On the Shoulder of the Giant: A Multi-Scale Mixed Reality Collaboration with 360 Video Sharing and Tangible Interaction

Thammathip Piumsomboon[1,2], Gun A. Lee[1], Andrew Irlitti[1], Barrett Ens[3], Bruce H. Thomas[1] and Mark Billinghurst[1]

[1]School of ITMS
University of South Australia
Mawson Lakes, SA, Australia
{gun.lee, andrew.irlitti, bruce.thomas, mark.billinghurst}@unisa.edu.au

[2] School of Product Design
University of Canterbury
Christchurch, New Zealand
tham.piumsomboon@canterbury.ac.nz

[3] Immersive Analytics Lab,
Monash University
Melbourne, VIC, Australia
barrett.ens@monash.edu

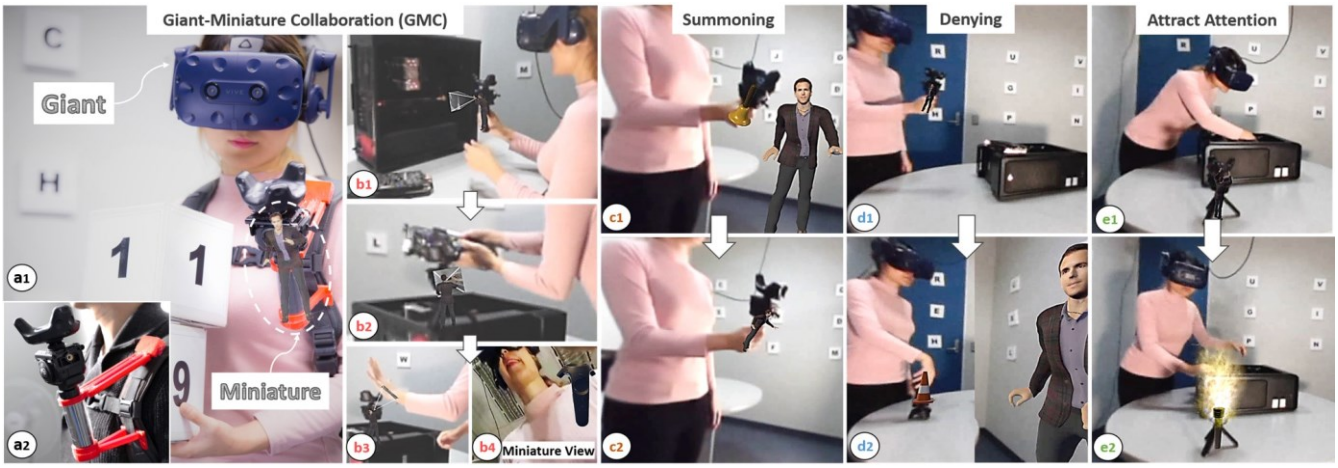

Figure 1: *Giant-Miniature Collaboration:* a1) Overview, a2) Protruded-shoulder-worn mount, b) GMC's use case scenario – PC assembly, c) *Giant-Miniature Interaction* – Summoning the remote VR user to *Miniature* mode, d) Denying access to the *Miniature* mode, e) The *Miniature* attracts attention of the *Giant* by changing his representation into a lit torch.

ABSTRACT

We propose a multi-scale Mixed Reality (MR) collaboration between the *Giant*, a local Augmented Reality user, and the *Miniature*, a remote Virtual Reality user, in *Giant-Miniature Collaboration (GMC)*. The *Miniature* is immersed in a 360-video shared by the *Giant* who can physically manipulate the *Miniature* through a tangible interface, a combined 360-camera with a 6 DOF tracker. We implemented a prototype system as a proof of concept and conducted a user study (n=24) comprising of four parts comparing: A) two types of virtual representations, B) three levels of *Miniature* control, C) three levels of 360-video view dependencies, and D) four 360-camera placement positions on the *Giant*. The results show users prefer a shoulder mounted camera view, while a view frustum with a complimentary avatar is a good visualization for the *Miniature* virtual representation. From the results, we give design recommendations and demonstrate an example *Giant-Miniature Interaction*.

CHI 2019, May 4-9, 2019, Glasgow, Scotland, UK.

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
DOI: https://doi.org/10.1145/3290605.3300458

CCS CONCEPTS

• Human-centered computing~Mixed/augmented reality • Human-centered computing~Computer supported cooperative work

KEYWORDS

Mixed Reality; remote collaboration; live panorama sharing; Wearable Interface; Tangible User Interface; multi-scale

ACM Reference format:
Thammathip Piumsomboon, Gun A. Lee, Andrew Irlitti, Barrett Ens, Bruce H. Thomas & Mark Billinghurst. 2019. On the Shoulder of the Giant: A Multi-Scale Mixed Reality Collaboration with 360 Video Sharing and Tangible Interaction. In *2019 CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland, UK*. ACM, New York, NY, USA. **17** pages. https://doi.org/10.1145/3290605.3300458

1 INTRODUCTION

Mixed Reality (MR) technologies have the potential to enhance collaboration beyond the limits of the physical world. One of those possibilities is through a multi-scale MR collaboration (MSMRC) [55, 57], which utilizes the capability of Virtual Reality (VR) to supports a multi-scale collaborative virtual environment (MCVE) [76] and extends it to the real world using Augmented Reality (AR) as shown by earlier research [5, 35, 57].

