similar way as described above, Walchshäusl's [Wal22] avatar displayed a 2D image of the user's head and shoulder. Additionally, it featured a virtual 3D nose placed within the user's field of view to help reduce motion sickness. Nevertheless, this element was omitted in GeoTravel as its appearance contrasted with the overall realism of the VE. The extraction of the head and shoulders was also changed. In both Schwann's [Sch21] and Walchshäusl's [Wal22] projects, the extraction process was implemented using an OpenCV plugin for Unity [Too], which was developed in C# and offered all major features of OpenCV 3. It is important to note that this plugin has not received updates or support since 2019, and as of the time of writing this thesis, the latest available version of OpenCV is 4.8.1 [tea]. To leverage the newest features and capabilities of OpenCV, its original C++ version 4.8.0 is employed in GeoTravel. However, since it is written in C++ and the scripting language used in Unity is C# [Tecb], exchanging data between the two languages is needed, a process commonly referred to as marshalling [Mica]. The exact details of the process are elaborated further in section 4.5. To summarize, the process involved defining a C++ API that exposed the extraction functionality. This API was subsequently compiled into a dynamic link library (DLL) and imported into Unity, where it could be accessed within the C# scripts. OpenCV has an implemented classifier to detect faces in images, which is used in the extraction function. It relies on Haar-like features [LM02] and the AdaBoost training algorithm [Has+09]. After detecting a face, an oval mask is generated around it. Given the assumption that headshot-style photos are uploaded to GeoTravel, the function also extracts the relative positions of the neck and shoulders, in addition to the head. Figure 4.3a illustrates the resulting image following the extraction process, which is then combined with the 3D model of the avatar. Sebastian Walchshäusl's avatar, as described in his work [Wal22], did not incorporate a walking stick, as players used only their hands for interaction with the environment. In GeoTravel, the primary method of interaction is through the use of the Touch controller, which is represented by the walking stick within the VE. Therefore, the right 3D hand of the avatar was adjusted to hold the cane. As it is described in section 4.5, three physical objects are incorporated into the VR environment, prompting users to hold them in their hands and interact with them. Thus, it is necessary to alternate between tracking the Touch Controller and hand tracking since the current Oculus XR plugin does not support simultaneous tracking of both. When the Touch controller is in use, the virtual avatar's left hand does not mirror the movement of the user's left hand; instead, it remains stationary at the side of the avatar's body. When hand tracking is active, the walking stick is deactivated, and the avatar once again replicates the movements of the user's hands (fig.4.3b). GeoTravel provides the option to completely deactivate the 3D avatar. In such instances, only the virtual hands remain visible, and if the Touch controller is being tracked, the right hand will also hold the walking stick. To enhance social interaction, the avatar's movements are transmitted over the network, allowing two players engaged in the same session to observe each other's actions. The reason for only using static images of the users' faces was made to avoid the need for a second external camera that would need to be moved around, considering that users have the freedom to turn in any direction to explore the VE.

(a) Third POV (b) First POV

Figure 4.3: The Virtual Avatar

4.4 Virtual Environment

GeoTravel was first designed as a VR travel application specifically for older adults. It included a virtual environment with three different locations, each featuring three famous landmarks. The users could visit München, exploring its Mariensäule, Fischbrunnen, and Neues Rathaus; Hamburg with its Speicherstadt, Justizforum, and Kehrwiedersteg; and Egypt with its Pyramiden, Khufu Schiff, and Sphynx. Choosing famous landmarks around the world serves two primary objectives: enabling seniors to virtually travel to unexplored destinations, and incorporating locations they may have previously visited, thereby stimulating their memories. Since GeoTravel is a project developed in the Technical University of Munich with a primarily German target audience in mind, Munich and Hamburg were incorporated as potentially visited locations to evoke memories, familiarity and positive feelings among seniors. Egypt introduced an element of excitement, offering a possibly unexplored destination to add the thrill of discovering new places, thereby motivating older adults to engage with the application. In their respective user studies, Williams [Wil21] and Schwan [Sch21] both noted the participants' expressed interest in visiting more locations. Among their wishes were natural settings, such as forests or beaches, as well as specific destinations such as England, Norway, Paris, and the USA. Schwan [Sch21] introduced a web-based search feature that retrieves images from Flickr [Fli] and allows users to explore any location of their choice. They have to simply write the name of the desired destination and downloadable images are displayed. However, a drawback of this method is that the majority of the images are of poor quality, making them unsuitable for the application. Following these observations, three new destinations were included in GeoTravel: London, Paris, and Turkey. In addition to the participants' preferences, London and Paris were chosen because high quality 360-degree images of their most famous landmarks were available online. Turkey, on the other hand, was included based on the discovery of three high-quality 360-degree videos, two of which captured captivating scenes of a forest and a beach.

In London, the user can visit Piccadilly Circus, London Eye, and Big Ben. The three landmarks in VR can be seen in figure 4.4. In Unity, the 360-degree images of the landmarks are mapped to a cubemap with a latitude-longitude layout. This cubemap is subsequently used to create a material for a Skybox. In each location, the respective Skybox is displayed, allowing the users to explore the landmark in VR. All three navigation methods explained in section 3.1 are supported in this VE. London is included as a location in the "Orte" section (3.2a) and the three landmarks can be visited through the "Sehenswürdigkeiten" section. In the VE, floating images of the landmarks are displayed and the 3D interaction system contains mini 3D models of the landmarks, which are illustarted on figure 4.4c. The selection of landmarks was guided by the image quality of the accessible 360-degree images, following guideline 1. No audio was included because of the static nature of the images. When incorporating London as a new location, the same issue regarding image selection was encountered, as described by Ternier et al. [Ter+12]. The landmarks, although among the most famous in London, were selected based on their image quality rather than their historical value or popularity among tourists.

Paris is the second new location that was integrated. In this city, the users can explore the Eiffel Tower, the Notre Dame Cathedral, and the Louvre, as it can be seen in figure 4.5. The VE was created in the same way as described for London. To introduce some variety to the experience, an evening 360-degree image of the Notre Dame Cathedral was chosen, as all other landmarks were captured during the day. Once again, the user has access to all three navigation methods. Both London and Paris can be explored together with a friend through the multiplayer mode.

Despite the static nature of the used 360-degree images, it is the author's belief that they would be successful in creating an immersive VE. Ternier et al. [Ter+12] and Albert et al. [Alb+14] utilized real-world images for their respective research topics and observed their positive impact on the target groups.

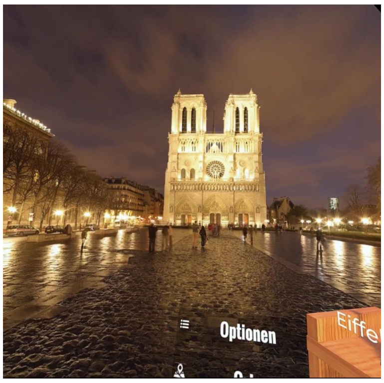
(a) Piccadilly Circus

(b) Big Ben

(c) London Eye

Figure 4.4: London in VR

(a) Eiffel Tower



(b) Notre Dame Cathedral

(c) Louvre

Figure 4.5: Paris in VR

Turkey is the last new location added to GeoTravel. It differs from the rest because instead of 360-degree images 360-degree videos were used. Following guideline 1, high-quality videos were incorporated, complemented by synchronized audio. In alignment with guideline 11 and the preferences voiced by participants in the user studies conducted by Williams [Wil21] and Schwan [Sch21], this location includes not only a historical landmark but also natural settings. The users have the opportunity to explore an ancient Roman amphitheater in Hierapolis, an enchanting forest with a river, and a serene beach resort. Snapshots from the videos can be seen in figure 4.6. To construct the virtual environment, a video player was added to the Unity scene. It processes the video clips of the landmarks and produces a render texture for each frame. This texture is subsequently used to generate a material for a Skybox, utilizing a panoramic shader. The resulting Skybox is presented to users, allowing them to fully immerse themselves in the environment. Furthermore, each video is accompanied by synchronized audio. As the original audio in the video clips was found to be insufficiently loud for older adults, it was extracted from the clips and its volume was amplified. The video of the Roman amphitheater is not particularly dynamic since it primarily focuses on the architectural design of the amphitheater. However, in the background, tourists can be seen walking around the landmark, taking photos of it and engaging in conversations. The audio captures their voices and together with the video creates an immersive VE, in which the user could observe the landmark in peace but still feel the presence of other tourists. In contrast to the relative stillness of the amphitheater, the forest setting provides a more dynamic environment. Its central part is the running river. In addition to it, different birds and butterflies can be seen flying around, contributing to the liveliness of the surroundings. The audio complements the visuals, capturing the sounds of the river, the rustling of the forest trees, and the different noises of animals. The beach resort creates a balance between the stillness of the amphitheater and the liveliness of the forest. It displays the sea with its small waves, the moving leaves of palm trees and people that are relaxing on sun loungers, or walking along the shore. The sound of the wind and sea can be heard in the audio and the splashes of people walking in the water contribute to the sense of realism in the virtual scene. In addition to enhancing the feeling of presence through high-quality 360-degree videos and synchronized audios, Turkey differs from the rest of the locations by not employing the 3D navigation system. This decision was made to maintain a less cluttered environment and preserve the scene's realism, as the appearance of the 3D navigation system was found to be too unrealistic. Nevertheless, the other two navigation methods are supported. In the multiplayer mode, in order to ensure that a client who joins the VR session of the host late is synchronized to the exact video frame the host is observing, the video frame number is transmitted over the network.

Following the observations of Karaosmanoglu et al. [Kar+21], which assert that 360-degree recordings of famous locations effectively generate captivating and immersive VEs that elicit positive emotions, cognitive stimulation, and memory recollection in patients with dementia, the author of this work hypothesizes that the 360-degree videos of Turkey integrated in GeoTravel could demonstrate similar results in older adults.

(a) Roman amphitheater (b) Forest (c) Beach resort

Figure 4.6: Turkey in VR

4.5 Object Integration and Interaction with AV

After examining the project update (sec.4.2), the integration of a new virtual avatar (sec.4.3) and virtual locations (sec.4.4), this section focuses on the primary research topic of the thesis - how AV technology is used for the integration of and interaction with real-world physical objects within GeoTravel. First, the communication with the OpenCV DLL is examined (sec.4.5.1). Subsequently, the camera calibration process and its objectives are explained (sec.4.5.2), followed by exploring the two implemented tracking methods — Aruco marker tracking (sec.4.5.3) and Vuforia tracking (sec.4.5.4). Finally, the three AV objects integrated into the VE are presented (sec.4.5.5).

4.5.1 OpenCV Dynamic Link Library

As it is briefly explained in section 4.3, to leverage the latest features and capabilities of OpenCV, the previously utilized in GeoTravel OpenCVSharp Unity plugin [Too] had to be replaced with a Dynamic Link Library (DLL) containing an API to functions implemented in C++ that use OpenCV 4.8.0 [tea]. Marshalling is necessary to enable the exchange of data between C#, the scripting language used in Unity, and the C++ code of OpenCV. A Dynamic Link Library (DLL) is a file that offers shared functions and resources to be called from a separate application [Micb]. In contrast to static linking, which loads the library into the application's memory at linking time and copies all the object code into the application during the build, dynamic linking loads the library either at load time or runtime. It only copies the necessary information needed by Windows at runtime to locate and load the DLL. The DLL created for GeoTravel and imported into Unity contains all the functionality needed for the detection and pose estimation of Aruco markers (4.5.3), the extraction of faces from a photo (4.3), and the cropping of a region of interest (RoI) in an image. To illustrate the complete process of invoking a function from the DLL within a C# script in Unity, the cropping of a RoI is examined. This functionality was initially introduced along with the marker detection to display a live video feed of the user's surroundings around a detected marker on a virtual table and was implemented in C#. After the integration of the OpenCV DLL, it was separated from the marker detection process as a standalone function that takes an image, two points to define the location of the region in the image, and a padding factor to specify the desired spread around the region. This function is then called with the coordinates of the top-left and bottom-right corners of the detected marker and crops a region around the marker with the specified padding. To achieve this, a mask is generated from the image and the RoI. In this mask, pixels within the RoI are assigned a value of 255, while pixels outside the RoI are set to 0. Subsequently, this mask is applied to the original image, adjusting the transparency values of the pixels. Pixels associated with a mask value of zero become completely transparent, ensuring that only the pixels within the RoI remain fully visible. Within Unity, a dedicated script that imports all the DLL functions invokes the RoI cropping function. The current video frame is passed to it as a reference to an array of pixels. The C++ code processes the pixels, and the modified result is then returned to Unity. Since OpenCV supports GPU acceleration to improve the performance of certain algorithms, the RoI cropping function also has a GPU variant. The decision to utilize the GPU was made to reduce the workload of the CPU and to accelerate the marker tracking. To use the GPU support, OpenCV provides a CUDA module specifically designed for NVIDIA GPUs that needs to be enabled when building the OpenCV source files. The cropping function utilizes GPU-accelerated functions implemented in the CUDA module to generate and apply the mask. Before these functions can be used, however, the video frame passed from Unity needs to be uploaded to the GPU and afterward, the final result has to be downloaded from the GPU. These additional steps along with the relative small size of the image might be the reason why a significant speedup was not observed for the cropping function. Unfortunately, as of the time of writing this thesis, no marker detection algorithm was found that leverages the GPU-acceleration of OpenCV. Therefore, it was decided that none of the other functions offered by the DLL would have a GPU variant. A snapshot from the live feed of the camera attached to the Oculus HMD around a detected marker is illustrated on figure 4.7.

