This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3271709

SceneFusion: Room-Scale Environmental Fusion for Efficient Traveling Between Separate Virtual Environments

Miao Wang, *Member, IEEE*, Yi-Jun Li, Jinchuan Shi, Frank Steinicke, *Member, IEEE*

Abstract—Traveling between scenes has become a major requirement for navigation in numerous virtual reality (VR) social platforms and game applications, allowing users to efficiently explore multiple virtual environments (VEs). To facilitate scene transition, prevalent techniques such as instant teleportation and virtual portals have been extensively adopted. However, these techniques exhibit limitations when there is a need for frequent travel between separate VEs, particularly within indoor environments, resulting in low efficiency. In this paper, we first analyze the design rationale for a novel navigation method supporting efficient travel between virtual indoor scenes. Based on the analysis, we introduce the SceneFusion technique that fuses separate virtual rooms into an integrated environment. SceneFusion enables users to perceive rich visual information from both rooms simultaneously, achieving high visual continuity and spatial awareness. While existing teleportation techniques passively transport users, SceneFusion allows users to actively access the fused environment using short-range locomotion techniques. User experiments confirmed that SceneFusion outperforms instant teleportation and virtual portal techniques in terms of efficiency, workload, and preference for both single-user exploration and multi-user collaboration tasks in separate VEs. Thus, SceneFusion presents an effective solution for seamless traveling between virtual indoor scenes.

Index Terms—Virtual Reality, Collaborative Virtual Environments, Scene Transition.

1 Introduction

Recently, the advancements in virtual reality (VR) technology have increased the interest in and usage of VR across various application domains, including entertainment, communication, work, training, simulation, and education [1]–[3]. With VR users can immerse themselves in virtual environments (VEs) and interact with 3D objects and VEs or communicate with remote peers or friends. However, the indoor physical environments in which most users experience VR applications are typically smaller than most VEs, which typically limits the natural movements of VR users. Furthermore, it can also be daunting for VR users to expend the same amount of physical energy to travel between distant virtual locations as they would in the physical world. To address these challenges, techniques for scene transition are often required [4], enabling users to conveniently travel between distant or separate virtual spaces by instantly or gradually changing their locations, while revealing new visuals and making previous objects fade and disappear. Despite being less realistic than real walking, existing scene transition techniques [4]–[6] are more efficient for traveling.

- *Miao Wang is with the State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing 100191, China, and also with Zhongguancun Laboratory, Beijing 100094, China. E-mail: miaow@buaa.edu.cn*
- *Yi-Jun Li, and Jinchuan Shi are with the State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing 100191, China. E-mail: {yaoling, shijinchuan}@buaa.edu.cn*
- *Frank Steinicke is with the Department of Informatics, Universität Hamburg, 22605 Hamburg, Germany. Email: frank.steinicke@uni-hamburg.de*
- *Corresponding author: Miao Wang.*

Manuscript received July, 2022

Target-based scene transition techniques, such as instant teleportation, are often utilized in VR applications to transport users to different VEs based on triggering events [4], [7]–[12]. Other popular approaches include transitioning through virtual portals [11], [13]–[19], commonly used in various VR applications (e.g., *Pavlov* [20], a shooting game, and *VRChat* [21], a virtual chatroom). These techniques enable virtual users to gather for discussion, task execution, or entertainment in a shared collaborative virtual environment (CVE) [22]. However, the spatial layout and ongoing events around the target arrival locations can be either invisible or partially occluded, which can differ from what users imagine before scene switching [23]. This abrupt scene context change can negatively impact VR users' visual cognition, introduce additional mental workload, and break the immersive experience during exploration and collaboration activities. Besides, switching environments will cause the original environment information to be lost. If users only need to meet temporarily and then return back to the previous environment to continue the exploration, this sudden environment change will also interrupt the continuity of the VR experience.

For many VR application scenarios, remote users are likely to collaborate or communicate with each other while preserving the original environment information. For example, in an architectural application, two users may need to explore separate virtual rooms and wish to share their layouts and furniture, or simply compare them side-by-side. Another use case is the integration of small factories into a larger one, where the workers gain more space for interactive activities and easy access to distributed equipment. Similar scenarios can occur in VR training, education,

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3271709

games, social events, and so on. Therefore, a high-efficiency technique with preservation of visual continuity and spatial awareness is crucial for users traveling between VEs.

In this paper, we focus on efficient traveling between virtual indoor spaces, where users frequently travel between separate virtual rooms during virtual exploration (e.g., navigation, searching, transportation). In contrast to scene transition techniques such as instance teleportation or virtual portals that locally connect the virtual spaces at a point, our motivation is to fuse the individual spaces and offer a shared space for efficient locomotion with engaging interaction and improved spatial awareness throughout the entire environments.

Based on this idea, we introduce the design rationale for the new scene fusion technique that emphasizes the importance of efficiency, visual plausibility, and versatility. With a small number of operations, users can easily fuse indoor VEs with different layouts to create a visually pleasant shared space with rich visual information. This enables them to collaborate with other users from a different virtual rooms and interact with virtual objects more efficient way. The proposed technique is realized through *SceneFusion*, which fuses separate virtual rooms into an integrated environment suitable for both single-user exploration and multi-user collaboration tasks. The key innovation involves (i) dividing, (ii) aligning, and (iii) stitching each virtual room, while maintaining structural and textural smoothness (Figure 1). As a result, separate rooms are fused, enabling users to actively travel to the other adjacent room through continuous locomotion across the fusion area instead of being passively teleported.

Experimental results confirm that our novel technique is more efficient, and comprehensively preferred in comparison with baseline teleportation techniques.

The main contributions of this work include:

- An analysis of existing scene transition techniques and design rationale of the new room-scale environmental fusion technique considering efficiency, visual plausibility, and versatility.
- The presentation of the SceneFusion technique that fuses separate VEs via structure stitching and texture blending, enabling users to actively travel between separate VEs with continuous visual change and retaining most visual information of both VEs.
- User studies involving single-user exploration and multi-user collaboration tasks show that SceneFusion significantly outperforms existing techniques such as instant teleportation and virtual portal techniques on efficiency, workload, and preference.

2 Related Work

2.1 Multi-User Collaboration in VR

CVEs [22] provide users with a natural and intuitive interaction experience [24], making remote collaboration more realistic than traditional video streaming conferences [12]. Therefore, CVEs are increasingly being leveraged by geographically distributed and co-located teams for collaborative work [22]. Collaboration tasks usually require collaborators to form a group to effectively solve problems, requiring virtual systems to provide users with opportunities to meet together.

According to the taxonomy from a recent survey [25], collaborators can connect with each other in various forms. From the synchronicity perspective, synchronous collaboration is commonly adopted for aiding user cooperation [12], [26]–[33], while asynchronous collaboration can serve as a supplement when team members are not available at the same time [34]–[37]. From the space distribution perspective, co-located collaboration [38]–[40] restricts all the collaborators into the same physical tracking area while remote collaboration offers a shared space for geographically-distributed users to meet and complete tasks [12], [41]–[45]. Besides, collaborators can access the shared space using a range of hardware devices and play different roles or possess different abilities [25], [28], [46]–[51].

One approach to offering a shared environment for collaboration is to relocate all participants to a new common virtual space, such as a virtual meeting room or classroom [37], [51]. Alternatively, remote VR users can be transported into the local AR user's space through 3D reconstruction techniques [7], [26], [29], [42], [52] and 360 panoramas [10], [12], [29], [41], [42]. For example, Stotko et al. [52] provided a consumer-grade client-server system that uses Marching Cubes [53] to reconstruct and update the local user's environment in real-time, enabling multiple remote clients to explore the environment simultaneously. Rhee et al. [12] proposed an asymmetric collaboration system that utilizes a 360° camera to capture the environment of a local AR host in real time, providing remote VR users with a high-fidelity telepresence experience. Nevertheless, after participants enter a new shared space, information about their original environments will be lost.

Thus, some alternative techniques aim to blend different spaces together to form a shared environment, capable of retaining enough information about the original rooms. Zhang et al. [54] proposed VirtualCube, a 3D video conference system that employs large-format screens to blend different spaces and maintain the correct mutual eye gaze contact between participants, providing the feeling of staying in the same room. Young et al. [55] proposed Mobileportation, which uses an RGBD sensor to incrementally reconstruct the 3D mesh of the local user's environment, enabling the remote user to freely explore the shared space using mobile devices, while preserving the original environment's information. Other methods [56]–[58] embed the VE into the indoor scene or automatically generate a VE according to the real space reference. Hartmann et al. [56] proposed Reality-Check, which utilizes real-time 3D reconstruction to embed the physical world inside the VE, allowing users to interact with the items and communicate with the people situated in a real environment while maintaining the sense of presence within the VE. Sra et al. [58] introduced the Oasis system, which uses inside-out tracking mobile devices to automatically generate immersive virtual reality environments based on the real indoor scene template, allowing users to interact with virtually generated objects and receive the passive haptics feedback [59] provided by the corresponding real objects. Some studies have also examined VEs generated by the open physical world [60]. Sayyad et al. [60] compared natural walking with instant teleportation in a wide-area

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3271709

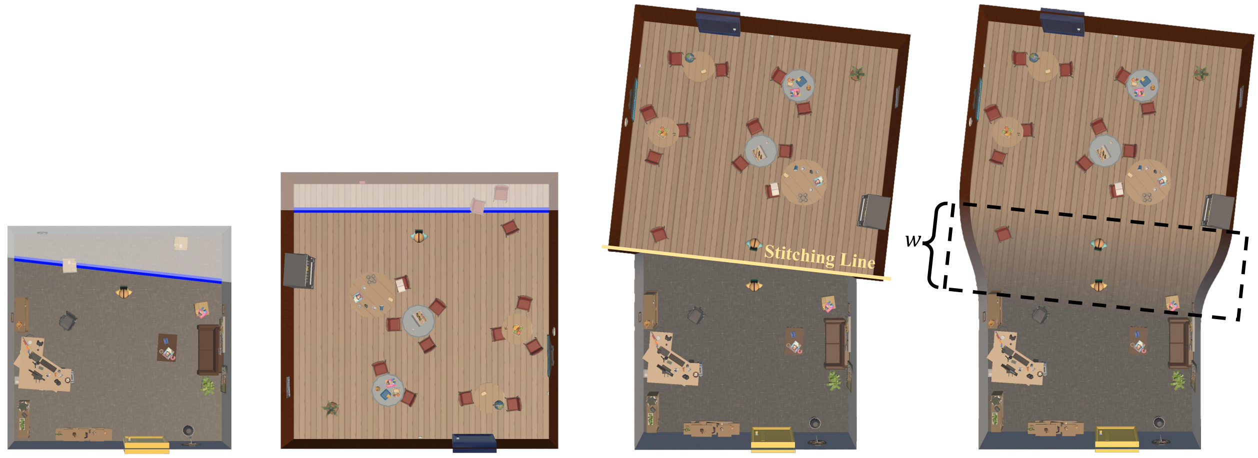


Fig. 1. Illustration of the room fusion process of our SceneFusion technique. (a) and (b) indicate two separate rooms, and a dividing line is generated manually or automatically in each room, which divides the room into two parts. Our technique retains the part where the user is currently standing and discards the other part (shown in translucent). Then two reserved parts will be aligned and stitched according to the centers of the dividing lines (shown in (c)). Finally, the cubic Bézier curve is applied to smooth the fusion area (shown in (d)). The dotted box indicates the fusion area.

VE generated from a $61\,m \times 26\,m$ area, showing the ability of natural walking to reduce simulator sickness.

