electronics

Article

How Do Background and Remote User Representations Affect Social Telepresence in Remote Collaboration?: A Study with Portal Display, a Head Pose-Responsive Video Teleconferencing System

Seongjun Kang [1], Gwangbin Kim [1], Kyung-Taek Lee [2] and SeungJun Kim [1,*]

[1] School of Integrated Technology, Gwangju Institute of Science and Technology, Gwangju 61005, Republic of Korea; ksjryan0728@gm.gist.ac.kr (S.K.); gwangbin@gm.gist.ac.kr (G.K.)
[2] Korea Electronics Technology Institute, Seongnam 13509, Republic of Korea; ktechlee@keti.re.kr
* Correspondence: seungjun@gist.ac.kr; Tel.: +82-62-715-5331

Abstract: This study presents Portal Display, a screen-based telepresence system that mediates the interaction between two distinct spaces, each using a single display system. The system synchronizes the users' viewpoint with their head position and orientation to provide stereoscopic vision through this single monitor. This research evaluates the impact of graphically rendered and video-streamed backgrounds and remote user representations on social telepresence, usability, and concentration during conversations and collaborative tasks. Our results indicate that the type of background has a negligible impact on these metrics. However, point cloud streaming of remote users significantly improves social telepresence, usability, and concentration compared with graphical avatars. This study implies that Portal Display can operate more efficiently by substituting the background with graphical rendering and focusing on higher-resolution 3D point cloud streaming for narrower regions for remote user representations. This configuration may be especially advantageous for applications where the remote user's background is not essential to the task, potentially enhancing social telepresence.

Keywords: human–computer interaction; social telepresence; telepresence system; video conference system; immersive display

Citation: Kang, S.; Kim, G.; Lee, K.-T.; Kim, S. How Do Background and Remote User Representations Affect Social Telepresence in Remote Collaboration?: A Study with Portal Display, a Head Pose-Responsive Video Teleconferencing System. *Electronics* **2023**, *12*, 4339. https://doi.org/10.3390/electronics12204339

Academic Editor: Dorota Kamińska

Received: 18 September 2023
Revised: 15 October 2023
Accepted: 17 October 2023
Published: 19 October 2023

1. Introduction

Video conferencing is essential for remote collaboration, offering the advantage of face-to-face interaction enriched with important nonverbal cues such as gestures and gaze, which are crucial for effective dialogue [1–3]. However, many video conferencing applications fall short in conveying a genuine sense of interconnectivity, particularly in terms of nonverbal cues. Another issue is that the streaming videos are often not responsive to the diverse viewpoints and environments of the current and remote users [4,5]. As a result, the precise orientation of gestures, gazes, and other spatial details such as the location, size, and direction of objects in the surroundings may be compromised [6]. This loss of spatial information can mislead remote users and may require them to engage in additional cognitive processes to understand the shared data. Furthermore, the disparity in spatial information can diminish the sense of eye contact among users, weakening social connections [7–9]. Therefore, maintaining "spatial faithfulness" is essential for fostering efficient and intuitive remote collaboration [10,11] in video telepresence settings.

Prior research on teleconferencing systems has predominantly centered on synchronizing two distinct physical spaces while retaining nonverbal information to create a sense of presence. Studies by Gaver et al. (1995) [12] and Nakanishi et al. [13] achieved this by aligning camera movements with users' head movements, thereby emulating natural

viewpoints. This feature allows users to navigate remote spaces as if they were physically present. Further research has employed fixed-depth camera arrays to capture both RGB and depth information, enabling the creation of point clouds for a three-dimensional representation of the environment [14–16]. However, these methods introduce challenges, such as the quality of the 3D spaces, which directly impacts user experience [17,18]. Imperfect or distorted spatial representations can disrupt the immersive experience and lead to misunderstandings during collaborative tasks.

Despite the advancements in creating spatially accurate virtual spaces, the use of multiple cameras introduces both economic and spatial challenges. This has led to innovative solutions aimed at minimizing equipment while maintaining quality. For example, Liu et al. used temporal information to supplement current frames with previous ones, thus requiring only a single camera [19]. Although this approach encountered some data losses in RGB-D information, it represented a significant attempt to reconstruct a 3D space with limited resources. However, the persistent challenge of data voids resulting from RGB-D data loss during point cloud streaming remains unresolved.

Alternatively, some studies have explored using graphical backgrounds and human avatars to address the challenges associated with streaming 3D reconstructed environments [20,21]. For example, Kauff et al. [20] and Tanger et al. [22] replaced the real background with a static virtual one, eliminating the need for depth cameras or extensive 3D scanning hardware. Jo et al. [21] used avatars instead of remote users in a virtual reality telepresence system. While these approaches seem to provide cost-effective and streamlined solutions regarding both hardware and software needs, the potential impact on the user experience remains uncertain. Comparative evaluations are needed to determine whether these graphical substitutes can provide an experience comparable to that of more complex systems.

To address the afore-described challenges and limitations associated with extant solutions, we herein present Portal Display, a screen-based video conferencing system fitted with a depth camera, designed to offer users an enhanced sense of spatial depth during video conferencing. Unlike prior systems employing multiple depth cameras, Portal Display achieves this immersive experience using only a single depth camera, making it both economically and spatially efficient. This study not only addresses economic and spatial constraints but also aims to compare the user experiences provided by different representation methods, emphasizing the significance of such comparisons in teleconferencing solutions. To this end, we pose the following research questions:

- RQ1: How does the type of remote user representation (point cloud streaming vs. graphical rendering) in Portal Display influence overall system usability, social telepresence, and concentration toward the remote user?
- RQ2: How does the type of remote user's background representation (point cloud streaming vs. graphical rendering) impact overall system usability, social telepresence, and concentration toward the remote user?
- RQ3: Do the types of remote user and background representation (point cloud streaming vs. graphical rendering) interact in impacting the overall usability, social telepresence, and concentration within the Portal Display system?

Figure 1 illustrates the comprehensive research design adopted for this study, encompassing Portal Display's technical design, experimental process, and subsequent analyses.

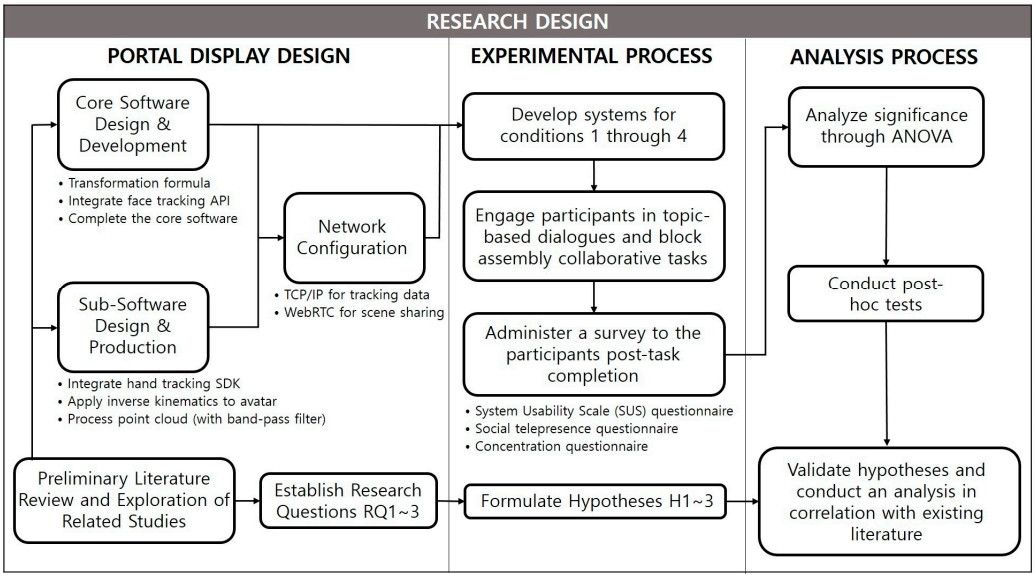


Figure 1. Comprehensive research design, highlighting the development, experimental, and analytical phases.

2. Research Hypotheses

Recent studies have provided valuable insights into the effectiveness of different representation methods in virtual environments. Yu et al. [23] emphasized the potential of point cloud representations in enhancing presence, behavior impression, and humanness in VR settings. Drawing from their findings, our first set of hypotheses (H1a–H1c) centers on the perceived advantages of point cloud representations over graphical renderings.

Kauff et al. [20] showcased the enhanced telepresence achieved when combining point cloud representations of remote users with virtual backgrounds. These insights informed our second set of hypotheses (H2a–H2c), which explore the influence of background representation on various user experience metrics.

In addition to these individual studies, there is a body of research suggesting that the combination of different methods for representing remote users and backgrounds does not detrimentally impact user experience [20–22]. This guided the formulation of our third primary hypothesis (H3).

- H1a: Point cloud representations of remote users enhance system usability more than graphical renderings.
- H1b: Point cloud representations of remote users enhance telepresence more than graphical renderings.
- H1c: Point cloud representations of remote users enhance user concentration more than graphical renderings.
- H2a: The influence of background representation (point cloud vs. graphical rendering) on system usability is minimal.
- H2b: The influence of background representation (point cloud vs. graphical rendering) on telepresence is minimal.
- H2c: The influence of background representation (point cloud vs. graphical rendering) on user concentration is minimal.
- H3: The interaction effect of different methods of representing remote users and backgrounds on user experience is negligible.

We tested these hypotheses by establishing an experimental setting for simple conversation and collaborative tasks. User experience, focusing on system usability, social telepresence, and concentration toward the remote user, was assessed via a questionnaire.

3. Portal Display

3.1. Stereoscopic Vision with a 2D Screen

Achieving a sense of depth on a 2D display requires aligning the 3D graphic scene with the shifting viewpoint, specifically, stereo disparity. For instance, during a meeting, the primary view will feature participants' frontal faces when they directly face each other, whereas profile views become more pronounced if perspectives are oriented toward the lateral sides. Thus, teleconferencing systems maintaining this stereo disparity can emulate the spatial interactions of physical environments.