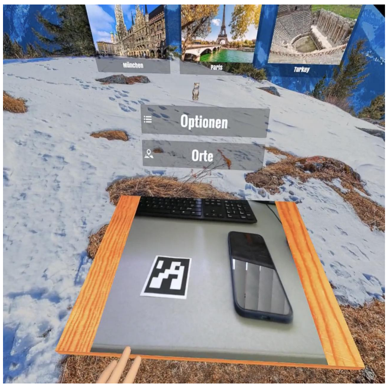


Figure 4.7: Result of the RoI cropping function

4.5.2 Camera Calibration

Achieving accurate tracking requires the use of high-quality cameras. However, cameras nowadays often involve a trade-off between image quality and price. The relatively cheap pinhole camera is a lensless device that has a small opening to pass the light rays, forming an inverted image on the image plane on the opposite side of the camera [The]. On figure 4.10a, the pinhole camera model can be seen. To map a 3D real-world object into the image plane, the calculation of the camera matrix is needed. It is the output of the camera calibration algorithm and it is calculated using the intrinsic and extrinsic camera parameters. The intrinsic parameters describe the internal characteristics of the camera, including the focal length, principal point, and skew coefficient. They define the intrinsic matrix, $\boldsymbol{K}$, where $[c_x,\ c_y]$ is the principal point, $(f_x,\ f_y)$ is the focal length in pixels, and s is the skew coefficient, which accounts for any non-orthogonality between the image axes.

$$\begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

Figure 4.8: Intrinsic matrix, $\boldsymbol{K}$

The extrinsic parameters define the camera's position, $\boldsymbol{t}$, and orientation, $\boldsymbol{R}$, in the 3D world and are represented by the matrix $\begin{bmatrix} R & t \end{bmatrix}$. The resulting camera matrix, $\boldsymbol{P}$, is equal to the product of the intrinsic matrix, $\boldsymbol{K}$, and the extrinsic matrix $\begin{bmatrix} R & t \end{bmatrix}$:

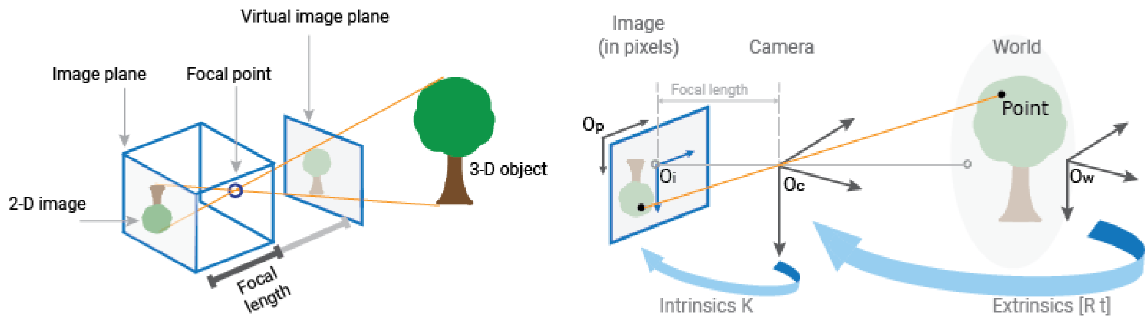
$$\mathrm{P} = \mathrm{K}\begin{bmatrix} R & t \end{bmatrix}$$

Figure 4.9: Camera matrix, $\boldsymbol{P}$

Since the pinhole camera does not have a lens, its camera matrix does not consider lens distortions. However, camera devices with lenses must address radial and tangential distortions in their models. The radial distortion is characterized by a warping effect that makes straight lines appear curved, and the tangential distortion is caused by parallel misalignment between the lens and the image plane [The]. Three coefficients, k_1, k_2, k_3, are used to account for the radial distortion, and two, p_1, p_2, for the tangential distortion correction.

Figure 4.10b illustrates the three different coordinate systems that are involved in the camera calibration process. A point in the real world is represented by a 3D coordinate in the world's coordinate system, denoted as O_w. When this point is multiplied with the extrinsic matrix, it is rigidly transformed into a point in the camera's 3D coordinate system, O_c. Further multiplication with the intrinsic matrix then projects this point from the 3D camera's coordinates into 2D image coordinates. The following equation is derived:

$\mathrm{W}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathrm{P}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$, where W is a scale factor, $\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$ is the image point, and $\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$ is the world



(a) Pinhole camera model (b) Camera calibration coordinate systems

Figure 4.10: Camera calibration, images taken from [The]

point.

GeoTravel employed camera calibration to achieve optimal tracking performance. OpenCV supports camera calibration with different calibration patterns, such as a black-white chessboard and circle boards [Gáb]. The calibration function implemented for GeoTravel utilized a black-white chessboard with dimensions 10x7. A total of eleven images of the chessboard were captured with the external camera from different viewpoints and used for the camera calibration. The process is outlined as follows. The chessboard pattern is detected in each image. Afterward, its corners are located and extracted with a function provided by OpenCV. The next step is establishing correspondences between the 3D world points of the chessboard, which are known, and the detected 2D image points of the corners. Using this correspondences, the **calibrateCamera()** function of OpenCV is called, which computes the camera's intrinsic matrix and distortion coefficients. The reprojection error is also returned to evaluate the quality of the calibration. Additionally, as a visual evaluation of the calibration quality, undistortion of the original images is performed with the computed parameters and the resulting images are displayed. All parameters obtained from the calibration process are stored in a file for further use.

For the tracking algorithm that is explained in section 4.5.3, the intrinsic matrix $\mathbf{K}$ returned from the calibration function is used, with the skew coefficient set to zero. From the returned radial distortion coefficients $\begin{bmatrix} k_1 & k_2 & k_3 \end{bmatrix}$, only k_1, k_2 are used because they are usually sufficient for the calibration [The]. Both tangential distortion coefficients p_1 and p_2 are employed. In the case that camera calibration is not performed and no parameters are provided, default values are used for the tracking. Specifically, c_x and c_y are computed by taking the half of the image width and height, respectively. Both f_x and f_y are set to the maximum of the image width and height, while the distortion coefficients are set to 0.

4.5.3 Aruco Marker Tracking

Aruco markers are a type of fiducial markers created by Garrido-Jurado et al. [Gar+14]. They are black and white square markers with a unique pattern, which are designed to be easily detectable by computer vision algorithms. These markers are often used for camera calibration, object tracking, and mixed reality applications. Their distinct pattern consists of a black border and an internal grid of black-and-white cells, which is used to construct a binary matrix. In this matrix, black cells are represented by 0, while 1 corresponds to white cells. The matrix dimension is determined by the size of the marker, meaning that a 6x6 marker has a black border and an internal 6x6 grid. A valid marker fulfills specific criteria for its binary matrix, which are established by the inclusion of the marker in a marker dictionary. A marker dictionary is a collection of a specific number of markers with the same size [Opea]. According to Fiala [Fia10], a good marker dictionary should address the following aspects:

- **The false positive rate** that corresponds to the false detection of a marker when there is none present.
- **The false negative rate** that calculates how many markers are not detected despite being present.
- **The intermarker confusion rate** that establishes the number of markers that are detected with a wrong id.
- **The dictionary size** that specifies how many markers are included in the dictionary.

Garrido-Jurado et al. [Gar+14] considered these criteria for their marker dictionaries and created an algorithm for an automatic generation of dictionaries with a given size, n, and number of markers, m. The binary matrix of their markers creates a tuple of n words of length n. The first two criteria are addressed by an error detection and correction method. To ensure a low intermarker confusion rate, the n binary words have a high number of bit transitions and their distance to all other markers' words in the dictionary is maintained below a specific threshold. Garrido-Jurado et al. [Gar+14] use the rotation-invariant Hamming distance as their distance function. The n words of a valid marker are then concatenated and an integer value is computed that is used as the marker's id.

The steps Garrido-Jurado et al. [Gar+14] define to detect a marker from a specific dictionary and estimate its pose are the following:

1. **Image segmentation:** First, the original image is converted to a gray-scale image. Then, adaptive thresholding is applied for better segmentation of objects from the background.
2. **Contour extraction and filtering:** Contour extraction is applied to the thresholded image and the identified contours are approximated to polygons. Only polygons with four corners are kept because the markers have a rectangular shape.
3. **Marker code extraction:** The filtered rectangular polygons undergo perspective removal by applying the homography transformation. Afterwards, the image is binarized and a regular grid of 0s and 1s is obtained. Since the markers must have a black border, all the bits of the grid border must be equal to 0. If that is not the case, the potential marker is rejected.

4. **Marker identification and error correction:** All markers with black borders are further analyzed. For each candidate, four ids are extracted from its binary grid for each possible rotation. If any of the ids is contained in the dictionary, the marker is considered as valid.

5. **Pose estimation:** The pose of a detected marker with respect to the camera can be estimated by minimizing the reprojection error of its corners. It computes the difference between the observed 2D image coordinates of the corners and their corresponding 3D coordinates when projected back onto the image plane. To minimize this error the rotation and translation parameters are iteratively adjusted until a desirable alignment is achieved.

OpenCV provides the Aruco module [Opea], which is built upon the ArUco Library developed by Garrido-Jurado et al. [Gar+]. This module includes functions for generating, detecting, and estimating the pose of Aruco markers. To avoid handling big integer values, the ids of the markers in the module correspond to the marker's index within the specified dictionary. For example, one of the predefined dictionaries consists of 250 markers of size 6x6. In this dictionary, the first marker is assigned the id 0, and the last one - the id 249. The detection of aruco markers and their pose estimation follow the steps discussed above.

As mentioned in section 4.5.1, the Aruco module is used in GeoTravel via the OpenCV DLL. The interaction between the DLL and Unity for the marker tracking is illustrated on figure 4.11. The OpenCV DLL contains a marker tracker API, which can be invoked from Unity. This API has a function for marshalling all the required information for the marker tracking provided by Unity, and the result of the marker detection and pose estimation returned by the MarkerTracker. Unity provides the live video feed from the camera attached to the HMD, along with the calculated camera calibration parameters, the marker size in meters, and the ids of single markers and aruco boards. Additionally, parameters for the marker poses and corners are specified, in which the result of the marker tracking is stored and returned to Unity. The **findAllArucoMarkers()** function of the MarkerTracker contains all the logic for detecting aruco markers and estimating their poses. First, the detection parameters are specified, such as the adaptive thresholding value before extracting the contours and the corner refinement method. Then, the marker dictionary is chosen. For GeoTravel, a corner refinement method is selected using subpixel accuracy and the predefined dictionary with 250 markers of size 4x4 is used. From this dictionary, a total of 7 markers are utilized - 2 single markers with ids 1 and 2, and 5 markers forming an aruco board with ids starting from 89 to 93. Aruco boards consist of multiple markers arranged in a desired 2D or 3D layout that provide a single pose for the whole board [Opeb]. Usually, 2D boards with a known relation between the markers are preferred, such as the gridboard [4.12a]. With aruco boards, the pose estimation is more accurate because more corners are used for minimizing the reprojection error. Furthermore, occlusions can be addressed since not all markers have to be detected to compute a pose. A 3D aruco cube is selected for GeoTravel to utilize these benefits. It can be seen on figure 4.12b and its application is further discussed in section 4.5.5. The origin of the aruco cube's coordinate system is in its center. Given the known marker size and the side size of the cube, determining the coordinates of the corners is straightforward. With these coordinates and the ids of the used markers, a board is defined in the **findAllArucoMarkers()** function. The next step is transforming the image provided by Unity, which consists of four channels (red, green, blue, and alpha), into an image without the alpha channel. This format is necessary for compatibility with the implemented **detectMarkers()** function in the Aruco module, which stores the ids of the detected markers and the coordinates of their corners. The retrieved ids from the **detectMarkers()** function are subsequently searched to check if single marker ids or board ids provided by Unity are present. For every single marker id and board id that is detected, the respective pose of the marker or board is computed. For that purpose, the Aruco module provides the **solvePnP()** function. It estimates the marker's pose in the camera's coordinate system by computing the rotation and translation vectors that minimize the reprojection error between the observed 2D image points of the corners and their corresponding projected 3D points onto the image plane. For precise reprojection, the **solvePnP()** function requires the camera's intrinsic matrix and distortion coefficients, which have been precomputed and provided through Unity.

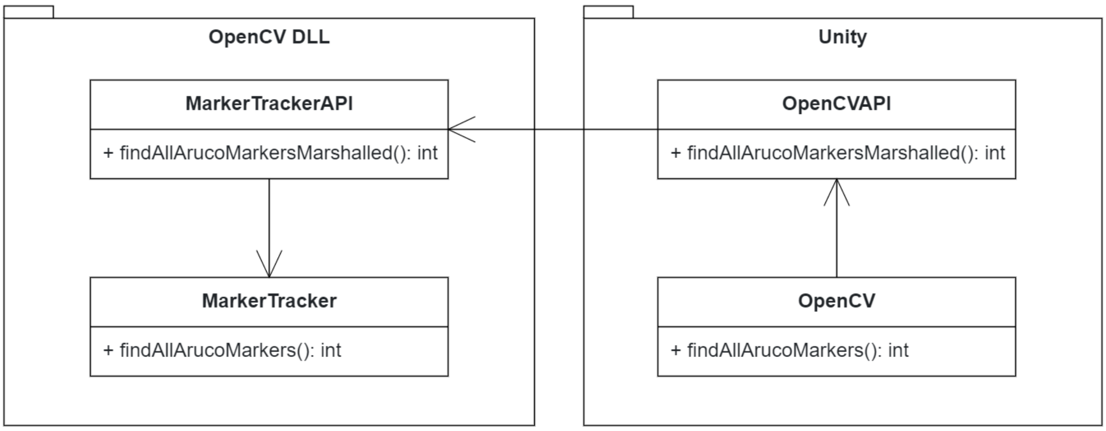
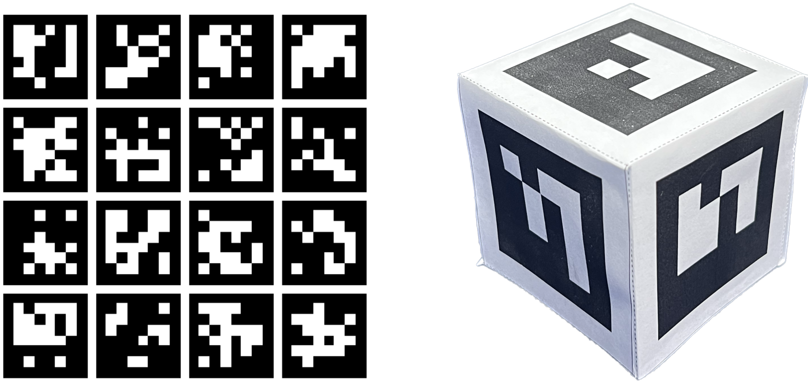


Figure 4.11: UML diagram of the marker tracking

To minimize pose estimation jitter, a Linear Kalman Filter is applied to all estimated marker poses in the **findAllArucoMarkers()** function in the DLL. The Linear Kalman Filter is a recursive algorithm that estimates the state of a discrete-time process from a series of noisy measurements [WB+95] and is included in the Aruco module. There are two key components to consider - the process model and the measurement model. The process model explains how the system's state changes over time while the measurement model describes how the current state of the system is connected to the observed measurements. Both models are linear and include terms to account for noise. For position and orientation tracking, the process state is defined by a 18x1 vector. It contains the tracked object's position (x, y, z), velocity $(\dot{x}, \dot{y}, \dot{z})$, acceleration $(\ddot{x}, \ddot{y}, \ddot{z})$, and its rotation represented by three Euler angles (roll, pitch, jaw) along with their first and second derivatives [Pen]. The process model computes an a priori estimate for the state vector in the next time step, which is then corrected by the measurement model to obtain a posteriori estimate. The a posteriori estimates of the marker



(a) Gridboard (b) Aruco cube

Figure 4.12: Aruco boards

poses are then converted into a 4x4 transformation matrix and marshalled to be accessed from Unity.