2.2 Long-distance Traveling Techniques

Traveling in the VE is a basic requirement for VR applications, and various traveling techniques have been developed and summarized by previous work [61]–[63]. Among them, walking-based locomotion techniques including walking-in-place [64]–[66], and redirected walking [67]–[70] can offer users an experience similar to natural walking and ensure the movement continuity, but they may not be efficient for long-distance traveling or scene switching. In some situations, VR users may need to travel between different virtual locations to complete tasks or meet friends, which could require significant time and effort if they were to walk directly to the location over a large distance.

To make users more efficiently travel to another scene, powerful and easy-to-use teleportation techniques are often required. Instant teleportation [4], [7]–[12], [71] is one of the most common teleportation techniques able to teleport the VR user to a new location instantaneously. Nevertheless, this sudden position change can interrupt the user's visual continuity and cause disorienting problems [72]. In multi-user collaboration tasks, instant teleportation can cause one collaborator to disappear and reappear frequently, making the other collaborator confused and hard to follow [11], [30].

It is practical to provide additional visual cues to enhance users' awareness of collaborators' position changes and activities in VR. For example, Thanyadit et al. [11] proposed substituted visualization techniques by considering the design requirements of time efficiency, traceability, intuitiveness, and recognizability. Piumsomboon et al. [73] introduced the Mini-Me technique, which utilizes an adaptive avatar to assist the local user in tracking the teleported remote collaborator.

Other than instant teleportation, an alternative way of scene transitioning is to use virtual portals [11], [13]–[19], which are doorways connecting two different virtual locations. The user can walk through the portal to reach the other virtual location without interruption and getting lost. This technique provides the user with a continuous world and the user can acquire part of the visual information of another room through the portals, which is helpful to maintain the user's orientation and sense of presence [14].

Besides, there are other techniques for traveling long distances. Bolte et al. [74] introduced the jumper metaphor for long-distance traveling, which allows the user to virtually jump to a predicted target location. This method can provide a smoother viewpoint transition during the teleportation process, which is more effective than physically walking and less disruptive than instant teleportation. For vertical navigation, Vasylevska et al. [75] proposed the elevator metaphor to make vertical movements between multi-level VEs more natural and realistic.

3 Design Rationale of SceneFusion

We focus on the problem of efficient exploration in virtual indoor scenes, where single or multiple users need to frequently travel between virtual rooms to complete certain tasks. In various VR applications such as training and games, teleportation is commonly employed for switching between scenes. While most existing teleportation techniques can rapidly teleport a user from one virtual room to another, the occlusion issue during teleportation makes it hard for the users to foresee the environment around the target teleportation point and keep the sense of direction, thus hindering the user experience. Besides, frequent scene switching using existing teleportation techniques inevitably imposes a heavy cognitive load on users, and the visual information of the original room will also be lost after teleportation. To overcome these issues, we introduce the design rationale of our SceneFusion technique from three aspects: efficiency, visual plausibility, and versatility.

3.1 Efficiency

Efficiency is a crucial metric for evaluating performance in VR [11], [23], [76], [77], which is closely related to factors

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3271709

such as time cost, frequency of operations, travel distance, and cognitive workload. In existing techniques where frequent scene switching is necessary, instant teleportation may be annoying due to the increased number of button clicks and position adjustments, leading to additional task workload and longer task completion time [11]. Another technique using portals requires the user to walk through the established virtual portal to enter the other room, and orients the user to the target position, which breaks the continuous walking flow and introduces additional moving distance. Our method aims to fuse the individual rooms and enable the user to walk in a large accessible area that connects both rooms, maintaining a continuous walking flow and potentially reducing the user's travel distance and frequency of unnecessary operations.

3.2 Visual Plausibility

Appropriate visual, auditory, and tactile feedback in VEs can improve task execution and increase user engagement [78]–[80]. In this work, we mainly focus on improving the visual plausibility of traveling in VEs, as the visual appearance has a significant impact on users' virtual experiences [81]–[83]. Directly stitching individual virtual rooms may result in structural or textural inconsistency due to differences in room size, shape, layout, and decoration [84], [85]. Moreover, exposure to VEs with inappropriate visual designs may cause motion sickness symptoms such as discomfort, headache, nausea, and fatigue when exploring VR [86]–[88]. Context awareness is also a key factor affecting the user's visual plausibility, and adequate context information serves as visual cues that facilitate better understanding and task execution within the VE [11], [89]–[91]. However, existing teleportation techniques limit accessibility to visual information, including remote environment layouts and ongoing events, and frequent scene switching may result in users losing track of other collaborators [11]. To maintain visual plausibility, we propose using the cubic Bézier curve for environmental fusion that smoothly stitches structures and blends textures of rooms, resulting in an aesthetically pleasing outcome [92]–[94].

3.3 Versatility

The technique should be adaptable to varying testing conditions [30], [31], [77], [95], [96]. For example, Weissker et al. [30], [31] proposed a controller-based short-distance teleportation approach that utilizes the group-forming concept. This method empowers a dominant user to teleport multiple users to a new location by simply clicking the controller buttons and rotating the controller, reducing the group's total manual effort. It is also applicable to different VEs, and can meet the requirement with up to teleporting 10 participants. In alignment with our goal, the novel environmental fusion technique should be able to adapt to diverse interior spaces with various attributes such as layouts, textures and sizes. The fused space should support various typical VR activities such as communication, locomotion, and exploration. Besides, it should also satisfy the collaboration requirements of multiple distributed users, specifically facilitating concurrent collaboration between two or more participants.

4 THE SCENEFUSION TECHNIQUE

We introduce the SceneFusion technique that fuses individual virtual rooms into a whole VE for continuous navigation and exploration. Instead of being passively teleported, the user can actively move to the originally separate areas via continuous locomotion techniques such as walking-in-place.

4.1 Preliminary

The SceneFusion technique is designed for both single-user exploration and multi-user collaboration in virtual indoor scenes, and is applicable to virtual reality activities such as virtual conferences or exhibitions. Although indoor rooms can be complex and diverse, some indoor spaces (e.g., spaceships) can even have curve walls, for simplicity, we only consider regular rectangular rooms in this work. Formally, given an enclosed virtual 3D room, virtual objects inside the room are divided into 5 categories: *floor*, *ceiling*, *wall*, *static object*, and *interactive object*. The *floor* and *ceiling* indicate the actual height of the virtual room, which we assume to be constant in this preliminary exploration. *Walls* outline the boundary of the enclosed space, and users are explicitly told not to cross through them to get to the outside area which is invalid for user activities. *Static objects* are stationary virtual objects such as furniture that cannot be moved by users, indicating the interior layout of the virtual room and affecting the user's walking path and behaviors during the task execution process. *Interactive objects* can be interactively manipulated by users to dynamically change their positions, orientations, or motion states for exploration purposes. To simplify the calculation, we use a top-view orthographic projection to create a 2D planar layout graph of the virtual room, as commonly used in [95], [97], [98]. The shape of a virtual room is represented by a sequence of 2D points in counter-clockwise order, with each *wall* represented by a segment connecting a pair of adjacent points. The location and geometry of each *static object* and *interactive object* are represented by a proxy bounding box.

4.2 Method

Given individual virtual rooms for navigation and exploration, the SceneFusion technique allows users to create a connected space for continuous exploration or face-to-face collaboration.

4.2.1 Line-Based Room Division

The core idea of our SceneFusion technique is to fuse separate virtual rooms (two rooms in our experiments) and form a smoothly integrated virtual room. However, individual virtual rooms are enclosed by walls, and direct fusion is not possible. To address this, our technique allows the user to divide the room into two parts by indicating a straight line on the floor. Objects on the user's side of the line are preserved, while on the other side, the *static objects* or *interactive objects* including those intersected by the line are split out and discarded, as shown in Figure 1 (a)(b). In single-user cases, the user can specify the dividing line and the room part to be reserved in advance using other approaches, such as entering each room space using instant teleportation or indicating the line according to the top thumbnail of each room.

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3271709

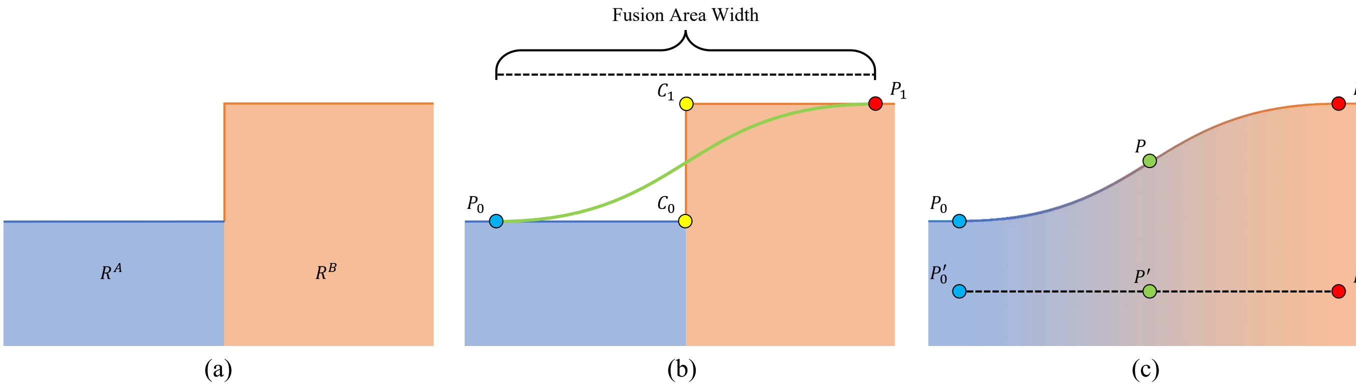


Fig. 2. Illustration of visual smooth optimization. (a) indicates two aligned rooms before optimization. (b) shows the Bézier curve formed by four control points. P_0 and P_1 are intersection points of the fusion area and room walls. C_0 and C_1 are points that control smoothness. (c) reveals the final optimized visual effects of our SceneFusion technique, where the texture of the part of the floor and wall in the fusion area is linearly blended along the fusion area width.

Normally, a straightforward approach to specifying the dividing line is by using the controller to draw a segment on the floor. This segment is then extended to intersect with the walls and form a straight dividing line. This line can be manually specified by the room designer or the user, or automatically generated using heuristic algorithms based on different objectives. For example, since providing only task-related information and clues is of great benefit to reducing users' cognitive demand [77], a dividing line can be generated in front of the user by retaining the most *interactive objects* meanwhile removing the most *static objects*.

4.2.2 Visual-Aware Room Fusion

After properly setting the dividing line in each room, room fusion is performed to facilitate efficient traveling while maintaining the visual richness and plausibility as much as possible. For users' convenience, separate users can verbally communicate before establishing the fusion area. Once the dividing line is appropriately positioned, the users can confirm the choice by clicking the grip button. After users on both sides confirm their choices, the fusion operation will begin automatically. To avoid sudden visual change, each user's field of view will gradually turn black and then become bright again after the fusion process is complete. In the fusion process, we first stitch the reserved regions by spatially aligning the centers of the two dividing lines (Figure 1(c)). Then, we smoothly stitch the walls and blend the textures of walls and floors within a w-meter fusion area across the stitching line, as illustrated in Figure 1(d).

Structure Stitching. Directly stitching the two reserved room regions (denoted as R^A and R^B respectively) by aligning the dividing lines can result in a hard-stitch effect with abrupt structure and visual changes (Figure 2 (a)). Furthermore, objects close to the wall or room corner may be occluded from the user's viewpoint. To avoid blocking the view, we employ a cubic Bézier curve to stitch each pair of the stitching wall structures smoothly.

Bézier curve is a parametric curve model defined by a set of control points, where two points are the ends of the curve and the other points determine the shape of the curve. A cubic Bézier curve is typically expressed as:

$$B(t) = P_0(1-t)^3 + 3C_0t(1-t)^2 + 3C_1t^2(1-t) + P_1t^3, \quad (1)$$

where $B(t), t \in [0,1]$ denotes any point located on the curve. P_0, P_1 represent the curve's endpoints, whereas C_0, C_1 correspond to the other control points.