Past research on MR collaboration has explored sharing live 2D video [12, 32], which offers only a limited view, and 3D reconstructed scenes [14, 19, 58, 68] that can be challenging to dynamically update and lack in detail with current technologies. In comparison, 360-video sharing offers omni-directional viewing capability that can be live streamed [30, 69]. Comparing to sharing 3D reconstruction, 360-videos require lower processing power and network bandwidth yet provide higher update rate and better-quality image when full 3D immersion is not crucial but possible with a stereo 360-camera. It also offers high mobility using minimal hardware setup.

A few researchers have begun investigating 360-video sharing for MR collaboration, by using a 360-camera, either head-worn [44] or backpack mounted [8], but these approaches make it impossible to directly see the collaborator's face. To our knowledge, there has been no work on 360-video sharing applied to MSMRC. This shortcoming has led us to create design requirements for a MSMRC that addresses these issues:

R1) The remote VR user can see the local AR user's face, action, and environment through a simple hardware setup.
R2) Offers mobility and flexibility for the local AR user to control the remote user's view.
R3) Provides an appropriate visual representation of the remote VR user so that the local AR user is aware of the attention and to improve the communication fidelity.

From these requirements, we have implemented our MSMRC prototype that offers two modes of collaboration, the *Conventional Scale Collaboration (CSC)*, with shared 3D reconstruction, and the *Giant-Miniature collaboration (GMC)*, through a 360-video with a 1:12 scale disparity. In this paper, we explore *GMC*, as an alternative metaphor that offers 360-video sharing and tangible interaction. We expect that this can benefit tasks that demands mobility and flexibility to support a room scale collaboration as well as better perspective of fine details, such as assembly of models or circuit of a PC (see Figure 1b). However, a 360-video-based approach also come with some challenges such as view dependency where the remote user's view is affected by the local user controlling the camera. Thus, we conducted a comprehensive study to investigate these characteristics of sharing a 360-video with an augmented visual presentation.

The main contributions of the paper are:

- A novel multi-scale MR collaborative system prototype supporting two modes of collaboration, *Giant-Miniature Collaboration* between a local AR user, as the *Giant*, and a remote VR user, as the *Miniature* through 360-video sharing and tangible interaction, and *Conventional Scale Collaboration* with a regular size avatar in a static 3D reconstructed space.
- Three novel examples of *Giant-Miniature Interaction*
- A four-part user study exploring *GMC* in terms of virtual representations, *Miniature* controls, 360-video view dependencies, and 360-camera placements.
- Discussions and Design recommendations based on the user study results.

2 RELATED WORK

2.1 Mixed Reality Remote Collaboration

Researchers have explored using MR [49] technology for enhancing remote collaboration [56] where the virtual representation of the physical task environment is captured and shared with a remote collaborator who views it using VR. The remote user then communicates back to the local user not only through verbal communication but also using visual communication cues shown on the AR interface on the local user's side.

Early works in MR remote collaboration used 3D models prepared in advance as virtual representations of the physical environment [53] while others explored sharing 3D reconstructions of physical scenes viewed by remote users on a desktop computer [68] or handheld devices [66]. More recently, researchers also experimented using a Head-Mounted Display (HMD) to collaborate in a

shared 3D reconstructed environment captured with a depth camera [14, 15, 41, 58, 70]. These systems visualized virtual hand gesture cues either captured with a depth camera [14, 66, 70] or represented by virtual hand models [41]. While these prior works using static 3D reconstructions can capture the spatial structure of the physical environment, they have limitations in updating dynamically changing scenes and the visual quality is typically worse than a video image.

Others have tried building MR remote collaboration systems based on sharing live video streams. Early prototypes explored sharing visual communication cues such as a pointer and drawing annotations overlaid onto shared live 2D video on a handheld device [13, 17-19, 33, 34] or using wearable interfaces [21, 24]. Further explorations included sharing eye gaze [3, 21, 43] or hand gestures [2, 24]. These prior works shared an egocentric viewpoint through a 2D video stream, so the remote user's view was dependent on the motion of the camera capturing and sharing the video. To overcome this limitation, researchers have investigated various approaches, such as saving key frames to view later [34, 51, 52], placing a camera fixed in the environment [12, 32], or mounting a camera on a remote controlled robotic platform [38, 40, 60, 63, 64] on a mobile robot [1, 39, 47, 48] or even a drone [23, 54, 55]. These approaches still have limitations with the field of view and delay in remote controlling the view with a mechanical system.

As an alternative, researchers also investigated sharing 360 panorama [59] to let the remote user freely look around, although it supports only changing the viewing direction but not position. Some of the early works simply shared static images [6] or added a live 2D video insert on a static panorama image [50, 62], and overlaid simple MR visual cues for collaboration, such as pointers or drawings. As 360 panorama cameras became more affordable (e.g., Ricoh Theta S, Insta360, Samsung Gear 360, etc.), social networking platforms, such as Facebook and YouTube, started supporting shared live streaming of 360 panorama video. Kasahara et al. [30] developed the Jack-in system where the sharer wore a custom built panorama camera on his or her head and a remote user watched the shared panorama in a HMD. Their research mostly focused on sharing the experience through live 360 panorama, but not on active interaction between the users using MR cues. To investigate challenges in collaboration over live 360 panorama, Tang et al. [69] mounted a 360 panorama camera on a backpack monopod, sharing the wearer's surroundings from an exocentric viewpoint with a remote viewer watching the shared panorama video on a tablet. They found the remote viewer had difficulties in communicating location and orientation information due to the lack of sharing gestures and other non-verbal communication cues.

