Augmented Virtual Teleportation for High-Fidelity Telecollaboration

Taehyun Rhee, Stephen Thompson, Daniel Medeiros, *Member, IEEE*, Rafael dos Anjos, and Andrew Chalmers

Fig. 1. Augmented Virtual Teleportation - an asymmetric platform for remote collaboration. From the left: a remote traveler wearing a VR HMD, the Mixed Reality collaboration space seen in their display, the space shown on an AR display, and the local host.

Abstract—Telecollaboration involves the teleportation of a remote collaborator to another real-world environment where their partner is located. The fidelity of the environment plays an important role for allowing corresponding spatial references in remote collaboration. We present a novel asymmetric platform, *Augmented Virtual Teleportation* (AVT), which provides high-fidelity telepresence of a remote VR user (*VR-Traveler*) into a real-world collaboration space to interact with a local AR user (*AR-Host*). AVT uses a 360° video camera (360-camera) that captures and live-streams the omni-directional scenes over a network. The remote VR-Traveler watching the video in a VR headset experiences live presence and co-presence in the real-world collaboration space. The VR-Traveler's movements are captured and transmitted to a 3D avatar overlaid onto the 360-camera which can be seen in the AR-Host's display. The visual and audio cues for each collaborator are synchronized in the Mixed Reality Collaboration space (MRC-space), where they can interactively edit virtual objects and collaborate in the real environment using the real objects as a reference. High fidelity, real-time rendering of virtual objects and seamless blending into the real scene allows for unique mixed reality use-case scenarios. Our working prototype has been tested with a user study to evaluate spatial presence, co-presence, and user satisfaction during telecollaboration. Possible applications of AVT are identified and proposed to guide future usage.

Index Terms—Telepresence, Collaboration, Real-time, Mixed Reality, 360° Panoramic Video

1 Introduction

There has been a growing interest in using virtual reality (VR) and augmented reality (AR) to provide fully virtual spaces or combined virtual and real-world spaces for people to interact and communicate. Unlike previous video conferencing or voice chat, VR and AR provides more natural interaction in remote collaboration by improving the sense of presence and co-presence to be closer to face-to-face meetings.

One important aspect in VR is to create a sense of presence in virtual spaces defined by computer-generated, three-dimensional, interactive real-time systems. In VR, the user's physical body stays in a "physical space", but their senses and feelings are transported into the "virtual space" [60]. In VR-based collaboration systems, users are presented in a shared virtual space where they can communicate and interact with virtual objects and avatars. A physical space can be recreated virtually with varying levels of fidelity - the measure of how exact real-world stimuli is reproduced [19]. A virtual environment with a high level of fidelity closely resembles its real-world counterpart. However, recreating a real-world space with high fidelity is a difficult task that often requires sophisticated setup for capturing, modeling, and realistic rendering.

Using AR for collaboration is a viable alternative where these variables play a smaller role. In AR, people are located in a real-world environment, where one is still able to communicate with the remote person and interact with virtual content [34, 46]. Recent attempts in this matter include projections of remote people onto walls [53] and realistic reconstructions of remote people and objects [49] that are visualized with see-through head mounted displays (HMDs). These scenarios allow for people to collaborate while maintaining their own personal space with a high sense of presence.

However, in practical use cases of teleconferencing (e.g. remote assistance, inspection and planning), the remote traveler needs to be "teleported" to a real-world environment where their collaborative partner is located. In this case, the fidelity of the environment plays a big role in allowing spatial references to be used accurately and providing a high sense of telepresence (i.e. the perception of presence within a physically remote or simulated site [14]) for the traveler. Using a combination of VR and AR in an *asymmetrical* setup is a possible solution, as shown through the Beaming project [28, 47, 62], providing a high sense of telepresence to a remote traveler collaborating with an AR user. Although the Beaming project describes a viable framework to design an asymmetrical system, no formal evaluation was conducted to test how sense of presence and co-presence compared between VR and AR users. Neither has the literature described or tested a cost-effective, asymmetrical system with high fidelity.

Creating a high-fidelity, real-world environment in VR through 3D

• *Taehyun Rhee is with Victoria University of Wellington. E-mail: taehyun.rhee@ecs.vuw.ac.nz.*
• *Stephen Thompson is with Victoria University of Wellington. E-mail: stephen.thompson@ecs.vuw.ac.nz.*
• *Daniel Medeiros is with Victoria University of Wellington. E-mail: daniel.medeiros@vuw.ac.nz.*
• *Rafael dos Anjos is with Victoria University of Wellington. E-mail: rafael.dosanjos@vuw.ac.nz.*
• *Andrew Chalmers is with Victoria University of Wellington. E-mail: Andrew.Chalmers@ecs.vuw.ac.nz.*

Manuscript received 10 Sept. 2019; accepted 5 Feb. 2020.
Date of publication 18 Feb. 2020; date of current version 27 Mar. 2020.
Digital Object Identifier no. 10.1109/TVCG.2020.2973065

reconstruction [46, 64] is challenging, thereby limiting a high sense of presence in live telecollaboration [73]. One recent viable solution to this problem is using 360° capturing cameras and videos (360-camera/video) in VR scenarios, which has been shown to provide a high sense of presence [37, 68]. Moreover, recent work has enabled virtual objects to be realistically blended into 360-videos [57], allowing users to have the illusion of interacting with content inside 360-video streams and enabling the design of novel types of interaction in 360-videos.

This paper presents a novel telecollaboration platform, "Augmented Virtual Teleportation" (AVT), that converges VR and AR in an asymmetric system. We implemented AVT by adapting a 360-camera, which is able to capture omni-directional scenes of the real-world collaboration space, and live-streaming it to a remote user. The remote user (VR-Traveler) wears a VR HMD and feels telepresence in the collaboration space with a wide field of regard from the 360-video, and co-presence with the local user with visual and audio cues through the networked connection. The local user in the collaboration space (AR-Host) sees the VR-Traveler's avatar through an AR display. The movements of the VR-Traveler's head and hands are captured by dedicated trackers, transferred to the collaboration space, and overlaid onto the position of the 360-camera in the AR-Host's view. The coordinates of the VR-Traveler and AR-Host are aligned in a mixed reality collaboration space (MRC-space) and supports synchronized visual and audio cues for collaboration. Then, both VR and AR users interactively add, delete, and manipulate virtual objects in the MRC-space while using the real-world objects as references. For high visual fidelity of MRC-space, we illuminate virtual objects using the 360-video as the real-world light source, and blend the virtual objects seamlessly into streamed real-world videos. We conducted a user study to evaluate presence and user satisfaction for remote collaboration using AVT. The results show that both remote and local users felt a high sense of spatial presence and co-presence with high user satisfaction in telecollaboration. Finally, we identify and propose applications that can be done using the AVT platform in order to provide effective collaboration for remote users. To the best of our knowledge, this is the first paper to present and evaluate high fidelity telecollaboration in a realistic mixed reality space with 360-videos, enabling natural remote collaborations using a face-to-face paradigm. The main contributions of our paper are summarized as follows:

- We present AVT, a novel telecollaboration platform to teleport a remote VR-Traveler into a mixed reality collaboration space (MRC-space) to interact with the AR-Host in another physical space.
- We developed AVT using a novel asymmetric setup converging VR and AR in a unified framework that provides synchronized visual and audio cues for both VR and AR users in telecollaboration.
- Our MRC-space provides a high fidelity collaboration space, where remote collaborators can interact with virtual objects using the real world as a reference.
- We evaluate AVT with a user study to measure spatial presence, co-presence, and user satisfaction. We also propose several scenarios for feasible applications.

2 Related Work

2.1 Presence and Co-presence

Presence, Telepresence: The VR experience is strongly connected to the sense of presence and telepresence [63]. Presence has been previously defined in multiple ways: the feeling of being physically present with the virtual objects [59], or as a combination of spatial presence, involvement and realism [58]. Ultimately, the definitions can be summarized as a complete feeling of "being there" [22] in the virtual environment.

Many factors can support or hinder one's feeling of presence. HMDs containing positional and rotational freedom provides a heightened sense of presence compared to traditional displays [11]. Factors such as mismatching visual and sensory information felt by the user can negatively affect the feeling of presence [72]. Regenbrecht et al. [56] were able to maintain a high sense of presence in Mixed Reality (MR) by decreasing the visual quality of the real environment. In doing so they could provide visual coherence in rendering between the real and virtual. On the other hand, it has been shown that increasing the realism of objects and environments can also increase the sense of presence [24, 52]. Preliminary results by Bouchard. et al. [8] suggests that people who perceive the virtual environment as a real place had a significant increase in feeling presence and telepresence.