Illustratively, we present the stitching algorithm for one pair of stitching walls, and the other pair of walls can be stitched correspondingly. We set the control point C_0 as the intersection of the stitching line and the wall of the smaller fusion area, and C_1 as the intersection of the stitching line and the wall of the larger fusion area. P_0 and P_1 are sampled along the stitching walls, each at $w/2$ meters from the stitching line respectively (Figure 2 (b)), where w is the fusion area width across the stitching line. The original wall segments $\overline{P_0C_0}$, $\overline{C_0C_1}$, and $\overline{C_1P_1}$ are then replaced by the generated Bézier curve segment between P_0 and P_1. As a result, the stitched wall structure smoothly transitions across the fusion area.

Texture Blending. To ensure a smooth visual transition, we further adopt a linear texture blending approach for both walls and floors within the fusion area. Regarding wall texture blending, for any point P on the Bézier curve P_0P_1, the blended color $c(P)$ is computed as:

$$c(P) = \frac{|PP_1|}{|P_0P_1|}c^A(P) + \frac{|P_0P|}{|P_0P_1|}c^B(P), \quad (2)$$

where $|PP_1|$, $|P_0P|$, $|P_0P_1|$ are lengths of curve segments represented by corresponding endpoints, $c^A(P)$ and $c^B(P)$ indicate the original wall texture colors of R^A and R^B at position P respectively.

The floor texture blending is similarly conducted, except that the color $c(P')$ is blended along a line segment $\overline{P'_0P'_1}$ across P' and perpendicular to the stitching line.

After wall stitching and texture blending, users can seamlessly reach the other room using short-range locomotion techniques. Note that SceneFusion supports both single-user exploration and multi-user collaboration tasks with even more than two users and rooms. Once a new room is generated through the environmental fusion of two individual rooms, it can be further fused with another room. Room fusion examples are provided in Figure 3 and also in the supplementary material.

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3271709

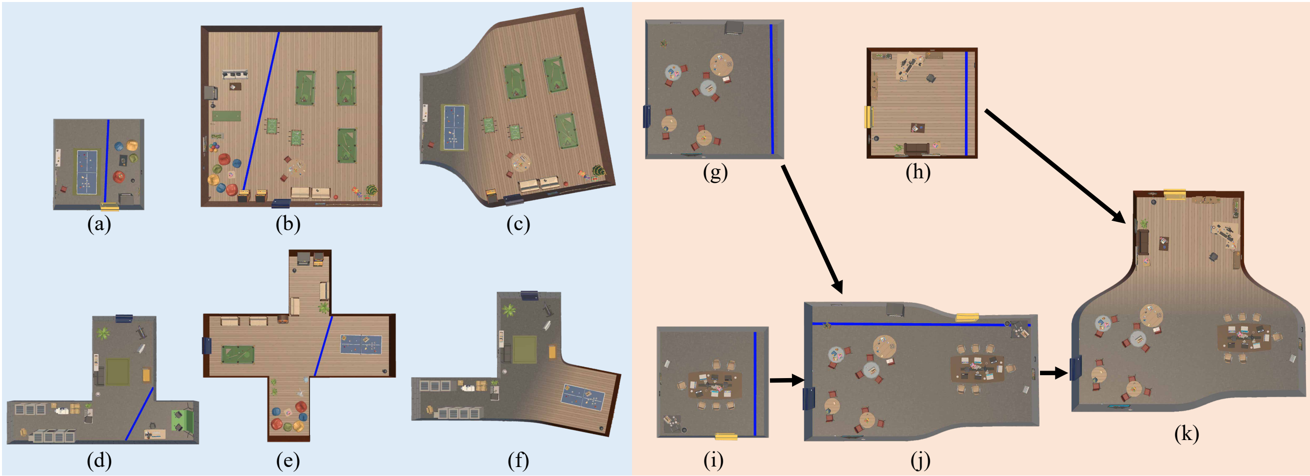


Fig. 3. More examples about scene fusion. (a) and (b) indicate a pair of regular rooms. (d) and (e) represent a pair of irregularly shaped rooms. (c) and (f) show the top views of the fused rooms by (a)(b) and (c)(d) respectively. The example of fusing three rooms is shown on the right. The intermediate space (j) is formed by fusing (g) and (i); (k) is formed by further fusing (h) and (j).

4.3 Locomotion Strategy

While SceneFusion is compatible with most types of locomotion strategies for short-range movement, we propose to choose the one that is consistent with the user's movement in the real world for ease of use. In comparison, walking-in-place technique is cost-effective and able to provide similar proprioceptive feedback and perceived physical strain as real walking [99]. Thus, we adopted walking-in-place as the locomotion technique and implemented it using tapping gesture in our experiments [65].

5 User Study Overview

We conducted a series of user studies to determine the key parameter of SceneFusion and evaluated its performance in both single- and multi-user application scenarios. Initially, we conducted a pilot study involving 12 users to experience environmental fusion results with candidate fusion area width values and determine the suitable value based on collected user feedback. The pilot study indicated an optimal width of $4\,m$ for the fusion region, and details can be found in the supplementary material. After that, we compared the SceneFusion technique with scene transition techniques including instant teleportation and virtual portals. We conducted user experiments in a single-user exploration task (Section 6), where separate environments are pre-connected with identically fixed teleportation points or fusion area positions between testing techniques. This allowed users to directly travel between individual environments without the need to interactively connect them. Furthermore, we evaluated the techniques in a multi-user collaboration task (Section 7), allowing the users to interactively specify the positions for scene transitioning, and providing a higher degree of interactive freedom with the full interaction process.

6 Study 1: Single-User Exploration

To assess the efficacy of the SceneFusion technique and compare it with existing approaches, we conducted a single-user experiment in which the participant had to travel between two virtual rooms to complete a pick-and-place task involving the search and transportation of a virtual object to a designated position. This experimental design necessitated the frequent referencing of layout information from both rooms in order to successfully execute the task.

6.1 Study Design

The study utilized a within-subjects design to compare the performance of three scene traveling techniques: instant teleportation (abbreviated as *IT*), virtual portal (abbreviated as *VP*), and SceneFusion (abbreviated as *SF*). The single-user pick-and-place task was employed to evaluate the efficacy of these techniques. Participants were tasked with searching for and picking up virtual props in one room, before proceeding to search for target positions to place the virtual props in another room, using the tested techniques as quickly as possible. To complete the task, the participant was required to frequently switch between rooms.

Techniques. To facilitate teleportation, a blue, luminous circle with a diameter of $1\,m$ was placed on the floor of each room as a reference point for the *IT* technique. Upon standing on the circle and pressing the grip button, the participant could immediately travel to the teleportation point in the other room. Although various modifications of this technique exist, such as allowing the user to teleport to the other room without considering their standing position, our experiment ensured fairness by requiring the user to stand inside the blue indicator to perform the teleportation operation. The *VP* technique employed a virtual portal measuring $1.5\,m \times 2.4\,m$ in each room that allowed the participant to view and access the other room. The *SF* technique involved fusing the individual rooms by clicking the grip button and then using the walking-in-place technique to navigate and complete tasks within the fused VE. To ensure fairness, the positions of the teleportation points in *IT*, virtual portals in *VP*, and dividing lines in *SF* were pre-defined and spatially aligned. Figure 4 displays the top and first-person views of the environments featuring the tested techniques.

Environments. The study employed 6 pairs of virtual rooms with varying furniture layouts, measuring $\langle 8 \times$

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3271709

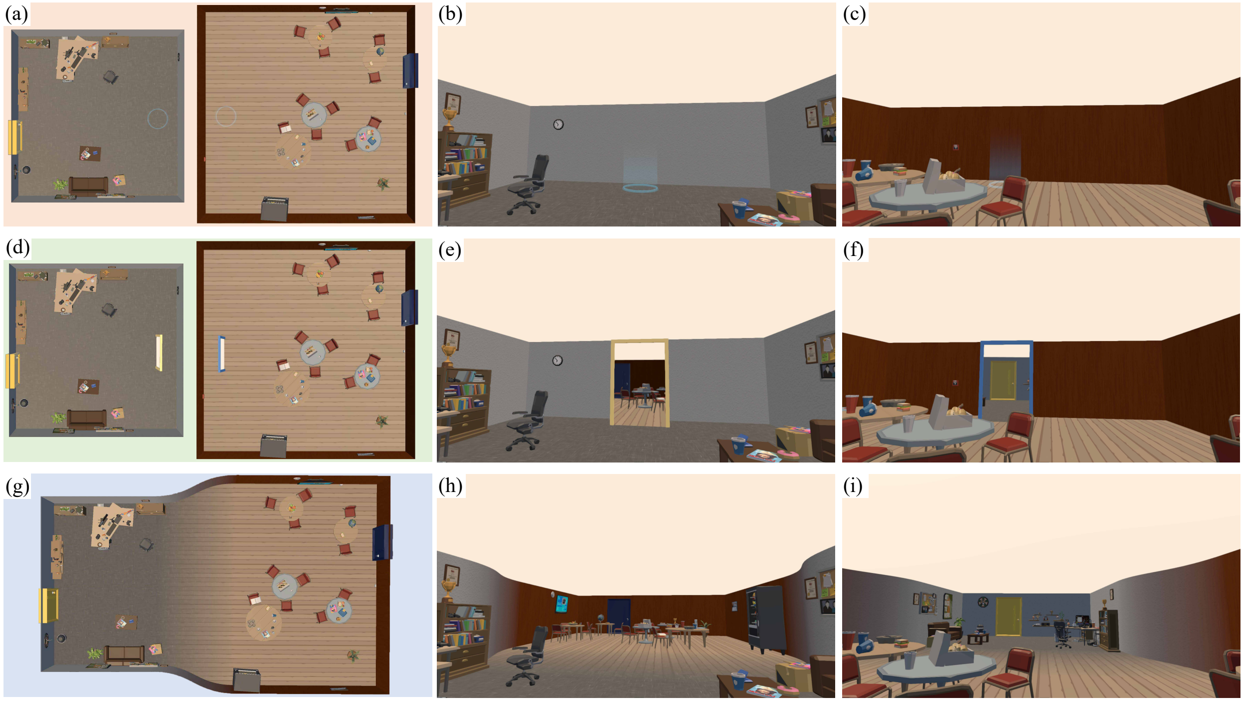


Fig. 4. Illustration of representative scene traveling techniques. From top to bottom: instant teleportation, virtual portal, and SceneFusion. Figures (a), (d), (g) represent the top view of room layouts. Figures (b), (e), (h) represent the first-person view in the smaller room, while figures (c), (f), (i) represent the first-person view from the larger room.

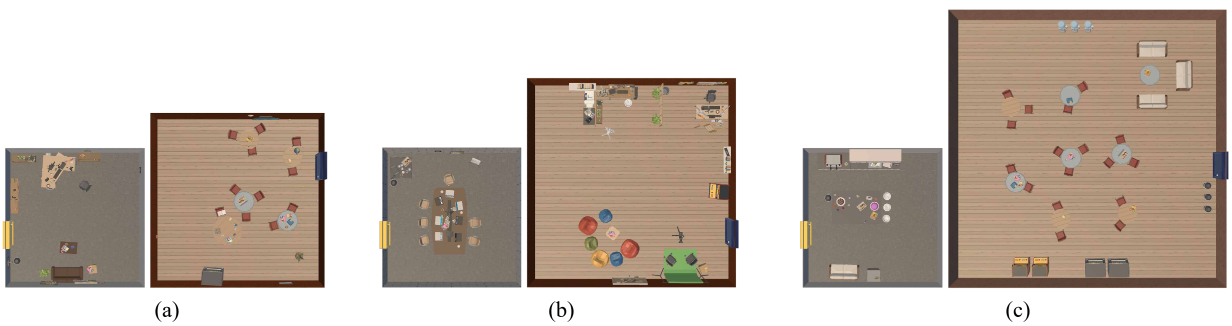


Fig. 5. Illustration of representative VEs used in Study 1. Figures (a), (b), (c) represent the top views of room layouts with sizes of $\langle 8 \times 8\,m^2, 10 \times 10\,m^2 \rangle$, $\langle 8 \times 8\,m^2, 12 \times 12\,m^2 \rangle$, and $\langle 8 \times 8\,m^2, 16 \times 16\,m^2 \rangle$ respectively.