In Unity, an OpenCVAPI script acts as the entry point to access the functions provided by the MarkerTrackerAPI. The **findAllArucoMarkersMarshalled()** function is invoked every frame, passing the live video feed and essential marker detection data to the DLL. The returned marker poses, however, require further transformation because OpenCV has a right-handed coordinate system and uses row-major order to store arrays and Unity utilizes a left-handed coordinate system and column-major order. Therefore, the obtained marker pose matrix is firstly transposed to align with Unity's column-major order. Then, the y-axis is flipped by negating the y-component of the translation vector and adjusting the rotation quaternion to switch to a left-handed coordinate system. Lastly, to account for the difference in the up-axis orientation between the MarkerTracker (z-axis as up) and Unity (y-axis as up), the resulting transformation matrix is rotated clockwise around the x-axis by 90 degrees. After these adjustments, the obtained marker poses can be used in GeoTravel. However, it is important to note that these poses are relative to the camera's coordinate system, as it is explained in step 5 of the marker detection process defined by Garrido-Jurado et al. [Gar+14]. The spatial relationship graph depicted in figure 4.13 illustrates the Unity Hierarchy used to account for this fact. In the virtual world of GeoTravel, the player is represented by the virtual avatar, as discussed in section 4.3. Since no teleportation or other navigation methods are supported, the avatar maintains a fixed position in the world, only moving its head and hands. This movement is tracked by the user's HMD. The video feed of the camera attached on top of the HMD is used for the marker tracking, therefore the obtained marker poses are relative to the camera's coordinate system. A virtual camera, representing the real-world camera, is introduced into the hierarchy at the same position as the physical camera. All utilized markers in GeoTravel are enlisted as children to the virtual camera object, and the marker poses obtained from the MarkerTracker are applied to their local transforms. Their further application is explained in section 4.5.5.

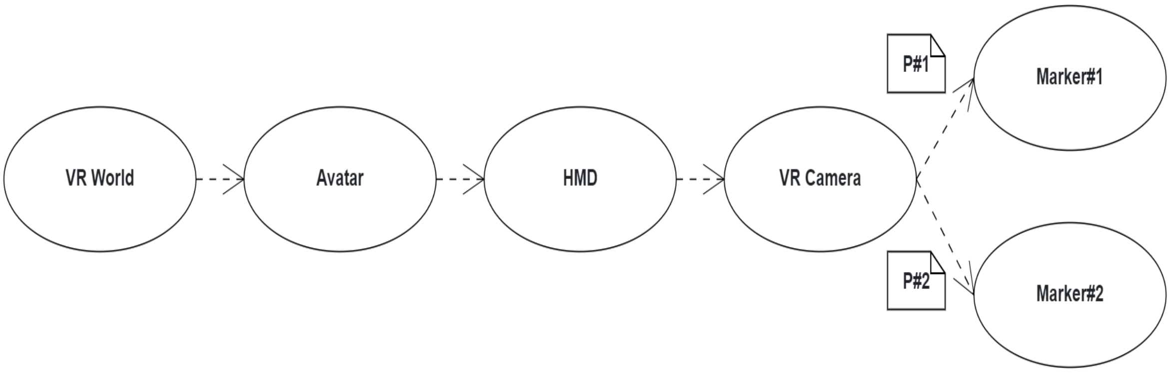


Figure 4.13: Spatial relationship graph

4.5.4 Vuforia Tracking

Aruco markers are designed to be attached to flat surfaces and their detection and accurate pose estimation assume they are planar. Aruco boards can have a 3D layout and be used with non-planar objects. However, it is important to note that while the overall board can be non-planar, the individual markers in the board must remain planar. This constraint presents challenges when attempting to use aruco markers to track objects with complex, non-planar shapes. Moreover, when using an aruco board, it is necessary to manually define the coordinates of all the corners of its 2D markers within the board's coordinate system, which is straightforward for a cube but more challenging for random 3D layouts. As a result, GeoTravel employs a second tracking approach using the Vuforia SDK. The Vuforia Engine supports the development of AR applications for multiple platforms, including iOS, Android, and Unity [PTCa]. It utilizes the latest advancements in computer vision to detect and track different objects in various environments. While initially designed for AR applications, the similarities between AR and AV make it suitable for integrating real-world objects into the VE. Vuforia offers six primary techniques for the detection and tracking of objects: [PTCb]:

- **Model targets:** use 3D digital models to recognize and track the shape of objects.
- **Area targets:** utilize 3D scans of large indoor environments.
- **Image targets:** extract natural features from 2D images and compare them against a known target resource database.
- **Multi targets:** are the equivalent of aruco boards, consisting of image targets.
- **Cylinder targets:** detect and track images with cylindrical or conical shape.
- **VuMarks:** allow for a customizable design of planar markers.

GeoTravel incorporates two image targets and one cylinder target for the integration of its AV objects. It was decided against using 3D digital models or 3D scans because they cannot be reused with multiple objects and their creation is more complex. Designing a valid VuMark also requires following multiple requirements, therefore it was omitted as a potential object tracking target.

4.5.5 Integrated Objects

Following an examination of the two object tracking methods used in GeoTravel, the subsequent section provides a detailed explanation of the AV objects integrated into the application, aligning with guidelines 6 and 12. Three AV objects are currently supported - a water bottle, a camera and a smartphone. Both the aruco marker tracker and the Vuforia tracker can successfully detect and track all of these objects. The real-world physical objects' pose is used in the VE to control the movement of their virtual representations. Only one object at a time can be interacted with to avoid overwhelming the user, following guideline 7. In the multiplayer mode, only the local user can see and manipulate the AV objects.

First, the AV water bottle was integrated into GeoTravel. Previously, the application could track a physical cup, and initially it was planned to simply substitute its tracking approach. However, upon consideration, it was decided to use a water bottle instead of a cup because it offers a safer and more convenient option for virtual interaction compared to cups that may be prone to spills. A simple 3D model of a water bottle was selected and imported into Unity. To match the visual appearance of real-world water bottles, its material was adjusted to be transparent. A liquid shader was created and applied to the model to simulate water. The simulated water follows the natural behavior of water to stay parallel to the ground because of the gravity acting on it. Therefore, if the virtual water bottle is rotated, the liquid inside it remains parallel to the virtual ground. To add more realism to the simulation, the water surface does not simply remain parallel but also wobbles according to the movement. The wobbling effect decreases over time and is the result of a sine wave distortion. The shader can simulate the appearance of any liquid by changing its colour and a light blue was selected to represent water. The volume of the liquid contained in the virtual bottle can be controlled as well to match the volume in the physical bottle. The virtual size of the water bottle is adjusted to match the physical measurements of the tracked real-world bottle. The position and rotation of the virtual bottle are controlled by the estimated pose of the respective tracking approach. The aruco marker tracker uses an aruco cube for the water bottle. The bottom of the cube is removed so that it could be positioned on top of the physical bottle. Inside Unity, an empty game object is created to represent the aruco cube. The virtual water bottle is then placed as its child, offsetting it to match the physical alignment of the cube and the bottle. If at least one of the cube's markers is detected by the marker tracker, the parent object is activated and the estimated pose is applied to it. If the aruco cube is no longer tracked, the parent game object's position and rotation are determined by the hand tracking from the HMD, depending on the user's handedness. If the user is left-handed, the virtual bottle would follow a predefined transform on the left virtual hand; otherwise, it would be on the right virtual hand. The Vuforia tracker detects a cylinder target for the water bottle. Its bottom is also removed so that it could be placed on top of the physical bottle. Similarly to the aruco marker tracker, a parent game object is activated upon identifying the target, and the virtual bottle is set as its child. The parent object adopts the pose of the cylinder target. If the target is lost, the hand tracking information is once again utilized. On figure 4.14, the virtual water bottle can be seen (4.14c), along with the physical setup of the bottle with the aruco cube (4.14a) and the cylinder target (4.14b).

(a) Aruco tracking (b) Vuforia tracking (c) Virtual water bottle

Figure 4.14: AV water bottle

The second integrated AV object was the camera. It was inspired from the work of Karaosmanoglu et al. [Kar+21] on "Memory Journalist VR" and the real-world camera they included, tracked by a Vive tracker. The main difference between the camera in "Memory Journalist VR" and the one in GeoTravel is that, in GeoTravel, the camera is tracked via 2D markers, which do not introduce any additional weight to the physical object, are not system specific and do not require additional investment. A limitation to this approach, however, is that the markers must be visible from the camera attached to the HMD. Furthermore, accurate marker detection and pose estimation depend on factors such as the physical properties of the material used to print the markers on, the camera characteristics, and the current lighting conditions and occlusions. These factors are not applicable to the Vive tracker or other standalone tracking devices. In GeoTravel, a defective Canon camera serves as a physical controller. Although it can no longer capture real-world images, it remains physically intact and ensures natural interaction. To simulate a realistic photo-capturing action in the VE, the physical camera button for taking photos was adapted to recognize presses and send signals to the application. To achieve this, Jonas Weigand connected a microcontroller to the button and positioned it where the camera batteries would be, ensuring it remains hidden from the user. The microcontroller uses Bluetooth Low Energy (BLE) wireless communication to establish a connection with GeoTravel and transmit the detected button presses. The BLE communication consists of two essential parts - the Attribute Protocol (ATT) and Generic Attribute Profile (GATT) [BL]. The ATT functions similarly to a client-server relationship, in which the server stores information and regulates it, and the client accesses it. For GeoTravel, the stored information is the state of the physical camera button, which is accessed by the application. The GATT groups the information in a logical order by creating a hierarchy of services and characteristics. A service contains a set of characteristics that define a specific functionality provided by the server. Each characteristic represents a distinct value crucial to the functionality. Currently, GeoTravel uses only one service with one characteristic - the image capture. The characteristic can have three values - one when the button is not pressed, a second for a gentle button press, and a third for a firm button press. The gentle button press indicates a photo capture, while a hard press is used for focusing the image, which does not play a role in the application since focusing is not required in the VE. The microcontroller of the camera is recognizable by its unique id. Both the service and the characteristic are also identifiable by 128-bit UUIDs. With the device id and the two UUIDs, a BLE connection can be established from a Unity script and the received data from the characteristic can be interpreted. A 3D virtual camera model is integrated into GeoTravel to represent the physical camera. Since no such model could be found that resembles the physical appearance of the real camera, a similar model is used in which the camera button is at the same place as its physical counterpart. The virtual model was scaled to match the measurements of the physical camera. This adjustment ensures that when the users hold the physical camera in their hands, the virtual model does not feel larger or smaller compared to its physical counterpart. The virtual camera also has a display, allowing the users to observe the VE before taking a picture of it. When the physical camera button is lightly pressed, the virtual camera's display briefly fades to white and a stuttering sound is played to create a more natural and immersive experience within the VE. All the pictures captured by users within the VE are saved and can be accessed after the VR session. Currently, these pictures cannot be viewed directly on the virtual camera because that would require detecting additional button presses on the physical camera to maintain a natural mapping of actions. For detecting and tracking the physical camera, complex aruco boards or Vuforia targets were ruled out for several reasons. Firstly, the physical camera is relatively small, making it challenging to attach complex 3D markers. Secondly, real-world interactions with cameras primarily involve observing their display and pressing the respective buttons, which does not necessitate turning the camera around. Therefore, simple 2D markers attached on top of the camera are utilized for both tracking approaches implemented in GeoTravel. They are positioned slightly in front of the camera, allowing the users to comfortably hold the device and have access to its buttons. The physical arrangement of the markers and the virtual camera model can be seen on figure 4.15. Handling the detection and loss of a marker is analogous to the process described for the AV water bottle.