Our proposed system addresses stereo disparity by transforming the 3D graphic environment and projecting it onto a flat display. This transformation is synchronized with the user's head position, as illustrated in Figure 2. Within the 3D engine environment, this involves a sequential composite linear transformation method, further detailed in Figure 3. Standard linear transformations in commonly used 3D engines like Unity, Unreal Engine, CryEngine, and Godot primarily consist of rotation and scaling operations. In our system, implemented via the Unity 3D engine, these transformations are updated at a consistent frequency of 60 Hz, ensuring real-time adjustments in both scaling and rotation. The practical implementation of these linear transformations within Unity can be seen in Supplementary Video S1. Figure 4 and Supplementary Video S2 showcase the intrinsic stereo disparity of the Portal Display, highlighting its dynamic adaptability in relation to the user's head position. The design intent behind Portal Display is to offer variable view sections based on user perspectives. To achieve precise head position tracking, we integrated the FaceTrackNoIR API, which allows us to determine the position of the user's eyes and use them as reference coordinates. Given that individuals may have a dominant eye, the linear transformation is centered on the offset of the midpoint between the two eyes. This approach was influenced by the Cyclops method introduced by Petkov [24], which simplifies the camera positioning requirements for stereoscopic display.

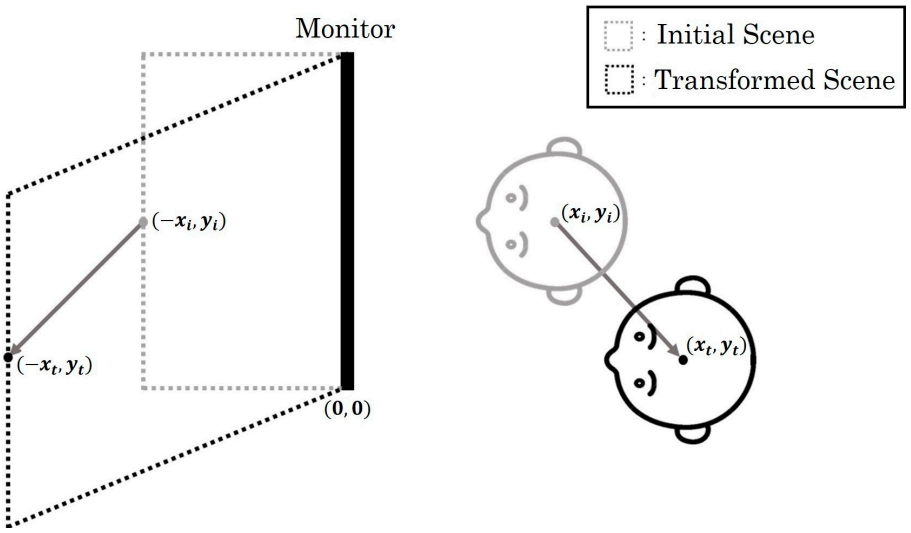
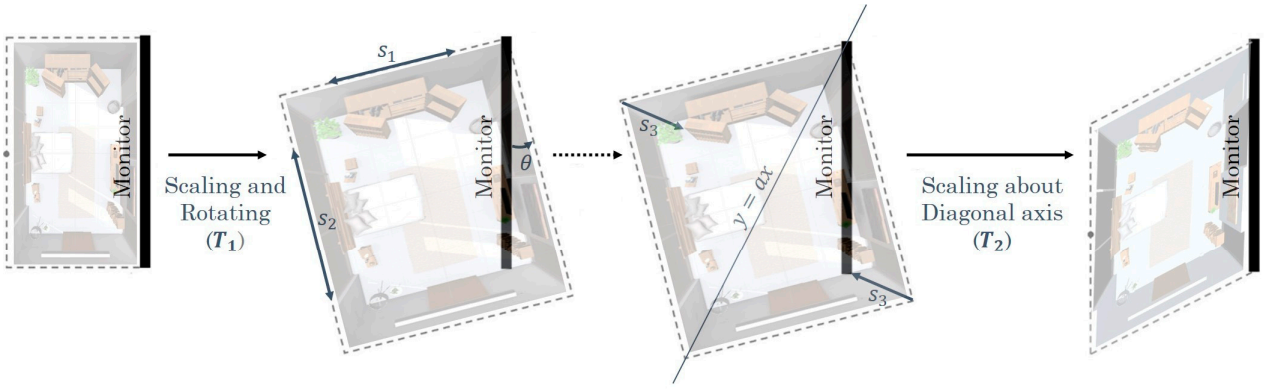


Figure 2. Linear transformation of the scene aligning symmetrically with the user's head position.



$$T_1 = \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \qquad T_2 = \begin{pmatrix} \cos(artan(a)) & -sin(-artan(a)) \\ sin(-artan(a)) & cos(-artan(a)) \end{pmatrix} \begin{pmatrix} s_3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(artan(a)) & -sin(artan(a)) \\ sin(artan(a)) & cos(artan(a)) \end{pmatrix}$$

Figure 3. Step-by-step composite linear transformation process within the 3D engine space.

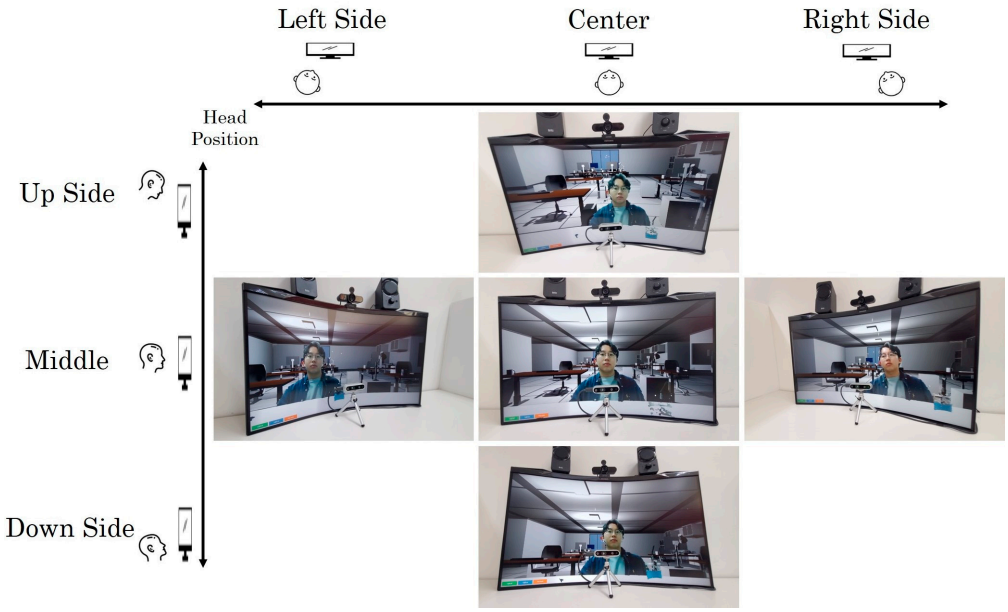


Figure 4. Changes in the stereoscopic vision of the scene according to the user's head position.

3.2. Representation Types of the Background

The Portal Display provides two modalities for background representation: point cloud streaming (depicting the actual environment) and graphical rendering (illustrating a virtual environment).

For the point cloud streamed background, real-time data are sourced from the Intel depth camera D435 (Intel Corporation, Santa Clara, CA, USA). These RGB-D data are subsequently integrated into Unity 3D via the RealSense SDK 2.0, resulting in a digital representation congruent with the physical environment (Figure 5a). Conversely, the graphically rendered background, while maintaining the space's inherent geometry, utilizes prerendered prefab models (Figure 5b). Both methods comprise spatial details and accommodate viewpoint transformations, thereby providing users with an immersive depth perception. Consequently, the Portal Display system is versatile, accommodating any environment primarily composed of 3D entities, regardless of the mode of background representation.

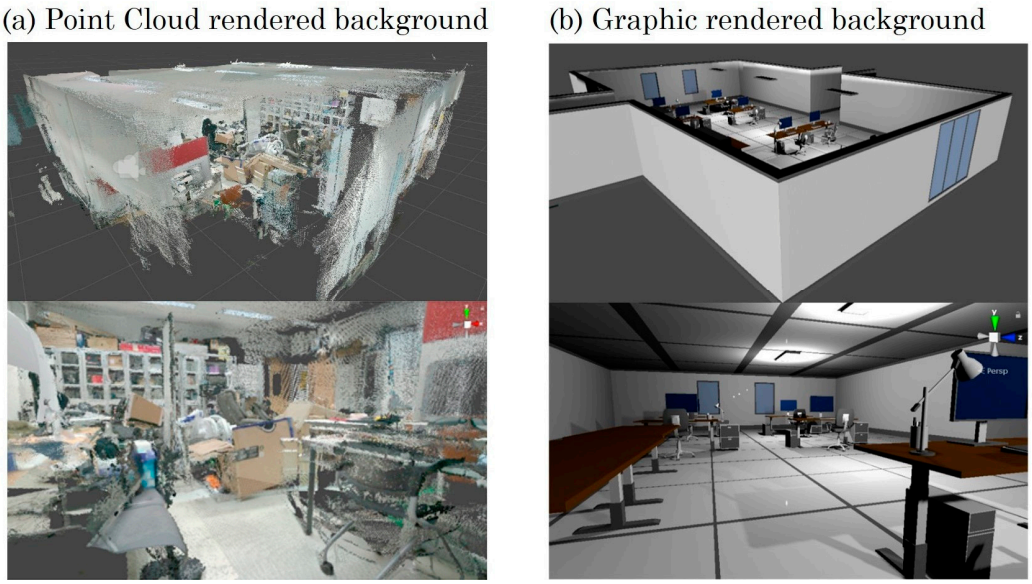


Figure 5. Indoor environment with (**a**) point cloud streaming and (**b**) graphical rendering.

3.3. Representation Types of the Remote User

The Portal Display provides two modalities for depicting remote users: point cloud streaming (Figure 6a) and graphical rendering (Figure 6b).