Most recently, researchers started looking into adding MR visual cues to support non-verbal communication in live 360 remote collaboration. Lee et al. [44] created a system that allows sharing of hand gestures and view awareness cues over live 360 panorama captured by a head-worn camera, enabling two-way non-verbal communication between a pair of users wearing AR or VR displays. Cai et al. [8] demonstrated a system which shows the remote user's virtual head and hands in a shared live 360 panorama captured from a backpack mounted 360-camera. Compared to this research, our work explores using a tangible representation of a remote collaborator in the context of multi-scale MR remote collaboration.

In terms of collaborative virtual environment (VE), Beck et al. [4] explored an immersive telepresence system that supported distributed groups of users to meet in a shared VE. At each site, the users and a small portion of the local interaction space was captured using a cluster of RGBD cameras. Compared to this work and the other research focusing on shared VE, our research emphasizes on sharing a room-scale or larger environmental information of the real-world task space that is relevant for the collaboration. In other words, the majority of the collaborative contents in our MR collaboration are real.

2.2 Multi-scale Remote Collaboration

Multi-scale Collaborative Virtual Environments (MCVE) support collaboration between multiple users at different scales. In one of the earliest examples, Zhang and Furnas [76] explored collaboration between city planners at different scales, one at a regular scale at street-level and another as a giant at city-scale, and showed that users can complement each other's actions (e.g. navigation and manipulation in VR) by taking advantage of working at different scales. Kopper et al. [37] and Fleury et al. [11] studied navigation techniques to help users interact and collaborate at different scales, while Le Chénéchal et al. [42] investigated co-manipulation in a MCVE. These works showed that multi-scale interfaces can be used for effective collaboration in VR.

While not as common as in VR, researchers have also explored applying multi-scale collaboration to MR systems that combine AR and VR technologies. Kiyokawa [35]

developed a system which allowed users to easily transition between VR and AR and collaborate across multiple scales. In the MagicBook [5], while a user views AR content overlaid on a physical book, another user could scale down and fly into the 3D virtual scene in VR so they can collaboratively explore the scene at different scales. These works focused on face to face collaboration, as opposed to recent research that emphasized remote collaboration for example [57] demonstrated multi-scale remote collaboration between an AR and a VR user. The AR user's environment was 3D reconstructed and shared with the VR user who could scale themselves up into a giant or down into a miniature for collaboration. Another example is [55] that introduced a concept of multi-scale mixed collaboration (MSMRC) and explored using an adaptive pair of cameras mounted on a drone to adjust the eye separation of the virtual camera and create an illusion of being a giant.

Compared to prior work using 3D modeled or reconstructed environment, our research is one of the first that applies the concept of multi-scale collaboration to a live 360 panorama-based MR remote collaboration system. We also combine the concept of multi-scale collaboration with a Tangible User Interface (TUI) [29] as we use a handheld 360 panorama camera as a tangible representation of the remote collaborator, superimposed with a virtual representation to indicate the remote user's attention and action to the local user. In contrast, past research on remote collaboration with TUI only focused on tabletop or planar workspaces [7, 9, 36, 46, 61].

3 MULTI-SCALE MR COLLABORATION

Past works demonstrated various implementations of multi-scale MR collaborative systems between a local AR user and a remote VR user [5, 35, 55, 57], yet none of them were sharing live 360 panorama of the collaborative environment. In this research, we have created a multi-scale MR collaborative system with two modes of collaboration: *Conventional Scale Collaboration (CSC)* mode, which shares a 3D reconstruction to the remote VR user at a regular scale (section 3.1), and *Giant-Miniature Collaboration (GMC)* mode, that shares a 360-video at different scale (section 3.2). The transition between the two modes can be made by the remote VR user using the controllers, or the local AR user through tangible interface in *Giant-Miniature Interaction* (section 3.5).

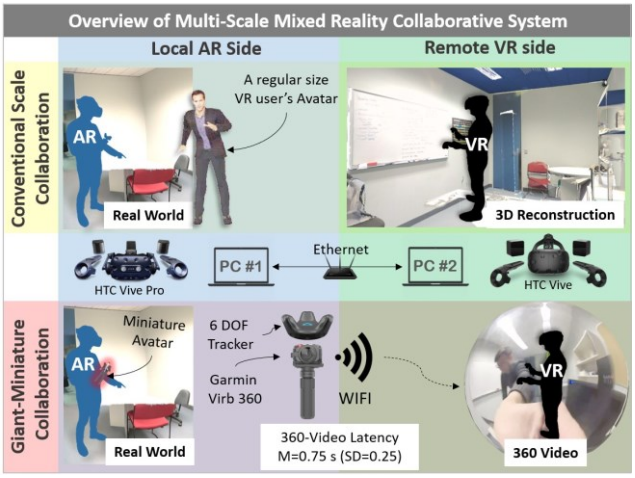

Figure 2: An overview of *GMC* prototype.