Lastly, the AV smartphone was integrated. Similarly to the AV camera, a single 2D aruco marker and an image target are used for the two tracking approaches, respectively. They are positioned on top of the smartphone to avoid any contact with its display, ensuring no interference with the user interaction. In the VE, a 3D model of a smartphone is incorporated, aligning with the dimensions of the physical smartphone used as a controller. The AV setup of the smartphone is depicted on figure 4.16. In GeoTravel, the AV camera is an essential component of the virtual travel experience. Its incorporation is based on the familiarity and widespread use of cameras among older adults. With the advancement of technology, however, smartphones have become an irreplaceable part of the travel experience, assuming the role of cameras for capturing and storing photos. Furthermore, researchers have observed the increasing influence of smartphones on seniors' daily lives and their feasibility even for those with cognitive and emotional difficulties [Bus+21; Ram+16]. Therefore, the AV smartphone in GeoTravel was designed to align with the latest advancements and provide users with the opportunity to capture and browse through their pictures. Unlike the AV camera, which connects to the application via BLE, the AV smartphone uses a MR framework called Ubi-Interact [Web+21]. Ubi-Interact is platform-independent and can be imported into Unity as a package and also used with Android and iOS devices. It allows for a WiFi connection to be established between the PC on which GeoTravel runs and the physical smartphone. On the Unity side, the connection is enabled once the online scene is loaded. For the physical smartphone, a separate application was developed. It starts with a connection panel where the user can input the IP address of the VR PC. After it successfully connects to the VR PC, it starts transmitting events, which are invoked when the user touches one of the application's buttons. In the VE of GeoTravel, the virtual smartphone is activated when its currently attached marker is detected. Since the virtual model's size is adjusted to the size of the physical smartphone and the offset between the smartphone and the marker is known, the user can see their virtual avatar holding the virtual smartphone in their hands. Its screen displays a background image, two icons, and the current time. The same layout of icons is displayed on the mobile app after the IP connection is established. Each icon contains a button, which is slightly larger than the icon itself to ensure that when the virtual avatar of the user presses the virtual icon of the smartphone, their real hands are aligned with the right mobile button and the respective event is triggered. If the estimated pose of the marker is imprecise or the marker is lost, misalignments could occur, disrupting the VR experience. The smartphone offers two functionalities. Pressing the camera icon opens the camera app. It loads a display where the VE can be observed, and pictures of it can be taken. A button below the display is responsible for taking an image. When pressed, the display fades to white similarly to the AV camera, accompanied by a distinct stuttering sound. Once the user has taken all the pictures they desire, they can close the camera by pressing the X button in the top right corner. All captured images can be viewed in the gallery app, accessible from the second icon. The gallery app opens by displaying the first image taken. The users can navigate between the images by pressing the next and previous buttons. The total number of images, along with the position of the currently displayed image, is shown at the bottom. The gallery can also be closed from the X button in the top. The captured images are only stored on the VR PC and cannot be viewed from the mobile application. However, the storage of images is not limited to those captured during the current VR session. The users can access images taken during their previous VR sessions as well. The choice to exclude a lock screen was made to limit the steps to access the two functionalities and provide a simple interaction experience for seniors. The camera and gallery apps can be seen on figure 4.17, along with the home screen.

(a) Aruco tracking

(b) Vuforia tracking

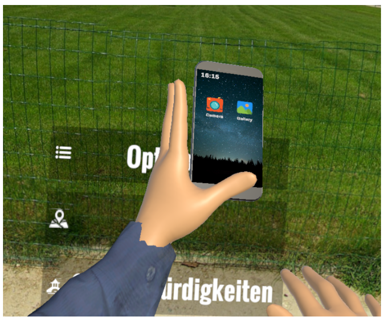
(c) Virtual camera

Figure 4.15: AV camera

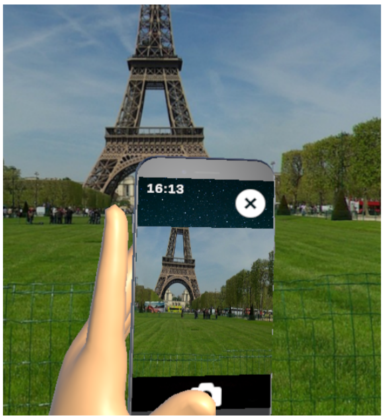
(a) Aruco tracking

(b) Vuforia tracking

Figure 4.16: AV smartphone



(a) Home screen



(b) Camera app

(c) Gallery app

Figure 4.17: AV smartphone apps

5 Evaluation

To evaluate the newly integrated aspects in GeoTravel and answer the research questions that are defined within this thesis, it is essential to conduct a user study. The application targets older people and was developed taking VR design guidelines for seniors into consideration. However, organising a user study with older adults at the time of writing this work was not possible due to an increase in COVID-19 and other viral cases. Therefore, the usability of GeoTravel was tested with students from the Technical University of Munich, asking them to evaluate the application from the perspective of seniors.

5.1 Setup

For the user study, it was decided to employ the Vuforia tracking approach to track the AV objects. In addition to the reasons for integrating Vuforia mentioned in section 4.5.4, two more could be outlined. An aruco marker relies on the detection of four points for tracking, namely its corners. Despite enabling the corner refinement method, there were instances where the corners of the aruco markers were not successfully detected, leading to occasional jitter. Aruco boards rely on more than four points for tracking, however, with fewer markers detected the estimated pose is more imprecise. Vuforia, on the other hand, extracts multiple points of interest from its targets that could be used for estimating the pose. Additionally, a drop in the application's performance was observed when enabling the aruco marker tracker. Since the camera video feed is passed from Unity to the OpenCV DLL and it is a known limitation of Unity that it only supports 30 FPS for webcam textures, a refactoring of the code is needed to enable higher frame rate. Such a refactoring was not yet implemented before conducting the user study. Therefore, to ensure optimal tracking of the AV objects during the user study, the Vuforia targets depicted in section 4.5.5 were used. They are the default targets offered by Vuforia for best detection. The original size of the image targets is 7cm by 12cm, while the cylinder target has a diameter of 6cm and a length of 12cm. Although these dimensions might seem large for the camera and smartphone, they were retained to create optimal tracking conditions. As a physical phone, Nokia 6.1 was employed to run the Android smartphone application. Each AV object had a completely virtual counterpart that used the hand tracking data from the HMD to stay in one of the virtual avatar's hands, based on the user's handedness. Participants could capture photos or open applications on the virtual camera and smartphone by pressing the corresponding virtual buttons using one of their avatar's index fingers. Participants were informed about possible motion sickness and the option to stop the study at any time. They then provided consent for the user study and the analysis of the collected data. Initially, the users were prompted to freely explore the VE of GeoTravel, visiting all the locations they wanted. Once they had explored to their satisfaction, they were introduced to the AV objects. Each participant received either one of the virtual objects or its AV counterpart. The order of interaction — whether they started with the purely virtual object or its AV counterpart — was randomized, as was the order of the objects. The goal of the user study was to answer the three research questions (1).

5.2 Study Outcome

A total of 17 students participated in the user study, of which 9 were female and 8 were male. Their age ranged from 19 to 33 years old, with the average age being 23.88 years old. They were asked about their previous VR experience: 17.6% never used VR before, 35.3% used it once, 41.2% interact with VR occasionally and 5.9% use VR regularly. Their app usage on smartphones was also evaluated. 17.6% of the participants reported using their smartphones for basic functionalities, such as checking the email or the calendar. 82.4% use additional apps, such as health-apps or games. Two standardized questionnaires were utilized to evaluate GeoTravel: IPQ [Sch03] and SUS [Bro95]. IPQ is used to measure the general sense of presence (PRES), the spatial presence (SP) - the participant's sense of being physically present in the VE, the involvement (INV) - the experienced level of engagement with the VE, and the perceived realism of the VE (REAL) [Mel+23]. The SUS score assesses the usability of the system. The IPQ results can be seen in table 5.1. The four subareas can receive scores ranging from 0 to 6, and a higher score indicates a more positive rating. The scores for PRES and SP were between 4.1 and 4.5, while lower scores for INV and REAL were observed. The lower score for INV could be explained by the given answers to the question "I still paid attention to the real environment". The main focus of this work is the integration and interaction with AV objects. Therefore, during the user study, the participants were asked to hold and manipulate real physical objects, maintaining a tangible connection with the real-world. The average results to the questions "How aware were you of the real world surroundings while navigating in the virtual world? (i.e. sounds, room temperature, etc.)?" and "I was not aware of my real environment. (i.e. sounds, room temperature, etc.)" were 3.59 and 3.47, respectively. Since most of the locations in the VE of GeoTravel are static 360-degree images without any accompanying audio, hearing any sounds form the real-world is unavoidable. Despite that, the virtual world was rated as captivating ($\mathbf{M} = 4.35, \sigma = 1.50$). In the area REAL, the participants gave very low scores to the question "The virtual world seemed more realistic than the real world" ($\mathbf{M} = 1.06, \sigma = 0.83$), which was to be expected since "imagining a more real world than our own is impossible", as one of the participants argued. Furthermore, both questions about how real the virtual world seemed received lower scores, which could be explained by two main reasons. Firstly, the static nature of the 360-degree images cannot simulate the dynamics of the real-world. Secondly, the virtual avatar cannot move freely within the VE, which was what most participants noted as a limitation. The general sense of presence (PRES) received positive ratings, with 88.2% of the participants assigning a score of 4 or higher. The main question that negatively influenced the SP area was "I felt like I was just perceiving pictures" ($\mathbf{M} = 2.71, \sigma = 1.49$). Despite incorporating three 360-degree videos into

GeoTravel, the majority of the locations still utilize static images. All of the participants agreed that the videos with synchronized audio were superior, indicating a potential improvement for the SP area. The SUS score for GeoTravel was **68.24**, which is interpreted as a "Good" usability score, however close to the border with the previous grade. Since the participants were asked to evaluate the usability of the application from the perspective of an older relative, the majority (65%) believed that their older relatives would need the help of a technical person to use the system. When asked about whether they would frequently want to use GeoTravel together with their older relatives, the results were average ($\mathbf{M} = 3.71, \sigma = 0.99$). To assess how the AV objects were perceived, some questions about their integration and interaction were asked with a Likert scale of 1-5 (strongly disagree-strongly agree). The AV bottle was perceived the least as a part of the virtual experience ($\mathbf{M} = 3.71, \sigma = 1.26$), followed by the smartphone ($\mathbf{M} = 4.00, \sigma = 1.17$), and the camera ($\mathbf{M} = 4.41, \sigma = 0.71$). Regarding how natural the interaction with the AV objects was, the smartphone offered the least natural interaction ($\mathbf{M} = 3.18, \sigma = 1.22$), followed by the bottle ($\mathbf{M} = 3.82, \sigma = 1.13$), and the camera ($\mathbf{M} = 4.11, \sigma = 0.98$). However, interacting with the AV objects was evaluated as more natural than with the purely virtual objects ($\mathbf{M} = 4.35, \sigma = 1.06$). When asked whether they could imagine their older relatives using the AV water bottle as a reminder to drink more water in the real-world, 76.4% of the participants gave a score of 3 or lower. The participants believed that the AV camera could be motivating for their older relatives to use GeoTravel more frequently ($\mathbf{M} = 3.76, \sigma = 1.16$). The smartphone scored similar results ($\mathbf{M} = 3.35, \sigma = 1.11$).

PRES	**SP**	**INV**	**REAL**
4.41	4.14	3.40	2.56

Table 5.1: IPQ results

6 Discussion

The results from the user study revealed important insights regarding the research questions of this thesis (1), despite not being conducted with the target group. All of the participants unanimously acknowledged that the AV objects offered a more natural interaction compared to the purely virtual objects, effectively confirming the second half of **RQ1**. The first part of **RQ1** received varying results for the three AV objects. The AV camera offered the most natural interaction, with a well-functioning button press detection and relatively stable tracking. A common wish among the participants was zooming before taking a picture and being able to see the captured images. Since the microcontroller currently tracks only one button on the physical camera, such actions are not supported. However, integrating them could further enhance the intuitiveness of the camera. The interaction with the AV water bottle was less natural to the participants. During the user study it was noted that the majority of them tried to open the bottle and drink from it. Such manipulation is not supported for two primary reasons. Firstly, both the aruco cube and the Vuforia cylinder target that are used for detection and tracking are placed on top of the physical bottle, hiding its cap. This choice is deliberate, as using a board or a target covering only the bottle's side without its top would restrict tracking from certain angles. Moreover, once the bottle is moved closer to the mouth, it is no longer visible to the external camera, which prevents its tracking. Secondly, attempting to open the physical bottle could result in spilling its liquid, especially in cases where the tracking is not perfect, posing a significant safety concern. Additionally, it is not possible to estimate the amount of water that could be consumed from the physical bottle, resulting in a potential mismatch between the simulated virtual water and its real-world counterpart. It is the author's belief that these limitations in the interaction, along with the noticeably more frequent tracking jitter compared to the camera, were responsible for the observed results. The AV smartphone obtained the lowest score for natural interaction. It was the hardest object to detect and track despite using one of Vuforia's default image targets. The physical smartphone also occluded a large part of the participants' hands, which resulted in poor hand tracking from the HMD. To minimize this effect, the participants were prompted to hold the smartphone in a way that was perceived as uncomfortable. Moreover, 82.4% of them indicated that they utilize their smartphones for advanced applications, which provide greater functionality and involve more complex interactions. The two implemented apps for the AV smartphone only support basic manipulation, potentially limiting the experience for advanced users. These observations could explain why interacting with the smartphone was not evaluated as very natural. **RQ2** also received mixed reviews. Rating the AV bottle as not being a significant part of the virtual experience could be explained by the limited interaction it offers. Additionally, it was not believed that it could motivate seniors to stay hydrated in the real-world ($\mathbf{M} = 2.82, \sigma = 1.01$). It was suggested by two of the participants to add textual instructions on the bottle or convey them through the virtual dog as an extra reminder to drink water. Both the camera and smartphone were viewed as potentially more motivating for older adults to engage with GeoTravel, even with the smartphone's interaction being somewhat limited in intuitiveness (camera: $\mathbf{M} = 3.71, \sigma = 1.16$, smartphone: $\mathbf{M} = 3.35, \sigma = 1.11$). When asked about their opinion, some of the participants believed that the camera would be more familiar and user-friendly for their older relatives. On the other hand, others thought that the additional functionality of the smartphone, namely the image gallery, would be more intriguing and enhance motivation. **RQ3** was unanimously confirmed by the participants. All of them found the 360-degree videos with synchronised audio more immersive and superior compared to the static images. Some of them suggested adding audio to the static locations as a potential improvement. One participant expressed their concern that a highly dynamic video with different sounds could overwhelm seniors. However, they still expressed a preference for videos. The SUS score of **68.24** was mainly negatively affected by the belief that GeoTravel might not be particularly user-friendly for older adults. Some comments noted during the user study mentioned that the participants' older relatives had never used a HMD, leading to uncertainty about its reception. Some expressed concerns that the HMD might be too heavy and uncomfortable for seniors, particularly with the mounted external camera on the top. One participant even thought that both the controller and the physical water bottle would be too heavy for a senior. The multiplayer option was not enabled during the user study because there was no available hardware to support it. Furthermore, GeoTravel was not designed for a younger, technology-oriented audience, offering only basic interaction and functionalities. These observations might explain why the prospect of using the application frequently with their older relatives was not particularly enticing for the participants.