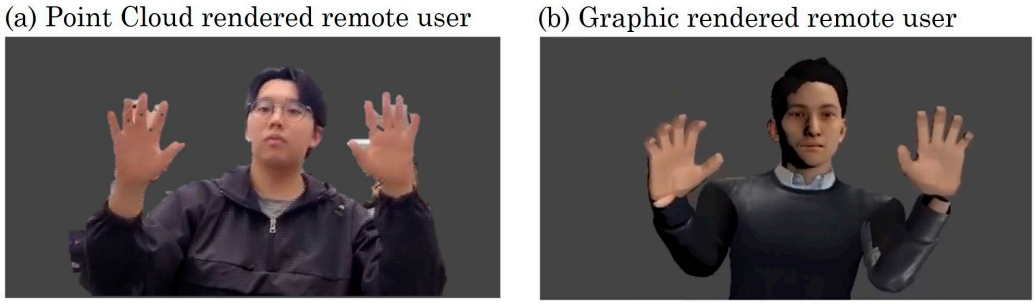


Figure 6. Remote users represented with (**a**) point cloud streaming and (**b**) graphical rendering.

For the point cloud streaming representation, the raw data from the depth camera are processed using band-pass filters. The filters refine the captured visuals by eliminating irrelevant video components, thus transmitting only enhanced images of the remote users to the local viewer (Figure 7 and Video S3). This filtering process facilitates differentiating remote users from their immediate backgrounds, allowing their seamless integration with a virtual backdrop, as detailed in Section 3.2.

However, the graphical rendering representation employs facial landmark detection algorithms and leap motion tracking. This setup primarily captures the kinematics of the user's upper body. We collected the head position and rotation values using the FaceTrackNoIR API (LamaJoy Software, Abbekerk, The Netherlands). based on webcam input, while the hand position and rotation values were acquired via the Leapmotion SDK. Using Unity engine's "Final IK" asset, the derived positions and rotations of these key joints, the head and both wrists, were then subjected to inverse kinematics within the feasible human upper body movement range. This ensured that the avatars in the system could convincingly emulate the natural movements of the actual users (Figures 6b and 8, and Video S3).

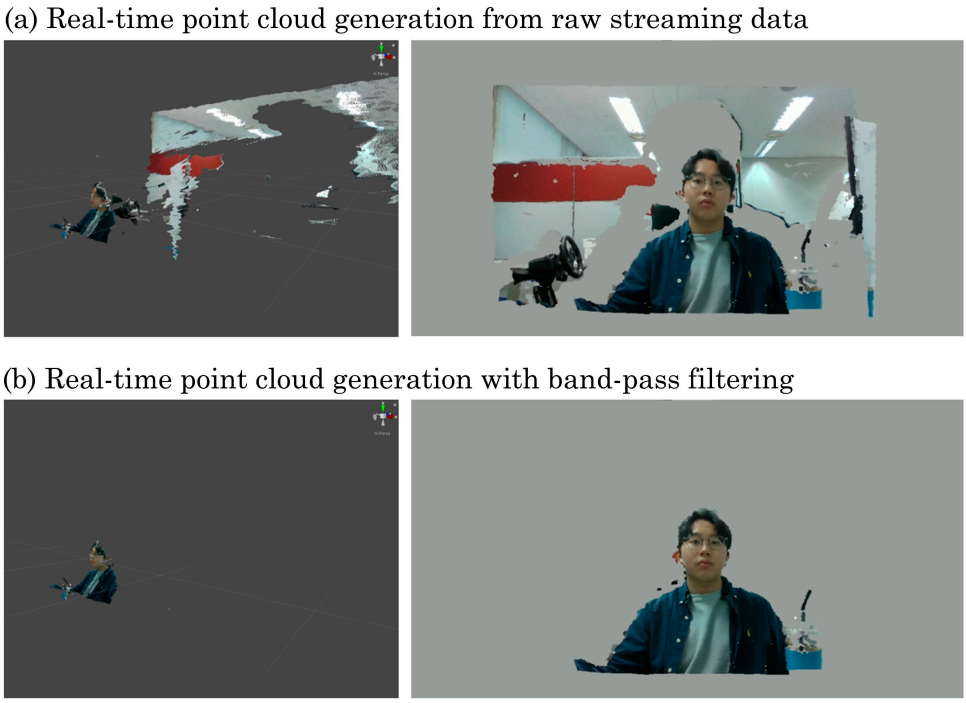


Figure 7. Point cloud streaming from a depth camera capturing a remote user, with background removed using band-pass filters.

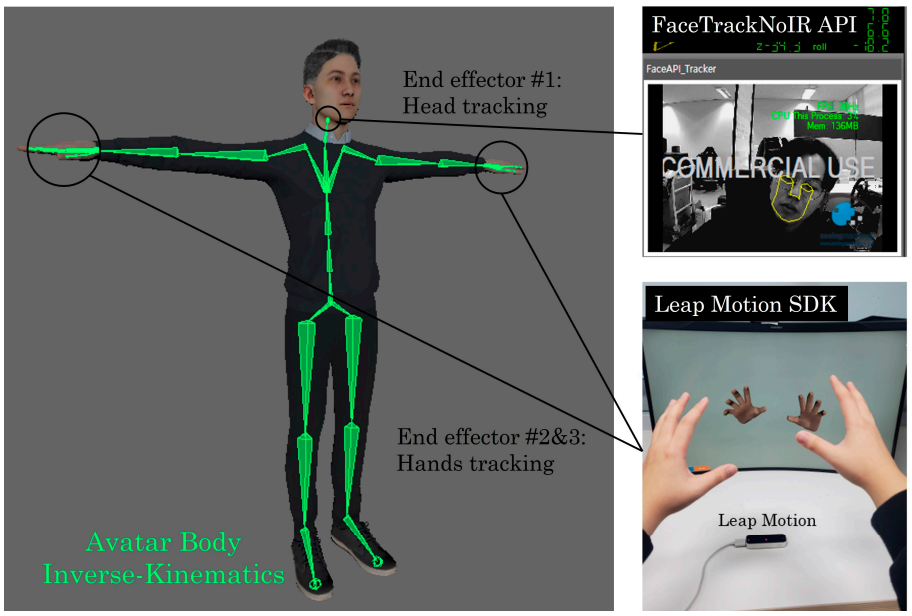


Figure 8. Inverse kinematics to transform the end-effector joint positions (facial landmarks and upper body) into avatar motions.

3.4. Networking and Setup between Different PCs

To facilitate teleconferencing by bridging two distinct spaces, both the point cloud streaming for scene/avatar representation and the head position data for stereoscopic disparity must be shared between remote users. Consequently, we implemented a server–client wireless network built on TCP/IP and P2P protocols, leveraging the capabilities of the WebRTC API.

Figure 9 outlines the network architecture underlying the Portal Display system. During the "Data Collection Process", the depth camera relays RGB-D information to the Unity engine on each user's PC using the Intel RealSense SDK 2.0. Simultaneously,

each PC's webcam collects user head position data, which are subsequently transmitted to computers in different locations via TCP/IP communication. In the "Portal Rendering Process", this amalgamation of RGB-D information and head position data constitutes the core of the Portal Display algorithm's application and rendering. Within the "Scene Streaming Process," the dynamically altered scene—reflecting real-time head pose data—is streamed to computers across different sites using WebRTC [25,26]. Consequently, the remote user's representation within the display and their ambient environment dynamically adapt to the local user's head movements.

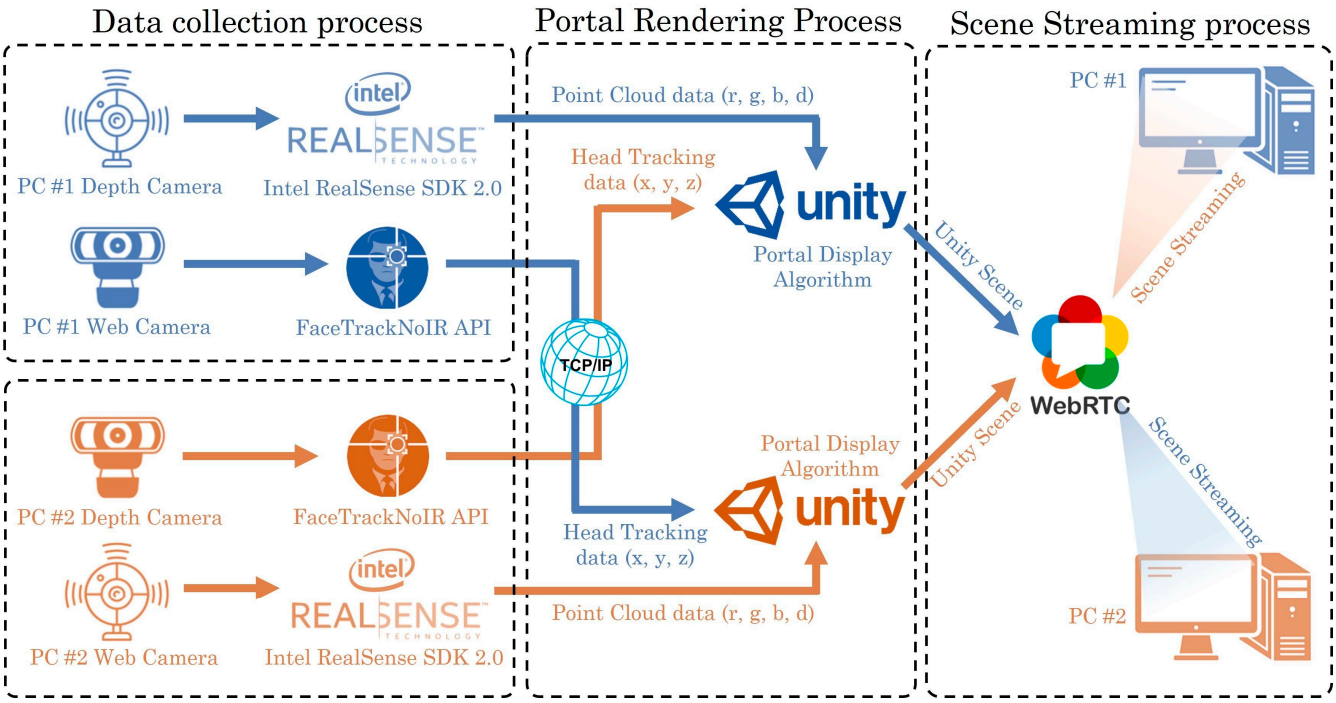


Figure 9. Network architecture for the Portal Display system, detailing data flow across processes.