Co-presence: Co-presence is the sense that someone is present with other people in a remote or shared environment [10, 20, 74]. People who are not co-located in the AR/MR space need to be represented by an avatar. The avatar appearance impacts not only presence but also co-presence in collaborative virtual environments [20]. Realistic human avatars are generally better for user-presence [23], but often require complex setups for reconstructing real people with high-fidelity [49]. The level of graphical fidelity of avatars are known to be dependant on both the task and perspective they are used. Abstract avatars are often preferred for tasks requiring precise object manipulations [44] and locomotion tasks [42]. Realistic avatars are also more susceptible to uncanny valley effects, due to both graphical and motion fidelity factors [39]. Another study found the difference between using partial body and full body was not significantly worse, but using motion generated from the user as opposed to predefined animation was better [32]. Prior work have used digitized photos and videos to act as reference points for collaborative activities for co-located and remote users [30].

2.2 360° panoramic Video

While traditional VR experiences use fully virtual environments, another option is the use of a 360° panoramic images [12] and videos [54] (360-video). 360-videos give users a high sense of telepresence in a captured real-world environment while providing the freedom to view the environment in any direction.

The use of 360-videos has been shown to provide a high sense of presence, leading the users to feel more emotionally attached to the experience [15, 16]. Using higher resolutions and higher quality encoding, especially when viewing with an HMD, helps with improving the sense of presence [69]. In the same study, it was found that having little to no motion was best for reducing nausea, however having a small amount of motion improves presence. Evidence suggests that viewing 360-video in a HMD provides better spatial awareness and enjoyment over traditional display systems [41]. Viewing stereoscopic 360-video has not been found to improve presence over monoscopic 360-video [5].

2.3 Mixed Reality Rendering

Matching the visual quality between virtual objects and the real world helps to improve the sense of presence experienced by users [56, 57]. Realistic lighting of virtual objects and seamless blending into real-world footage are important for high-fidelity MR.

Debevec et. al. [13] presented differential rendering for achieving visually convincing composition. Geometric representation of the real word scene is required for mutual inter-reflections (shadowing, color bleeding) between the virtual and real objects. It can be reconstructed using RGB-D cameras (e.g. image depth [21, 38], spatial mapping [9]), camera motion [50, 67], or dedicated 3D scanners such as LIDAR.

Image based lighting (IBL) uses a high dynamic range 360° image of the scene as a light source [13]. Recent work in this field has tackled the problems of IBL using boosted dynamic range images [25], realistic shadowing and light estimation [57], and reflections from 360 video [67]. These approaches were proven to contribute to a high sense of presence in MR scenarios, which is suitable for teleconferencing scenarios.

2.4 Collaboration and Communication Systems

Research has started exploring collaboration between AR and VR users following the introduction of video and optical see-through HMDs [43].

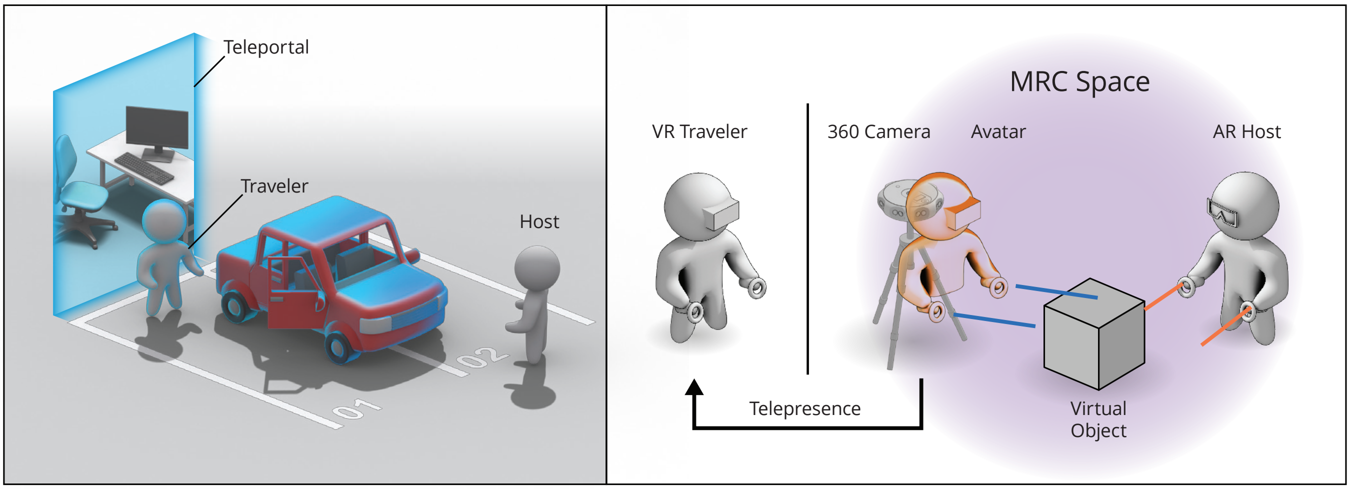


Fig. 2. Overview; A concept of remote collaboration using teleportation (left), and our AVT platform using an asymmetric VR-AR setup (right).

In this context, virtual objects can be used as instructions to enable collaboration in remote assistance scenarios [48]. This has been shown to be be effective for sharing context-related information in AR, and improving the situational awareness of remote colleagues [40]. Studies have supported this idea, showing that AR is especially helpful in a collaborative setting [4]. 3D reconstruction of the local scene can also be sent to the remote user to provide 6 degrees of freedom (DOFs) to explore the local scene [1,64]. Microsoft Research created a system that live-streams a 3D reconstruction of people in a remote space to the local host wearing a HoloLens device [49], giving the local user the illusion of seeing the remote user. However effective, this system requires high-end hardware in order to process and stream the reconstruction in real-time. Kolkmeier et al. [29] extended it to an asymmetric setup providing a shared collaboration space represented by a 3D point-cloud captured from RGB-D sensors. The visual quality of the 3D representation is low in the remote traveler's view and limits their presence in the shared space. This can cause a problem in remote collaboration when the visual quality of the room plays an important role. For these contexts, previous work suggests that 360-video is a better option for a cost-effective high-fidelity solution [68], providing higher levels of presence in collaborative scenarios.

Wearable cameras live-streaming to remote users can be used to help with communication [6, 26]. Velamkayala et al. [70] found that video conferencing using the HoloLens improves collaboration when comparing to mobile devices. Live streaming video of the local scene along with corresponding depth can be sent from to the remote user [3, 17, 18, 61]. A potential limitation of wearable cameras is that the view of the remote users can become fully or partially dependent on the local user. View independence between the local and remote user improves the speed of collaboration [65] and provides freedom for the remote user [27]. The use of 360-camera to share the viewpoint of a user can provide effective collaboration, as the remote user can explore the augmented environment using rotational movements [55, 68].

Recent research has explored using live-streaming 360-cameras to bring remote users into the local space. Remote users are able to view the local space through a HMD, mobile smartphone device [66], or even by navigating a web browser based interface for viewing and interacting with the 360-video. Lee et al. [37] proposed a system where the remote user can view the local scene through a live-streaming 360-video. A device was placed next to the remote user to provide 4D effects (e.g. wind or heat) for improved sense of presence.

A different system [36] was proposed for collaboration where the local user could wear a see-through AR display along with a 360° live streaming camera. The remote user could view the 360° live stream and annotate parts of the video. Interaction with the scene was limited to overlaid information and the remote user's view was partially coupled to the local user. Another prototype tested different avatar visualisations and the best placement of the 360-camera in a multi-scale setup [55]. A study was conducted exploring collaboration using a 360-video compared to 3D reconstructed geometry [68]. It concluded that the captured high fidelity 360-video provides better collaboration and presence compared to a static render of the real geometry. A common drawback of these works is that only the remote person is assessed by the user evaluation and an actor is used to direct the experience locally.

The concept of asymmetric collaboration was explored in the Beaming project [28, 47, 62], where the "teleporting/beaming" of one user to a different environment was shown to allow simple communication between users in a shared environment. Other works have explored this paradigm in the following years, but suffer from low visual fidelity of the environment or super-imposed objects [29, 56], which may hinder the sense of presence. Furthermore, since collaborators sometimes share their viewing position [55, 68], they mostly rely on pointing gestures instead of using complex body language cues (posture, gaze, etc.). AVT overcomes these limitations by providing face-to-face communication in a high-fidelity MRC-space.

3 AVT Overview

The concept of telecollaboration requires that a host in a specific physical location sets a teleportal. Then, the traveler in a different physical location teleports through the teleportal to the host's space and uses their tools for collaboration. Once there, they are able to look around the physical environment, communicate with the host, and perform collaborative tasks (See Figure 2, left).