3.1 Conventional Scale Collaboration (CSC)

Conventional Scale Collaboration (CSC) mode is an MR collaboration with shared static 3D reconstruction at a regular scale similar to those demonstrated in [56, 58] (Figure 2). The user on one side is represented by an avatar to their collaborator on the other side. The VR user is also immersed in a reconstructed physical environment of the AR user. The local AR space was scanned using the Faro Focus^{M} 70 [10], a high precision 3D scanner (error ±3mm, range 0.6-70m). An example scan result is shown in the upper-right picture of Figure 2. The reconstructed model was pre-loaded on the VR system and calibrated using a VR controller to align the model's origin to the real world.

3.2 Giant-Miniature Collaboration (GMC)

Giant-Miniature Collaboration (GMC) mode is an MR collaboration through shared live 360 panorama video (Figure 2). In *GMC*, the two users are in different scales, and the remote VR user, the *Miniature*, sees a live 360-video from the local AR user, the *Giant*. The *Miniature*'s avatar appears tiny and attaches to the 360-camera, which the *Giant* can control as a TUI. The *Miniature*'s avatar measured 15 cm for the *Giant* and the *Giant* appeared 12× larger (~22 m or ~72 ft) to the *Miniature* in the 360-video. The 15cm avatar's size matches the height of the 360-camera including its handle. The avatar's eye (VR user's virtual camera) aligns to the camera sensor and the avatar's feet touch the ground when the camera is placed on a surface with the handle used as a tripod. The measurement to match the size was physically done using a ruler placed next to tripod/360camera.

3.3 System Overview

3.3.1 Hardware Overview

Local AR user side (*Giant* in *GMC*) included an HTC Vive Pro with two 2nd generation Lighthouse sensors [26] driven by a Windows 10 desktop computer (Intel Core i7-6700K at 4.0 GHz, 16 GB RAM, and NVIDIA GeForce GTX 1070), one Garmin VIRB 360 camera [16], and one Vive Tracker (2018 model) [28] with a custom-made camera mount. We used the video see-through (VST) mode of the HTC Vive Pro's stereo camera (captured in VGA@90Hz) for the AR experience.

Remote VR user side (Miniature in *GMC*) included an HTC Vive with two 1st generation Lighthouse sensors [25] driven by a Windows 10 laptop computer (Intel Core i7-6700HQ at 2.6 GHz, 16 GB RAM, and NVIDIA GeForce GTX 1070).

Networking - The two sides were networked through a router (D-Link DSL-2888A) using a 100MB Ethernet connection and the Garmin VIRB 360 camera was connected to the same network using 2.4 GHz WIFI (802.11b). The two sides communicated through voice over IP using an HMD's built-in microphone and a headphone. The average latency of the 360-video stream was measured as 0.75 second (SD=0.25).

3.3.2 Software Overview

Development Tools - We developed our software using the Unity game engine version 2017.3.0f3 [71] with SteamVR for Unity [72] and Vive SRWorks SDK for VST mode [27] (AR side).

Virtual Representation - Based on past research on virtual characters, it was found that realistic virtual human was favorable for remote collaboration [75]. Therefore, we chose to use a realistic avatar in the *CSC* mode. However, we were not aware of any past survey that elicited user preferences of a *Miniature* avatar. Hence, we conducted an online survey (n=32, 3 females, mean age 31.4, SD=8.3) similar to [20] through the online forums (Reddit) in AR/MR related channels. We compared three options: a realistic looking, a semi-cartoonish, and a cartoonish avatar. We found that almost half of the participants preferred the realistic avatar and so we used the same avatar design to represent the VR user in both *CSC* and *GMC* modes.

3.4 Giant-Miniature Interaction

The transition between the two modes can be made by the remote VR user using the controllers, or the local AR user through tangible interaction. In this section, we demonstrate latter interaction. We developed three *Giant-Miniature Interaction (GMI)* techniques to showcase *GMC* (Figure 1). The *Giant* could use the camera as a tangible interface to perform different actions with the *Miniature*. The *Miniature* could also change the appearance to communicate with the *Giant*. Several interactions could be supported:

3.4.1 ***Summoning*** – The Giant can summon the remote VR user to enter the *Miniature* mode (from *CSC* to *GMC*) by waving the camera like a bell (Figure 1-c1). After a couple of waves, a virtual summoning bell appears which is then replaced by the Miniature (Figure 1-c2). On the VR side, the VE showing a 3D reconstruction is replaced by a 360-video sphere.

3.4.2 ***Denying*** – For privacy, the *Giant* can flip the camera upside down, removing the VR user from the *GMC* mode (Figure 1-d1) and teleporting him back to *CSC* mode (Figure 1-d2), a virtual cone appears indicating that the remote VR user could no longer enter the *Miniature* mode.

3.4.3 ***Attracting Attention*** – The *Giant* could lose attention of the *Miniature* (Figure 1-e1), in this circumstance, the *Miniature* could attract attention of the *Giant* by changing his/her appearance as a virtual torch on fire (Figure 1-e2). The *Giant's* view could dim down and only lit the *Miniature* area, forcing the *Giant* to pay attention to the *Miniature*.