Figure 10 visually represents the Portal Display's deployment across two separate settings. Augment immersion could be obtained by employing desks with matching colors and designs in each location, fostering an illusion of mutual presence across a shared desk. While this setup is recommended to improve the user experience, it remains optional. Notably, analogous techniques leveraging desk properties to craft shared space illusions have been documented in other teleconferencing systems [4,12,14,27].

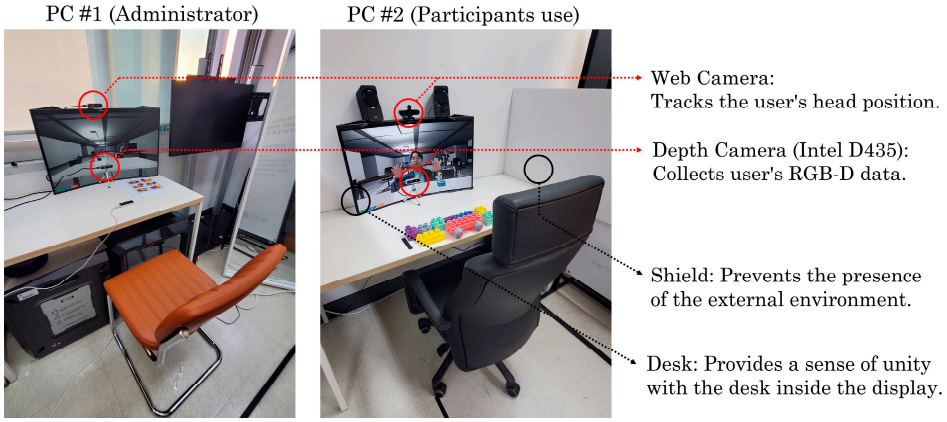


Figure 10. Portal Display setup in two separate environments.

4. Study Procedures

4.1. Task Design

We evaluated our system through tasks that incorporated both a simple dialogue and block assembly collaboration with a remote participant. Referring to Tanaka et al. [28], who assessed a physical embodied telepresence system where the participants engaged in questioning and answering specific electronic device-related problems and solutions, we crafted a similar dialogue approach. The participants were engaged in a 5 min conversation on predetermined topics, including culture, education, and content in virtual worlds, adhering to a standardized protocol (Figure A1). Inspired by Zillner et al. [29], who tested the utility of a digital whiteboard by having participants design and rearrange furniture on a shared blueprint, we directed our participants to assemble specific car models using blocks (Figure A2). During this task, both verbal and nonverbal cues, including gestures such as pointing, were employed to guide the remote participant. Previously, Onishi et al. [30] and Kim et al. [31] highlighted the importance of gestures such as eye-gaze for communication and collaboration; our block assembly task was also intentionally designed to authentically represent nonverbal cues and diverse viewpoints. By guiding the participants to utilize their eyes and fingers for selecting and assembling blocks of particular colors and shapes, we sought to demonstrate the effectiveness of Portal Display in affording a nuanced sense of spatial orientation, interactivity, and genuine telepresence.

4.2. Analysis Strategy

The participants undertook tasks in four conditions (Figure 11), defined by two background variations and two remote user representation methods. To counteract any learning effects or biases arising from the sequence in which the conditions were presented, we employed a Latin-square counterbalanced order. This design ensured that each condition appeared exactly once in every position, thereby mitigating order effects, such as fatigue or familiarity, that could potentially influence the participants' responses.

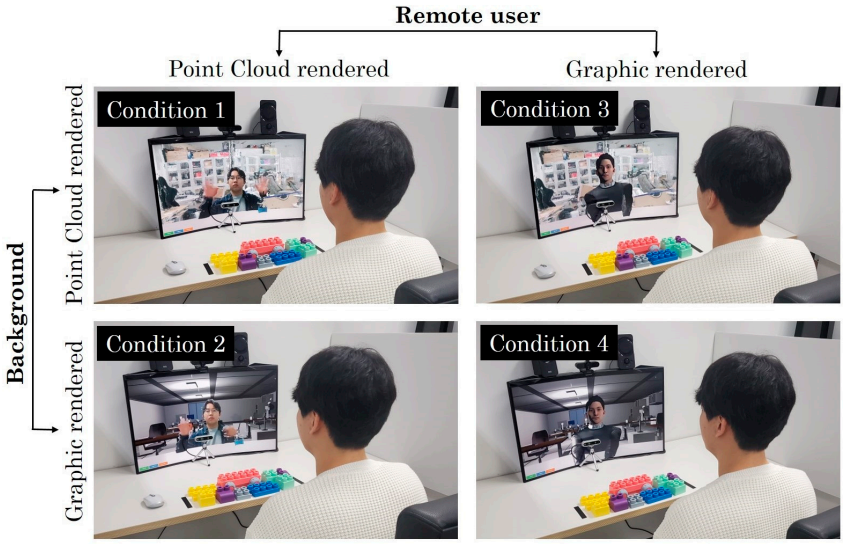


Figure 11. Four different conditions (two backgrounds × two remote users) used in the experiment.

4.3. Protocol

After each task, the participants evaluated system usability using the System Usability Scale (SUS) questionnaire, which is a reliable tool for gauging system usability, consisting of 10 items that reflect users' impressions on the ease and satisfaction of system use. Each item is rated on a 5-point scale, yielding a final SUS score ranging from 0 to 100. Scores above 68 typically indicate above-average usability. Additionally, we gauged social telepresence perception for each setting by using a questionnaire by Nakanishi et al. [13,32]. We also included questions assessing the participants' concentration during the session. This

comprehensive approach provided insights into the effects of each condition on usability, telepresence, and focus. Feedback for social telepresence and concentration was captured on a 1–7 Likert scale, where 1 signified strong disagreement and 7 strong agreement (Figure 12).

Usability Questionnaires (SUS)
Q1. I think that I would like to use this system frequently.
Q2. I found the system unnecessarily complex.
Q3. I thought the system was easy to use.
Q4. I think that I would need the support of a technical person to be able to use this system.
Q5. I found the various functions in this system were well integrated.
Q6. I thought there was too much inconsistency in this system.
Q7. I would imagine that most people would learn to use this system very quickly.
Q8. I found the system very cumbersome to use.
Q9. I felt very confident using the system.
Q10. I needed to learn a lot of things before I could get going with this system.

Social Telepresence Questionnaires
Q1. It was as if I was facing the other person directly.
Q2. I felt like I was talking directly to the other person.
Q3. I felt like I was in the same room as the other person.

Concentration on remote user Questionnaires
Q1. I was able to fully concentrate on the remote user.
Q2. I was able to focus on the remote user without being distracted by the background.

Figure 12. Usability, social telepresence, and concentration questions in the user questionnaire.

5. Study Results

We recruited 15 participants (9 males and 6 females) aged 23.73 years on average (mean M = 23.73, standard deviation SD = 1.35). All statistical analyses were performed using JASP software version 0.16.3. After conducting normality tests for skewness and kurtosis, we determined that all the data were normally distributed. We used two-way repeated measures analysis of variance (ANOVA) to verify the main effect of variables and the Bonferroni test for post hoc analyses.

5.1. System Usability Scale (SUS)

Based on the adjective categories associated with raw SUS scores as defined by Brooke, J. [33], all four conditions were rated as having acceptable usability ranges (Condition 1: M = 82.17, SD = 11.53; Condition 2: M = 80.50, SD = 14.18; Condition 3: M = 79.00, SD = 13.69; Condition 4: M = 75.00, SD = 14.11) (Figure 13).

From the ANOVA, there appeared to be an influence of the representation method of remote users on usability (F-statistic F (1, 14) = 11.109, significance level p = 0.005) (Table 1). The point cloud streamed conditions generally scored higher in SUS than the graphically rendered conditions. However, when examining the two levels of backgrounds, either point cloud streamed or graphically rendered, no significant difference in SUS scores was observed (F (1, 14) = 0.286, p = 0.601) (Table 1). Furthermore, the interaction between the independent variables, namely background types and remote user types, did not yield any significant effect on usability (F (1, 14) = 0.508, p = 0.488) (Table 1).

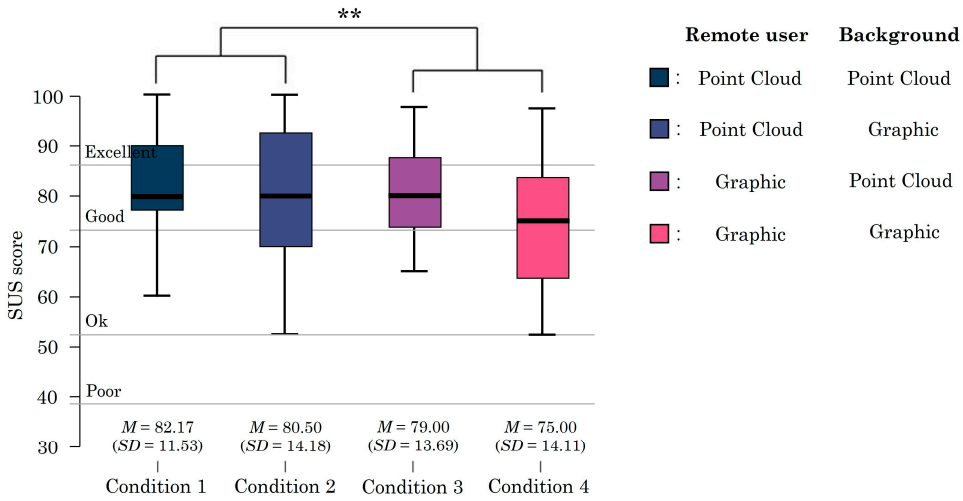


Figure 13. Results of SUS score (Condition 1: PCD remote user + PCD background, Condition 2: PCD remote user + graphic background, Condition 3: graphic remote user + PCD background, Condition 4: graphic remote user + graphic background) (** $p < 0.01$).