$8\,m^2, 10 \times 10\,m^2 \rangle$, $\langle 8 \times 8\,m^2, 12 \times 12\,m^2 \rangle$, and $\langle 8 \times 8\,m^2, 16 \times 16\,m^2 \rangle$ (see Figure 5 and supplementary material). For each pair of virtual rooms, we specified 9 candidate virtual props along with corresponding target positions to perform the pick-and-place task. We used the Latin Square approach and randomly selected 3 virtual props for a task trial.

Hypotheses. We made four hypotheses:

H1(a): The efficiency of *SF* is the highest. This is inferred from the evidence that *SF* offers the user a broader view of the environment in the other room, pre-cueing virtual props and target positions. Higher efficiency can be reflected by shorter average task completion time and walking distance.

H1(b): The workload for *SF* is the lowest. With *SF*, users spend less effort searching for props or target positions and traveling between rooms.

H1(c): The presence with *SF* is the highest. This is due to the increased visual cues provided by fusing two separate rooms, which reduces confusion for the user.

H1(d): The preference for using *SF* is the highest. This is because *SF* is easy to understand and use, and the fused environment enhances participants' ability to perceive visual information about their surroundings.

Apparatus and Participants. The experiment was carried out in a laboratory having a physical tracking area of $6m \times 6m$. The participant wore an HTC Vive Pro Eye HMD, which had a combined resolution of 2880×1600 (110° FoV) at a 90 Hz refresh rate. Two hand-held controllers were used to interact with the VEs, and two Vive trackers were attached to the participant's feet to capture the motion of walking-in-place. The system was implemented on a PC equipped with an Intel Core i7 processor, 32GB RAM, and a GeForce RTX 2080Ti GPU using Unity3D (v2019.4). The virtual scene was created by using the public assets of *Room Building Starter Kit* and *POLYGON Office*.

We recruited 20 participants (12 males, 8 females) with an age range of 21 to 27 years ($mean \pm std = 23.85 \pm 1.35$)

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3271709

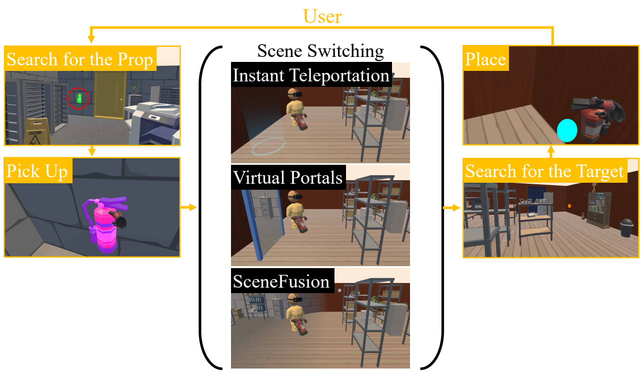


Fig. 6. Task execution process of Study 1. The user was required to search for the virtual prop which would appear in a random room, and deliver the prop to its target position in the other room.

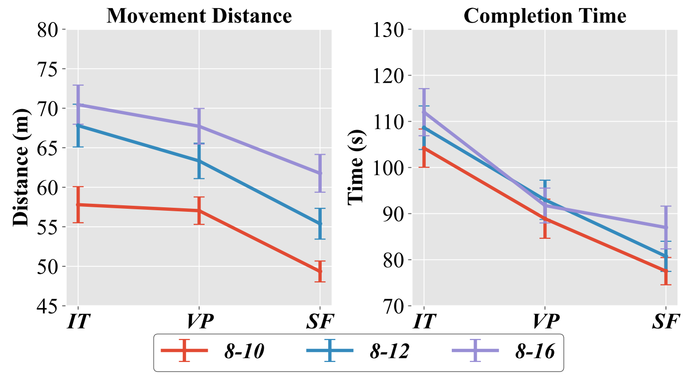


Fig. 7. The figure shows the movement distance and completion time of different techniques in various environments. The standard errors are represented by the error bars. The lower bounds of the figure do not start at 0 intentionally to improve its legibility.

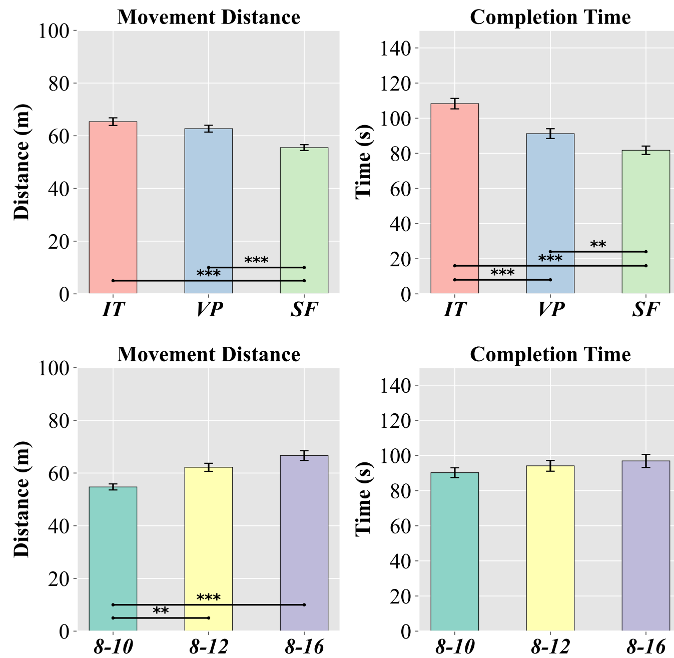


Fig. 8. The main effect of TECHNIQUE and ENVIRONMENT on movement distance and completion time. Significant differences are marked with $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. Error bars indicate the standard errors.

from our university for this study. All participants had normal or corrected-to-normal vision, and only two had no prior experience with VR.

Task and Procedure. When participants arrived at the laboratory, they were asked to sign an IRB consent form and complete a demographic questionnaire. Next, they received instructions on how to operate the VR devices and perform walking-in-place correctly. Following this, participants were given a training phase consisting of 3 trials with the *IT*, *VP*, and *SF* techniques used in a fixed order, following the principle of increasing complexity to help participants better understand the experiment's goal. Each trial required participants to complete 3 rounds of the pick-and-place task: finding a flashing virtual prop, picking it up by holding down the controller trigger, locating the flashing ball target position in the other room, and placing the virtual prop onto the target position (see Figure 6). When a virtual prop was correctly placed, another one then started flashing, until all of the 3 props in a trial were successfully found and placed.

In the testing phase, each participant was asked to complete 3 (techniques) $\times$ 6 (environments) formal trials, with the wall and floor textures of the rooms being randomly selected for each participant. The trials were organized into 6 environment groups, and each group included 3 trials utilizing different techniques. The order of both environments and techniques was randomized and counterbalanced using the Latin Square approach. Prior to each trial, participants were asked to complete a Pre-SSQ questionnaire. Completion time and walking distance were recorded for each trial, with movement being counted only when the participant was walking in place. Teleportation distance was not considered in the statistical analysis to measure the effort required to complete the task. Following each trial, participants completed Post-SSQ [100], Raw TLX [101], IPQ [102], and preference questionnaires. The entire experiment lasted approximately 90 minutes and concluded with an interview to gather feedback on the different techniques.

6.2 Results

We collected 360 valid trials (20 participants $\times$ 3 techniques $\times$ 6 environments). The environments were categorized into 3 types based on their size: $\langle 8 \times 8\,m^2, 10 \times 10\,m^2 \rangle$ (denoted as *8-10*), $\langle 8 \times 8\,m^2, 12 \times 12\,m^2 \rangle$ (denoted as *8-12*), and $\langle 8 \times 8\,m^2, 16 \times 16\,m^2 \rangle$ (denoted as *8-16*). Kolmogorov-Smirnov tests confirmed normal distribution of the movement distance and task completion time data. Accordingly, we performed repeated measures ANOVA (RM-ANOVA) tests ($\alpha = 0.05$) and Bonferroni-adjusted post-hoc tests. In case of violation of the sphericity assumption, Greenhouse-Geisser corrections were applied. The means and standard errors of movement distance and completion time of each technique in different environments are illustrated in Figure 7. Furthermore, the main effects of TECHNIQUE and ENVIRONMENT on movement distance and completion time are depicted in Figure 8. To analyze the non-normally distributed data, we employed Non-parametric Friedman tests. Post-hoc analysis was conducted using Wilcoxon signed-rank tests ($\alpha = 0.05$) with Bonferroni correction, if significant effects were observed. We solely focus on the main

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3271709

and interaction effects related to the factor TECHNIQUE, as our objective is to evaluate the performance of different techniques under different environment setups.

Movement Distance. No interaction effect between TECHNIQUE × ENVIRONMENT was found ($F(4,156) = .486, p = .746, \eta^2 = .012$). A significant main effect of TECHNIQUE on the movement distance ($F(2,78) = 17.835, p < .001, \eta^2 = .314$) was found. *SF* had a significantly shorter movement distance than *IT* ($-9.852\,m, p < .001$) and *VP* ($-7.199\,m, p < .001$).

Completion Time. There was no interaction effect between TECHNIQUE × ENVIRONMENT ($F(4,156) = .415, p = .798, \eta^2 = .011$). A significant main effect of TECHNIQUE on the completion time was found ($F(2,78) = 50.224, p < .001, \eta^2 = .563$). Participants using *SF* spent significantly less time than those using *IT* ($-26.533\,s, p < .001$) and *VP* ($-9.467\,s, p = .001$). And *VP* is significantly less time-consuming than *IT* ($-17.066\,s, p < .001$).

User Preference. We conducted a non-parametric Friedman test to analyze the data from 7-Likert questionnaires of user preference on TECHNIQUE. The results revealed that TECHNIQUE had significant effects on the user preference ($\chi^2(2) = 21.443, p < .001$). *SF* had significantly higher scores than *IT* ($+1.800, p < .001$) and *VP* ($+1.500, p = .001$). Besides, there was no significant difference found between *IT* and *VP* ($-.300, p = .385$).

Workload. We conducted non-parametric Friedman tests under all measures of Raw-TLX, because Kolmogorov-Smirnov tests revealed that the data was not normally distributed. It showed TECHNIQUE-ENVIRONMENT had significant effects on all measures ($\chi^2(8) = 67.116, p < .001$ for mental, $\chi^2(8) = 41.449, p < .001$ for physical, $\chi^2(8) = 68.689, p < .001$ for temporal, $\chi^2(8) = 72.493, p < .001$ for effort, $\chi^2(8) = 61.141, p < .001$ for performance, $\chi^2(8) = 86.072, p < .001$ for frustration, $\chi^2(8) = 80.047, p < .001$ for overall). Figure 9 shows means and standard errors of Raw TLX questionnaire scores on all measures. On the measures of physical and effort physical, effort and overall, *SF-8-10*, *SF-8-12* and *SF-8-16* got significantly lower scores than *IT-8-10* ($-11.825, p = .002$ for physical, $-14.775, p < .001$ for effort, $-12.999, p < .001$ for overall), *IT-8-12* ($-13.100, p < .001$ for physical, $-15.200, p < .001$ for effort, $-13.817, p < .001$ for overall) and *IT-8-16* ($-10.650, p < .001$ for physical, $-11.725, p < .001$ for effort, $-11.758, p < .001$ for overall) respectively. In addition, the effort and overall scores of *TF-8-10* were significantly lower than that of *VP-8-10* ($-11.200, p < .001$ for effort, $-9.225, p = .001$ for overall).