7 Future Work

Some of the feedback given by the participants in the user study highlighted important aspects for GeoTravel that could be addressed in the future. According to some participants, the vibration of the controller activated when the walking stick touched one of the mini models in the 3D interaction system was insufficient. It was suggested to add a visual indicator, for example a red highlight around the selected model. A similar recommendation was made for the navigation menu. Highlighting the current location in green was suggested to assist seniors in better orienting themselves in the VE. A textual instruction attached to the AV water bottle could better illustrate its purpose to remind older adults to stay hydrated. Additionally, the virtual dog could serve as a visual reminder by barking at the bottle. The zooming option requested for the AV camera could certainly be extended to the smartphone as well. The microcontroller of the physical camera could be adjusted to also track the zoom button. For the smartphone, a zooming gesture could be enabled, however, its intuitiveness would need to be tested with seniors. As an alternative, a simple button could be integrated that would allow zooming in and out with a fixed amount. Some participants attempted to take a selfie with the smartphone, which hid the smartphone's marker from the external camera's view. Even though taking selfies might not be particularly important to seniors, it could motivate their grandchildren to use GeoTravel with them. The aruco marker would need to be replaced with an aruco board and the Vuforia image target with a multi target. A gallery for the AV camera was also proposed, which could be integrated by adjusting the microcontroller. The smartphone's gallery could benefit from zooming the pictures. Two participants expressed a sensation of floating in the virtual environment, which was the result of how some of the 360-degree images/videos were captured and the nature of Unity's Skyboxes. Finding high quality 360-degree images/videos can be challenging, and, consequently, some locations indeed create a floating feeling or make the virtual avatar appear larger than the depicted people. When the ground in the image/video is relatively flat, it could be simulated by UV mapping the texture onto a plane [arf]. For a better solution, additional research into this issue is required. The appearance of the virtual avatar was only commented by one of the participants, who found it not realistic enough. Additionally, they suggested a female avatar since currently only a generic male avatar is incorporated. In terms of realism, chapter 2 explored some existing solutions. Enabling communication between two players over the network could also contribute to a better social interaction within the application. To overcome Unity's limitation with webcam textures, the existing communication between Unity and the OpenCV DLL could be refactored. Opening the video from the DLL and transferring it to Unity could potentially result in higher frame rates and more stable marker tracking. To improve the interaction with the AV smartphone, the hybrid tracking solution developed by Klinker et al. [Eic+] could be used.

8 Conclusion

The Virtual Reality travel application for older adults, GeoTravel, was improved with more immersive virtual environments created with 360-degree videos with synchronized audio, and a dynamic virtual avatar. Two tracking approaches and three AV objects were integrated to support natural interaction and create a rich sensory experience. Even though a user study with the target group was not possible, the evaluation with students provided valuable insights. The 360-degree videos were preferred over the static 360-degree images. The AV objects offered more intuitive interaction than their completely virtual counterparts. The AV camera was the best perceived AV object and found to be suitable for seniors. Despite some tracking difficulties and non-intuitve interaction, GeoTravel received a Good" (SUS) usability score of **68.24**, however, a future evaluation with the target group is necessary.

List of Figures

2.1 The three I's of VR, taken from [BC03] 10
2.2 User Interface in VR, adapted from [SC03] 11
2.3 Milgram's mixed reality continuum, adapted from [Mil+95] 15
2.4 Left: Real life environment. Right: Virtual Reality Environment. Taken from [ZRB21] 17
2.5 AV-based prototype for social eating among solitary older adults, taken from [KBN19] 19
2.6 AV-based videoconferencing system, taken from [Reg+04] 20
2.7 VR field trip, taken from [BLW18] 22
2.8 VE for early detection of navigational impairment in AD and MCI, taken from [CSD08] 30
2.9 VR Building spatial navigation exercise, taken from [WM16] 31
2.10 The hardware setup (a), the VE and virtual camera (b), taken from [Kar+21] . 33
2.11 Memory Journalist VR, taken from [Kar+21] 34

3.1 Virtual button on walking stick, taken from [Sch21] 38
3.2 Navigation methods 39
3.3 Images taken from [Sch21] 40

4.1 The offline scene 42
4.2 Recently visited places 43
4.3 The Virtual Avatar 45
4.4 London in VR 47
4.5 Paris in VR 47
4.6 Turkey in VR 49
4.7 Result of the RoI cropping function 50
4.8 Intrinsic matrix, $\boldsymbol{K}$ 51
4.9 Camera matrix, $\boldsymbol{P}$ 51
4.10 Camera calibration, images taken from [The] 52
4.11 UML diagram of the marker tracking 55
4.12 Aruco boards 56
4.13 Spatial relationship graph 57
4.14 AV water bottle 59
4.15 AV camera 62
4.16 AV smartphone 62
4.17 AV smartphone apps 62

List of Tables

2.1 AV Applications Summary 26
2.2 VR Research Summary 34

5.1 IPQ results 65

Bibliography

[RR] H. Ritchie and M. Roser. *Age Structure*. URL: https://ourworldindata.org/age-structure (visited on 11/05/2023).

[Wai04] L. Waite. "The Demographic Faces of the Elderly". In: *Population and development review* 30 (Jan. 2004), pp. 3–16.

[Qia+22] G. Qiao, L. Ding, K. Xiang, B. Prideaux, and J. Xu. "Understanding the Value of Tourism to Seniors' Health and Positive Aging". In: *International Journal of Environmental Research and Public Health* 19.3 (2022), p. 1476. ISSN: 1660-4601. DOI: 10.3390/ijerph19031476.

[AJ11] Y.-j. Ahn and M. Janke. "Motivations and Benefits of the Travel Experiences of Older Adults". In: *Educational Gerontology - EDUC GERONTOL* 37 (Aug. 2011), pp. 653–673. DOI: 10.1080/03601271003716010.

[PS15] I. S. Patricia L. Mokhtarian and M. E. Singer. "What Moves Us? An Interdisciplinary Exploration of Reasons for Traveling". In: *Transport Reviews* 35.3 (2015), pp. 250–274. DOI: 10.1080/01441647.2015.1013076. eprint: https://doi.org/10.1080/01441647.2015.1013076. URL: https://doi.org/10.1080/01441647.2015.1013076.

[CW09] E. Y. Cornwell and L. J. Waite. "Social Disconnectedness, Perceived Isolation, and Health among Older Adults". In: *Journal of Health and Social Behavior* 50.1 (2009), pp. 31–48. ISSN: 0022-1465. DOI: 10.1177/002214650905000103. URL: https://dx.doi.org/10.1177/002214650905000103.

[Mil+95] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. "Augmented reality: a class of displays on the reality-virtuality continuum". In: *Telemanipulator and Telepresence Technologies*. Ed. by H. Das. Vol. 2351. International Society for Optics and Photonics. SPIE, 1995, pp. 282–292. DOI: 10.1117/12.197321. URL: https://doi.org/10.1117/12.197321.

[BC04] E. Brown and P. Cairns. "A grounded investigation of game immersion". In: 2004. DOI: 10.1145/985921.986048.

[EM05] L. Ermi and F. Mäyrä. "Fundamental Components of the Gameplay Experience: Analysing Immersion." In: Jan. 2005.

[Ste92] J. Steuer. "Defining Virtual Reality: Dimensions Determining Telepresence". In: *Journal of Communication* 42.4 (1992), pp. 73–93. ISSN: 0021-9916. DOI: 10.1111/j.1460-2466.1992.tb00812.x.

[AF17] Z. Akhtar and T. Falk. "Audio-Visual Multimedia Quality Assessment: A Comprehensive Survey". In: *IEEE Access* PP (Sept. 2017), pp. 1–1. DOI: 10.1109/ACCESS.2017.2750918.

[Csi14] M. Csikszentmihalyi. *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*. Apr. 2014, pp. 1–298. ISBN: 978-94-017-9087-1. DOI: 10.1007/978-94-017-9088-8.

[Bas+17] A. S. Bastos, R. F. Gomes, C. C. Dos Santos, and J. G. Rodrigues Maia. "Assessing the Experience of Immersion in Electronic Games". In: 2017. DOI: 10.1109/svr.2017.27.

[HC06] N. Haywood and P. Cairns. "Engagement with an Interactive Museum Exhibit". In: Jan. 2006, pp. 113–129. ISBN: 978-1-84628-192-1. DOI: 10.1007/1-84628-249-7_8.

[Ban15] J. Banks. "Object, Me, Symbiote, Other: A social typology of player-avatar relationships". In: *First Monday* 20.2 (Feb. 2015). DOI: 10.5210/fm.v20i2.5433. URL: https://firstmonday.org/ojs/index.php/fm/article/view/5433.

[Ban+19] J. Banks, N. D. Bowman, J.-H. T. Lin, D. Pietschmann, and J. A. Wasserman. "The common player-avatar interaction scale (cPAX): Expansion and cross-language validation". In: *International Journal of Human-Computer Studies* 129 (2019), pp. 64–73. ISSN: 1071-5819. DOI: https://doi.org/10.1016/j.ijhcs.2019.03.003. URL: https://www.sciencedirect.com/science/article/pii/S1071581918304610.

[HIT15] S. Hillier, M. Immink, and D. Thewlis. "Assessing Proprioception: A Systematic Review of Possibilities". In: *Neurorehabilitation and Neural Repair* 29.10 (2015). PMID: 25712470, pp. 933–949. DOI: 10.1177/1545968315573055. eprint: https://doi.org/10.1177/1545968315573055. URL: https://doi.org/10.1177/1545968315573055.

[Shi08] E. Shinkle. "Video games, emotion and the six senses". In: *Media, Culture & Society* 30.6 (2008), pp. 907–915. DOI: 10.1177/0163443708096810. eprint: https://doi.org/10.1177/0163443708096810. URL: https://doi.org/10.1177/0163443708096810.

[Sla03] M. Slater. "A Note on Presence Terminology". In: *Presence Connect* 3 (Jan. 2003).

[BM07] D. A. Bowman and R. P. Mcmahan. "Virtual Reality: How Much Immersion Is Enough?" In: *Computer* 40.7 (2007), pp. 36–43. ISSN: 0018-9162. DOI: 10.1109/mc.2007.257.

[Mcm03] A. Mcmahan. "Immersion, engagement, and presence: A method for analyzing 3-D video games". In: *The Video Game Theory Reader* (Jan. 2003), pp. 67–86.

[DS00] N. Durlach and M. Slater. "Presence in Shared Virtual Environments and Virtual Togetherness". In: *Presence: Teleoperators and Virtual Environments* 9.2 (Apr. 2000), pp. 214–217. DOI: 10.1162/105474600566736. eprint: https://direct.mit.edu/pvar/article-pdf/9/2/214/1623414/105474600566736.pdf. URL: https://doi.org/10.1162/105474600566736.

[BHB03] F. Biocca, C. Harms, and J. K. Burgoon. "Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria". In: *Presence: Teleoperators and Virtual Environments* 12.5 (Oct. 2003), pp. 456–480. DOI: 10.1162/105474603322761270. eprint: https://direct.mit.edu/pvar/article-pdf/12/5/456/1623957/105474603322761270.pdf. URL: https://doi.org/10.1162/105474603322761270.

[SW97] 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.

[Sto95] A. Stone. *The War of Desire and Technology at the Close of the Mechanical Age*. MIT Press, 1995. ISBN: 9780262193627. URL: https://books.google.li/books?id=JEawQgAACAAJ.

[BG99] J. Bolter and R. Grusin. *Remediation: Understanding New Media*. MIT Press, 1999. ISBN: 9780262024525. URL: https://books.google.de/books?id=AQvsnQEACAAJ.

[Pro+95] J. D. Prothero, H. G. Hoffman, D. E. Parker, I. Thomas A. Furness, and M. J. Wells. "Foreground/Background Manipulations Affect Presence". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 39.21 (1995), pp. 1410–1414. DOI: 10.1177/154193129503902111. eprint: https://doi.org/10.1177/154193129503902111. URL: https://doi.org/10.1177/154193129503902111.

[Fen99] C. Fencott. "Presence and the content of Virtual Environments". In: *2nd International Workshop on Presence* (1999).

[Kru91a] M. Krueger. *Artificial Reality II*. Polar Research. Addison-Wesley, 1991. ISBN: 9780201522600. URL: https://books.google.de/books?id=gO4YAQAAIAAJ.

[Coa92] G. Coates. "Program from Invisible Site—a virtual sho, a multimedia performance work presented by George Coates". In: Performance Works, San Francisco, CA, 1992.

[Gre92] P. Greenbaum. "The lawnmower man. Film and video". In: 9 (3), pp. 58-62., 1992.

[Kru91b] M. W. Krueger. "Artificial reality (2nd ed.)" In: 9 (3), pp. 58-62.: Reading, MA: Addison-Wesley., 1991.

[Gib14] J. J. Gibson. "The Ecological Approach to Visual Perception". In: (2014). DOI: 10.4324/9781315740218.