We propose Augmented Virtual Teleportation (AVT), a novel asymmetric platform converging VR and AR experiences in a mixed reality framework to realize telecollaboration. The concept of teleportation is implemented by live streaming 360-video to a VR HMD worn by the remote traveler (VR-Traveler) to feel telepresence. They feel co-located with the local host and experience a high sense of presence in the captured environment. The VR-Traveler's 3D avatar is then transmitted through the network and overlaid on top of the 360-camera shown in the AR display of the local host (AR-Host). The view of the VR-Traveler and AR-Host are aligned in the MRC-space where they can manipulate virtual objects with synchronized visual and audio cues, such as they would in face-to-face interaction (See Figure 2, right).

4 AVT System

We implemented the AVT system using the Unity 3D game engine with a client-server setup for networking. A 360-camera located with the AR-Host live captures and streams the 360-video to a local server, where the 360-video is then transmitted to the remote client PC for the remote VR-Traveler to view. The client-server network was implemented over a LAN setup using UDP as shown in Figure 3. In our system, the AR-Host is the network host, storing the true state of the system and virtual objects, sharing this information to the remote client to maintain a coherent system state.

4.1 Asymmetric VR AR Framework

VR-Traveler views the live 360-video streamed from the server to the remote client. The VR-Traveler sees the video through a VR HMD providing a wide field of regard with 3-DOF view rotation. The AR-Host who set the teleportal is visible in the live-streaming video with included audio cues. The VR-Traveler's head and hand motions are captured by the HMD and controllers. The positions and orientations are streamed over the network to the AR-Host for updating the position, rotation, and pose of the avatar.

AR-Host views the real-world collaboration space through an AR capable device. In our implementation we used a video see-through HMD created by attaching a ZED Mini camera to the front of a VR HMD. The ZED Mini Unity plugin was used to integrate the ZED Mini into Unity, ensuring the MRC-Space is viewed correctly through the ZED Mini camera. This setup enables a wide field of view for the AR-Host (compared to state-of-the-art see-through HMD [33, 45]) to explore the MRC-Space with 6-DOF. A 3D avatar of the VR-Traveler appears to the AR-Host rendered as a partial body (Figure 4), using the transferred traveler's head and hand motions to animate the 3D avatar. We apply inverse kinematics to the arms to generate natural movement. The avatar is shaded as a glowing outline to avoid possible uncanny valley from realistic shading [32].

Remote and local users are provided with their own personal menu interface. It can only be interacted with by the individual, appearing as only a simple textured quad to the other user. Personal menus are used for scene calibration, avatar appearance adjustment, navigation and creating shared windows (See Section 5.4).

4.2 Mixed Reality Collaboration Space (MRC-space)

4.2.1 360-video live streaming

The 360-camera is placed in the real-world collaboration space at eye height and is connected to the server PC. We use OBS Studio to encode the stitched video from the 360-camera in H264 encoding set for low latency streaming. The video is streamed over a local network to the client PC using an Nginx server, with its RTMP module, running on the AR-Host PC. Video streaming and playback in Unity3D is handled using the Vive Media Decoder plugin, a wrapper for the FFmpeg library. We have also modified the plugin by enabling additional options provided by FFmpeg to optimize the plugin for lower latency.

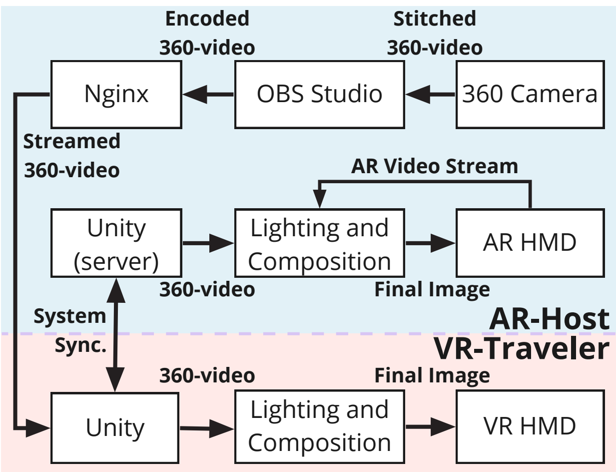


Fig. 3. System data flow: The 360-camera located with the AR-Host streams video to the VR-Traveler. Updated information is synced between the AR-Host (Server) and VR-Traveler.

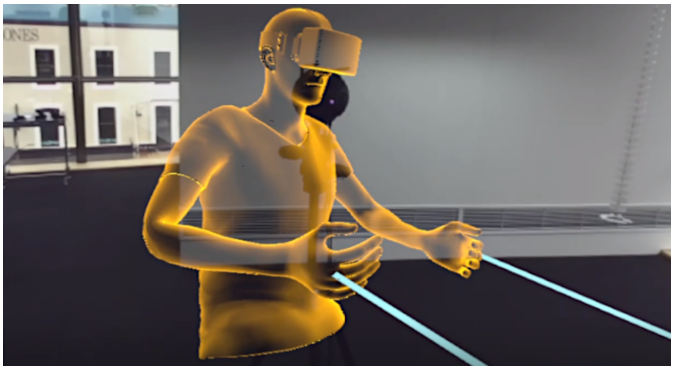


Fig. 4. VR-Traveler's avatar shown from the AR-Host's point of view.

4.2.2 Space Calibration

Coordinates of AR-Host and VR-Traveler need to be aligned in the collaboration space. We used the center of the 360-camera as the origin of the MRC-space, where the VR-Travelers eyes are positioned to view the real-world surroundings. The AR-Host manually annotates the origin and orientation of the MRC-space while aligning a 3D mock-up of the 360-camera into the physical 360-camera. The VR-Traveler can make additional refinements to the origin and orientation of the mock-up 360-camera for better alignment between the VR-Traveler and AR-Host. Once set, the VR-Traveler's position and orientation is aligned with the virtual 360-camera and the AR-Host's coordinates. This process can be automated by attaching additional markers or sensors on the 360-camera, but users often prefer the manual fine tuning.

AR HMDs often contain depth sensing for geometry reconstruction. We take advantage of this feature by using the reconstructed geometry for occlusion handling for physical simulation and mixed reality lighting. The AR-Host can access the geometry reconstruction through their personal menu. The 360-video is projected onto geometry received from the AR-Host so that virtual objects can feel grounded in the scene rather than floating in front of a monoscopic 360-video. An example of reconstructed planes are shown in Figure 5(b).

4.2.3 Mixed Reality Collaboration

Users can visually communicate through head gaze and pointing with their controllers. Color coded 3D rays projecting from the controllers help to clarify where the users are pointing, The 360-video also provides full body language communication of the AR-Host to the VR-Traveler. Verbal communication is achieved through a VoIP dedicated channel.

The AR-Host and VR-Traveler can interact with virtual objects for synchronous collaboration. We implemented physics simulation for the virtual objects using the Unity3D Physics Engine. AVT supports asynchronous object manipulation in the MRC-space where each user takes turns to manipulate the shared object. During user-to-object interactions, physics for the object is simulated on the local machine of the user in order to provide smoother, low-latency interaction for better precision when manipulating the object. Otherwise, the physics are simulated on the AR-Host server and synced with the VR-Traveler client, creating the synchronized MRC-space.

4.2.4 Lighting and Composition

Corresponding illumination between the virtual objects and the real-world surroundings is important for enabling a high-fidelity mixed reality space. To achieve this, we use the 360-video on both users' PCs as the real-world environment map for lighting the virtual objects. We extract important lighting details from the 360-video for high frequency lighting to cast shadows [57], implement image based lighting (IBL) for

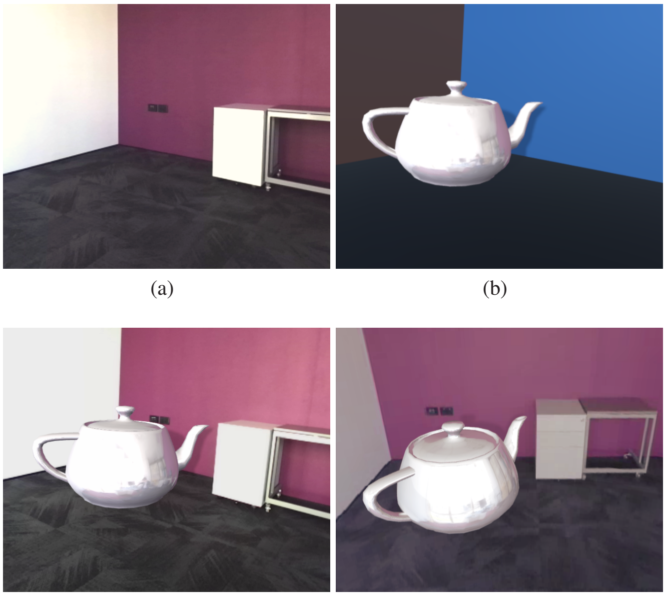


Fig. 5. Lighting and composition: From the input video in AR HMD (a), we reconstruct planes (b). 3D objects are then lit using the 360-video, providing consistent illumination and blending shown in AR (c) and VR (d) point of views.

ambient lighting [25], and finally composite the rendering with shadows using differential rendering. Our method has been implemented in Unity3D and the real-world light updates are applied to the MRC-space in real-time to provide synchronized illumination cues for both VR-Traveler and AR-Host.