No significant effects were found in the post-hoc tests between techniques on *Presence* or *Simulator Sickness*.

6.3 Interview Findings

Open comments were collected from interviews to analyze the effect of TECHNIQUE. The findings can be summarized in the following aspects:

Advantages and Disadvantages of SceneFusion. Most participants (16/20) thought SceneFusion was the most helpful technique for traveling between the rooms. Some of them believed that it helped them gain broader views while finishing the tasks. P5 said: *"The advantage of SceneFusion is that the field of vision is wide. Because the objects to be picked up randomly appeared in these rooms, you can find the target just by turning around, which greatly reduced the time to search for the target."* P7 commented: *"After fusing rooms, the vision was wider, which made it easier to find items directly."* The comments provide evidence that our SceneFusion technique is effective in addressing the occlusion issue between rooms and can enhance the user's exploration experience by providing additional visual cues.

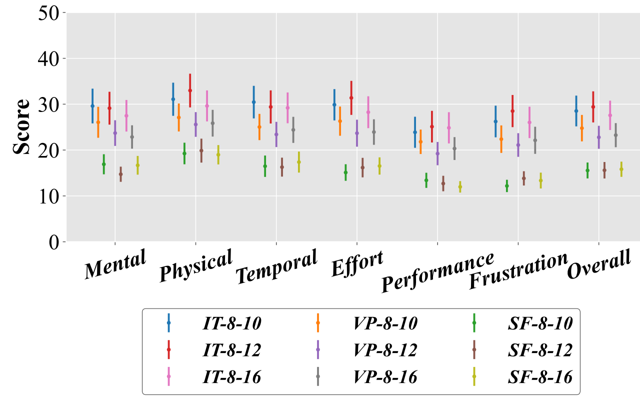


Fig. 9. Means and standard errors of Raw TLX questionnaire scores on all measures in Study 1.

In addition, others thought SceneFusion was more natural and easier to use. P4: *" SceneFusion reduced my memory burden, visually presented all information related to the task, and allowed me to complete the task in the shortest time with the least operations"*. P6: *"Because this method is the most natural. We completed the tasks in a complete scene, and there was no need to switch scenes frequently in the process"* The received comments substantiate that the fused environment can heighten context awareness, provide a seamless and natural exploration experience, and reduce the amount of unnecessary walking.

However, 4 participants held the opinion that SceneFusion made scenes too large. P12 and P17 commented: *"The scenes were larger after being fused, so it was not easy to explore."* It indicates that the provision of visual cues in the fused room may impede the users' ability to concentrate.

Comments on IT and VP. Some participants preferred the virtual portal technique due to its intuitive usability and lower operational demands. P13: *"Using the virtual portal technique, you can preview the layout of the other room, which was more natural."* P9, P18:*"You do not need to divide rooms or press to teleport with the virtual portal."*

Others preferred instant teleportation because it kept the original scenes: P7: *"Instant teleportation did not modify the scene layout, so users can fully explore the original scenes."* P1: *"Instant teleportation was similar to the transmission in video games, so it is interesting to use."* These opinions indicate that certain users did not place significant importance on the continuity of the travel process, instead prioritizing methods that preserve the original room layout during exploration.

6.4 Discussion

The results of the study showed that SceneFusion outperformed instant teleportation ($-9.852\,m, -26.533\,s$) and the virtual portal technique ($-7.199\,m, -9.467\,s$) in terms of

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3271709

movement distance and time cost, and the differences were significant at a 5% level of significance, providing support for **H1(a)**. This could be attributed to SceneFusion's larger accessible area, which enables participants to move across the fusion area with more freedom. In contrast, instant teleportation and virtual portal require participants to reach or pass through specific locations to teleport, leading to a longer distance. Furthermore, SceneFusion provides more visual context, allowing participants to plan their path more efficiently and reduce unnecessary exploration.

Results from the Raw TLX questionnaires indicated that SceneFusion had the lowest scores in all measures. In the interview findings, some participants (P5, P7) reported that SceneFusion provided a stronger sense of the room layout simultaneously (wider vision, more visual information, etc.), which was helpful for quickly finding the virtual prop and target position. Furthermore, some participants (P4, P9, P18) reported that when using instant teleportation, they had to adjust their standing position and press the controller grip button, as they were unable to teleport themselves if out of the circle region. This additional adjustment increased the required time cost and workload. Thus, **H1(b)** was supported. Despite the majority of positive comments, some participants preferred to explore separate rooms rather than the fused environment, as they believed that the larger fused environment made it more difficult to concentrate.

The preference questionnaires showed that our SceneFusion technique received significantly higher scores compared to instant teleportation $(+1.800, p < .001)$ and virtual portal $(+1.500, p = .001)$. Additionally, the majority of participants (16/20) found SceneFusion to be the most useful technique. We believe that the fused environment provided participants with an easier way to understand the layout of each individual room, while also reducing the disruption caused by the teleportation process. This allowed participants to focus more on their tasks and resulted in a more immersive experience. Therefore, **H1(d)** was confirmed. The IPQ scores surprisingly revealed no significant difference in presence between techniques. This result may be attributed to the fact that participants evaluated presence based on factors such as lighting, materials, and layouts of the environments, where TECHNIQUE had only a minor impact. Therefore, rejecting **H1(c)** was justifiable.

7 STUDY 2: MULTI-USER COLLABORATION

Study 1 showed that the SceneFusion technique was more efficient, labor-saving, and preferred than alternatives in a single-user task. In Study 2, we aimed to evaluate the performance of different techniques in a virtual collaboration task. To establish a connection between individual rooms before executing the pick-and-place task, participants were allowed to indicate the positions and rotations of each method preset. As there was no interaction effect between ENVIRONMENT and TECHNIQUE found in Study 1, we reused the virtual rooms of size $\langle 8 \times 8\, m^2, 10 \times 10\, m^2 \rangle$. To provide more practical virtual environments for interactive technique presetting, we added other virtual furniture to fill in the original blanks, as shown in Figure 10.

(a) (b)

Fig. 10. Room layouts used for test trials of Study 2. (a) and (b) show the top views of the two rooms respectively.

7.1 Collaboration Study Design

In this study, we evaluated the collaborative performance of different techniques in virtual environments where two participants, denoted as *User A* and *User B*, in separate virtual rooms, denoted as *Room A* and *Room B*, were asked to complete a pick-and-place task as quickly as possible. Initially, virtual props were placed in *Room A*, and *User A* was asked to find, pick up, and deliver them one by one to *User B*. Subsequently, *User B* had to find target positions in *Room B* for placement of the delivered props in succession. Participants were allowed to walk and meet anywhere in the accessible area. Before executing the pick-and-place task, participants were required to use their controllers to specify scene connections (teleportation points, virtual portals, or dividing lines) in individual rooms.

Hypotheses. In this study, we formulated four hypotheses similar to Study 1:

H2(a): *SF* is the most efficient technique compared to *IT* and *VP*, as it offers a larger accessible area for traveling between rooms, which enhances partner, virtual prop, and target position awareness. Efficiency was measured using the average completion time and average walking distance.

H2(b): The workload associated with *SF* is the smallest, as it enables participants to expend less effort in searching for props or target positions and reduces the workload of teleportation between rooms.

H2(c): The presence with *SF* is the highest, as it provides more visual cues by fusing two separate rooms, making users more aware of room layouts and partner activities.

H2(d): Participants' preference for *SF* is the highest because it is easy to use and understandable, and the fused environment enables them to perceive visual information about their surroundings more easily, improving their spatial understanding.

Apparatus. The experimental setup in Study 2 was similar to that of Study 1, with the addition of an extra VR device for the second participant. To separate the two users, a screen was placed in the middle of the tracking space. As the users were in the same physical space, they were able to converse without requiring any remote devices.

Participants. This collaborative study recruited 12 pairs of participants, comprising 11 males and 13 females aged between 22 to 32 years $(mean \pm std = 25.13 \pm 2.13)$ from our university. The distribution of the pairs was as follows:

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3271709

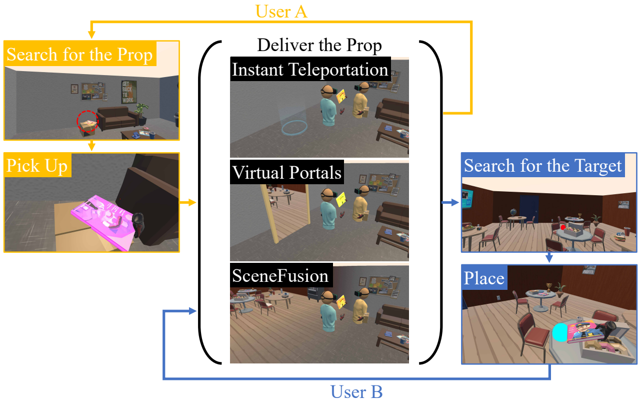


Fig. 11. Task execution process of the multi-user collaboration study. After building the connection between rooms, *User A* was required to search for the virtual prop in *Room A* and deliver the item to *User B*, while *User B* was required to receive the item from *User A* and place it in the correct target position.

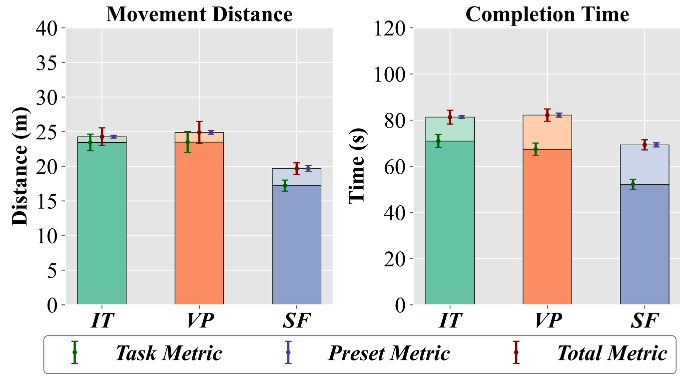


Fig. 12. The means and standard errors of the movement distance and completion time in Study 2. In the bars, darker colors represent task movement distances or task completion times, and lighter colors represent preset movement distances or preset completion times. The whole bar represents the total movement distance or total completion time. Standard errors are marked in various colors corresponding to the types of metrics. Error bars in different colors mean standard errors of different kinds of metrics.

4 female-only, 3 male-only, and 5 mixed-gender pairs. All participants had normal or corrected-to-normal vision and previous experience with virtual reality (VR). The familiarity between teammates was assessed on a 7-point Likert scale ranging from 1 (never met before) to 7 (best friends or lovers), yielding $mean \pm std = 4.54 \pm 2.67$.

Task and Procedure. Upon arrival at the laboratory in pairs, participants were informed of the experimental objective and asked to sign an IRB consent form. They were instructed to complete a demographic questionnaire and subsequently received training on operating the VR devices and utilizing walking-in-place locomotion. The training comprised 6 trials conducted in the order of $\langle IT, VP, SF, IT, VP, SF \rangle$, with the intent of facilitating participant familiarity with the techniques. For the first 3 trials, one participant acted as *User A*, and the other acted as *User B*, and their roles were exchanged for the latter 3 trials. At the beginning of each trial, participants had to establish a connection between the rooms using the given method for collaboration, and this connection was not allowed to be changed during task execution. For *IT*, either controller could be used to set the position of the virtual point, which was fixed upon release of the trigger button. For *VP* and *SF*, either controller could be used to set the position of the virtual portal and dividing line, with wrist rotation used to control their orientation. All 3 source props and target destinations were highlighted until the connection was established to assist participants in making better decisions. During each trial, the pairs of participants worked collaboratively to complete 3 rounds of the pick-and-place task: *User A* searched in *Room A* for the highlighted virtual prop, picked it up, and delivered it to *User B*, who then searched in *Room B* for the highlighted target position and placed the virtual prop onto it (Figure 11). A trial was considered complete after 3 successful pick-and-place rounds. The goal was to complete the tasks as quickly as possible.