[Bro+10] M. Brown, A. Kehoe, J. Kirakowski, and I. Pitt. "Beyond the Gamepad: HCI and Game Controller Design and Evaluation". In: Dec. 2010, pp. 197–219. ISBN: 978-3-319-15984-3. DOI: 10.1007/978-1-84882-963-3_12.

[PV18] J. Perret and E. Vander Poorten. "Touching Virtual Reality: a Review of Haptic Gloves". In: June 2018.

[FIO21] C. Flavián, S. Ibáñez-Sánchez, and C. Orús. "The influence of scent on virtual reality experiences: The role of aroma-content congruence". In: *Journal of Business Research* 123 (2021), pp. 289–301. ISSN: 0148-2963. DOI: https://doi.org/10.1016/j.jbusres.2020.09.036. URL: https://www.sciencedirect.com/science/article/pii/S0148296320306093.

[Aba+14] A. Abaza, M. A. Harrison, T. Bourlai, and A. Ross. "Design and evaluation of photometric image quality measures for effective face recognition". In: *IET Biometrics* 3.4 (2014), pp. 314–324. DOI: https://doi.org/10.1049/iet-bmt.2014.0022. eprint: https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-bmt.2014.0022. URL: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-bmt.2014.0022.

[Man86] J. M. Mandler. "On the comprehension of temporal order". In: *Language and Cognitive Processes* 1.4 (1986), pp. 309–320. ISSN: 0169-0965. DOI: 10.1080/01690968608404680.

[Pis] E. Piskunov. *2D vs 3D Games: Can You Tell The Difference?* URL: https://ilogos.biz/2d-vs-3d-games-5-biggest-differences-between-2d-and-3d-games/ (visited on 09/04/2023).

[Bom] G. Bomb. *options Menu*. URL: https://www.giantbomb.com/options-menu/3015-6644/ (visited on 09/04/2023).

[FH98] K. Fitz and B. Holloway. "Lemur - A Tool for Timbre Manipulation". In: (Nov. 1998).

[Din] D. Dinesh. *Is Frequency Of Light Constant: When, Why, How And Detailed Facts*. URL: https://lambdageeks.com/is-frequency-of-light-constant/?utm_content=cmp-true (visited on 07/10/2023).

[Ska+11] P. Skalski, R. Tamborini, A. Shelton, M. Buncher, and P. Lindmark. "Mapping the road to fun: Natural video game controllers, presence, and game enjoyment". In: *New Media & Society* 13 (Mar. 2011), pp. 224–242. DOI: 10.1177/1461444810370949.

[BC03] G. Burdea and P. Coiffet. *Virtual Reality Technology*. IEEE Press. Wiley, 2003. ISBN: 9780471360896. URL: https://books.google.de/books?id=0xWgPZbcz4AC.

[SC03] W. Sherman and A. Craig. *Understanding Virtual Reality: Interface, Application, and Design*. Morgan Kaufmann series in computer graphics and geometric modeling. Elsevier Science, 2003. ISBN: 9781558603530. URL: https://books.google.de/books?id=7ssd%5C_st8kjMC.

[Gir18] C. Girvan. "What is a virtual world? Definition and classification". In: *Educational Technology Research and Development* 66 (Feb. 2018). DOI: 10.1007/s11423-018-9577-y.

[CKK19] V. Clay, P. König, and S. U. König. "Eye tracking in virtual reality". In: *Journal of Eye Movement Research* 12.1 (2019). ISSN: 1995-8692. DOI: 10.16910/jemr.12.1.3.

[Cas+19] P. Caserman, A. Garcia-Agundez, R. Konrad, S. Göbel, and R. Steinmetz. "Real-time body tracking in virtual reality using a Vive tracker". In: *Virtual Reality* 23.2 (2019), pp. 155–168. ISSN: 1359-4338. DOI: 10.1007/s10055-018-0374-z.

[Ben+01] S. Benford, C. Greenhalgh, T. Rodden, and J. Pycock. "Collaborative Virtual Environments". In: *Commun. ACM* 44.7 (July 2001), pp. 79–85. ISSN: 0001-0782. DOI: 10.1145/379300.379322. URL: https://doi.org/10.1145/379300.379322.

[CS98] E. F. Churchill and D. Snowdon. "Collaborative virtual environments: An introductory review of issues and systems". In: *Virtual Reality* 3.1 (1998), pp. 3–15. ISSN: 1359-4338. DOI: 10.1007/bf01409793.

[Hag96] O. Hagsand. "Interactive multiuser VEs in the DIVE system". In: *Multimedia, IEEE* 3 (Feb. 1996), pp. 30–39. DOI: 10.1109/93.486702.

[Sno+96] D. Snowdon, C. Greenhalgh, S. Benford, A. Bullock, and C. Brown. "A review of distributed architectures for networked virtual reality". In: *Virtual Reality* 2.1 (1996), pp. 155–175. ISSN: 1359-4338. DOI: 10.1007/bf02534449.

[MAB92] K. Meyer, H. L. Applewhite, and F. A. Biocca. "A Survey of Position Trackers". In: *Presence: Teleoperators and Virtual Environments* 1.2 (May 1992), pp. 173–200. DOI: 10.1162/pres.1992.1.2.173. eprint: https://direct.mit.edu/pvar/article-pdf/1/2/173/1622423/pres.1992.1.2.173.pdf. URL: https://doi.org/10.1162/pres.1992.1.2.173.

[Ama17] A. Amamra. "Smooth head tracking for virtual reality applications". In: *Signal, Image and Video Processing* 11.3 (2017), pp. 479–486. ISSN: 1863-1703. DOI: 10.1007/s11760-016-0984-4.

[Mar+20] E. Markopoulos, P. Markopoulos, N. Laivuori, C. Moridis, and M. Luimula. "Finger tracking and hand recognition technologies in virtual reality maritime safety training applications". In: *2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. 2020, pp. 000251–000258. DOI: 10.1109/CogInfoCom50765.2020.9237915.

[Son] Sony. *PlayStation Move motion controller*. URL: https://www.playstation.com/en-bg/accessories/playstation-move-motion-controller/ (visited on 09/13/2023).

[Hin+94] K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell. "Passive real-world interface props for neurosurgical visualization". In: 1994. DOI: 10.1145/191666.191821.

[JR19] N. Jain and S. Rastogi. "SPEECH RECOGNITION SYSTEMS – A COMPREHENSIVE STUDY OF CONCEPTS AND MECHANISM". In: *Acta Informatica Malaysia* 3.1 (2019), pp. 01–03. ISSN: 2521-0874. DOI: 10.26480/aim.01.2019.01.03.

[Pet16] H. Petkar. "A Review of Challenges in Automatic Speech Recognition". In: *International Journal of Computer Applications* 151.3 (2016), pp. 23–26. ISSN: 0975-8887. DOI: 10.5120/ijca2016911706.

[Ali+20] S. Ali, R. Alauldeen, A. Ruaa, and R. Khamees. "What is Client-Server System: Architecture, Issues and Challenge of Client -Server System (Review)". In: (Feb. 2020). DOI: 10.5281/zenodo.3673071.

[PSP93] R. Pausch, M. Shackelford, and D. Proffitt. "A user study comparing head-mounted and stationary displays". In: *Proceedings of 1993 IEEE Research Properties in Virtual Reality Symposium*. 1993, pp. 41–45. DOI: 10.1109/VRAIS.1993.378265.

[KS13] B. Kress and T. Starner. "A review of head-mounted displays (HMD) technologies and applications for consumer electronics". In: *Photonic Applications for Aerospace, Commercial, and Harsh Environments IV*. Ed. by A. A. Kazemi, B. C. Kress, and S. Thibault. Vol. 8720. International Society for Optics and Photonics. SPIE, 2013, 87200A. DOI: 10.1117/12.2015654. URL: https://doi.org/10.1117/12.2015654.

[WS06] D. Wagner and D. Schmalstieg. "Handheld Augmented Reality Displays". In: Apr. 2006, pp. 321–321. DOI: 10.1109/VR.2006.67.

[CKY20] E. Chang, H.-T. Kim, and B. Yoo. "Virtual Reality Sickness: A Review of Causes and Measurements". In: *International Journal of Human-Computer Interaction* 36 (July 2020), pp. 1–25. DOI: 10.1080/10447318.2020.1778351.

[Wyk] A. Wykes. *Everything you need to know about spatial audio in headphones*. URL: https://www.soundguys.com/surround-sound-headphones-guide-49389/ (visited on 09/13/2023).

[CMH08] V. Chouvardas, A. Miliou, and M. Hatalis. "Tactile displays: Overview and recent advances". In: *Displays* 29.3 (2008), pp. 185–194. ISSN: 0141-9382. DOI: https://doi.org/10.1016/j.displa.2007.07.003. URL: https://www.sciencedirect.com/science/article/pii/S0141938207000613.

[Sut68] I. E. Sutherland. "A head-mounted three dimensional display". In: 1968. DOI: 10.1145/1476589.1476686.

[Kru83] M. Krueger. *Artificial Reality*. Addison-Wesley, 1983. ISBN: 9780201047653. URL: https://books.google.bg/books?id=EYNQAAAAMAAJ.

[CM92] T. Caudell and D. Mizell. "Augmented reality: an application of heads-up display technology to manual manufacturing processes". In: 1992. DOI: 10.1109/hicss.1992.183317.

[Azu97] R. T. Azuma. "A Survey of Augmented Reality". In: *Presence: Teleoperators and Virtual Environments* 6.4 (1997), pp. 355–385. ISSN: 1054-7460. DOI: 10.1162/pres.1997.6.4.355.

[BCL15] M. Billinghurst, A. Clark, and G. Lee. "A Survey of Augmented Reality". In: *Foundations and Trends® in Human–Computer Interaction* 8.2-3 (2015), pp. 73–272. ISSN: 1551-3955. DOI: 10.1561/1100000049. URL: https://dx.doi.org/10.1561/1100000049.

[ZRB21] D. A. Zuniga Gonzalez, D. Richards, and A. A. Bilgin. "Making it Real: A Study of Augmented Virtuality on Presence and Enhanced Benefits of Study Stress Reduction Sessions". In: *International Journal of Human-Computer Studies* 147 (2021), p. 102579. ISSN: 1071-5819. DOI: https://doi.org/10.1016/j.ijhcs.2020.102579. URL: https://www.sciencedirect.com/science/article/pii/S1071581920301816.

[Kor+19] D. Korsgaard, T. Bjørner, P. Sørensen, and J. Bruun-Pedersen. "Older adults eating together in a virtual living room: opportunities and limitations of eating in augmented virtuality". In: Sept. 2019, pp. 168–176. ISBN: 978-1-4503-7166-7. DOI: 10.1145/3335082.3335093.

[Reg+04] H. Regenbrecht, T. Lum, P. Kohler, C. Ott, M. Wagner, W. Wilke, and E. Mueller. "Using Augmented Virtuality for Remote Collaboration". In: *Presence: Teleoperators and Virtual Environments* 13.3 (June 2004), pp. 338–354. DOI: 10.1162/1054746041422334. eprint: https://direct.mit.edu/pvar/article-pdf/13/3/338/1624093/1054746041422334.pdf. URL: https://doi.org/10.1162/1054746041422334.

[CKM83] S. Cohen, T. Kamarck, and R. Mermelstein. "A Global Measure of Perceived Stress". In: *Journal of Health and Social Behavior* 24.4 (1983), pp. 385–396. ISSN: 00221465. URL: http://www.jstor.org/stable/2136404 (visited on 09/13/2023).

[KØG04] H. H. Keller, T. Østbye, and R. Goy. "Nutritional Risk Predicts Quality of Life in Elderly Community-Living Canadians". In: *The Journals of Gerontology: Series A* 59.1 (Jan. 2004), pp. M68–M74. ISSN: 1079-5006. DOI: 10.1093/gerona/59.1.M68. eprint: https://academic.oup.com/biomedgerontology/article-pdf/59/1/M68/9732988/M68.pdf. URL: https://doi.org/10.1093/gerona/59.1.M68.

[San+20] Z. I. Santini, P. E. Jose, E. York Cornwell, A. Koyanagi, L. Nielsen, C. Hinrichsen, C. Meilstrup, K. R. Madsen, and V. Koushede. "Social disconnectedness, perceived isolation, and symptoms of depression and anxiety among older Americans (NSHAP): a longitudinal mediation analysis". In: *The Lancet Public Health* 5.1 (2020), e62–e70. ISSN: 2468-2667. DOI: 10.1016/s2468-2667(19)30230-0.

[De 90] J. M. De Castro. "Social facilitation of duration and size but not rate of the spontaneous meal intake of humans". In: *Physiology & Behavior* 47.6 (1990), pp. 1129–1135. ISSN: 0031-9384. DOI: https://doi.org/10.1016/0031-9384(90)90363-9. URL: https://www.sciencedirect.com/science/article/pii/0031938490903639.

[23] *Human Aspects of IT for the Aged Population: 9th International Conference, ITAP 2023, Held as Part of the 25th HCI International Conference, HCII 2023, Copenhagen, Denmark, July 23–28, 2023, Proceedings, Part II*. Copenhagen, Denmark: Springer-Verlag, 2023. ISBN: 978-3-031-34916-4.

[GTM12] C. Grevet, A. Tang, and E. Mynatt. “Eating Alone, Together: New Forms of Commensality”. In: *Proceedings of the 2012 ACM International Conference on Supporting Group Work*. GROUP ’12. Sanibel Island, Florida, USA: Association for Computing Machinery, 2012, pp. 103–106. ISBN: 9781450314862. DOI: 10.1145/2389176.2389192. URL: https://doi.org/10.1145/2389176.2389192.