Additionally, reconstructed geometry by the AR-Host using depth sensors attached into their HMD is used to reconstruct 3D planes for receiving shadows from virtual objects in the real-environment. Synchronized positions and orientations of the objects in the MRC-space ensure that both VR-Traveler and AR-Host are properly displayed for virtual objects as shown in Figure 5.

4.3 Applications

We propose three application scenarios of AVT that are implemented as working prototypes as shown in Figure 7 and the supporting video.

"Design review": When reviewing the design layout of a room, remote collaborators require a clear understanding of the surrounding environment, high-fidelity rendering of 3D virtual objects, and consistent lighting of the virtual objects with the surroundings. Figure 7(a) shows our implementation of a scenario where a remote designer teleports to the clients room, moving 3D objects to different locations. They are able to verify how the furniture interacts with the rooms surroundings with real lighting conditions, while communicating with the client.

"Distance training": Face-to-face communication promotes engagement in distance training tasks. Figure 7(b) shows a student performing a task in their laboratory, benefiting from the augmented presence of their instructor who can supervise the usage of the physical and virtual laboratory equipment.

"Remote assistance": AVT allows independent visualization for the remote user and the local user, which can be useful in remote assistance scenarios. Figure 7(c) shows a scenario where a remote user is co-located with a local user. The remote user is able to monitor risks (such as forgetting a step in a recipe) which may not noticed by the local user.

5 User Study

To evaluate the AVT platform, we conducted a user study that measured presence, co-presence and user satisfaction of telecollaboration for both the VR-travelers and AR-Host. Differently from previous works that used actors for local participants, we designed our user study to evaluate whether our system can guarantee similar conditions for both groups of users for face-to-face remote collaboration. In the following sub-sections we describe the system setup, the participants, the tasks performed, and the metrics used in our user study.

5.1 System Setup

Two separate setups for the traveler and host were implemented for our user study. Both users achieved consistent 11ms frame-rates to match the 90Hz refresh rate of the HMDs. We used a LAN network wired with 1Gbps Cat6 Ethernet cable for the 360-video streaming and system networking.

For the VR-Traveler, we used an Oculus Rift CV1 HMD setup for visualization. This HMD has a resolution of 1080x1200 per eye, a refresh rate of 90 Hz and 110 degrees of diagonal field of view (FOV). The Rift also provides six degrees of freedom (6DoF) head movement, which is enabled by two Oculus Rift Constellation sensors. To capture user input, we used two tracked Oculus Touch controllers to provide natural user interaction. The VR-Traveler used a laptop with an Intel Core I7-6700HQ CPU, Nvidia GTX 980M, and 16GB RAM.

For the AR-Host, we used a Vive Pro with a ZED Mini camera attached to the front to provide high resolution video pass-through. The ZED Mini camera operates at 1280x720 resolution per eye at 60Hz with 60ms of latency. The AR-Host used a computer with an Intel Xeon Gold 5122 CPU, Nvidia Quadro P5000 GPU, and 32GB RAM.

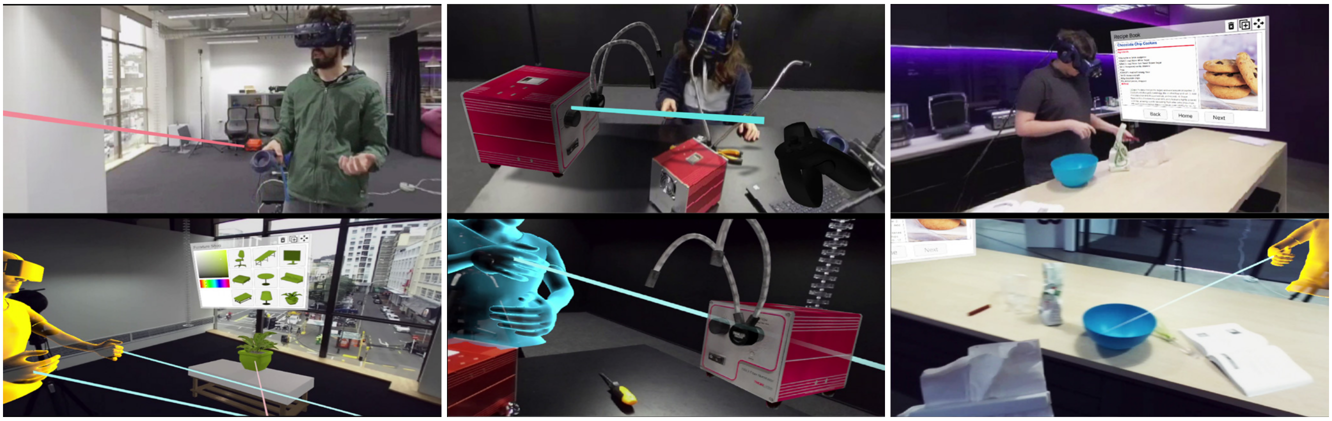
To live stream the real-world surroundings of the AR-Host to VR-traveler, we used a 360-camera mounted on top of a tripod and positioned according to the height of the VR-Traveler to improve their feeling of presence and co-presence. Optimal balance among minimal network latency, smooth frame-rate, and high visual quality is required for high-fidelity telecollaboration. We tested three different 360-cameras (Ricoh Theta V, Samsung 360 Round, Insta360 Pro 2) with respect to latency and visual quality (e.g., resolution, dynamic range, seamless stitching) to find the best suited option. For image quality, the Insta360 Pro 2 and Samsung 360 Round both provide similar levels of quality and dynamic range. The Samsung 360 Round provides the best stitching, whereas the Insta360 Pro 2 and Ricoh Theta V linearly blend seams together causing ghosting artifacts. The outputs from the three cameras can be seen in Figure 6.

Camera	Ricoh Theta V		Insta360 Pro 2		Samsung 360 Round	
Resolution	HD	4K	HD	4K	2K	4K
Preview	0.5s	0.7s	0.6s	0.6s	1.7s	1.7s
Unity	0.6s	0.8s	1.2s	1.2s	1.7s	10.5s
FFplay	0.8s	1.2s	1.5s	1.4s	2.1s	2.1s

Table 1. Approximate latency at different stages of the pipeline for each camera.

We recorded the approximate latency when the stitched video is first viewable (OBS Studio for the Ricoh Theta and proprietary software for the Insta360 Pro 2 and Samsung 360 Round) and when viewed in Unity by the VR-Traveler. We also test the latency on the VR-Traveler's laptop with FFplay, the media player that ships with FFmpeg, for reference. In each case the video is streamed at 6Mbps. Our optimized video decoder in Unity was able to stream with minimal latency in most cases, however it could become unstable and perform poorly for larger bit-rates and resolutions (as seen in the 4K Samsung 360 Round compared to the latency displayed in FFplay). Based on our test shown in Table 1, we chose the Insta360 Pro 2 for our user test setup, with HD resolution for improved stability during the user study testing.

Fig. 6. Image comparison between 360-cameras: from the left, a 360-video frame captured by Ricoh Theta V, Insta360 Pro 2, and Samsung 360 Round.



(a) Design review (b) Assisted training (c) Remote assistance

Fig. 7. Applications of AVT: The VR-Traveler's view (top) and the AR-Host's view (bottom).

5.2 Methodology

Our user study was conducted using a between-subjects design in order to compare the aforementioned metrics between the two setups and groups of participants. We can summarize our hypothesis as follows:

H_0: Both VR-Travelers and AR-Hosts have equivalent levels of social presence and both are high, when using the AVT system.

H_1: Both VR-Travelers and AR-Hosts have equivalent levels of spatial presence and both are high, when using the AVT system.

The study was conducted in a controlled environment of two physically separated rooms. Each pair of participants, located in each of the two rooms, performed two tasks that lasted an average of 40 minutes. For our user study, we recruited 40 participants for the user study with ages varying from 18 to 64 years old (average age was 28 years,SD=8.78). Participants were randomly assigned into two groups, where one experienced the system as an AR-Host and the other, the VR-traveler. Most of the users had previous experience with VR systems or video-games.

We greeted the participants, presented them with a short description of the study and the tasks they were going to perform and the consent form. Once participants had provided informed consent they were asked to fill a pre-test questionnaire to collect users' demographic and baseline information on their previous experience of using VR and AR. We randomly assigned the places where the participants experienced the asymmetric AVT setup in either VR (traveler) or AR (host). MRC-space calibration was performed prior to the participants starting the study.