4 USER STUDY

We conducted a four-part user study to examine *GMC* and evaluate the three design requirements given in the introduction section. Part A of the study evaluated two virtual representations of the *Miniature*, an avatar and a frustum, which corresponded to the requirement R3. Part B examined the level of control that the *Giant* has over the *Miniature*, and Part C compared the effects of view dependencies of 360-video on the *Miniature*, these two

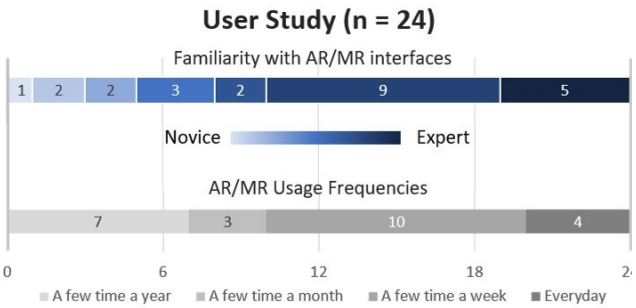

Figure 3: User Study - demographic results.

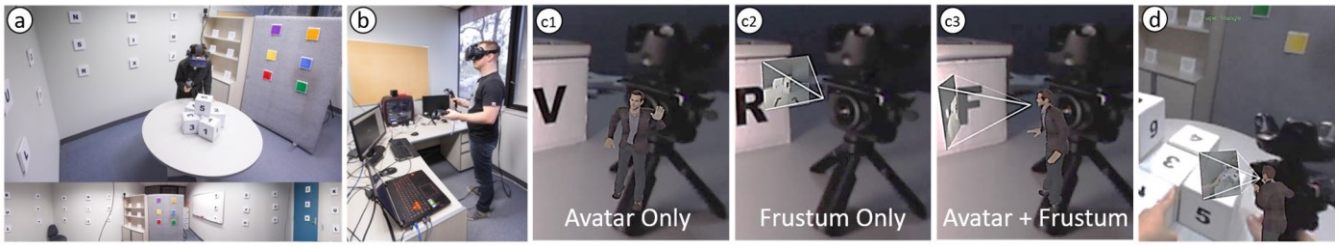

Figure 4: Study setup: a) Participant's space, b) actor's space, c) conditions in Part A, d) *giant*'s perspective during the trial.

parts assessed requirement R2. Part D compared four different mounting positions of the 360-camera on the *Giant*, which examined requirement R1. In Parts A and B, the participants were in the role of the *Giant*, collaborating with an actor, who was playing the *Miniature*. In Part C and D, the participants experienced the *Miniature* role. For Part C, 360-videos and 6 DOF tracking data were pre-recorded within the experimental space to ensure the consistency of experience across participants. Finally, in Part D, 360-videos were pre-recorded from both indoor and outdoor environment simulating a house inspection scenario.

4.1 Participants

We recruited 24 participants (5 female, mean age of 29.0 years, SD=6.4) from the University of South Australia. We sought participants with some experience with AR or VR interfaces to reduce the effect of novelty and for the potential insight from their experiences. Their familiarity with AR/MR, measured on a 7-point Likert scale, from 1 for a novice to 7 for an expert, yielded M=5.08 (SD=1.75) (Figure 3d). The frequencies of use were at least few times a year (29%) if not more than few times a month (71%) (see Figure 3e).

4.2 Setup

Our experimental space was setup in a room (4.9m by 2.9m), divided by a tall cabinet and a partition into two spaces, the participant's space (2.9m by 2.9m) and the actor's space (2 m by 2.9 m). For every part of the study, the participants always performed their tasks in the participant's space (Figure 4a) regardless of their roles and filled out the questionnaires on a computer in the actor's space (Figure 4b). Both sides used an HMD with a headphone and a built-in microphone for voice over IP communication. Videos of the trials were recorded using a DSLR camera on the participant's side, and screen recorded on the actor's side.

The room was normally lit with 8 fluorescent tubes. However, this posed a problem with the HTC Vive Pro's stereo camera (captured in VGA@90Hz) in the VST mode flickering image caused by the unmatched AC power frequency. For Part A and B that required VST mode, we replaced the AC lightings with 3 battery-powered DC light sources to eliminate the flicker (one 1500 Lumen, and two 800 Lumen all white lights).

The experimental tasks required physical props, including 27 Styrofoam tiles of letters (A-Z and '&'), 6 numbered cardboard boxes, and paper cut-out tiles in 6 shapes and 6 colors.

4.3 Procedure

The study was conducted in the order of Part B → A → C → D. This was to ensure better understanding of the tasks by the participants, as Parts A and B were easier than Part C. This also minimized the potential effects of simulator sickness on Parts A and B from viewing 360-videos in Parts C and D.

After an introduction to the study, the participants signed a consent form and completed a demographic questionnaire. The researcher explained all the equipment involved in the study. The participants were then asked to try on the HTC Vive Pro HMD and fitted the headset. At the beginning of each part, the researcher explained the purpose of the study and the participant's role in completing the tasks. The user study took approximately 90 minutes to complete for each participant. Further details on each part of the study will be given in the following sections.

4.4 Study Part A – Virtual Representations

We were interested in comparing an abstract virtual representation, such as a view frustum which was found to be an effective representation [56], and a potential alternative of an avatar for the *Giant* to understand the *Miniature*'s attention. Our implementation of a view frustum provides a video overlay, allowing the *Giant* to see what the *Miniature* sees (see Figure 4-c2). In addition to the view frustum, the *Giant* could see the virtual models of the VR controllers in the video, representing the *Miniature*'s hands. Therefore, any hand movement could

be observed in this condition too but from the *Miniature*'s ego-centric perspective.