Table 1. Within-subjects' effects on usability (** $p < 0.01$).

Case	Sum of Squares	df	Mean Square	F	*p*
Remote user	555.104	1	555.104	11.109	0.005 **
Residuals	699.583	14	49.970		
Background	12.604	1	12.604	0.286	0.601
Residuals	617.083	14	44.077		
Remote user × Background	37.604	1	37.604	0.508	0.488
Residuals	1035.833	14	73.988		

Post hoc analysis was conducted to delve deeper into the observed trends, and the results are presented in Table 2. Interestingly, post hoc analysis revealed no significant differences between any specific pairs of conditions regarding usability (Table 2). This suggests that, although there is an overarching trend pointing toward better usability with the point cloud streaming method for representing remote users, pinpointed comparisons between specific conditions did not solidify this finding. It is also noteworthy that the type of background representation did not have a considerable impact on usability (Figure 13).

Table 2. Post hoc comparisons of usability.

		Mean Difference	SE	t	p_{bonf}
PCD remote user, PCD background	Graphic remote user, PCD background	4.500	2.875	1.565	0.775
	PCD remote user, Graphic background	−0.667	2.806	−0.238	1.000
	Graphic remote user, Graphic background	7.000	2.504	2.796	0.056
Graphic remote user, PCD background	PCD remote user, Graphic background	−5.167	2.504	−2.063	0.291
	Graphic remote user, Graphic background	2.500	2.806	0.891	1.000
PCD remote user, Graphic background	Graphic remote user, Graphic background	7.667	2.875	2.667	0.077

5.2. Social Telepresence

In our study, the social telepresence responses, as indicated in Figure 12, were measured on a scale from 1 to 7. A score between 5 and 7 was considered a "positive" indication that participants experienced a strong sense of social telepresence. Conversely, a score between 1 and 3 was interpreted as a "negative" reflection, suggesting a lack of or reduced sense of social telepresence. Using this categorization, Conditions 1 and 2, representing remote users through point cloud streaming, received positive ratings for social telepresence (Condition 1: M = 5.49, SD = 1.08; Condition 2: M = 5.13, SD = 1.10). In contrast, Conditions 3 and 4, which featured graphically rendered remote users, were given negative evaluations (Condition 3: M = 3.98, SD = 1.34; Condition 4: M = 3.00, SD = 1.10) (Figure 14).

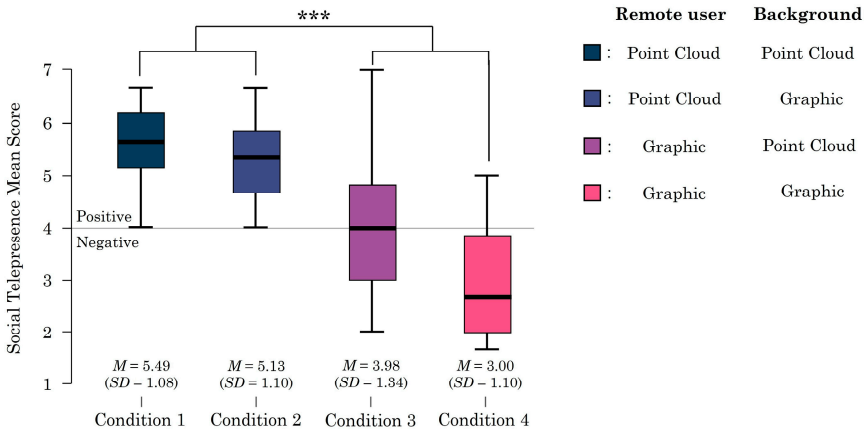


Figure 14. Results of social telepresence (Condition 1: PCD remote user + PCD background, Condition 2: PCD remote user + graphic background, Condition 3: graphic remote user + PCD background, Condition 4: graphic remote user + graphic background) (*** $p < 0.001$).

Based on the ANOVA, several insights emerged. When examining the representation method of remote users, the point cloud streamed conditions exhibited significantly higher social telepresence than the graphically rendered conditions (F (1, 14) = 49.807, $p < 0.001$) (Table 3). In the case of the representation method of the background, the point cloud streamed conditions showed significantly higher social telepresence than the graphically rendered conditions (F (1, 14) = 12.263, p = 0.004) (Table 3). Furthermore, the interaction between the two independent variables, namely background types and remote user types, did not demonstrate a significant effect on social telepresence (F (1, 14) = 3.940, p = 0.067) (Table 3).

Table 3. Within-subjects' effects on social telepresence (** $p < 0.01$, *** $p < 0.001$).

Case	Sum of Squares	df	Mean Square	F	*p*
Remote user	49.807	1	49.807	25.696	<0.001 ***
Residuals	27.137	14	1.938		
Background	6.667	1	6.667	12.263	0.004 **
Residuals	7.611	14	0.544		
Remote user × Background	1.452	1	1.452	3.940	0.067
Residuals	5.159	14	0.369		

Post hoc analysis was conducted to delve deeper into the observed trends, and the results are presented in Table 4. The results indicated significant differences between Conditions 1 and 3 (t (14) = 3.853, p = 0.006), between Conditions 1 and 4 (t (14) = 6.119, $p < 0.001$), and between Conditions 2 and 4 (t (14) = 5.440, $p < 0.001$) (Table 4). These differences highlight that the point cloud streamed user conditions yielded significantly higher social telepresence than the graphically rendered user conditions. Interestingly, there was also a significant difference observed between Conditions 3 and 4 (t (14) = 3.965,

$p = 0.003$) (Table 4). This suggests that within the graphic remote user conditions, using a point cloud background resulted in a significantly increased social telepresence compared with using a graphic background.

Table 4. Post hoc comparisons of social telepresence (** $p < 0.01$, *** $p < 0.001$).

		Mean Difference	SE	t	p_{bonf}
PCD remote user, PCD background	Graphic remote user, PCD background	1.511	0.392	3.853	0.006 **
	PCD remote user, Graphic background	0.356	0.247	1.442	0.965
	Graphic remote user, Graphic background	2.489	0.407	6.119	<0.001 ***
Graphic remote user, PCD background	PCD remote user, Graphic background	−1.156	0.407	−2.841	0.058
	Graphic remote user, Graphic background	0.978	0.247	3.965	0.003 **
PCD remote user, Graphic background	Graphic remote user, Graphic background	2.133	0.392	5.440	<0.001 ***

5.3. Concentration on Remote User

All four conditions were evaluated in terms of focus on the remote user (Condition 1: $M = 5.35$, $SD = 0.88$; Condition 2: $M = 5.55$, $SD = 1.09$; Condition 3: $M = 4.25$, $SD = 1.18$; Condition 4: $M = 5.20$, $SD = 0.89$) (Figure 15).

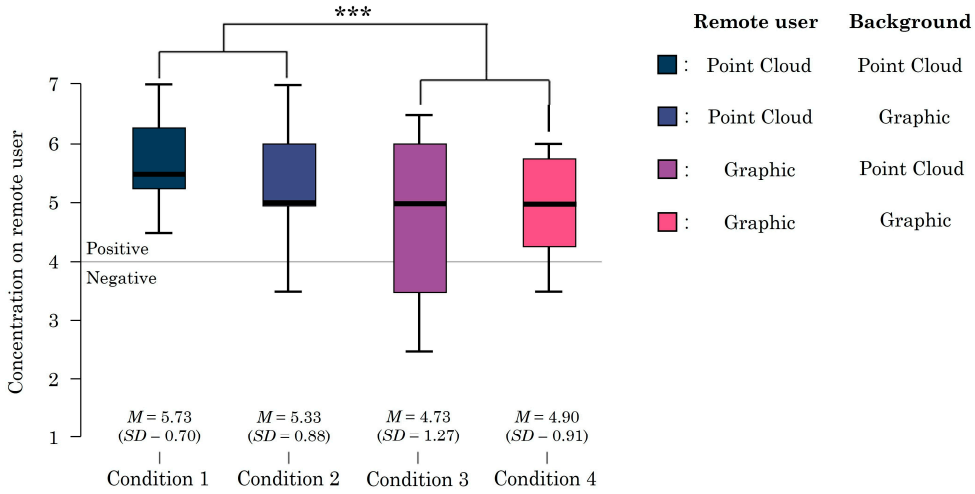


Figure 15. Results of remote user concentration (Condition 1: PCD remote user + PCD background, Condition 2: PCD remote user + graphic background, Condition 3: graphic remote user + PCD background, Condition 4: graphic remote user + graphic background) (*** $p < 0.001$).

Based on the ANOVA, several insights emerged. According to the examination of the representation method of remote users, the point cloud streamed conditions resulted in significantly better concentration than the graphically rendered conditions did (F (1, 14) = 23.405, $p < 0.001$) (Table 5). However, when examining the two levels of backgrounds, either point cloud streamed or graphically rendered, no significant difference in concentration was observed (F (1, 14) = 0.227, $p = 0.641$) (Table 5). Furthermore, the interaction between the independent variables, namely background types and remote user types, did not yield any significant effect on usability (F (1, 14) = 1.364, $p = 0.262$) (Table 5).

Table 5. Within-subjects' effects on concentration (*** $p < 0.001$).

Case	Sum of Squares	df	Mean Square	F	p
Remote user	7.704	1	7.704	23.405	<0.001 ***
Residuals	4.608	14	0.329		
Background	0.204	1	0.204	0.227	0.641
Residuals	12.608	14	0.901		
Remote user × Background	1.204	1	1.204	1.364	0.262
Residuals	12.358	14	0.883		

Post hoc analysis was conducted to delve deeper into the observed trends, and the results are presented in Table 6. The results indicated significant differences between Conditions 1 and 3 (t (14) = 3.518, p = 0.011), and between Conditions 1 and 4 (t (14) = 2.910, p = 0.047) (Table 6). These differences highlight that the point cloud streamed user conditions yielded significantly better concentration than the graphically rendered user conditions. This result implies that the concentration on the remote user increases when the remote user is represented with the point cloud streaming method as opposed to the graphical rendering method. Moreover, the representation type of the background did not significantly affect the concentration on the remote user (Figure 15).

Table 6. Post hoc comparisons of concentration (** $p < 0.01$).