Users were asked to perform the training task first before the main task. After completing the main task we asked each participant to fill out a post-test questionnaire to assess user preferences, sense of presence, social presence, and spatial presence.

5.3 Collaboration Task

We conducted two tasks for the user study, each task required two participants (VR-Traveler and AR-Host) placed in two different rooms.

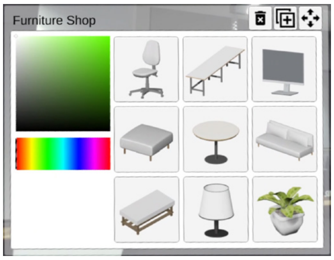
Fig. 8. Training Task (left) and Main Task (right).

Training task: To familiarize themselves with the task and system, each participant was asked to perform a training scenario where they could interact freely with the virtual objects in the MRC-space together with their remote partner. The task and setup was similar to the main task, but with a simpler configuration. The training task was to place virtual 3D furniture into the MRC-space while collaborating with a remote partner (Figure 8, left). Nine furniture models were given and users had to choose a color, then select and place them in the MRC-space. Test data was not recorded during the task. The activity allowed participants to get familiar with the system setup,interfaces and controls. Participants were given roughly 10 minutes before moving onto the main task.

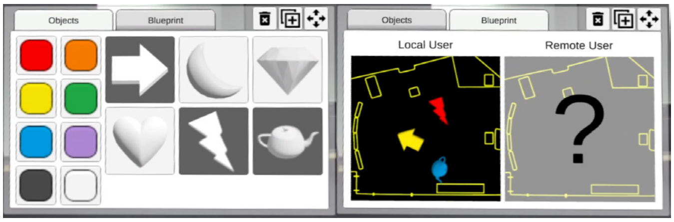
Main task: The main task was designed to encourage mutual communication between participants. For this task, the two participants were assigned as an operator and an instructor. The operator was asked to place three objects guided by the remote instructor, using a layout blueprint illustrating what the objects are and where to place them. Only the instructor can see the blueprint to guide layouts. Figure 8 (right) shows an example.

The roles of the operator and instructor were randomly and evenly assigned to place three virtual objects. The six objects were: arrow, lightning bolt, diamond, crescent, heart, and teapot. During the task, the operator could only operate their assigned objects; setting the color, position, and orientation as guided by the instructor. The instructor can share verbal, pointing, gaze cues during the collaboration. The task is completed once the participants were satisfied with the placement of all six objects (three per person).

5.4 User Interfaces



(a) Training Task



(b) Main Task

Fig. 9. User interface for the user study tasks.

Users were provided similar but slightly different interfaces for the two tasks as shown in Figure 9(a). A shared window was provided for the training task, which included objects listed for the 3D furniture and a color editor. Once a color and furniture item were selected, participants could manipulate the furniture (translating and rotating) using controllers.

A similar interface was provided for the main task, however the color editor had been simplified to 8 colors in the shared window for user convenience. Some objects were grayed out so that participants could only create their assigned objects as the operator. A new tab was included for viewing a blueprint of where the virtual objects should be placed in the room. In addition, users could view a copy of their own blueprint in their personal menus.

5.5 Metrics

To evaluate presence, co-presence and effectiveness of collaboration of the AVT system, we used the following metrics:

Spatial Presence: Spatial Presence deals with how people feel to be spatially present in a location, and is related to the "concept of being there", also known as presence. To measure spatial presence in our setup, we used the MEC-SPQ questionnaire [71], which divides spatial presence in eight dimensions, consisting of 5-likert scale questions. For our study, we focused on two: *"sense of self-location"*, which is the feeling of being physically located in an environment and *"spatial situational model" (SSM)*, which relates to how people process spatial cues and interpret them related to the space.

Social Presence: This concept is an important aspect of collaboration and measures how people feel together in the same environment, even when remotely located. To measure social presence, we used the known SoPQ (Spatial Presence Questionnaire) [7], which divides social presence in six sub-categories, each of which containing a variable amount of 7-likert scale questions. In our user study, we chose three of them : *"Co-presence (CoP)"*, which is the feeling that people are located in the same place of the other participant; *"Attentional Allocation (AA)"*: which is the amount of attention one user give to the other users; and *"Perceived Message Understanding (PMU)"*, which measures how well a user understands the message being given by the other participant.

User Preferences: We included an additional 7-likert scale questionnaire in our user study to include aspects not covered by the above two questionnaires. This questionnaire focused on aspects about the system, users' preferences, and actions. We also included open questions to receive additional feedback from users about the system. The questions are summarized in Table 2.

6 Results

Figure 1 shows a snapshot of our AVT system captured in both VR-Traveler and AR-Host views. To evaluate AVT, we performed an user study and measured qualitative user study metrics. System performance metrics are measured in section 5.1, and few feasible applications are introduced in section 4.3.

6.1 User Evaluation

To evaluate the user perspective of AVT in terms of social presence, spatial presence and overall user preferences, we collected subjective data during our user study in the form of questionnaires. For the questionnaires, we first checked whether the groups showed statistical difference by using the average spatial presence and social presence as the input to the Mann-Whitney U test, with statistical difference suggesting a lack of statistical significance. Then, we performed equivalence tests to guarantee similarity between the two groups - the VR-Travelers and AR-Hosts. For that, we used two-one-sided t-test (TOST) analysis to look for equivalent results between our two groups (Table 3), similarly to [35]. we used 1 and -1 as the Upper and Lower bounds, as our Likert-scale is a discrete variable. We also conducted an additional statistical test using the Wilcoxon Signed Ranks test to assess statistical significance between our metrics and the neutral line and high-score line, ensuring high-scores for both groups of users using our system. For each test we used 4 as the median for the 7-point Likert scale in the overall questionnaires and spatial presence metrics, and 3 in spatial presence (since this metric uses a 5-Likert scale), as the neutral value. A median value of 5 was used for spatial presence and overall questionnaires and 3.5 out of 5 for the spatial presence questionnaire, as the high-score line.

By doing these statistical tests, we can ensure that both groups are equivalent and elicit a high subjective feeling of spatial presence, social presence and overall preferences.

6.1.1 Social Presence

Measured social presence metrics are shown in Figure 10. In our results, both the VR-Traveler and AR-Host maintained a high level of co-presence , as seen by our neutral score-line and high-score line statistical test (Table 3). We also found similar scores for both AA and PMU, as seen by the equivalence test that provided statistically significant results. Regarding social presence and its sub-components, no statistical differences between the traveler and host were found in co-presence, perceived message understanding, and attention allocation, which confirms hypothesis H0.

6.1.2 Spatial Presence

Figure 11 shows the results of spatial presence. The VR-Traveler and AR-Host both felt very high levels of self-location (as supported by our neutral-line and high-score line test, Table 3), believed that the environment did not contain errors (q1), and had a good spatial reference of the environment (spatial situation model). No statistical significance was found between the VR-Traveler and the AR-Host, when comparing spatial presence and its sub-components, and both distributions are equivalent as pointed by the TOST test(Table 3), which confirms our hypothesis H1. Interestingly, in SSM, we noticed a slight improvement from the VR-Traveler over the AR-Host, but with no statistical significance.

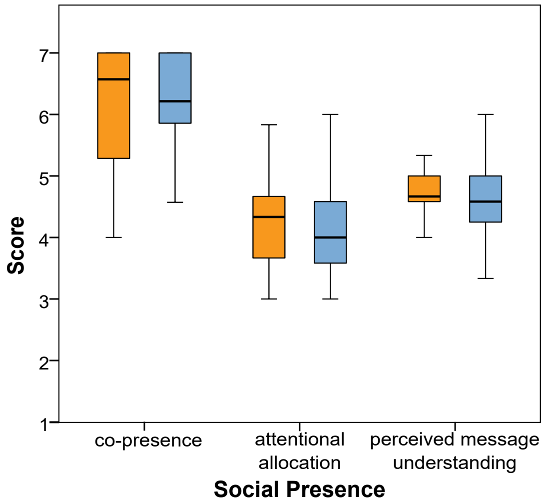


Fig. 10. The Social Presence results (Boxplot) evaluated in three dimensions: co-presence, attentional allocation (AA) and perceived-message understanding (PMU). Orange represents the AR-Host and blue represents the VR-Traveler.