In this first part of the study, we investigated the effects of an avatar and a view frustum on the ability of the *Giant* to understand the *Miniature*'s attention on an object in a room scale collaborative environment. This involved the usage of basic non-verbal communication cues, e.g. gaze and gestures, when the *Miniature* is directing the *Giant* to find an object of interest. We were most interested in the usefulness of each virtual representation for indicating the *Miniature*'s attention, therefore, we did not incorporate any manipulation task to reduce the complexity of this part of the study.

4.4.1 Study Design. Part A was a within-subject 2×2 factorial design where the two independent variables were the presence/absence of an avatar or a view frustum. The combination yielded 4 conditions, *No Visualization (NV)* as the baseline, *Frustum Only (FO)*, *Avatar Only (AO)*, and *Avatar + Frustum (AF)*. The order of conditions was counterbalanced between participants. Dependent variables included task completion time and number of errors. Our subjective measures were *Networked Mind Measure of Social Presence* questionnaire [22] (on *Co-Presence* and *Perceived Message Understanding* subscales), Subjective *Mental Effort Question (SMEQ)* [65], task difficulty using the *Single Easement Question (SEQ)* [65], and user preferences. Social Presence, SMEQ, and task difficulty were collected after each condition and user preference was collected at the end of Part A.

4.4.2 Tasks. The goal of this study was for the participants, as the *Giant*, to observe the virtual representations of the *Miniature* for collaboration in addition to the baseline verbal communication. The task for the participants was to guess the object of interest that the actor was randomly assigned to find in each trial. The object could be either a letter (on one of the three walls), a number (on a table), a color (on a partition) or a shape (on a cabinet). The actor was not allowed to directly tell the participants what the object was, and they had to search the room together and find the object of interest as fast as they could. In addition to randomization of the object to find, the physical props were randomly shuffled in each session to ensure minimal learning effect by the actor. This task simulates a situation when it is difficult to verbally describe the object of interest in the scene to the collaborator e.g. a workspace full of similar items. In all conditions, our actor consistently acted the same way by looking, pointing, and gesturing toward the object of interest utilizing their body language, a natural behavior to assist verbal communication.

During the trial, the participants, as the *Giant*, held the 360-camera representing the *Miniature* with an overlaid virtual representation (except for the no visualization condition). In this study, the *Giant* could only control the *Miniature*'s position but not its orientation. This is a condition that we call *Independent* (more details in Part B). They were directed by the actor, as the *Miniature*, who had the information of the object to look for. In case of no visualization, the participant could not see any visual representation of the *Miniature*. The participants could ask the actor, any binary questions that the actor could answer in "Yes/No" or "True/False". A wrong guess or a binary question asked was counted as an error. The participants were given two training trials for each condition. There were six experimental trials in total for each condition, with 3 random letters, 1 number, 1 color, and 1 shape object. This part took about 20 minutes.

4.4.3 Hypotheses. Our hypotheses for this part were:

A1. An avatar or a view frustum being present lowers the subjective mental effort and task difficulty.
A2. An avatar or a view frustum being present improves the user performance.
A3. An avatar or a view frustum being present increases *Aggregated Social Presence score* (*AsoP*), in terms of the *Co-Presence* (*CoP*) and *Perceived Message Understanding* (*Msg*) subscales.
A4. Participants prefer seeing both virtual representations of the *Miniature* (i.e. *Avatar + Frustum* condition).

4.4.4 Results. When dependent variables did not comply with normality assumptions, a Repeated-measures ANOVA with the Aligned Rank Transform (ART) was used for nonparametric factorial analysis [74]. For post-hoc pairwise comparisons, we used Wilcoxon signed-rank tests with Bonferroni correction (the p-value is adjusted while the alpha level is kept at .05). Figure 5 presents the results.

4.4.4.1 Performance. We used the aggregated task completion time (*Tct*) and aggregated number of errors (*Errors*) from 6 trials combined resulting in 24 datapoints per condition for each variable. The Shapiro-Wilk test indicated that our data significantly deviated from a normal distribution (*Tct* – W=.79, p<.0001, *Errors* – W=.71, p<.0001), so the ART was applied. We found significant main effects of both *avatar* (*Tct* - $F_{1,69}$=70.37, p<.0001, *Errors* - $F_{1,69}$=294.8, p<.0001) and *frustum* (*Tct* - $F_{1,69}$=60.55, p<.0001, *Errors* - $F_{1,69}$=296.56, p<.0001). There were