In the testing phase, each pair of participants completed 3 trials of the pick-and-place task, with a total of 6 trials (3 techniques × 2 roles). The wall and floor textures were randomly assigned to each pair of participants. After finishing the first 3 trials with the techniques in a random order, the teammates switched roles to finish the next 3 trials with the techniques in a different random order. Prior to each trial, each participant completed a Pre-SSQ questionnaire. Completion time, walking distance, and task execution were recorded per trial. Post-SSQ, Raw TLX, and IPQ questionnaires were completed by each participant after each trial. A 7-point Likert scale questionnaire was completed by each participant after finishing all trials to determine preferences. Participants also provided feedback on the techniques through an interview. The experiment lasted approximately 50 minutes.

7.2 Results

72 valid trials (12 participant pairs × 3 techniques × 2 roles) were collected. Normal distribution was confirmed for movement distance and task completion time via Kolmogorov-Smirnov tests, leading to the use of RM-ANOVA tests with Bonferroni-adjusted post-hoc testing. Greenhouse-Geisser corrections were applied when sphericity assumptions were violated. Non-parametric Friedman tests were employed for other non-normally distributed data. Wilcoxon signed-rank tests with Bonferroni correction were conducted for post-hoc analysis upon discovering significant effects ($\alpha = 0.05$). Figure 12 shows the means and standard errors of the movement distance and completion time. Table 1 shows the main effect of TECHNIQUE and the results of post-hoc tests on movement distance and completion time.

Movement Distance. There was no interaction effect of TECHNIQUE × ROLE on the preset movement distance ($F(2, 46) = .060, p = .942, \eta^2 = .003$), the task movement distance ($F(2, 46) = .327, p = .723, \eta^2 = .014$) and the total movement distance ($F(1.577, 36.262) = .251, p = .726, \eta^2 = .011$). We found significant main effects of TECHNIQUE on the preset distance ($F(1.394, 32.063) = 9.434, p = .002, \eta^2 = .291$), the task distance ($F(1.515, 34.851) = 15.403, p < .001, \eta^2 = .401$) and the total distance ($F(1.415, 32.534) = 9.475, p < .001, \eta^2 = .292$). ROLE had no significant main effect on the metrics.

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3271709

TABLE 1
In Study 2, we compared the performance of different methods on movement distance and completion time. The results are presented using different colors, where each color represents a specific method. The color of the value indicates which method performs better under that metric. Significant differences are denoted by asterisks: $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$.

Metric	SF vs. IT	SF vs. VP	IT vs. VP
Preset Distance	1.641m***	1.053m	-0.581m
Task Distance	-6.238m***	-6.288m***	-0.050m
Total Distance	-4.597m***	-5.236m**	-0.639m
Preset Time	6.712s***	2.308s*	-4.404s***
Task Time	-18.733s***	-15.205s***	3.528s
Total Time	-12.022s**	-12.897s***	-0.876s

Completion Time. No interaction effect of TECHNIQUE × ROLE was found on the preset completion time ($F(2, 46) = .871, p = .425, \eta^2 = .037$), the task completion time ($F(1.405, 32.311) = .042, p = .959, \eta^2 = .002$) and the total completion time ($F(1.380, 31.729) = .000, p = .998, \eta^2 = .000$). It was revealed that TECHNIQUE had a significant main effect on the preset time ($F(2, 46) = 28.127, p < .001, \eta^2 = .550$), the task time ($F(1.295, 29.795) = 20.900, p < .001, \eta^2 = .476$) and the total time ($F(1.205, 27.709) = 10.490, p = .002, \eta^2 = .313$). There was no significant main effect of ROLE on the metrics.

User Preference. The non-parametric Friedman tests were used to analyze the data from self-made 7-Likert questionnaires of preference. Tests showed that TECHNIQUE had significant main effects on preference ($\chi^2(2) = 34.135, p < .001$). The post-hoc tests revealed that *SF* had significantly higher scores than *IT* ($+1.417, p < .001$) and *VP* ($+1.392, p < .001$). Although *VP* got higher scores than *IT*, there was no significant difference between them ($+.021, p = .942$).

Workload. Non-parametric Friedman tests under all metrics of Raw-TLX showed that TECHNIQUE-ROLE had significant effects on the physical ($\chi^2(5) = 25.882, p < .001$), temporal ($\chi^2(5) = 27.854, p < .001$), effort ($\chi^2(5) = 20.620, p = .001$), performance ($\chi^2(5) = 18.238, p = .003$), frustration ($\chi^2(5) = 22.696, p < .001$) overall ($\chi^2(5) = 21.915, p = .001$). Figure 13 shows means and standard errors of Raw TLX questionnaire scores on all measures.

No significant effects were found in the post-hoc tests between techniques on *Presence* and *Simulator Sickness*.

7.3 Interview Findings

In the gathered open comments, SceneFusion was found to be the preferred choice of almost all participants (20 out of 24) for enhancing collaboration and communication in virtual scenes. P6 commented: *"SceneFusion facilitated communication and collaboration between users. In the fused room, the users can observe each other's positions, making it easier to hand over objects and plan walking routes."* P23 said: *"SceneFusion gave the user a larger view, reducing unnecessary movement and facilitating collaborative work."* The comments indicate that the wide fusion area of SceneFusion provides users with rich visual information from both rooms.

Furthermore, some users found SceneFusion an intuitive tool that minimized unnecessary travel. P3 said: *"SceneFusion can adjust the size of the space to the user's needs and remove the space that is not needed. This reduced the distance the user has to travel during the tasks."* P20 said: *"Tasks with SceneFusion proceeded naturally as if they were in the same room."* These comments showed that users found that our SceneFusion technique can remove redundant parts of the space and make the movement more efficient. The fused environment also makes them feel more intuitive and natural.

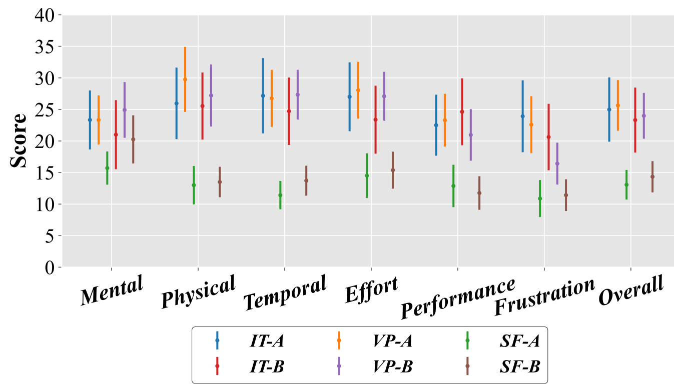


Fig. 13. Means and standard errors of Raw TLX questionnaire scores on all measures in Study 2.

However, some participants thought that SceneFusion decreased the efficiency of completing the tasks. P14 argued: *"Setting the dividing lines in the SceneFusion costs too much time, making the manipulation less efficient."* Although the manipulation of the SceneFusion consumed more time than the others, objective data showed that it can save much more time during the tasks. P17 said: *"SceneFusion changed the layouts of the rooms, which might confuse users in some complex rooms."* The comment indicates that layout changes made by SceneFusion may hinder the efficiency of task completion. Hence, it is imperative to retain as many visual features of the original rooms as possible during room fusion.

A few users preferred the virtual portal or instant teleportation. P9 explained: *"The virtual portal, which is similar to the doors in the real life, made me feel more immersive like I'm in the real world."* P15 said: *"I think instant teleportation was very easy and intuitive to use."* User feedback revealed that although SceneFusion benefited virtual collaboration and communication, it might decrease the immersion level.

7.4 Discussion

Three hypotheses, namely **H2(a)**, **H2(b)**, and **H2(d)**, were fully supported by our experimental results, while **H2(c)** was rejected, consistent with the results of Study 1. SceneFusion achieved the shortest total movement distance and the least total time cost compared to instant teleportation ($-4.597\,m$, $-12.022\,s$) and the virtual portal technique ($-5.236\,m$, $-12.897\,s$), despite having the longest preset movement distance and largest preset time cost. This finding indicates that although the preset operation of SceneFusion is more complex, it can significantly save task completion time and movement distance. The large accessible area around the fusion area provided by SceneFusion enables participants to have more freedom to choose the location for virtual prop delivery, while other techniques require both participants to stay in the same room. Additionally,

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3271709

SceneFusion can remove unnecessary space according to user needs, which leads to less travel for users.

SceneFusion was found to reduce the workload on all measures, as indicated by the Raw TLX results. In addition to shorter walking distances and more continuous locomotion, SceneFusion's richer visual context enables collaborators to locate one another and communicate face-to-face. In contrast, with instant teleportation, collaborators cannot see each other when they are in separate rooms, which sometimes leads to missed connections and additional teleportations. With the virtual portal technique, collaborators must decide who goes through the portal when they arrive at the same time, leading to hesitation or even embarrassment.

Based on the results and user feedback, SceneFusion was the most preferred technique due to its ability to remove wall occlusions and seamlessly blend two separate rooms, enabling collaborators to interact more freely and conveniently. However, some participants disliked SceneFusion due to its perceived unreality and complexity of use, which may be related to task types and user personalities.

Presence was not significantly affected by the technique used, consistent with the findings of Study 1.

8 LIMITATIONS

There are also several limitations in our work. First, we designed SceneFusion based on 2D planar maps and only tested its performance in virtual rooms with conventional rectangle shapes, which are the prevailing geometry for indoor rooms. Nevertheless, several indoor rooms can be complex and possess a diverse range of geometry, such as curved walls, circular corridors, and duplex rooms. Therefore, further exploration is required for extending SceneFusion to accommodate rooms with atypical and complex geometry. Besides, it is also imperative to investigate how to enhance the method's performance in outdoor environments.

Another limitation is that we reserved obvious blank areas in each room for generating dividing lines. The purpose was to retain all interactive objects in the fused room. However, the setting of these lines must be appropriate, otherwise, major interactive objects could be split out, or the fusion area may be too narrow to pass through, thus reducing subsequent VR exploration or collaboration efficiency. One possible solution is to automatically adjust the dividing lines based on user operations and allow users to specify the important props in advance. If some important props are considered to be discarded, the system can rearrange them to proper positions in the new shared space. Further research needs to be conducted to explore related approaches and methods.

Moreover, optimization of the fusion area width was not the primary objective of our work. In the pilot study, we only assessed subjective feelings of four discrete fusion area widths in two typical individual rooms, and the results revealed a tendency that users prefer longer widths. One possible solution is to use the whole length of the wall to create the fusion area, which can also handle rooms with shallow depths. But for rooms with large size differences, this strategy can profoundly alter the geometries of the original rooms and considerably affect visual plausibility. Therefore, the fusion area width should be restricted within a certain range that may be influenced by the size, geometry, or structure of individual rooms, which warrants further investigation.

Furthermore, our user studies focused only on evaluating the performance of SceneFusion with one or two users. Theoretically, SceneFusion can facilitate the collaboration of more than two users by iteratively fusing the new shared space with another individual room one at a time. However, the efficient completion of this process requires further exploration in future research.

9 CONCLUSION AND FUTURE WORK

In this paper, we propose SceneFusion, a novel technique for traveling between scenes that enables both single-user exploration and multi-user collaboration in individual VEs. SceneFusion is designed to optimize efficiency, visual plausibility, and versatility.

With SceneFusion, VR users can merge virtual rooms that were originally separated into a cohesive environment by blending the geometric structure and visual textures in the fusion area, providing a seamless visual context during active locomotion and communication. In contrast to commonly used teleportation techniques such as instant teleportation and virtual portal technique, SceneFusion is more efficient, as it reduces task completion time, walking distance, and human workload. Additionally, SceneFusion is user-friendly and aids in understanding the context of both environments and the activities of other collaborators.