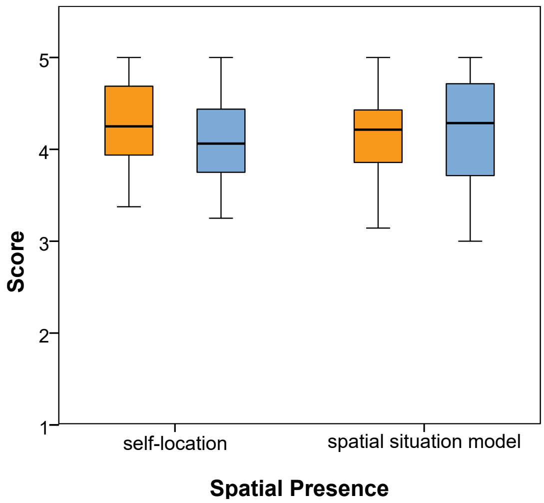


Fig. 11. Spatial Presence results (Boxplot) analysed by two metrics: Self-location and spatial situational model (SSM). Orange represents the AR-Host and blue represents the VR-Traveler.

6.1.3 User Preferences

Overall preference scores of our user test were high for both VR-Traveler and AR-Host, as shown in Table 2. Both users found it easy to understand when people used pointing cues (q2 and q4) and communicated using voice (q3). People also found it easy to create, select and move objects and to perform both tasks (q5-q8). They also found it easy to collaborate with their partner (q10-q14). We found a slight tendency towards the VR-Traveler to lead the interaction, except for task 1 (q15), where we found that the AR-Host felt the VR-Traveler contributed more to placing furniture in the room, with statistical significance (U=134.5, p=0.045). Regarding equivalence, most of the statements returned statistical significance, with the exception of q4, regarding easiness to communicate with visual elements; q6, about object selection; and finally q16 and q17, regarding perceived talking time.

6.2 Discussion

Social presence and Spatial presence: Users in both remote and local environments felt similarly high levels of spatial presence and social presence during our study; where "high" refers to the metric used on Vorderer et al. [71] for social and spatial presence measures, and a median value of 5 and above for our 7-likert scale and additional neutral and high-lines statistical test(Table 3). This confirms both our hypothesis H0 and H1. A high sense of self-location was sensed by both remote and local users, so we can state that both participants can feel "there" in the MRC-space while also feeling a high sense of connectedness with each other. Collaborators felt a similar sense of co-presence in both remote and local setups, which may enable effective collaboration as they feel present in the same physical environment. This enables them to perform the tasks clearly, as indicated by the PMU measure, while being aware of the presence of the remote user, as shown by the attention allocation (AA) measure. Regarding that, VR-Travelers stated in some cases, that they often forgot that they were in their physical room. We can say that the high-fidelity of the representation of the room provided by the 360-camera had an important impact in this aspect, as the same users clearly indicated that in the post-test interview. People also reacted favorably to the virtual object blending, as stated by the user preferences questionnaire. This indicate that our AVT platform is feasible to provide high-fidelity telecollaboration in both the AR-Host and VR-Traveler.

Collaboration and Communication: Collaborators in both setups easily communicated and collaborated with their remote partner. People stated that the use of voice and visual components helped them to indicate their intended behavior to their remote partner. The basic functions in our system were easily understood by both users, which contributed to the perceived task easiness. Participants also stated a high sense of contribution to solve the task, but when evaluating the level of their partners, AR-Hosts felt that their partner contributed a bit more. This denotes a tendency to the VR-Traveler to lead the interaction, which differs from previous work regarding asymmetric setups [51]. This was essentially noticed on task 1 (training) (q15), where people could freely communicate with each other in order to furnish the physical room with virtual objects. In this task, AR-Hosts

Question	VR	AR
q1: I felt that the virtual objects were situated in the same environment as the background	5(1.75)	5.5(2)
q2: It was easy to understand the other user when they were referring to an object/location	6(2)	6(1)
q3: It was easy to communicate using voice	7(1)	7(0.75)
q4: It was easy to communicate using visual elements (e.g. pointers, gestures)	6(2)	7(2)
q5: It was easy to create objects	6.5(1)	7(1)
q6: It was easy to select objects	6.5(1.75)	6(2)
q7: It was easy to move objects	5(2)	5.5(3)
q8: It was easy to perform the task 1 (training)	6(1)	6.5(1.75)
q9: It was easy to perform the task 2 (main)	6(1)	5.5(2.75)
q10: It was easy to collaborate with the other user	6(1)	6(1)
q11: I felt that I was easy to collaborate with	6(2)	6(1)
q12: I was active in solving task 2 (main)	7(1)	7(1)
q13: My partner was active in solving task 2 (main)	7(0)	7(1)
q14: I contributed to placing furniture	7(1.75)	6(1)
q15: My partner contributed to placing furniture*	7(0.75)	6(1)
q16: My partner talked more than me	4(1)	4(2.5)
q17: I talked more than my partner	4(2.75)	4(2)

Table 2. User preference results in 7-likert scale: Results are presented as Median(Interquartile range). * Indicates statistical significance.

	neutral		high		equiv.	
	VR (p)	AR (p)	VR(p)	AR(p)	DF	p-value
q1	0.006*	0.006*	0.713	0.467	37.077	0.016*
q2	<0.001*	0.002*	0.001*	0.014*	29.452	0.030*
q3	<0.001*	<0.001*	<0.001	<0.001*	27.941	0.001*
q4	<0.001*	0.001*	0.001*	0.098	34.075	0.080
q5	<0.001*	<0.001*	<0.001*	0.002*	37.333	0.003*
q6	<0.001*	<0.001*	0.025*	0.001*	33.475	0.061
q7	0.006*	0.002*	0.310	0.495	34.998	0.037*
q8	<0.001*	0.000*	0.001*	0.003*	35.460	0.023*
q9	0.001*	0.001*	0.144	0.020*	37.864	0.014*
q10	<0.001*	<0.001*	<0.001*	0.002*	30.803	0.005*
q11	<0.001*	0.001*	0.001*	0.023*	34.680	0.043*
q12	<0.001*	<0.001*	<0.001*	<0.001*	37.110	<0.001*
q13	<0.001*	<0.001*	<0.001*	<0.001*	37.840	<0.001*
q14	<0.001*	<0.001*	<0.001*	0.015*	30.092	0.010*
q15	<0.001*	<0.001*	<0.001*	<0.001*	37.837	0.005*
q16	0.371	0.322	0.053	0.002*	37.900	0.232
q17	0.590	0.525	0.033*	0.009*	36.611	0.116
Social Presence						
aa	<0.001*	<0.001*	0.001*	0.001*	37.873	<0.001*
cp	<0.001*	<0.001*	<0.001*	0.001*	33.042	0.003*
pmu	<0.001*	<0.001*	0.003*	0.003*	29.274	<0.001*
Spatial Presence						
sl	<0.001*	<0.001*	<0.001*	<0.001*	34.498	0.007*
ssm	<0.001*	<0.001*	0.001*	<0.001*	37.999	0.001*

Table 3. Results from the Wilcoxon Signed Ranks Test for comparing the results between the neutral line and the high-score line (5 for spatial presence and overall questionnaires and 3.5 for social presence). And, the results from both the Two One-sided paired t-test for equivalence and between AR-Hosts and VR-Travelers for the preferences, spatial and social presence metrics. We assumed 95% confidence with -1 and 1 as Lower and Upper bounds. AA (Attentional Allocation), CP(Co-presence), PMU(Perceived Message Understanding), SL(Self-Location) and SSM (Spatial Situational Model). * indicates statistical significance.

saw the VR-Travelers as contributing more to furnishing the room rather than themselves.

Other findings: AR-Hosts stated that they felt that their partner's avatar representation alongside the voice and visual elements enabled them to feel that their partner was in the same room. In our setup, the visual quality and expressiveness of the avatars differ between the AR-Host and the VR-Traveler. Although AR-Hosts see their partner's abstract avatar, their social presence ratings are higher (but not statistically different) compared to the VR-Traveler's score of seeing the real-captured partner in the 360-video. Users also stated in the post-test interview that it would be beneficial to move around in the environment, but also that the static position of the VR-Traveler did not hinder their ability to feel "there" and to collaborate with their partner.

7 Limitations and Future Requirements

While our work presents a working prototype of the teleportation concept with some feasible applications (as outlined in Section 4.3), it can be further improved to support more complex scenarios. In this section, we outline current limitations and requirements for realising such applications.

Latency: 360-cameras provides a challenge for real-time collaboration with low latency. We chose the Insta360 Pro 2 to balance the visual quality with latency. The entire setup had 1.2s of latency; with 0.6s required to record and stitch the footage and 0.6s required to encode, stream and decode the video. We did not sync the video with the rest of the system to reduce the impact of latency on communication and collaboration. However, this causes problems with the audio and user actions not matching the video. Latency from the 360-video streaming is expected to decrease as technical limitations improve. Advanced camera technology will be able to allow faster stitching without compromising visual quality. Better encoding standards for live-streaming 360-video will also reduce latency.