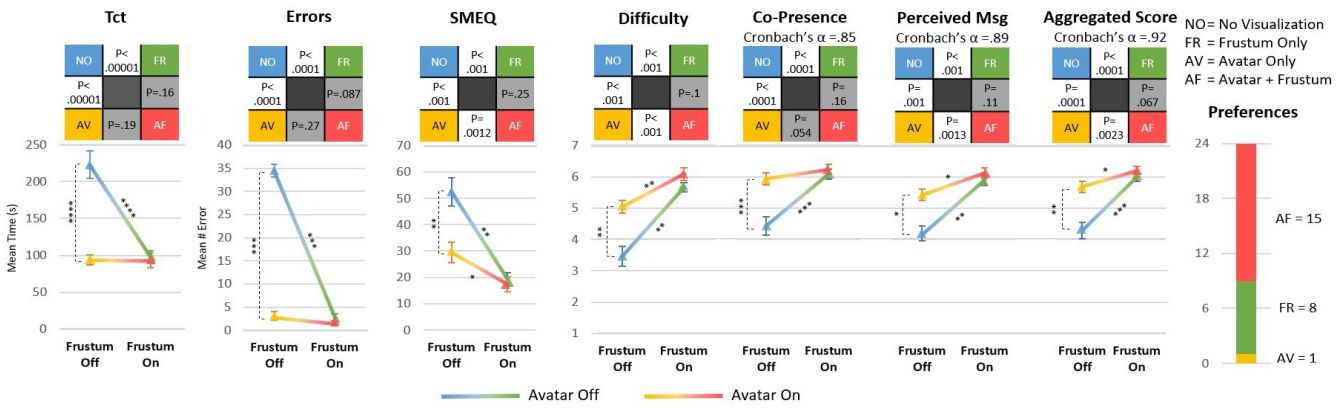

Figure 5: Results from Part A showing factorial plots to visualize interaction between the *Frustum* and the *Avatar* (**=p<.00001, ***=p<.0001, **=p<.001, and *<0.05). Error bars represent standard errors.**

significant interaction effects of *avatar* × *frustum* on both measures (*Tct* - $F_{1,69}$=51.42, p<.0001, *Errors* - $F_{1,69}$=285.23, p<.0001). Pairwise comparisons gave a significant difference only between *NV-AO*, and *NV-FO* for both *Tct* and *Errors*.

4.4.4.2 Subjective Mental Effort and Task Difficulty. We found significant main effects of both *avatar* (*SME* - $F_{1,69}$=18.64, p<.0001, *Difficulty* - $F_{1,69}$=24.78, p<.0001) and *frustum* (*SME* - $F_{1,69}$=55.43, p<.0001, *Difficulty* - $F_{1,69}$=59.86, p<.0001). Significant interaction effects were found for *avatar* × *frustum* on both measures (*SME* - $F_{1,69}$=13.02, p=.0005, *Difficulty* - $F_{1,69}$=7.02, p=.01). Pairwise comparisons gave a significant difference between *NV-AO*, *NV-FO*, and *AO-AF* for both *SME* and *Difficulty*.

4.4.4.3 Social Presence. An internal consistency test yielded good results (*Cronbach's α* > 0.92) for both subscales, *Co-Presence (CoP)* and *Perceived Message Understanding (Msg)*, and the overall *Aggregated Social Presence (ASoP)* score (see Figure 5). Again, we found significant main effects of *avatar* (*CoP* - $F_{1,69}$=29.96, p<.0001, *Msg* - $F_{1,69}$=19.91, p<.0001, *ASoP* - $F_{1,69}$=30.76, p<.0001) and *frustum* (*CoP* - $F_{1,69}$=38.05, p<.0001, *Msg* - $F_{1,69}$=43.88, p<.0001, *ASoP* - $F_{1,69}$=51.27, p<.0001). Significant interaction effects were also found (*CoP* - $F_{1,69}$=19.48, p<.0001, *Msg* - $F_{1,69}$=9.52, p=.003, *ASoP* - $F_{1,69}$=17.31, p<.0001). Pairwise comparisons gave a significant difference between *NV-AO* and *NV-FO* for *CoP*, and *NV-AO*, *NV-FO*, and *AO-AF* for *Msg* and *ASoP*.

4.4.4.4 User Preferences. Most of the participants preferred *AF* condition (62.5%) followed by the *FO* condition (33.3%). A Chi-squared goodness of fit test yielded a significant difference against random choice, $\chi^2(3) = 24.33$, p<.0001.

4.4.5 Discussion. Our results provide strong evidences to support all four hypotheses (A1, A2, A3, and A4), which showed that having virtual representations, either an avatar or a frustum, significantly improved the overall experience of the GMC and the remote collaboration through a 360-video for that matter. Having no visualization limits the ability of the local user to notice the remote collaborator's attention.

4.4.5.1 ***Miniaturized Virtual Representation.*** Past research has shown that sharing of hand gestures and view awareness cues [44], or virtual head and hands [8], enabled two-way non-verbal communication. Our findings also support this claim, however, we extend their results and show that a miniaturized virtual representation can also improve non-verbal communication despite being more compact and occupying less display space.

4.4.5.2 ***Avatar vs Frustum.*** For this asymmetric task, where the Miniature was guiding the Giant, we found that there were no significant difference with or without an avatar when there was a view frustum present. In contrast, the absent of a view frustum had a significant effect in terms of increased subjective mental effort, higher task difficulty, and reductions in Perceived Message Understanding and overall Social Presence. These results support the claim that a frustum is an effective awareness cue [56]. This finding leads us to believe that a view frustum (with a video overlay) might be a better virtual representation than an avatar to indicate the remote user's attention. Participants said:

"I feel like the frustum is the easiest way of telling what the subject is doing. The avatar is helpful, but I feel is made redundant by the frustum." – P5.