[KBN19] D. Korsgaard, T. Bjøner, and N. C. Nilsson. “Where would you like to eat? A formative evaluation of mixed-reality solitary meals in virtual environments for older adults with mobility impairments who live alone”. In: *Food Research International* 117 (2019). Special issue on "Virtual reality and food: Applications in sensory and consumer science”, pp. 30–39. ISSN: 0963-9969. DOI: https://doi.org/10.1016/j.foodres.2018.02.051. URL: https://www.sciencedirect.com/science/article/pii/S0963996918301443.

[Tru+21] P. Truong, K. Hölttä-Otto, P. Becerril, R. Turtiainen, and S. Siltanen. “Multi-User Virtual Reality for Remote Collaboration in Construction Projects: A Case Study with High-Rise Elevator Machine Room Planning”. In: *Electronics* 10.22 (2021). ISSN: 2079-9292. DOI: 10.3390/electronics10222806. URL: https://www.mdpi.com/2079-9292/10/22/2806.

[LD97] V. Lehner and T. DeFanti. “Distributed virtual reality: supporting remote collaboration in vehicle design”. In: *IEEE Computer Graphics and Applications* 17.2 (1997), pp. 13–17. DOI: 10.1109/38.574654.

[BLW18] C. W. Borst, N. G. Lipari, and J. W. Woodworth. “Teacher-Guided Educational VR: Assessment of Live and Prerecorded Teachers Guiding Virtual Field Trips”. In: *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2018, pp. 467–474. DOI: 10.1109/VR.2018.8448286.

[Vel+23] S. Vellingiri, R. P. Mcmahan, V. Johnson, and B. Prabhakaran. “An augmented virtuality system facilitating learning through nature walk”. In: *Multimedia Tools and Applications* 82.1 (2023), pp. 1553–1564. ISSN: 1380-7501. DOI: 10.1007/s11042-022-13379-w.

[Ter+12] S. Ternier, R. Klemke, M. Kalz, P. Ulzen, and M. Specht. “AR Learn: Augmented reality meets augmented virtuality”. In: *Journal of Cheminformatics - J Cheminf* 18 (Jan. 2012).

[Alb+14] A. Albert, M. Hallowell, B. Kleiner, A. Chen, and M. Golparvar-Fard. “Enhancing Construction Hazard Recognition with High-Fidelity Augmented Virtuality”. In: *Journal of Construction Engineering and Management* 140 (Apr. 2014). DOI: 10.1061/(ASCE)CO.1943-7862.0000860.

[YAB18] E. Yükseltürk, S. Altiok, and Z. Başer. “Using game-based learning with kinect technology in Foreign Language education course”. In: *Educational Technology and Society* 21 (Jan. 2018), pp. 159–173.

[Jon+08] A. Jones, J. E. Swan, G. Singh, and E. Kolstad. "The Effects of Virtual Reality, Augmented Reality, and Motion Parallax on Egocentric Depth Perception". In: *2008 IEEE Virtual Reality Conference*. 2008, pp. 267–268. DOI: 10.1109/VR.2008.4480794.

[Shi17] D.-H. Shin. "The role of affordance in the experience of virtual reality learning: Technological and affective affordances in virtual reality". In: *Telematics and Informatics* 34.8 (2017), pp. 1826–1836. ISSN: 0736-5853. DOI: https://doi.org/10.1016/j.tele.2017.05.013. URL: https://www.sciencedirect.com/science/article/pii/S0736585317301223.

[NSN16] N. C. Nilsson, S. Serafin, and R. Nordahl. "Walking in Place Through Virtual Worlds". In: *Human-Computer Interaction. Interaction Platforms and Techniques*. Ed. by M. Kurosu. Cham: Springer International Publishing, 2016, pp. 37–48. ISBN: 978-3-319-39516-6.

[Mor+14] C. Morato, K. N. Kaipa, B. Zhao, and S. K. Gupta. "Toward Safe Human Robot Collaboration by Using Multiple Kinects Based Real-Time Human Tracking". In: *Journal of Computing and Information Science in Engineering* 14.1 (Jan. 2014), p. 011006. ISSN: 1530-9827. DOI: 10.1115/1.4025810. eprint: https://asmedigitalcollection.asme.org/computingengineering/article-pdf/14/1/011006/6099649/jcise_014_01_011006.pdf. URL: https://doi.org/10.1115/1.4025810.

[Wei+15] T. Wei, B. Lee, Y. Qiao, A. Kitsikidis, K. Dimitropoulos, and N. Grammalidis. "Experimental study of skeleton tracking abilities from microsoft kinect non-frontal views". In: *2015 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*. 2015, pp. 1–4. DOI: 10.1109/3DTV.2015.7169367.

[AHK13] A. Albert, M. Hallowell, and B. Kleiner. "Enhancing Construction Hazard Recognition and Communication with Energy-Based Cognitive Mnemonics and Safety Meeting Maturity Model: Multiple Baseline Study". In: *Journal of Construction Engineering and Management* 140 (Jan. 2013). DOI: 10.1061/(ASCE)CO.1943-7862.0000790.

[CS06] G. Carter and S. Smith. "Safety Hazard Identification on Construction Projects". In: *Journal of Construction Engineering and Management-asce - J CONSTR ENG MANAGE-ASCE* 132 (Feb. 2006). DOI: 10.1061/(ASCE)0733-9364(2006)132:2(197).

[CCZ03] S. Cavusgil, R. Calantone, and Y. Zhao. "Tacit Knowledge Transfer and Firm Innovation Capability". In: *Journal of Business and Industrial Marketing* 18 (Feb. 2003), pp. 6–21. DOI: 10.1108/08858620310458615.

[Scr+10] T. E. Scruggs, M. A. Mastropieri, S. L. Berkeley, and L. Marshak. "Mnemonic Strategies: Evidence-Based Practice and Practice-Based Evidence". In: *Intervention in School and Clinic* 46.2 (2010), pp. 79–86. DOI: 10.1177/1053451210374985.

eprint: https://doi.org/10.1177/1053451210374985. URL: https://doi.org/10.1177/1053451210374985.

[DAJ11] D. Djaouti, J. Alvarez, and J. Jessel. "Classifying Serious Games: The G/P/S Model". In: 2011.

[Cha10] D. Charsky. "From Edutainment to Serious Games: A Change in the Use of Game Characteristics". In: *Games and Culture* 5.2 (2010), pp. 177–198. DOI: 10.1177/1555412009354727. eprint: https://doi.org/10.1177/1555412009354727. URL: https://doi.org/10.1177/1555412009354727.

[GAD02] R. Garris, R. Ahlers, and J. E. Driskell. "Games, Motivation, and Learning: A Research and Practice Model". In: *Simulation & Gaming* 33.4 (2002), pp. 441–467. DOI: 10.1177/1046878102238607. eprint: https://doi.org/10.1177/1046878102238607. URL: https://doi.org/10.1177/1046878102238607.

[LKH19] L. N. Lee, M. J. Kim, and W. J. Hwang. "Potential of Augmented Reality and Virtual Reality Technologies to Promote Wellbeing in Older Adults". In: *Applied Sciences* 9.17 (2019). ISSN: 2076-3417. DOI: 10.3390/app9173556. URL: https://www.mdpi.com/2076-3417/9/17/3556.

[Hug+17a] S. Hughes, K. Warren-Norton, P. Spadafora, and L. E. Tsotsos. "Supporting Optimal Aging through the Innovative Use of Virtual Reality Technology". In: *Multimodal Technologies and Interaction* 1.4 (2017). ISSN: 2414-4088. DOI: 10.3390/mti1040023. URL: https://www.mdpi.com/2414-4088/1/4/23.

[Lee+03] J. H. Lee, J. Ku, W. Cho, W. Y. Hahn, I. Y. Kim, S.-M. Lee, Y. Kang, D. Y. Kim, T. Yu, B. K. Wiederhold, M. D. Wiederhold, and S. I. Kim. "A Virtual Reality System for the Assessment and Rehabilitation of the Activities of Daily Living". In: *CyberPsychology & Behavior* 6.4 (2003). PMID: 14511450, pp. 383–388. DOI: 10.1089/109493103322278763. eprint: https://doi.org/10.1089/109493103322278763. URL: https://doi.org/10.1089/109493103322278763.

[CSD08] L. A. Cushman, K. Stein, and C. J. Duffy. "Detecting navigational deficits in cognitive aging and Alzheimer disease using virtual reality". In: *Neurology* 71.12 (2008), pp. 888–895. ISSN: 0028-3878. DOI: 10.1212/01.wnl.0000326262.67613.fe. URL: https://dx.doi.org/10.1212/01.wnl.0000326262.67613.fe.

[Kar+21] S. Karaosmanoglu, S. Rings, L. Kruse, C. Stein, and F. Steinicke. "Lessons Learned from a Human-Centered Design of an Immersive Exergame for People with Dementia". In: *Proceedings of the ACM on Human-Computer Interaction* 5 (Oct. 2021), pp. 1–27. DOI: 10.1145/3474679.

[Hug+17b] S. Hughes, K. Warren-Norton, P. Spadafora, and L. Tsotsos. "Supporting Optimal Aging through the Innovative Use of Virtual Reality Technology". In: *Multimodal Technologies and Interaction* 1.4 (2017), p. 23. ISSN: 2414-4088. DOI: 10.3390/mti1040023. URL: https://dx.doi.org/10.3390/mti1040023.

[BAG22] R. S. Baragash, H. Aldowah, and S. Ghazal. "Virtual and augmented reality applications to improve older adults' quality of life: A systematic mapping review and future directions". In: *DIGITAL HEALTH* 8 (2022), p. 20552076221132099. DOI: 10.1177/20552076221132099. eprint: https://doi.org/10.1177/20552076221132099. URL: https://doi.org/10.1177/20552076221132099.

[Has+21] M. Hassandra, E. Galanis, A. Hatzigeorgiadis, M. Goudas, C. Mouzakidis, E. M. Karathanasi, N. Petridou, M. Tsolaki, P. Zikas, G. Evangelou, and et al. "A Virtual Reality App for Physical and Cognitive Training of Older People With Mild Cognitive Impairment: Mixed Methods Feasibility Study". In: *JMIR Serious Games* 9.1 (2021), e24170. ISSN: 2291-9279. DOI: 10.2196/24170.

[COM20] M. Carrasco, N. Ortiz-Maqués, and S. Martínez-Rodríguez. "Playing with Nintendo Wii Sports: Impact on Physical Activity, Perceived Health and Cognitive Functioning of a Group of Community-Dwelling Older Adults". In: *Activities, Adaptation & Aging* 44.2 (2020), pp. 119–131. DOI: 10.1080/01924788.2019.1595261. eprint: https://doi.org/10.1080/01924788.2019.1595261. URL: https://doi.org/10.1080/01924788.2019.1595261.

[Kur+07] M. M. Kurtz, E. Baker, G. D. Pearlson, and R. S. Astur. "A Virtual Reality Apartment as a Measure of Medication Management Skills in Patients With Schizophrenia: A Pilot Study". In: *Schizophrenia Bulletin* 33.5 (2007), pp. 1162–1170. ISSN: 0586-7614. DOI: 10.1093/schbul/sbl039. URL: https://dx.doi.org/10.1093/schbul/sbl039.

[Fol+18] D. A. Foloppe, P. Richard, T. Yamaguchi, F. Etcharry-Bouyx, and P. Allain. "The potential of virtual reality-based training to enhance the functional autonomy of Alzheimer's disease patients in cooking activities: A single case study". In: *Neuropsychological Rehabilitation* 28.5 (2018). PMID: 26480838, pp. 709–733. DOI: 10.1080/09602011.2015.1094394. eprint: https://doi.org/10.1080/09602011.2015.1094394. URL: https://doi.org/10.1080/09602011.2015.1094394.

[Opt+10] G. Optale, C. Urgesi, V. Busato, S. Marin, L. Piron, K. Priftis, L. Gamberini, S. Capodieci, and A. Bordin. "Controlling Memory Impairment in Elderly Adults Using Virtual Reality Memory Training: A Randomized Controlled Pilot Study". In: *Neurorehabilitation and Neural Repair* 24.4 (2010). PMID: 19934445, pp. 348–357. DOI: 10.1177/1545968309353328. eprint: https://doi.org/10.1177/1545968309353328. URL: https://doi.org/10.1177/1545968309353328.

[Ija+22] K. Ijaz, T. T. M. Tran, A. B. Kocaballi, R. A. Calvo, S. Berkovsky, and N. Ahmadpour. "Design Considerations for Immersive Virtual Reality Applications for Older Adults: A Scoping Review". In: *Multimodal Technologies and Interaction* 6.7 (2022), p. 60. ISSN: 2414-4088. DOI: 10.3390/mti6070060.

[Por05] E. J. Porter. "Wearing and Using Personal Emergency". In: *Journal of Gerontological Nursing* 31.10 (2005), pp. 26–33. DOI: 10.3928/0098-9134-20051001-07. eprint: https://journals.healio.com/doi/pdf/10.3928/0098-9134-20051001-07.

URL: https://journals.healio.com/doi/abs/10.3928/0098-9134-20051001-07.

[Gol06] J. F. Golding. “Motion sickness susceptibility”. In: *Autonomic Neuroscience* 129.1 (2006). Nausea and Vomiting: An Interdisciplinary Approach, pp. 67–76. ISSN: 1566-0702. DOI: https://doi.org/10.1016/j.autneu.2006.07.019. URL: https://www.sciencedirect.com/science/article/pii/S1566070206002128.

[Eis+18] M. Eisapour, S. Cao, L. Domenicucci, and J. Boger. “Virtual Reality Exergames for People Living with Dementia Based on Exercise Therapy Best Practices”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 62.1 (2018), pp. 528–532. DOI: 10.1177/1541931218621120. eprint: https://doi.org/10.1177/1541931218621120. URL: https://doi.org/10.1177/1541931218621120.

[YC21] R. W. L. Yu and A. H. S. Chan. “Meta-analysis of the effects of game types and devices on older adults-video game interaction: Implications for video game training on cognition”. In: *Applied Ergonomics* 96 (2021), p. 103477. ISSN: 0003-6870. DOI: https://doi.org/10.1016/j.apergo.2021.103477. URL: https://www.sciencedirect.com/science/article/pii/S0003687021001241.