Avatar: The avatar representation of the VR-Traveler is an abstract outline of a person. Although we found a similar sense of presence for both the VR and AR users, a more detailed avatar could allow remote users to be more expressive in their non-verbal communication through body language and gaze cues, which is necessary for certain applications [2]. Future work could aim to improve the avatar quality for the VR-Traveler [31, 49] for richer communication through body language. Moreover, realistic avatars can be further explored such that the avatar can better match the high fidelity components of the environment. This could be especially important in remote training scenarios to promote higher levels of engagement between both users.

Disparity and Locomotion: Our setup is based on a static 360-camera capturing monoscopic 360-video and, therefore, users have limited depth disparity and freedom to move around. The ability to move around the high fidelity, live streamed environment is beneficial in remote assistance scenarios. This allows the remote user to supervise the task with more accurate risk assessment. Recent advances in capturing and modeling technology allow potential navigation inside panoramic video, but they are still very costly, have limited range of movement, and require offline post processing. Advancements on 360-camera technology could mitigate this problem, and our platform could be extended to adapt to such technological advancements.

Others: We used manual calibration for aligning coordinates in MRC-space. This process could be simplified by automating the calibration process by adding sensors and markers to the top of the 360-camera. More sophisticated algorithms for fine tuning could lead to better accuracy, reducing disparity in the scene when viewed by different users. We used a video pass-through HMD for the AR-Host to provide a wide field of view of the MRC-space as well as a consistent viewing experience between VR and AR users. However, our system can adapt to see-through HMDs when required. We tested AVT using a single user setup. Supporting multiple users can be a possible extension to enable group teleconferencing.

8 Conclusion

In this paper, we presented AVT, a novel platform that enables a high fidelity remote collaboration shared space. AVT uses a live streaming 360-camera to provide remote VR-Travelers a high sense of telepresence and co-presence, as well as footage of a mixed reality collaboration space with consistent real-world lighting to blend virtual objects. Our working prototype was evaluated in a user study to find how well the system could provide remote collaboration, communication, social presence and spatial presence among users. The study found that users felt a high level of telepresence in remote collaboration, co-presence with their partner, and had high satisfaction with the system. Our user study also showed that participants in different spaces with an asymmetric setup felt similar levels of presence and co-presence. Our platform provides a basic framework for AVT that contains several possible extensions for future work.

Acknowledgments

This project was supported by the Entrepreneurial University Program funded by TEC, New Zealand. We thank Ian Loh for his help to record and edit the video, and the CMIC staff for their valuable feedback and suggestions.

References

[1] M. Adcock, S. Anderson, and B. Thomas. Remotefusion: real time depth camera fusion for remote collaboration on physical tasks. In *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, pp. 235–242. ACM, 2013.

[2] R. K. d. Anjos, M. Sousa, D. Mendes, D. Medeiros, M. Billinghurst, C. Anslow, and J. Jorge. Adventures in hologram space: Exploring the design space of eye-to-eye volumetric telepresence. In *25th ACM Symposium*

on Virtual Reality Software and Technology, VRST 19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3359996.3364244

[3] D. Anton, G. Kurillo, and R. Bajcsy. User experience and interaction performance in 2d/3d telecollaboration. *Future Generation Computer Systems*, 82:77–88, 2018.

[4] D. Aschenbrenner, F. Leutert, A. Çençen, J. Verlinden, K. Schilling, M. E. Latoschik, and S. Lukosch. Comparing human factors for augmented reality supported single and cooperative repair operations of industrial robots. *Frontiers in Robotics and AI*, 6:37, 2019.

[5] M. Bessa, M. Melo, D. Narciso, L. Barbosa, and J. Vasconcelos-Raposo. Does 3d 360 video enhance user's vr experience?: An evaluation study. In *Proceedings of the XVII International Conference on Human Computer Interaction*, p. 16. ACM, 2016.

[6] M. Billinghurst, A. Nassani, and C. Reichherzer. Social panoramas: using wearable computers to share experiences. In *SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applications*, p. 25. ACM, 2014.

[7] F. Biocca, C. Harms, and J. Gregg. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence, Philadelphia, PA*, pp. 1–9, 2001.

[8] S. Bouchard, S. Dumoulin, J. Talbot, A.-A. Ledoux, J. Phillips, J. Monthuy-Blanc, G. Labonté-Chartrand, G. Robillard, M. Cantamesse, and P. Renaud. Manipulating subjective realism and its impact on presence: Preliminary results on feasibility and neuroanatomical correlates. *Interacting with Computers*, 24(4):227–236, 2012.

[9] I. Brilakis, H. Fathi, and A. Rashidi. Progressive 3d reconstruction of infrastructure with videogrammetry. *Automation in Construction*, 20(7):884–895, 2011.

[10] S. T. Bulu. Place presence, social presence, co-presence, and satisfaction in virtual worlds. *Computers & Education*, 58(1):154–161, 2012.

[11] F. Buttussi and L. Chittaro. Effects of different types of virtual reality display on presence and learning in a safety training scenario. *IEEE transactions on visualization and computer graphics*, 24(2):1063–1076, 2017.

[12] S. E. Chen. Quicktime vr: An image-based approach to virtual environment navigation. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pp. 29–38. ACM, 1995.

[13] P. Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pp. 189–198. ACM, 1998.

[14] J. V. Draper, D. B. Kaber, and J. M. Usher. Telepresence. *Human factors*, 40(3):354–375, 1998.

[15] D. Fonseca and M. Kraus. A comparison of head-mounted and hand-held displays for 360 videos with focus on attitude and behavior change. In *Proceedings of the 20th International Academic Mindtrek Conference*, pp. 287–296. ACM, 2016.

[16] J. D. Fraustino, J. Y. Lee, S. Y. Lee, and H. Ahn. Effects of 360 video on attitudes toward disaster communication: Mediating and moderating roles of spatial presence and prior disaster media involvement. *Public Relations Review*, 2018.

[17] L. Gao, H. Bai, G. Lee, and M. Billinghurst. An oriented point-cloud view for mr remote collaboration. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications*, p. 8. ACM, 2016.

[18] L. Gao, H. Bai, R. Lindeman, and M. Billinghurst. Static local environment capturing and sharing for mr remote collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, p. 17. ACM, 2017.

[19] S. J. Gerathewohl. Fidelity of simulation and transfer of training: a review of the problem. Technical report, Civil Aerospace Medical Institute, 1969.

[20] M. Gerhard, D. Moore, and D. Hobbs. Embodiment and copresence in collaborative interfaces. *International Journal of Human-Computer Studies*, 61(4):453–480, 2004.

[21] L. Gruber, J. Ventura, and D. Schmalstieg. Image-space illumination for augmented reality in dynamic environments. In *Virtual Reality (VR), 2015 IEEE*, pp. 127–134. IEEE, 2015.

[22] C. Heeter. Being there: The subjective experience of presence. *Presence: Teleoperators & Virtual Environments*, 1(2):262–271, 1992.

[23] P. Heidicker, E. Langbehn, and F. Steinicke. Influence of avatar appearance on presence in social vr. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 233–234. IEEE, 2017.

[24] J. S. Hvass, O. Larsen, K. B. Vendelbo, N. C. Nilsson, R. Nordahl, and S. Serafin. The effect of geometric realism on presence in a virtual reality game. In *Virtual Reality (VR), 2017 IEEE*, pp. 339–340. IEEE, 2017.

[25] T. Iorns and T. Rhee. Real-time image based lighting for 360-degree panoramic video. In *Pacific-Rim Symposium on Image and Video Technology*, pp. 139–151. Springer, 2015.

[26] S. Kasahara, M. Ando, K. Suganuma, and J. Rekimoto. Parallel eyes: Exploring human capability and behaviors with paralleled first person view sharing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 1561–1572. ACM, 2016.

[27] S. Kim, M. Billinghurst, and G. Lee. The effect of collaboration styles and view independence on video-mediated remote collaboration. *Computer Supported Cooperative Work (CSCW)*, 27(3-6):569–607, 2018.

[28] S. Kishore, X. Navarro, E. Dominguez, N. De La Peña, and M. Slater. Beaming into the news: a system for and case study of tele-immersive journalism. *IEEE computer graphics and applications*, 2016.

[29] J. Kolkmeier, E. Harmsen, S. Giesselink, D. Reidsma, M. Theune, and D. Heylen. With a little help from a holographic friend: The openimpress mixed reality telepresence toolkit for remote collaboration systems. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, VRST '18, pp. 26:1–26:11. ACM, New York, NY, USA, 2018. doi: 10.1145/3281505.3281542

[30] A. Kunert, A. Kulik, S. Beck, and B. Froehlich. Photoportals: shared references in space and time. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pp. 1388–1399, 2014.