		Mean Difference	SE	t	p_{bonf}
PCD remote user, PCD background	Graphic remote user, PCD background	1.00	0.284	3.518	0.011 **
	PCD remote user, Graphic background	0.400	0.345	1.160	1.000
	Graphic remote user, Graphic background	0.833	0.286	2.910	0.047 **
Graphic remote user, PCD background	PCD remote user, Graphic background	−0.600	0.286	−2.095	0.284
	Graphic remote user, Graphic background	−0.167	0.345	−0.483	1.000
PCD remote user, Graphic background	Graphic remote user, Graphic background	0.433	0.284	1.5250	0.845

Having presented the results in detail, we now turn our attention to the primary hypotheses that drove this study. Table 7 summarizes the results concerning these hypotheses, providing a concise reference for understanding the overarching findings.

Table 7. Summary of hypothesis results.

Hypothesis	Status
H1a: Point cloud representations of remote users enhance system usability more than graphical renderings.	Confirmed
H1b: Point cloud representations of remote users enhance telepresence more than graphical renderings.	Confirmed
H1c: Point cloud representations of remote users enhance user concentration more than graphical renderings.	Confirmed
H2a: The influence of background representation (point cloud vs. graphical rendering) on system usability is minimal.	Confirmed
H2b: The influence of background representation (point cloud vs. graphical rendering) on telepresence is minimal.	Rejected
H2c: The influence of background representation (point cloud vs. graphical rendering) on user concentration is minimal.	Confirmed
H3: The interaction effect of different methods of representing remote users and backgrounds on user experience is negligible.	Confirmed

6. Discussion

6.1. Proposed Linear Transformation Matrices for Typical 3D Engines and Their Extension to Telepresence Systems

We devised a specific development mechanism for a system that conveys a sense of depth to flat displays based on user head movement. Earlier research has explored depth perception improvements in flat displays using a Wii remote controller and based on head tracking [34] and systems offering immersive experiences to users in a room-shaped projection environment based on head tracking [35–37]. Although these studies provided solid contributions, they did not clearly outline the transformation matrices essential for conveying depth, presenting a hurdle for integration in typical 3D engines. This study elucidated a composite linear transformation algorithm optimized for prevalent 3D engines, including Unity, Unreal Engine, CryEngine, and Godot. Because this method was developed to facilitate the transformation of the entire virtual space within the 3D engine based on the user's head position, it enables the rendering of any 3D object with a depth effect. However, it remains a child object in the linear transformation domain. The relevant feature of our algorithm is its potential to demystify the complexities of stereoscopic flat displays for the broader research and developer community.

Capitalizing on the multifaceted capabilities of Portal Display, this study broadened its application to a telepresence system employing a singular depth camera. The initiative by Google Research, known as Project Starline, harnesses an array of depth cameras to engineer a telepresence system, rendering users from different locations in a three-dimensional presence on a flat screen [38]. Their method convincingly mirrors users from various spatial settings, such as in immediate proximity. In contrast, our iteration utilizes a singular depth camera for the flat screen-based telepresence system. The point clouds streamed in real time are anchored as child objects within the linear transformation domain. Such clouds undergo a linear transformation, offering multifaceted viewpoints. The inherent design mitigates the requirement for an ensemble of depth cameras to acquire spatial data from diverse points. However, when juxtaposed with a setup employing multiple depth cameras, the spatial intelligence developed is somewhat truncated, potentially compromising realism (this is further elaborated on in Section 6.5). Referencing [39], the integration of depth and computer vision into traditional video processing suggests that mobile video conferencing could evolve toward a more immersive experience. Thus, our proposed method, adaptable to mobile devices like smartphones with a single depth camera, alludes to the potential of enhancing teleconferencing experiences irrespective of the setting or circumstance.

6.2. Streaming Remote Users with Point Clouds Improves Usability, Telepresence, and Concentration

In our experimental setup, representing remote users with point cloud streaming was rated to provide better usability, social telepresence, and concentration. Nowak et al. [40] reported that the realism of the representation increases the social presence of the interacting person. According to the field of the psychology of perception, individuals inherently prefer anthropomorphic (human-like) representations during interactions, suggesting an innate preference for realism and lifelike interactivity (Zinchenko et al.) [41]. This psychological insight reinforces our findings, highlighting the preference for point cloud-centric representations which, due to their increased verisimilitude, fostered a heightened sense of social telepresence, in contrast to their graphical counterparts.

In a related study, Yu et al. [23] compared user experience between an avatar represented graphically in an HMD system and a user representation based on point cloud streaming. Their research underlined that point cloud-centric representations created an elevated sense of social telepresence. This aligns with our results despite the different telepresence technologies used in each study (HMD vs. flat display).

Further emphasizing the importance of detail in representations, Kang et al. [42] noted that while medium facial detail (mid-LOD) preserved social presence and cut costs, low detail (low-LOD) diminished social presence due to poor emotional portrayal in augmented

reality (AR). This underscores the significance of facial expressions and emotional understanding in representing users, particularly in collaborative scenarios. Drawing parallels with our study, the enhanced social telepresence observed in our results when users were represented with point clouds could be attributed to the realistic facial context that point clouds provide. Given the importance of facial cues in AR environments, as highlighted by Kang et al., it is plausible that the realistic portrayal of faces using point clouds plays a pivotal role in enhancing social telepresence in our setup.

While point cloud streaming enriches the realism of the remote user representation, potential pitfalls may occur. For example, it might introduce noise, such as jitteriness or gaps, elements absent in a sleek graphical portrayal. Zhang et al. [43] highlighted that this noise in 3D models crafted from point clouds could detrimentally influence user experience. Thus, these imperfections can influence user experiences. However, our findings confirm that, even considering potential pitfalls, the point cloud streaming approach outperformed the graphical method in delivering a superior user experience in terms of usability, social telepresence, and concentration. Consequently, we support adopting point cloud streaming as the go-to method for depicting remote users in flat-panel display-centric social telepresence platforms, even if minor noise issues are present.

6.3. Comparative Impact of Point Cloud Streaming for Background and Remote Users on Telepresence

Utilizing point cloud streaming for background depiction increases the sense of social telepresence compared with utilizing graphical representation. However, this effect is less pronounced than when adopting point cloud streaming for representing remote users. On average, the enhancement in social telepresence due to the background representation accounted for roughly 20% of the improvements generated by the remote user representation. These observations suggest that employing point cloud streaming for either the background or remote users augments social telepresence relative to graphical representations. However, the contribution is more pronounced with remote users. Our findings highlight that, within the framework of a flat-panel display-based social telepresence system, the verisimilitude provided by point cloud streaming for remote users exerts a more substantial influence on social telepresence than the background representation.

Jo et al. [44] used an HMD telepresence system to deduce that backgrounds based on point cloud streaming provided an improved sense of presence compared with graphical virtual backgrounds. The significance of our study lies in its ability to mirror such results within the paradigm of a flat-panel display-based immersive telepresence mechanism. Furthermore, Jo et al. [44] indicated that the chosen background did not affect the trust level toward remote users. Similarly, our findings showed that concentration on remote users remained consistent, regardless of the background employed. This result suggests that while backgrounds can modulate perceptions of realism, such as presence and social telepresence, they exert negligible influence on the perceived credibility and focus of remote users.

Additionally, our investigation uncovered only marginal differences in system usability when juxtaposing the two background representation methods (i.e., point cloud streaming vs. graphical rendering) within our telepresence system. Also, in broader video conferencing contexts like Zoom, Lee et al. [45] found that the use of virtual backgrounds did not significantly impact the overall usability of the system. Hence, virtual graphics can be seamlessly integrated as teleconferencing backgrounds without hampering system usability or diverting users' attention from remote participants. Opting for virtual graphics over point cloud streaming can result in significant computational advantages in terms of fiscal and technical facets. Such computational thriftiness becomes pivotal when considering bit rate enhancements in conferencing systems where network performance is paramount. Numerous contemporary studies [46–48] have embarked on optimizing point cloud datasets to alleviate networking complexities. Nonetheless, where there is a pressing need to heighten physical presence, employing point cloud streams for backgrounds can

be advantageous. This emphasizes that teleconferencing designs might need tailoring depending on their core objectives.

6.4. Insignificant Interaction Effects between Background Types and Remote User Representation Types

To the best of our knowledge, previous research [20–22] has not statistically evaluated the interaction effects between the background of a telepresence system and the chosen user representation method. In this study, we statistically examined whether such an interaction effect exists between these factors. Our findings revealed an insignificant interaction effect across combinations of the two background types (point cloud streaming vs. graphical rendering) and the two remote user representation types (point cloud streaming vs. graphical rendering). This suggests that the observed enhancements in social telepresence, usability, and concentration, achieved by utilizing point cloud streaming for user representation over graphic-based representation, are independent of the selected background mode whether point cloud streamed or graphically rendered.

Consequently, the choice of background can be flexibly configured to either point cloud streaming or graphic-based rendering, depending on system specifications. Notably, regardless of the background choice, the benefits provided by streaming the remote user via point cloud remain intact across various facets. Our results further imply that the confluence of background and remote user representation does not limit user experience in terms of social telepresence, usability, or concentration. This aligns with prior studies [20,21,44], which either integrated a realistically rendered remote user into a virtual background or juxtaposed a virtually depicted remote user against a realistic background.

6.5. Limitations and Future Works

Our study had several limitations, which can be considered in subsequent research. First, our system relied on a singular depth camera, which resulted in obscured instances or missing RGB-D data from remote users. This limitation might be alleviated by amalgamating data from multiple depth cameras or employing temporal information strategies to compensate for current frame data deficiencies, drawing on data from preceding frames—similar to the method elucidated by Liu et al. [19]. Moreover, our current system architecture pivots on TCP/IP and WebRTC, constraining networking to circumscribed environments connected to a shared Wi-Fi network. To transcend this constraint, future adaptations could deploy a TURN server, possibly leveraging coTURN, facilitating communication via public IP addresses external to the immediate network.