[PT12] T. P. Pham and Y.-L. Theng. “Game Controllers for Older Adults: Experimental Study on Gameplay Experiences and Preferences”. In: *Proceedings of the International Conference on the Foundations of Digital Games*. FDG ’12. Raleigh, North Carolina: Association for Computing Machinery, 2012, pp. 284–285. ISBN: 9781450313339. DOI: 10.1145/2282338.2282401. URL: https://doi.org/10.1145/2282338.2282401.

[Bak+19] S. Baker, R. M. Kelly, J. Waycott, R. Carrasco, T. Hoang, F. Batchelor, E. Ozanne, B. Dow, J. Warburton, and F. Vetere. “Interrogating Social Virtual Reality as a Communication Medium for Older Adults”. In: *Proceedings of the ACM on Human-Computer Interaction* 3.CSCW (2019), pp. 1–24. ISSN: 2573-0142. DOI: 10.1145/3359251.

[Die00] E. Diener. “Subjective well-being: The science of happiness and a proposal for a national index.” In: *American psychologist* 55.1 (2000), p. 34.

[DP] C. for Disease Control and Prevention. *Traumatic Brain Injury & Concussion*. URL: https://www.cdc.gov/traumaticbraininjury/data/index.html (visited on 08/26/2023).

[Ass] A. Association. *Alzheimer’s Disease Facts and Figures*. URL: https://www.alz.org/alzheimers-dementia/facts-figures (visited on 09/17/2023).

[SFT06] N. Silverstein, G. Flaherty, and T. Tobin. *Dementia and Wandering Behavior: Concern for the Lost Elder*. Springer Publishing Company, 2006. ISBN: 9780826196835. URL: https://books.google.bg/books?id=vPK6lijpje8C.

[LO06] I. Lancu and A. Olmer. “The Mini-Mental State Examination – an up-to-date review”. In: *Harefuah* 145 (Sept. 2006), pp. 687–90, 701.

[WM16] P. J. White and Z. Moussavi. "Neurocognitive Treatment for a Patient with Alzheimer's Disease Using a Virtual Reality Navigational Environment". In: *Journal of Experimental Neuroscience* 10 (2016). PMID: 27840579, JEN.S40827. DOI: 10.4137/JEN.S40827. eprint: https://doi.org/10.4137/JEN.S40827. URL: https://doi.org/10.4137/JEN.S40827.

[MM14] a. biyagowi ahmad, D. Mohaddes Khorassani, and Z. Moussavi. "Design and Application of a Novel Virtual Reality Navigational Technology (VRNChair)". In: *Journal of Experimental Neuroscience* 8 (Mar. 2014), pp. 7–14. DOI: 10.4137/JEN.S13448.

[Van+18] J. Van Santen, R.-M. Dröes, M. Holstege, O. B. Henkemans, A. Van Rijn, R. De Vries, A. Van Straten, and F. Meiland. "Effects of Exergaming in People with Dementia: Results of a Systematic Literature Review". In: *Journal of Alzheimer's Disease* 63.2 (2018), pp. 741–760. ISSN: 1387-2877. DOI: 10.3233/jad-170667.

[Pad+17] K. Padala, P. Padala, S. Lensing, R. Dennis, M. Bopp, P. Roberson, and D. Sullivan. "Home-Based Exercise Program Improves Balance and Fear of Falling in Community-Dwelling Older Adults with Mild Alzheimer's Disease: A Pilot Study". In: *Journal of Alzheimer's Disease* 59 (June 2017), pp. 1–10. DOI: 10.3233/JAD-170120.

[Unb+18] D. Unbehaun, D. D. Vaziri, K. Aal, R. Wieching, P. Tolmie, and V. Wulf. "Exploring the Potential of Exergames to Affect the Social and Daily Life of People with Dementia and Their Caregivers". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–15. ISBN: 9781450356206. DOI: 10.1145/3173574.3173636. URL: https://doi.org/10.1145/3173574.3173636.

[Löv+08] H. Lövheim, P.-O. Sandman, S. Karlsson, and Y. Gustafson. "Behavioral and psychological symptoms of dementia in relation to level of cognitive impairment". In: *International Psychogeriatrics* 20.4 (2008), pp. 777–789. ISSN: 1041-6102. DOI: 10.1017/s1041610208006777.

[Org] W. H. Organisation. *Dementia*. URL: https://www.who.int/news-room/fact-sheets/detail/dementia (visited on 08/27/2023).

[HAO04] P. Heyn, B. Abreu, and K. Ottenbacher. "The effects of exercise training on elderly persons with cognitive impairment and dementia: A meta-analysis1". In: *Archives of physical medicine and rehabilitation* 85 (Nov. 2004), pp. 1694–704. DOI: 10.1016/j.apmr.2004.03.019.

[SG04] K. A. Schutzer and B. Graves. "Barriers and motivations to exercise in older adults". In: *Preventive Medicine* 39.5 (2004), pp. 1056–1061. ISSN: 0091-7435. DOI: https://doi.org/10.1016/j.ypmed.2004.04.003. URL: https://www.sciencedirect.com/science/article/pii/S0091743504002026.

[Tab+19] L. Tabbaa, C. S. Ang, V. Rose, P. Siriaraya, I. Stewart, K. G. Jenkins, and M. Matsangidou. "Bring the Outside In". In: 2019. DOI: 10.1145/3290605.3300466.

[MAG01] M. MAGUIRE. "Methods to support human-centred design". In: *International Journal of Human-Computer Studies* 55.4 (2001), pp. 587–634. ISSN: 1071-5819. DOI: https://doi.org/10.1006/ijhc.2001.0503. URL: https://www.sciencedirect.com/science/article/pii/S1071581901905038.

[Viv] Vive. *Vive Tracker*. URL: https://www.vive.com/us/accessory/tracker3/ (visited on 08/28/2023).

[Nas20] G. Nassif. "Development of an Elderly Friendly User Interface by Adopting Controller-based Navigation in a Virtual Reality Travel Application". MA thesis. Technische Universität München, Nov. 2020.

[Wil21] A. Williams. "Enabling Seniors to Interact in Virtual Reality". BA thesis. Technische Universität München, Apr. 2021.

[Sch21] J. Schwan. "Companion Guidance and Interactive Components for an Elderly-Friendly Virtual Reality Travel Application". BA thesis. Technische Universität München, Aug. 2021.

[Teca] U. Technologies. *Unity*. URL: https://unity.com/ (visited on 09/25/2023).

[Man] U. Manual. *About the Oculus XR Plugin*. URL: https://docs.unity3d.com/Packages/com.unity.xr.oculus@1.2/manual/index.html (visited on 09/28/2023).

[Too] P. P. Tools. *OpenCV plus Unity*. URL: https://assetstore.unity.com/packages/tools/integration/opencv-plus-unity-85928 (visited on 09/28/2023).

[Met] Meta. *Meta Quest 2*. URL: https://www.meta.com/quest/products/quest-2/tech-specs/#tech-specs (visited on 09/26/2023).

[Log] Logitecg. *Logitech C270 HD Webcam*. URL: https://www.logitech.com/en-us/products/webcams/c270-hd-webcam.960-000694.html (visited on 09/26/2023).

[Mir] Mirror. *Mirror*. URL: https://github.com/MirrorNetworking/Mirror#made-with-mirror (visited on 09/26/2023).

[Wal22] S. Walchshäusl. "Generating an Environment for Socializing Between Elderly in a VR Supermarket". MA thesis. Technische Universität München, 2022.

[tea] O. team. *OpenCV: Open Source Computer Vision Library*. URL: https://github.com/opencv/opencv/tree/4.8.1 (visited on 09/28/2023).

[Tecb] U. Technologies. *Coding in C-sharp in Unity for Beginners*. URL: https://unity.com/how-to/learning-c-sharp-unity-beginners (visited on 09/28/2023).

[Mica] Microsoft. *Overview of Marshaling in C++/CLI*. URL: https://learn.microsoft.com/en-us/cpp/dotnet/overview-of-marshaling-in-cpp?view=msvc-170 (visited on 09/28/2023).

[LM02] R. Lienhart and J. Maydt. "An extended set of Haar-like features for rapid object detection". In: *Proceedings. International Conference on Image Processing*. Vol. 1. 2002, pp. I–I. DOI: 10.1109/ICIP.2002.1038171.

[Has+09] T. Hastie, S. Rosset, J. Zhu, and H. Zou. "Multi-class AdaBoost". In: *Statistics and Its Interface* 2.3 (2009), pp. 349–360. ISSN: 1938-7989. DOI: 10.4310/sii.2009.v2.n3.a8.

[Fli] Flickr. *Flickr*. URL: https://www.flickr.com/ (visited on 09/29/2023).

[Micb] Microsoft. *Walkthrough: Create and use your own Dynamic Link Library (C++)*. URL: https://learn.microsoft.com/en-us/cpp/build/walkthrough-creating-and-using-a-dynamic-link-library-cpp?view=msvc-170 (visited on 10/02/2023).

[The] I. The MathWorks. *What Is Camera Calibration?* URL: https://www.mathworks.com/help/vision/ug/camera-calibration.html (visited on 10/02/2023).

[Gáb] B. Gábor. *Camera calibration With OpenCV*. URL: https://docs.opencv.org/4.x/d4/d94/tutorial_camera_calibration.html (visited on 10/03/2023).

[Gar+14] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez. "Automatic generation and detection of highly reliable fiducial markers under occlusion". In: *Pattern Recognition* 47 (June 2014), pp. 2280–2292. DOI: 10.1016/j.patcog.2014.01.005.

[Opea] OpenCV. *Markers and Dictionaries*. URL: https://docs.opencv.org/4.8.0/d5/dae/tutorial_aruco_detection.html (visited on 10/12/2023).

[Fia10] M. Fiala. "Designing Highly Reliable Fiducial Markers". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.7 (2010), pp. 1317–1324. DOI: 10.1109/TPAMI.2009.146.

[Gar+] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez. *ArUco: a minimal library for Augmented Reality applications based on OpenCV*. URL: https://www.uco.es/investiga/grupos/ava/portfolio/aruco/ (visited on 10/26/2023).

[Opeb] OpenCV. *Detection of ArUco Boards*. URL: https://docs.opencv.org/4.x/db/da9/tutorial_aruco_board_detection.html (visited on 10/26/2023).

[WB+95] G. Welch, G. Bishop, et al. "An introduction to the Kalman filter". In: (1995).

[Pen] K. Pentenrieder. *Sensor Fusion using the Kalman Filter*. URL: https://campar.in.tum.de/Chair/KalmanFilter (visited on 10/28/2023).

[PTCa] PTC. *Vuforia: Market-Leading Enterprise AR*. URL: https://www.ptc.com/en/products/vuforia (visited on 10/29/2023).

[PTCb] PTC. *Vuforia Engine Features and Functionality for AR App Development*. URL: https://www.ptc.com/en/products/vuforia/vuforia-engine/ar-app-development (visited on 10/29/2023).

[BL] M. BL. *Bluetooth low energy Characteristics, a beginner's tutorial*. URL: https://devzone.nordicsemi.com/guides/short-range-guides/b/bluetooth-low-energy/posts/ble-characteristics-a-beginners-tutorial (visited on 11/03/2023).

[Bus+21] P. A. Busch, G. I. Hausvik, O. K. Ropstad, and D. Pettersen. “Smartphone usage among older adults”. In: *Computers in Human Behavior* 121 (2021), p. 106783. ISSN: 0747-5632. DOI: https://doi.org/10.1016/j.chb.2021.106783. URL: https://www.sciencedirect.com/science/article/pii/S0747563221001060.

[Ram+16] A. T. Ramsey, J. L. Wetherell, D. D. Colin Depp, and E. Lenze. “Feasibility and Acceptability of Smartphone Assessment in Older Adults with Cognitive and Emotional Difficulties”. In: *Journal of Technology in Human Services* 34.2 (2016). PMID: 27683018, pp. 209–223. DOI: 10.1080/15228835.2016.1170649. eprint: https://www.tandfonline.com/doi/pdf/10.1080/15228835.2016.1170649. URL: https://www.tandfonline.com/doi/abs/10.1080/15228835.2016.1170649.

[Web+21] S. Weber, D. Dyrda, M. Ludwig, and G. Klinker. “Ubi-Interact”. In: *MobiQuitous 2020 - 17th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. MobiQuitous ’20. Darmstadt, Germany: Association for Computing Machinery, 2021, pp. 291–300. ISBN: 9781450388405. DOI: 10.1145/3448891.3448924. URL: https://doi.org/10.1145/3448891.3448924.

[Sch03] T. Schubert. “The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness”. In: *Zeitschrift für Medienpsychologie* 15 (Apr. 2003), pp. 69–71. DOI: 10.1026//1617-6383.15.2.69.

[Bro95] J. Brooke. “SUS: A quick and dirty usability scale”. In: *Usability Eval. Ind.* 189 (Nov. 1995).

[Mel+23] M. Melo, G. Gonçalves, J. Vasconcelos-Raposo, and M. Bessa. “How Much Presence is Enough? Qualitative Scales for Interpreting the Igroup Presence Questionnaire Score”. In: *IEEE Access* 11 (2023), pp. 24675–24685. ISSN: 2169-3536. DOI: 10.1109/access.2023.3254892.

[arf] arfish. *How do I add a "Floor" into a VR SkyBox*. URL: https://forum.unity.com/threads/how-do-i-add-a-floor-into-a-vr-skybox.1269971/ (visited on 11/12/2023).

[Eic+] C. Eichhorn, D. A. Plecher, T. Mesmer, L. Leder, T. Simecek, N. Boukadida, and G. Klinker. “Shopping in between Realities - Using an Augmented Virtuality Smartphone in a Virtual Supermarket”. In: ().