Our SceneFusion technique can be particularly valuable when users require a connection between individual rooms, and most of the visual information of the rooms needs to be preserved. Apart from the single- or multi-user pick-and-place tasks conducted in our experiment, SceneFusion has the potential to benefit other VR applications, such as VR interior design. By fusing two separate virtual rooms, the layouts and decoration details of both rooms can be visualized, and users can immerse themselves in the fused environments, discuss and compare differences in interior designs, and make interactive edits. For educational, training, or entertainment purposes, this technique can be applied multiple times to fuse areas of interest from more virtual rooms by defining appropriate dividing lines.

In future work, we aim to enhance the SceneFusion technique by proposing automated techniques for determining dividing lines to reduce human labor and optimize visual effects. Currently, we have conducted extensive user experiments on single- and double-user pick-and-place tasks. More complicated studies with grouped users in other kinds of VR tasks are regarded as future works. Furthermore, exploring the adaptation of this novel technique to open-world VEs and asynchronous collaboration tasks is also worth investigating.

ACKNOWLEDGMENTS

This work was directly supported by the National Natural Science Foundation of China (Project Number: 61932003) and the Fundamental Research Funds for the Central Universities. Frank Steinicke was supported by funds from the BMBF, BMWi, DFG, and EU.

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3271709

REFERENCES

[1] L. Freina and M. Ott, "A literature review on immersive virtual reality in education: state of the art and perspectives," in *The international scientific conference elearning and software for education*, vol. 1, no. 133, 2015, pp. 10–1007.

[2] M. Slater and M. V. Sanchez-Vives, "Enhancing our lives with immersive virtual reality," *Frontiers in Robotics and AI*, vol. 3, p. 74, 2016.

[3] M. Wang, X.-Q. Lyu, Y.-J. Li, and F.-L. Zhang, "Vr content creation and exploration with deep learning: A survey," *Computational Visual Media*, vol. 6, pp. 3–28, 2020.

[4] K. Rahimi, C. Banigan, and E. D. Ragan, "Scene transitions and teleportation in virtual reality and the implications for spatial awareness and sickness," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 6, pp. 2273–2287, 2018.

[5] K. R. Moghadam and E. D. Ragan, "Towards understanding scene transition techniques in immersive 360 movies and cinematic experiences," in *2017 IEEE Virtual Reality (VR)*. IEEE, 2017, pp. 375–376.

[6] M. Husung and E. Langbehn, "Of portals and orbs: An evaluation of scene transition techniques for virtual reality," in *Proceedings of Mensch und Computer 2019*, 2019, pp. 245–254.

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

[8] L. Men, N. Bryan-Kinns, A. S. Hassard, and Z. Ma, "The impact of transitions on user experience in virtual reality," in *2017 IEEE virtual reality (VR)*. IEEE, 2017, pp. 285–286.

[9] A. MacQuarrie and A. Steed, "The effect of transition type in multi-view 360 media," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 4, pp. 1564–1573, 2018.

[10] J. Young, T. Langlotz, M. Cook, S. Mills, and H. Regenbrecht, "Immersive telepresence and remote collaboration using mobile and wearable devices," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 5, pp. 1908–1918, 2019.

[11] S. Thanyadit, P. Punpongsanon, T. Piumsomboon, and T.-C. Pong, "Substituting teleportation visualization for collaborative virtual environments," in *Symposium on Spatial User Interaction*, 2020, pp. 1–2.

[12] T. Rhee, S. Thompson, D. Medeiros, R. Dos Anjos, and A. Chalmers, "Augmented virtual teleportation for high-fidelity telecollaboration," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 5, pp. 1923–1933, 2020.

[13] K. Kiyokawa and H. Takemura, "A tunnel window and its variations: Seamless teleportation techniques in a virtual environment," in *HCI International*. Citeseer, 2005.

[14] G. Bruder, F. Steinicke, and K. H. Hinrichs, "Arch-explore: A natural user interface for immersive architectural walkthroughs," in *2009 IEEE Symposium on 3D User Interfaces*. IEEE, 2009, pp. 75–82.

[15] S. Freitag, D. Rausch, and T. Kuhlen, "Reorientation in virtual environments using interactive portals," in *2014 IEEE symposium on 3D user interfaces (3DUI)*. IEEE, 2014, pp. 119–122.

[16] P.-H. Han, C.-E. Hsieh, Y.-S. Chen, J.-C. Hsiao, K.-C. Lee, S.-F. Ko, K.-W. Chen, C.-H. Chou, and Y.-P. Hung, "Aoes: enhancing teleportation experience in immersive environment with mid-air haptics," in *ACM SIGGRAPH 2017 Emerging Technologies*, 2017, pp. 1–2.

[17] J. Liu, H. Parekh, M. Al-Zayer, and E. Folmer, "Increasing walking in vr using redirected teleportation," in *Proceedings of the 31st annual ACM symposium on user interface software and technology*, 2018, pp. 521–529.

[18] R. Soret, A.-M. Montes-Solano, C. Manzini, V. Peysakhovich, and E. F. Fabre, "Pushing open the door to reality: On facilitating the transitions from virtual to real environments," *Applied Ergonomics*, vol. 97, p. 103535, 2021.

[19] L. Wang, Y. Liu, X. Liu, and J. Wu, "Automatic virtual portals placement for efficient vr navigation," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 2022, pp. 628–629.

[20] V. Games, "Pavlov vr," https://store.steampowered.com/app/555160/Pavlov_VR/, accessed: 2021-09-03.

[21] V. Inc, "Vrchat," https://hello.vrchat.com/, accessed: 2021-09-03.

[22] E. F. Churchill and D. Snowdon, "Collaborative virtual environments: an introductory review of issues and systems," *virtual reality*, vol. 3, no. 1, pp. 3–15, 1998.

[23] C. Li, H. Huang, J.-M. Lien, and L.-F. Yu, "Synthesizing scene-aware virtual reality teleport graphs," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 6, pp. 1–15, 2021.

[24] D. Zeltzer, "Autonomy, interaction, and presence," *Presence: Teleoperators & Virtual Environments*, vol. 1, no. 1, pp. 127–132, 1992.

[25] M. Sereno, X. Wang, L. Besançon, M. J. McGuffin, and T. Isenberg, "Collaborative work in augmented reality: A survey," *IEEE Transactions on Visualization and Computer Graphics*, 2020.

[26] T. Piumsomboon, Y. Lee, G. Lee, and M. Billinghurst, "Covar: a collaborative virtual and augmented reality system for remote collaboration," in *SIGGRAPH Asia 2017 Emerging Technologies*, 2017, pp. 1–2.

[27] B. Yoon, H.-i. Kim, G. A. Lee, M. Billinghurst, and W. Woo, "The effect of avatar appearance on social presence in an augmented reality remote collaboration," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 547–556.

[28] A. Kunert, T. Weissker, B. Froehlich, and A. Kulik, "Multi-window 3d interaction for collaborative virtual reality," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 11, pp. 3271–3284, 2019.

[29] T. Teo, L. Lawrence, G. A. Lee, M. Billinghurst, and M. Adcock, "Mixed reality remote collaboration combining 360 video and 3d reconstruction," in *Proceedings of the 2019 CHI conference on human factors in computing systems*, 2019, pp. 1–14.

[30] T. Weissker, P. Bimberg, and B. Froehlich, "Getting there together: Group navigation in distributed virtual environments," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 5, pp. 1860–1870, 2020.

[31] T. Weissker and B. Froehlich, "Group navigation for guided tours in distributed virtual environments," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 5, pp. 2524–2534, 2021.

[32] L. Wang, X. Liu, and X. Li, "Vr collaborative object manipulation based on viewpoint quality," in *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2021, pp. 60–68.

[33] S.-Z. Xu, J.-H. Liu, M. Wang, F.-L. Zhang, and S.-H. Zhang, "Multi-user redirected walking in separate physical spaces for online vr scenarios," *IEEE Transactions on Visualization and Computer Graphics*, vol. Early Access, pp. 1–11, 2023.

[34] A. Irlitti, R. T. Smith, S. Von Itzstein, M. Billinghurst, and B. H. Thomas, "Challenges for asynchronous collaboration in augmented reality," in *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. IEEE, 2016, pp. 31–35.

[35] K. Chow, C. Coyiuto, C. Nguyen, and D. Yoon, "Challenges and design considerations for multimodal asynchronous collaboration in vr," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, pp. 1–24, 2019.

[36] G. A. Lee, S. Ahn, W. Hoff, and M. Billinghurst, "Enhancing first-person view task instruction videos with augmented reality cues," in *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2020, pp. 498–508.

[37] M.-Y. Liao, C.-Y. Sung, H.-C. Wang, and W.-C. Lin, "Virtual classmates: Embodying historical learners' messages as learning companions in a vr classroom through comment mapping," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 163–171.

[38] T. Weissker, A. Kulik, and B. Froehlich, "Multi-ray jumping: comprehensible group navigation for collocated users in immersive virtual reality," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 136–144.

[39] J. G. Grandi, H. G. Debarba, and A. Maciel, "Characterizing asymmetric collaborative interactions in virtual and augmented realities," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 127–135.

[40] T. Wells and S. Houben, "Collabar–investigating the mediating role of mobile ar interfaces on co-located group collaboration," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–13.

[41] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst, "A user study on mr remote collaboration using live 360 video," in *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2018, pp. 153–164.

[42] T. Teo, A. F. Hayati, G. A. Lee, M. Billinghurst, and M. Adcock, "A technique for mixed reality remote collaboration using 360

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3271709

panoramas in 3d reconstructed scenes," in *25th ACM Symposium on Virtual Reality Software and Technology*, 2019, pp. 1–11.

[43] T.-W. Hsu, M.-H. Tsai, S. V. Babu, P.-H. Hsu, H.-M. Chang, W.-C. Lin, and J.-H. Chuang, "Design and initial evaluation of a vr based immersive and interactive architectural design discussion system," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2020, pp. 363–371.

[44] Z. He, R. Du, and K. Perlin, "Collabovr: A reconfigurable framework for creative collaboration in virtual reality," in *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2020, pp. 542–554.

[45] H. Bai, P. Sasikumar, J. Yang, and M. Billinghurst, "A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing," in *Proceedings of the 2020 CHI conference on human factors in computing systems*, 2020, pp. 1–13.

[46] M. Agrawala, A. C. Beers, I. McDowall, B. Fröhlich, M. Bolas, and P. Hanrahan, "The two-user responsive workbench: support for collaboration through individual views of a shared space," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, 1997, pp. 327–332.

[47] T. T. H. Nguyen, T. Duval, and C. Fleury, "Guiding techniques for collaborative exploration in multi-scale shared virtual environments," in *GRAPP International Conference on Computer Graphics Theory and Applications*, 2013, pp. 327–336.

[48] J. Li, H. Deng, and P. Michalatos, "Catescape: an asymmetrical multiplatform game connecting virtual, augmented and physical world," in *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play*, 2017, pp. 585–590.

[49] J. Gugenheimer, E. Stemasov, J. Frommel, and E. Rukzio, "Sharevr: Enabling co-located experiences for virtual reality between hmd and non-hmd users," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 4021–4033.

[50] Y.-D. Malinov, D. E. Millard, and T. Bloun, "Stuckinspace: Exploring the difference between two different mediums of play in a multi-modal virtual reality game," in *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2021, pp. 501–510.

[51] D. Schott, P. Saalfeld, G. Schmidt, F. Joeres, C. Boedecker, F. Huettl, H. Lang, T. Huber, B. Preim, and C. Hansen, "A vr/ar environment for multi-user liver anatomy education," in *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2021, pp. 296–305.