Another confounding factor pertains to the caliber of graphical avatars enlisted for user experience appraisal. These avatars were not designed to mirror the facial expressions of remote users. Augmenting user experience could entail harnessing face motion capture of remote users, yielding avatars that faithfully emulate actual facial expressions. Subsequent research might also weigh the user experience relative to the avatar's verisimilitude, assessing, for instance, if facial expressions are rendered with fidelity or if textures are rendered with realism. This nuanced exploration would yield a richer understanding of avatars' influence on user experience. Jo et al. [44] measured user experiences with remote users in an HMD-based telepresence environment using a virtual avatar accurately reflecting facial expressions and a 3D-scanned lifelike avatar. Interestingly, no significant difference was observed between the virtual and 3D-scanned lifelike avatars. They concluded that the precise reflection of facial expressions and behaviors exerts a more crucial influence. The disparity between their findings and ours highlights the need for further investigations into the influence of the fidelity of an avatar's behavioral and visual expressions on presence.

7. Conclusions

By tracking the user's face and synchronizing the graphic space with the user's head movement, we could create the illusion of depth on a flat-panel display. Based on this, we developed a screen-based teleconferencing system called "Portal Display" that provides

multiple configurations for the background and remote user representation, including point cloud streaming and graphical rendering. As our system allows stereoscopic teleconferencing with graphical and streamed backgrounds and remote user representation using a single depth camera, we believe this system can be implemented using mobile devices equipped with a built-in depth camera. We conducted a user study comparing usability, social telepresence, and concentration when provided with different background and remote user representation configurations. The results showed that the background representation type did not significantly affect usability and concentration for remote users, presenting possibilities for a computationally efficient method by replacing the actual background with graphical rendering. For a remote user's representation, the point cloud streaming method is recommended as it substantially improves social telepresence. We suggest that the advantages of the graphical background revealed in our study, such as in user concentration, will guide future research in resolving the technical and cost constraints of the existing screen-based teleconferencing system.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/electronics12204339/s1, Video S1: Real-time Implementation of Composite Linear Transformations in Unity. Video S2: Dynamic Stereo Disparity Adaptation in Portal Display. Video S3: Implementing Two Representations of Remote Users: Point Clouded vs. Graphic Rendered.

Author Contributions: Conceptualization, S.K. (Seongjun Kang) and G.K.; methodology, S.K. (Seongjun Kang); software, S.K. (Seongjun Kang); validation, S.K. (Seongjun Kang), G.K. and S.K. (SeungJun Kim); formal analysis, S.K. (Seongjun Kang); investigation, S.K. (Seongjun Kang); resources, S.K. (Seongjun Kang); data curation, S.K. (Seongjun Kang); writing—original draft preparation, S.K. (Seongjun Kang) and G.K.; writing—review and editing, G.K., K.-T.L. and S.K. (SeungJun Kim); visualization, S.K. (Seongjun Kang); supervision, S.K. (SeungJun Kim); project administration, S.K. (SeungJun Kim). All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Technology Innovation Program (20018295, Meta-human: a virtual cooperation platform for a specialized industrial services) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea).

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of the Gwangju Institute of Science and Technology.

Informed Consent Statement: Informed consent was obtained from all subjects involved in this study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1

This appendix presents the scripts designed to emulate a conversational scenario between users located in separate spaces through the Portal Display system. Participants were provided with six conversational topics of similar length and plot. The scenario was replicated by prompting participants to respond to these topics.

Topic 1: Virtual tourism.
Q. This time, I would like to ask your opinion on "virtual tourism," one of the possible activities in the virtual world. (...omitted...) Would it be useful if VR virtual tourism, in which individual avatars look like the real world beyond these metabus tourism contents, proceeded?
A. (Experiment's answer)
Q. (Simple question)
A. (Experiment's answer)
Q. Oh, I see. Thank you for your good opinion. Let's move on to the next experiment.

Topic 2: Virtual Education.
Q. This time, I would like to ask your opinion on "virtual education," one of the possible activities in the virtual world. (...omitted...) Would it be useful if VR virtual classes were held where individual avatars looked like the real world beyond non-face-to-face lectures?
A. (Experiment's answer)
Q. (Simple question)
A. (Experiment's answer)
Q. Oh, I see. Thank you for your good opinion. Let's move on to the next experiment.

Topic 3: Virtual Game.
Q. This time, I would like to ask your opinion on "virtual game," one of the possible activities in the virtual world. (...omitted...) Would it be useful if VR virtual games were released where individual avatars look like the real world beyond metabus games?
A. (Experiment's answer)
Q. (Simple question)
A. (Experiment's answer)
Q. Oh, I see. Thank you for your good opinion. Let's move on to the next experiment.

Topic 4: A virtual concert.
Q. This time, I would like to ask your opinion on "virtual concert," one of the possible activities in the virtual world. (... omitted...) Would it be useful if there was a VR virtual concert where individual avatars looked like the real world beyond this virtual fan signing event?
A. (Experiment's answer)
Q. (Simple question)
A. (Experiment's answer)
Q. Oh, I see. Thank you for your good opinion. Let's move on to the next experiment.

Topic 5: Virtual shopping.
Q. This time, I would like to ask your opinion on "virtual shopping," one of the possible activities in the virtual world. (...omitted...) Would it be useful to say that VR shopping, where individual avatars look like the real world, goes beyond Internet shopping?
A. (Experiment's answer)
Q. (Simple question)
A. (Experiment's answer)
Q. Oh, I see. Thank you for your good opinion. Let's move on to the next experiment.

Topic 6: A virtual meeting.
Q. This time, I would like to ask your opinion on "virtual meeting," one of the possible activities in the virtual world. (...omitted...) Would it be useful if VR meetings where individual avatars look like the real world beyond video conferences such as ZOOM were held in the future?
A. (Experiment's answer)
Q. (Simple question)
A. (Experiment's answer)
Q. Oh, I see. Thank you for your good opinion. Let's move on to the next experiment.

Figure A1. Scripts for topics such as culture, education, and content in virtual worlds.

Appendix A.2

This appendix introduces the car models crafted to simulate a collaborative scenario between users in different locations using the Portal Display system. Participants were given six models with comparable assembly complexities. The collaborative scenario was reenacted by guiding participants both verbally and with hand gestures to assemble the blocks.

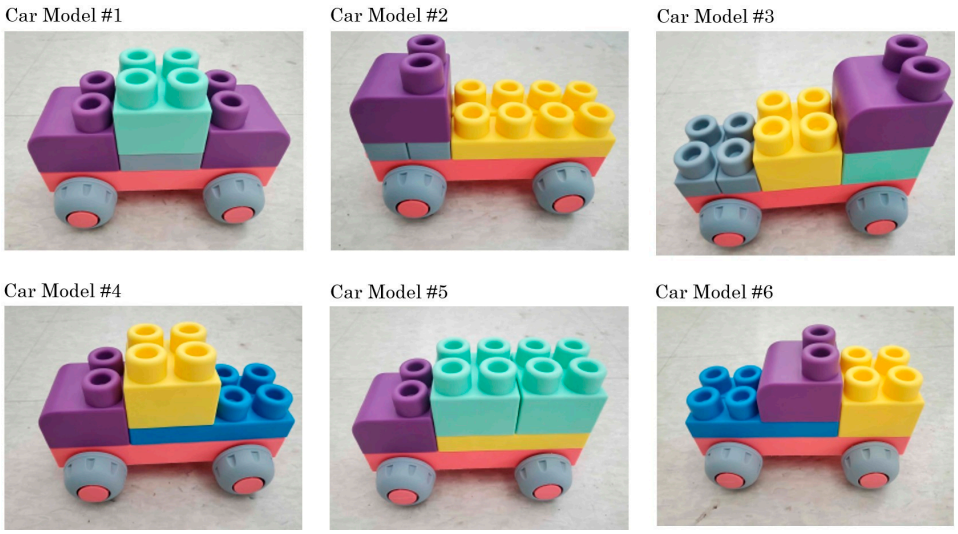


Figure A2. Car block models with similar levels of difficulty.