[31] M. E. Latoschik, F. Kern, J.-P. Stauffert, A. Bartl, M. Botsch, and J.-L. Lugrin. Not alone here?! scalability and user experience of embodied ambient crowds in distributed social virtual reality. *IEEE transactions on visualization and computer graphics*, 25(5):2134–2144, 2019.

[32] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, p. 39. ACM, 2017.

[33] M. Leap. Home — magic leap. `https://www.magicleap.com/`. (Accessed on 05/01/2019).

[34] M. Leap. Welcome to the shared reality — magic leap. `https://creator.magicleap.com/learn/guides/leap-welcome-to-shared`. (Accessed on 05/01/2019).

[35] C. Lee, G. A. Rincon, G. Meyer, T. Hllerer, and D. A. Bowman. The effects of visual realism on search tasks in mixed reality simulation. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):547–556, April 2013. doi: 10.1109/TVCG.2013.41

[36] 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)*, pp. 153–164. IEEE, 2018.

[37] H. Lee, G. Ha, S. Lee, and S. Kim. A mixed reality tele-presence platform to exchange emotion and sensory information based on mpeg-v standard. In *Virtual Reality (VR), 2017 IEEE*, pp. 349–350. IEEE, 2017.

[38] P. Lensing and W. Broll. Instant indirect illumination for dynamic mixed reality scenes. In *Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on*, pp. 109–118. IEEE, 2012.

[39] J.-L. Lugrin, M. Wiedemann, D. Bieberstein, and M. E. Latoschik. Influence of avatar realism on stressful situation in vr. *2015 IEEE Virtual Reality (VR)*, pp. 227–228, 2015.

[40] S. Lukosch, H. Lukosch, D. Datcu, and M. Cidota. Providing information on the spot: Using augmented reality for situational awareness in the security domain. *Computer Supported Cooperative Work (CSCW)*, 24(6):613–664, Dec 2015. doi: 10.1007/s10606-015-9235-4

[41] A. MacQuarrie and A. Steed. Cinematic virtual reality: Evaluating the effect of display type on the viewing experience for panoramic video. In *2017 IEEE Virtual Reality (VR)*, pp. 45–54. IEEE, 2017.

[42] D. Medeiros, R. K. dos Anjos, D. Mendes, J. a. M. Pereira, A. Raposo, and J. Jorge. Keep my head on my shoulders!: Why third-person is bad for navigation in vr. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, VRST '18, pp. 16:1–16:10. ACM, New York, NY, USA, 2018. doi: 10.1145/3281505.3281511

[43] D. Medeiros, M. Sousa, D. Mendes, A. Raposo, and J. Jorge. Perceiving depth: optical versus video see-through. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pp. 237–240. ACM, 2016.

[44] D. Mendes, D. Medeiros, M. Sousa, R. Ferreira, A. Raposo, A. Ferreira, and J. Jorge. Mid-air modeling with boolean operations in vr. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 154–157. IEEE,

2017.

[45] Microsoft. Microsoft hololens — mixed reality technology for business. https://www.microsoft.com/en-us/hololens. (Accessed on 05/01/2019).

[46] Microsoft. Shared experiences in mixed reality - mixed reality — microsoft docs. https://docs.microsoft.com/en-us/windows/mixed-reality/shared-experiences-in-mixed-reality. (Accessed on 05/01/2019).

[47] A. Nagendran, A. Steed, B. Kelly, and Y. Pan. Symmetric telepresence using robotic humanoid surrogates. *Computer Animation and Virtual Worlds*, 26(3-4):271–280, 2015.

[48] O. Oda, C. Elvezio, M. Sukan, S. Feiner, and B. Tversky. Virtual replicas for remote assistance in virtual and augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pp. 405–415. ACM, 2015.

[49] 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*, pp. 741–754. ACM, 2016.

[50] A. Pagani and D. Stricker. Structure from motion using full spherical panoramic cameras. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 375–382. IEEE, 2011.

[51] Y. Pan, D. Sinclair, and K. Mitchell. Empowerment and embodiment for collaborative mixed reality systems. *Computer Animation and Virtual Worlds*, 29(3-4):e1838, 2018.

[52] P. J. Pardo, M. I. Suero, and Á. L. Pérez. Correlation between perception of color, shadows, and surface textures and the realism of a scene in virtual reality. *JOSA A*, 35(4):B130–B135, 2018.

[53] T. Pejsa, J. Kantor, H. Benko, E. Ofek, and A. Wilson. Room2room: Enabling life-size telepresence in a projected augmented reality environment. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pp. 1716–1725. ACM, New York, NY, USA, 2016. doi: 10.1145/2818048.2819965

[54] T. Pintaric, U. Neumann, and A. Rizzo. Immersive panoramic video. In *Proceedings of the 8th ACM International Conference on Multimedia*, vol. 493, 2000.

[55] T. Piumsomboon, G. A. Lee, A. Irlitti, B. Ens, B. H. Thomas, and M. Billinghurst. On the shoulder of the giant: A multi-scale mixed reality collaboration with 360 video sharing and tangible interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 228. ACM, 2019.

[56] H. Regenbrecht, K. Meng, A. Reepen, S. Beck, and T. Langlotz. Mixed voxel reality: Presence and embodiment in low fidelity, visually coherent, mixed reality environments. In *Mixed and Augmented Reality (ISMAR), 2017 IEEE International Symposium on*, pp. 90–99. IEEE, 2017.

[57] T. Rhee, L. Petikam, B. Allen, and A. Chalmers. Mr360: Mixed reality rendering for 360 panoramic videos. *IEEE Transactions on Visualization & Computer Graphics*, pp. 1379–1388, 2017.

[58] T. Schubert, F. Friedmann, and H. Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10(3):266–281, 2001.

[59] T. B. Sheridan. Musings on telepresence and virtual presence. *Presence: Teleoperators & Virtual Environments*, 1(1):120–126, 1992.

[60] M. Slater, M. Usoh, and A. Steed. Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 3(2):130–144, 1994.

[61] R. S. Sodhi, B. R. Jones, D. Forsyth, B. P. Bailey, and G. Maciocci. Bethere: 3d mobile collaboration with spatial input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 179–188. ACM, 2013.

[62] A. Steed, W. Steptoe, W. Oyekoya, F. Pece, T. Weyrich, J. Kautz, D. Friedman, A. Peer, M. Solazzi, F. Tecchia, M. Bergamasco, and M. Slater. Beaming: An asymmetric telepresence system. *IEEE Computer Graphics and Applications*, 32(6):10–17, Nov 2012. doi: 10.1109/MCG.2012.110

[63] J. Steuer. Defining virtual reality: Dimensions determining telepresence. *Journal of communication*, 42(4):73–93, 1992.

[64] 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. *arXiv preprint arXiv:1805.03709*, 2018.

[65] M. Tait and M. Billinghurst. The effect of view independence in a collaborative ar system. *Computer Supported Cooperative Work (CSCW)*, 24(6):563–589, Dec 2015. doi: 10.1007/s10606-015-9231-8

[66] A. Tang, O. Fakourfar, C. Neustaedter, and S. Bateman. Collaboration in 360 videochat: Challenges and opportunities. Technical report, University of Calgary, 2017.

[67] J. Tarko, J. Tompkin, and C. Richardt. Omnimr: Omnidirectional mixed reality with spatially-varying environment reflections from moving 360 video cameras. In *IEEE Conference on Virtual Reality and 3D User Interfaces*, 2019.

[68] 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*, p. 201. ACM, 2019.

[69] H. T. Tran, N. P. Ngoc, C. T. Pham, Y. J. Jung, and T. C. Thang. A subjective study on qoe of 360 video for vr communication. In *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, pp. 1–6. IEEE, 2017.

[70] E. R. Velamkayala, M. V. Zambrano, and H. Li. Effects of hololens in collaboration: A case in navigation tasks. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 61, pp. 2110–2114. SAGE Publications Sage CA: Los Angeles, CA, 2017.

[71] P. Vorderer, W. Wirth, F. R. Gouveia, F. Biocca, T. Saari, L. Jäncke, S. Böcking, H. Schramm, A. Gysbers, T. Hartmann, et al. Mec spatial presence questionnaire. *Retrieved Sept*, 18:2015, 2004.

[72] S. Weech, S. Kenny, and M. Barnett-Cowan. Presence and cybersickness in virtual reality are negatively related: a review. *Frontiers in psychology*, 10:158, 2019.

[73] R. Wolff, D. J. Roberts, A. Steed, and O. Otto. A review of telecollaboration technologies with respect to closely coupled collaboration. *International Journal of Computer Applications in Technology*, 29(1):11–26, 2007.

[74] S. Zhao. Toward a taxonomy of copresence. *Presence: Teleoperators & Virtual Environments*, 12(5):445–455, 2003.