[52] P. Stotko, S. Krumpen, M. B. Hullin, M. Weinmann, and R. Klein, "Slamcast: Large-scale, real-time 3d reconstruction and streaming for immersive multi-client live telepresence," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 5, pp. 2102–2112, 2019.

[53] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," *ACM siggraph computer graphics*, vol. 21, no. 4, pp. 163–169, 1987.

[54] Y. Zhang, J. Yang, Z. Liu, R. Wang, G. Chen, X. Tong, and B. Guo, "Virtualcube: An immersive 3d video communication system," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 5, pp. 2146–2156, 2022.

[55] J. Young, T. Langlotz, S. Mills, and H. Regenbrecht, "Mobileportation: Nomadic telepresence for mobile devices," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 2, pp. 1–16, 2020.

[56] J. Hartmann, C. Holz, E. Ofek, and A. D. Wilson, "Realitycheck: Blending virtual environments with situated physical reality," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–12.

[57] L.-P. Cheng, E. Ofek, C. Holz, and A. D. Wilson, "Vroamer: generating on-the-fly vr experiences while walking inside large, unknown real-world building environments," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 359–366.

[58] M. Sra, S. Garrido-Jurado, and P. Maes, "Oasis: Procedurally generated social virtual spaces from 3d scanned real spaces," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 12, pp. 3174–3187, 2017.

[59] B. E. Insko, *Passive haptics significantly enhances virtual environments*. The University of North Carolina at Chapel Hill, 2001.

[60] E. Sayyad, M. Sra, and T. Höllerer, "Walking and teleportation in wide-area virtual reality experiences," in *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2020, pp. 608–617.

[61] C. Boletsis, "The new era of virtual reality locomotion: A systematic literature review of techniques and a proposed typology," *Multimodal Technologies and Interaction*, vol. 1, no. 4, p. 24, 2017.

[62] M. Al Zayer, P. MacNeilage, and E. Folmer, "Virtual locomotion: a survey," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 6, pp. 2315–2334, 2018.

[63] M. Di Luca, H. Seifi, S. Egan, and M. Gonzalez-Franco, "Locomotion vault: the extra mile in analyzing vr locomotion techniques," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–10.

[64] J. D. Wendt, M. C. Whitton, and F. P. Brooks, "Gud wip: Gait-understanding-driven walking-in-place," in *2010 IEEE Virtual Reality Conference (VR)*. IEEE, 2010, pp. 51–58.

[65] N. C. Nilsson, S. Serafin, M. H. Laursen, K. S. Pedersen, E. Sikström, and R. Nordahl, "Tapping-in-place: Increasing the naturalness of immersive walking-in-place locomotion through novel gestural input," in *2013 IEEE symposium on 3D user interfaces (3DUI)*. IEEE, 2013, pp. 31–38.

[66] S. Hanson, R. A. Paris, H. A. Adams, and B. Bodenheimer, "Improving walking in place methods with individualization and deep networks," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 367–376.

[67] S. Razzaque, *Redirected walking*. The University of North Carolina at Chapel Hill, 2005.

[68] F. Steinicke, G. Bruder, J. Jerald, H. Frenz, and M. Lappe, "Estimation of detection thresholds for redirected walking techniques," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 1, pp. 17–27, 2009.

[69] E. Langbehn, P. Lubos, and F. Steinicke, "Evaluation of locomotion techniques for room-scale vr: Joystick, teleportation, and redirected walking," in *Proceedings of the Virtual Reality International Conference-Laval Virtual*, 2018, pp. 1–9.

[70] Y.-J. Li, F. Steinicke, and M. Wang, "A comprehensive review of redirected walking techniques: Taxonomy, methods, and future directions," *Journal of Computer Science and Technology*, vol. 37, no. 3, pp. 561–583, 2022.

[71] A. Prithul, I. B. Adhanom, and E. Folmer, "Teleportation in virtual reality; a mini-review," *Frontiers in Virtual Reality*, p. 138, 2021.

[72] D. A. Bowman, D. Koller, and L. F. Hodges, "Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques," in *Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality*. IEEE, 1997, pp. 45–52.

[73] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst, "Mini-me: An adaptive avatar for mixed reality remote collaboration," in *Proceedings of the 2018 CHI conference on human factors in computing systems*, 2018, pp. 1–13.

[74] B. Bolte, F. Steinicke, and G. Bruder, "The jumper metaphor: an effective navigation technique for immersive display setups," in *Proceedings of Virtual Reality International Conference*, vol. 1, no. 2, 2011.

[75] K. Vasylevska, H. Kaufmann, and V. Khrystyna, "Influence of vertical navigation metaphors on presence," in *Challenging Presence-Proceedings of 15th International Conference on Presence (ISPR 2014)*, 2014, pp. 205–212.

[76] S. Benford, C. Greenhalgh, T. Rodden, and J. Pycock, "Collaborative virtual environments," *Communications of the ACM*, vol. 44, no. 7, pp. 79–85, 2001.

[77] S. Thanyadit, P. Punpongsanon, and T.-C. Pong, "Efficient information sharing techniques between workers of heterogeneous tasks in 3d cve," *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, no. CSCW, pp. 1–19, 2018.

[78] G. C. Burdea, "Haptic feedback for virtual reality," in *Virtual reality and prototyping workshop*, vol. 2. Citeseer, 1999, pp. 17–29.

[79] M. Zyda, "From visual simulation to virtual reality to games," *Computer*, vol. 38, no. 9, pp. 25–32, 2005.

[80] C. Summers and M. Jesse, "Creating immersive and aesthetic auditory spaces in virtual reality," in *2017 IEEE 3rd VR workshop on sonic interactions for virtual environments (SIVE)*. IEEE, 2017, pp. 1–6.

[81] T. Lokki and M. Grohn, "Navigation with auditory cues in a virtual environment," *IEEE MultiMedia*, vol. 12, no. 2, pp. 80–86, 2005.

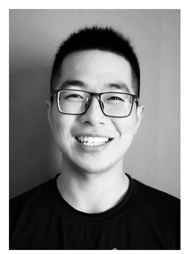
[82] L. Kruse, E. Langbehn, and F. Steinicke, "I can see on my feet while walking: Sensitivity to translation gains with visible feet,"

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2023.3271709

in *2018 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE, 2018, pp. 305–312.
[83] D. Kim, J. Kim, J.-e. Shin, B. Yoon, J. Lee, and W. Woo, "Effects of virtual room size and objects on relative translation gain thresholds in redirected walking," in *2022 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE, 2022, pp. 379–388.
[84] J. Guo, V. Vidal, I. Cheng, A. Basu, A. Baskurt, and G. Lavoue, "Subjective and objective visual quality assessment of textured 3d meshes," *ACM Transactions on Applied Perception (TAP)*, vol. 14, no. 2, pp. 1–20, 2016.
[85] Y. Nehmé, F. Dupont, J.-P. Farrugia, P. Le Callet, and G. Lavoué, "Visual quality of 3d meshes with diffuse colors in virtual reality: Subjective and objective evaluation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 3, pp. 2202–2219, 2020.
[86] F. L. Kooi and A. Toet, "Visual comfort of binocular and 3d displays," *Displays*, vol. 25, no. 2-3, pp. 99–108, 2004.
[87] Y. J. Jung, H. Sohn, S.-i. Lee, and Y. M. Ro, "Visual comfort improvement in stereoscopic 3d displays using perceptually plausible assessment metric of visual comfort," *IEEE Transactions on Consumer Electronics*, vol. 60, no. 1, pp. 1–9, 2014.
[88] Y. Zhang and Z. Wang, "Towards visual comfort: disciplines on the scene structure design for vr contents," in *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, 2018, pp. 190–196.
[89] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles, "Towards a better understanding of context and context-awareness," in *International symposium on handheld and ubiquitous computing*. Springer, 1999, pp. 304–307.
[90] T. T. Huyen Nguyen and T. Duval, "A survey of communication and awareness in collaborative virtual environments," in *2014 International Workshop on Collaborative Virtual Environments (3DCVE)*, 2014, pp. 1–8.
[91] J. Grubert, T. Langlotz, S. Zollmann, and H. Regenbrecht, "Towards pervasive augmented reality: Context-awareness in augmented reality," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 6, pp. 1706–1724, 2016.
[92] D. J. Walton and D. S. Meek, "Planar g2 transition between two circles with a fair cubic bézier curve," *Computer-Aided Design*, vol. 31, no. 14, pp. 857–866, 1999.
[93] N. Yoshida and T. Saito, "Interactive aesthetic curve segments," *The Visual Computer*, vol. 22, no. 9, pp. 896–905, 2006.
[94] H. N. Fitter, A. B. Pandey, D. D. Patel, and J. M. Mistry, "A review on approaches for handling bezier curves in cad for manufacturing," *Procedia Engineering*, vol. 97, pp. 1155–1166, 2014.
[95] Z.-C. Dong, X.-M. Fu, C. Zhang, K. Wu, and L. Liu, "Smooth assembled mappings for large-scale real walking," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 6, pp. 1–13, 2017.
[96] D. Yu, Q. Zhou, J. Newn, T. Dingler, E. Velloso, and J. Goncalves, "Fully-occluded target selection in virtual reality," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 12, pp. 3402–3413, 2020.
[97] Q. Sun, L.-Y. Wei, and A. Kaufman, "Mapping virtual and physical reality," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, pp. 1–12, 2016.
[98] M. Keshavarzi, A. Y. Yang, W. Ko, and L. Caldas, "Optimization and manipulation of contextual mutual spaces for multi-user virtual and augmented reality interaction," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2020, pp. 353–362.
[99] J. Feasel, M. C. Whitton, and J. D. Wendt, "Llcm-wip: Low-latency, continuous-motion walking-in-place," in *2008 IEEE symposium on 3D user interfaces*. IEEE, 2008, pp. 97–104.
[100] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *The international journal of aviation psychology*, vol. 3, no. 3, pp. 203–220, 1993.
[101] S. G. Hart, "Nasa-task load index (nasa-tlx); 20 years later," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 9. Sage publications Sage CA: Los Angeles, CA, 2006, pp. 904–908.
[102] T. Schubert, F. Friedmann, and H. Regenbrecht, "The experience of presence: Factor analytic insights," *Presence: Teleoperators & Virtual Environments*, vol. 10, no. 3, pp. 266–281, 2001.

Miao Wang is an Associate Professor with the State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, China. He is also an adjunct researcher with Zhongguancun Laboratory, Beijing, China. His research interests include Virtual Reality, Computer Graphics and Visual Computing. During 2016-2018, he did postdoc research in Visual Computing at Tsinghua University. He received his Ph.D. degree from Tsinghua University in 2016. He serves as a program committee member of IEEE VR and ISMAR conferences. He is a member of IEEE, ACM and Asiagraphics.

Yi-Jun Li is a Ph.D student at Beihang University, Beijing, China. He obtained his bachelor degree from Beihang University in 2019. His major research interest is in virtual reality and computer graphics and now he is working on research topics about panoramic video exploration, redirected walking and virtual collaboration.



Jinchuan Shi is a graduate student with the State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, China. He received the bachelor degree in Beihang University in 2022. His research interests includes Virtual Reality, Computer Graphics and Visual Computing.

Frank Steinicke is a professor for Human-Computer Interaction at the Department of Informatics, Universität Hamburg, Germany. His research is driven by understanding the human perceptual, cognitive and motor abilities and limitations in order to reform the interaction as well as the experience in computer-mediated realities. He received the IEEE VGTC VR Technical Achievement Award in 2023 and was inducted into the IEEE VR Academy in the same year. He is chair of the steering committee of the ACM SUI Symposium, and co-chair of the steering committee of GI SIG VR/AR. Furthermore, he is a member of the editorial boards of IEEE TVCG as well as Frontiers Section on Virtual Reality and Human Behaviour.