References

1. Daly-Jones, O.; Monk, A.; Watts, L. Some advantages of video conferencing over high-quality audio conferencing: Fluency and awareness of attentional focus. *Int. J. Hum.-Comput. Stud.* **1998**, *49*, 21–58. [CrossRef]
2. Junuzovic, S.; Inkpen, K.; Tang, J.; Sedlins, M.; Fisher, K. To see or not to see: A study comparing four-way avatar, video, and audio conferencing for work. In Proceedings of the 2012 ACM International Conference on Supporting Group Work, Sanibel Island, FL, USA, 27–31 October 2012; pp. 31–34.
3. Maloney, D.; Freeman, G.; Wohn, D.Y. "Talking without a Voice" Understanding Non-verbal Communication in Social Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* **2020**, *4*, 1–25. [CrossRef]
4. Nguyen, D.; Canny, J. Multiview: Spatially faithful group video conferencing. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Portland, OR, USA, 2–7 April 2005; pp. 799–808.
5. Adeboye, D. *Exploring the Effect of Spatial Faithfulness on Group Decision-Making*; Technical Report 952; University of Cambridge, Computer Laboratory: Cambridge, UK, 2020.
6. Wang, X.; Love, P.E.; Kim, M.J.; Wang, W. Mutual awareness in collaborative design: An Augmented Reality integrated telepresence system. *Comput. Ind.* **2014**, *65*, 314–324. [CrossRef]
7. Kuster, C.; Popa, T.; Bazin, J.-C.; Gotsman, C.; Gross, M. Gaze correction for home video conferencing. *ACM Trans. Graph.* **2012**, *31*, 1–6. [CrossRef]
8. Avrahami, D.; van Everdingen, E.; Marlow, J. Supporting Multitasking in Video Conferencing using Gaze Tracking and On-Screen Activity Detection. In Proceedings of the 21st International Conference on Intelligent User Interfaces, Sonoma, CA, USA, 7–10 March 2016; pp. 130–134.
9. Neureiter, K.; Murer, M.; Fuchsberger, V.; Tscheligi, M. Hand and eyes: How eye contact is linked to gestures in video conferencing. In Proceedings of the CHI'13 Extended Abstracts on Human Factors in Computing Systems, Paris, France, 27 April–2 May 2013; pp. 127–132.
10. Nguyen, D.T.; Canny, J. Multiview: Improving trust in group video conferencing through spatial faithfulness. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 28 April–3 May 2007; pp. 1465–1474.
11. Wang, W.; Wang, X.; Wang, R. A Spatial Faithful Cooperative System Based on Mixed Presence Groupware Model. In *Cooperative Design, Visualization, and Engineering, Proceedings of the International Conference on Cooperative Design, Visualization and Engineering, Luxembourg, 20–23 September 2009*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 269–275.
12. Gaver, W.W.; Smets, G.; Overbeeke, K. A virtual window on media space. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 7–11 May 1995; pp. 257–264.
13. Nakanishi, H.; Murakami, Y.; Kato, K. Movable cameras enhance social telepresence in media spaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, 4–9 April 2009; pp. 433–442.
14. Mulligan, J.; Zabulis, X.; Kelshikar, N.; Daniilidis, K. Stereo-based environment scanning for immersive telepresence. *IEEE Trans. Circuits Syst. Video Technol.* **2004**, *14*, 304–320. [CrossRef]
15. Maimone, A.; Fuchs, H. A first look at a telepresence system with room-sized real-time 3d capture and life-sized tracked display wall. In Proceedings of the ICAT, Osaka, Japan, 28–30 November 2011; pp. 4–9.
16. Dou, M.; Shi, Y.; Frahm, J.M.; Fuchs, H.; Mauchly, B.; Marathe, M. Room-sized informal telepresence system. In Proceedings of the 2012 IEEE Virtual Reality Workshops (VRW), Costa Mesa, CA, USA, 4–8 March 2012; pp. 15–18.
17. Desai, K.; Raghuraman, S.; Jin, R.; Prabhakaran, B. QoE studies on interactive 3D tele-immersion. In Proceedings of the 2017 IEEE International Symposium on Multimedia (ISM), Taichung, Taiwan, 11–13 December 2017; pp. 130–137.
18. Ebrahimi, T.; Alexiou, E.; Fonseca, T.A.; de Queiroz, R.L.; Torlig, E.M. A novel methodology for quality assessment of voxelized point clouds. In Proceedings of the Applications of Digital Image Processing XLI, San Diego, CA, USA, 19–23 August 2018; Volume 10752, pp. 174–190.
19. Liu, S.; Chou, P.A.; Zhang, C.; Zhang, Z.; Chen, C.W. Virtual view reconstruction using temporal information. In Proceedings of the 2012 IEEE International Conference on Multimedia and Expo, Melbourne, VIC, Australia, 9–13 July 2012; pp. 115–120.
20. Kauff, P.; Schreer, O. An immersive 3D video-conferencing system using shared virtual team user environments. In Proceedings of the 4th International Conference on Collaborative Virtual Environments, Bonn, Germany, 30 September–2 October 2002; pp. 105–112.
21. Jo, D.; Kim, K.H.; Kim, G.J. Effects of avatar and background representation forms to co-presence in mixed reality (MR) tele-conference systems. In Proceedings of the SIGGRAPH ASIA 2016 Virtual Reality meets Physical Reality: Modelling and Simulating Virtual Humans and Environments, Macau, China, 5–8 December 2016; pp. 1–4.
22. Tanger, R.; Kauff, P.; Schreer, O.; Pavy, D.; Louis Dit Picard, S.; Saugis, G. Team collaboration mixing immersive video conferencing with shared virtual 3D objects. In *Signals and Communication Technology, Proceedings of the Distributed Cooperative Laboratories: Networking, Instrumentation, and Measurements*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 3–10.
23. Yu, K.; Gorbachev, G.; Eck, U.; Pankratz, F.; Navab, N.; Roth, D. Avatars for Teleconsultation: Effects of Avatar Embodiment Techniques on User Perception in 3D Asymmetric Telepresence. *IEEE Trans. Vis. Comput. Graph.* **2021**, *27*, 4129–4139. [CrossRef] [PubMed]

24. Petkov, E. Generation of Stereo Images in 3D Graphics Applications for Stereoscopic and Nonstereoscopic Displays. In *Computer Science and Technologies*; University of Varna: Varna, Bulgaria, 2012; p. 47.
25. Vápeník, R.; Michalko, M.; Janitor, J.; Jakab, F. Secured web oriented videoconferencing system for educational purposes using WebRTC technology. In Proceedings of the 2014 IEEE 12th IEEE International Conference on Emerging eLearning Technologies and Applications (ICETA), Stary Smokovec, Slovakia, 4–5 December 2014; pp. 495–500.
26. Ryskeldiev, B.; Cohen, M.; Herder, J. StreamSpace: Pervasive Mixed Reality Telepresence for Remote Collaboration on Mobile Devices. *J. Inf. Process.* **2018**, *26*, 177–185. [CrossRef]
27. Wen, W.-C.; Towles, H.; Nyland, L.; Welch, G.; Fuchs, H. Toward a compelling sensation of telepresence: Demonstrating a portal to a distant (static) office. In Proceedings of the Visualization 2000. VIS 2000 (Cat. No. 00CH37145), Salt Lake City, UT, USA, 8–13 October 2000; pp. 327–333.
28. Tanaka, K.; Nakanishi, H.; Ishiguro, H. Physical embodiment can produce robot operator's pseudo presence. *Front. ICT* **2015**, *2*, 8. [CrossRef]
29. Zillner, J.; Rhemann, C.; Izadi, S.; Haller, M. 3D-board: A whole-body remote collaborative whiteboard. In Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, Honolulu, HI, USA, 5–8 October 2014; pp. 471–479.
30. Onishi, Y.; Tanaka, K.; Nakanishi, H. Embodiment of video-mediated communication enhances social telepresence. In Proceedings of the Fourth International Conference on Human Agent Interaction, Singapore, 4–7 October 2016; pp. 171–178.
31. Kim, K.; Bolton, J.; Girouard, A.; Cooperstock, J.; Vertegaal, R. Telehuman: Effects of 3d perspective on gaze and pose estimation with a life-size cylindrical telepresence pod. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, TX, USA, 5–10 May 2012; pp. 2531–2540.
32. Nakanishi, H.; Murakami, Y.; Nogami, D.; Ishiguro, H. Minimum movement matters: Impact of robot-mounted cameras on social telepresence. In Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, San Diego, CA, USA, 8–12 November 2008; pp. 303–312.
33. Brooke, J. SUS: A retrospective. *J. Usability Stud.* **2013**, *8*, 29–40.
34. Lee, J.C. Hacking the Nintendo Wii Remote. *IEEE Pervasive Comput.* **2008**, *7*, 39–45. [CrossRef]
35. Cruz-Neira, C.; Sandin, D.J.; DeFanti, T.A.; Kenyon, R.V.; Hart, J.C. The CAVE: Audio visual experience automatic virtual environment. *Commun. ACM* **1992**, *35*, 64–72. [CrossRef]
36. Manjrekar, S.; Sandilya, S.; Bhosale, D.; Kanchi, S.; Pitkar, A.; Gondhalekar, M. CAVE: An emerging immersive technology–a review. In Proceedings of the 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Cambridge, UK, 26–28 March 2014; pp. 131–136.
37. Cruz-Neira, C.; Sandin, D.J.; DeFanti, T.A. Surround-screen projection-based virtual reality: The design and implementation of the CAVE. In *Seminal Graphics Papers: Pushing the Boundaries*; Association for Computing Machinery: New York, NY, USA, 2023; Volume 2, pp. 51–58.
38. Lawrence, J.; Goldman, D.B.; Achar, S.; Blascovich, G.M.; Desloge, J.G.; Fortes, T.; Gomez, E.M.; Häberling, S.; Hoppe, H.; Huibers, A.; et al. Project Starline: A high-fidelity telepresence system. *ACM Trans. Graph.* **2021**, *40*, 242. [CrossRef]
39. Caviedes, J.E.; Wu, S.L. Combining computer vision and video processing to achieve immersive mobile videoconferencing. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 2467–2471.
40. Nowak, K.L.; Biocca, F. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence Teleoper. Virtual Environ.* **2003**, *12*, 481–494. [CrossRef]
41. Zinchenko, Y.P.; Kovalev, A.I.; Menshikova, G.; Shaigerova, L.A. Postnonclassical methodology and application of virtual reality technologies in social research. *Psychol. Russ. State Art* **2015**, *8*, 60–71. [CrossRef]
42. Kang, S.; Yoon, B.; Kim, B.; Woo, W. Effects of Avatar Face Level of Detail Control on Social Presence in Augmented Reality Remote Collaboration. In Proceedings of the 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Singapore, 17–21 October 2022; pp. 763–767.
43. Zhang, J.; Huang, W.; Zhu, X.; Hwang, J.N. A subjective quality evaluation for 3D point cloud models. In Proceedings of the 2014 International Conference on Audio, Language and Image Processing, Shanghai, China, 7–9 July 2014; pp. 827–831.
44. Jo, D.; Kim, K.H.; Kim, G.J. Effects of avatar and background types on users' co-presence and trust for mixed reality-based teleconference systems. In Proceedings of the 30th Conference on Computer Animation and Social Agents, Seoul, Republic of Korea, 22–24 May 2017; pp. 27–36.
45. Lee, M.; Park, W.; Lee, S.; Lee, S. Distracting moments in videoconferencing: A look back at the pandemic period. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 29 April–5 May 2022; pp. 1–21.
46. Subramanyam, S.; Viola, I.; Hanjalic, A.; Cesar, P. User centered adaptive streaming of dynamic point clouds with low complexity tiling. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 3669–3677.

47. Wang, Z.R.; Yang, C.G.; Dai, S.L. A Fast Compression Framework Based on 3D Point Cloud Data for Telepresence. *Int. J. Autom. Comput.* **2020**, *17*, 855–866. [CrossRef]
48. Van Holland, L.; Stotko, P.; Krumpen, S.; Klein, R.; Weinmann, M. Efficient 3D Reconstruction, Streaming and Visualization of Static and Dynamic Scene Parts for Multi-client Live-telepresence in Large-scale Environments. *arXiv* **2022**, arXiv:2211.14310.