"Frustum was the most helpful and least distracting. It's very obvious where the remote user is looking. In the avatar conditions, you mostly see the back of the avatar anyway, so many social benefits are lost." – P13

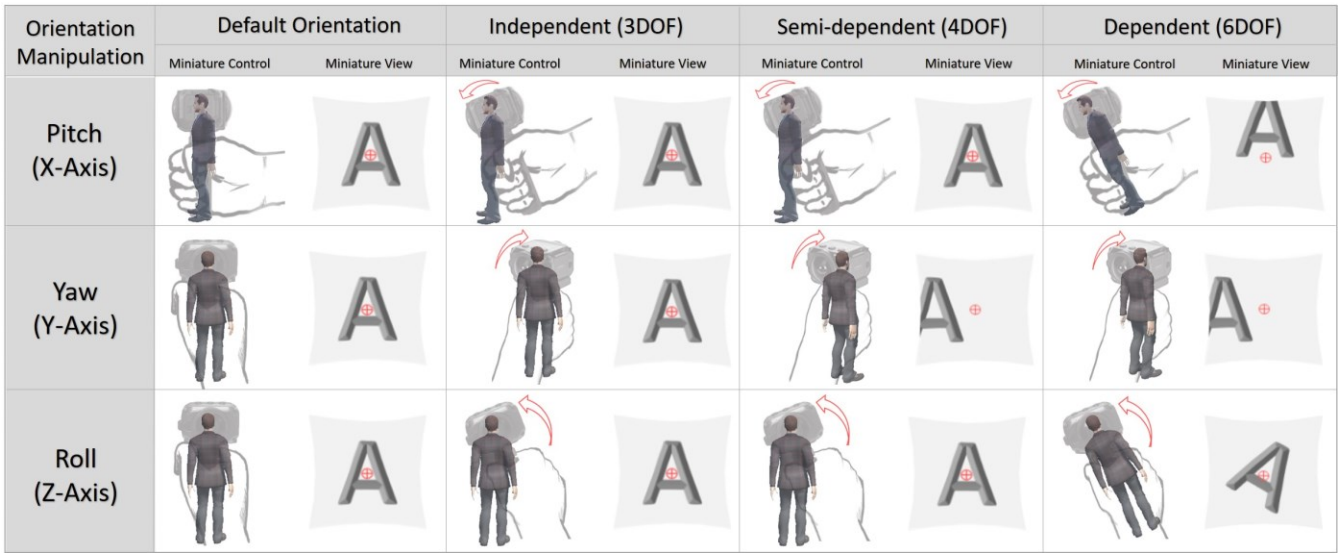

Figure 6: Three conditions in Part B and C, in each cell, the left image depicts the *Miniature* control through manipulation of the tangible interface, and the right illustrates the corresponding view of the *Miniature* due to the manipulation.

However, this is limited to the fact that the task focused on asymmetric roles with the remote user being an expert, and also being a simple search task without any manipulation involved. However, we have fulfilled our goal of comparing the two representations for their ability to indicate the *Miniature*'s attention, which is the scope of this paper. In terms of guiding manipulation task, a follow up study is required to compare an avatar and a frustum.

4.4.5.3 ***Avatar + Frustum.*** In terms of user preferences, most participants preferred having both virtual representations for offering all the complementary cues. Participants said:

"This combination gives the participant all the information. We understand the physical gestures of the avatar, the focus of their point of view and what they see." – P11

"It was the easiest to work with. I could see the avatar and the frustum at the same time and could cross-reference them for the correct guess." – P24

Nevertheless, we note that some of the participants found having both representations distracting. They do not need to be visible at all time and the users could easily choose the most appropriate representation depending on one's intent and what the collaboration entails.

4.5 Study Part B – Miniature Control

Part A gave us some insights into designing the virtual representation of the *Miniature* for the *Giant* to visualize. In Part B, we compared different levels of control of the *Miniature* by the *Giant*. The aims were to better design an appropriate level of control and to potentially create better interaction techniques for *GMC*. Once more, the participants took on the *Giant* role. The actor represented the *Miniature* with an avatar representation.

4.5.1 Study Design. Part B was a within-subject design where we investigated three levels of *Miniature* control (Figure 6): *Independent* (3DOF – position only), *Semi-dependent* (4DOF – position + yaw rotation), *Dependent* (6DOF – position + orientation). In the *Independent* condition, participants, as the Giant, could move but not rotate the Miniature avatar. In the *Semi-dependent* condition, the *Giant* could move and rotate the avatar left and right, around the ground vector. In the *Dependent* condition, the *Giant* had a full control of the *Miniature's* avatar. The order of conditions was counterbalanced. Our objective variables were task completion time and number of errors. Our subjective measures were *Social Presence* (on *Attention Allocation*, *Perceived Message Understanding*, and *Perceived Behavioral Interdependence* subscales), *SMEQ*, task difficulty and user preferences. Social Presence, SMEQ, and task difficulty were collected in each condition and user preference was collected at the end of Part B.

4.5.2 Tasks. The goal of this study was for the *Giant* to direct the *Miniature*'s attention to an object of interest, which was the opposite of Part A where the *Giant* was mostly on the receiving end of the communication. The task was similar to Part A and the same set of objects were used. But this time, the participants, as the *Giant*, was assigned a random object and the *Miniature* (the actor) had to guess what the object of interest was. The

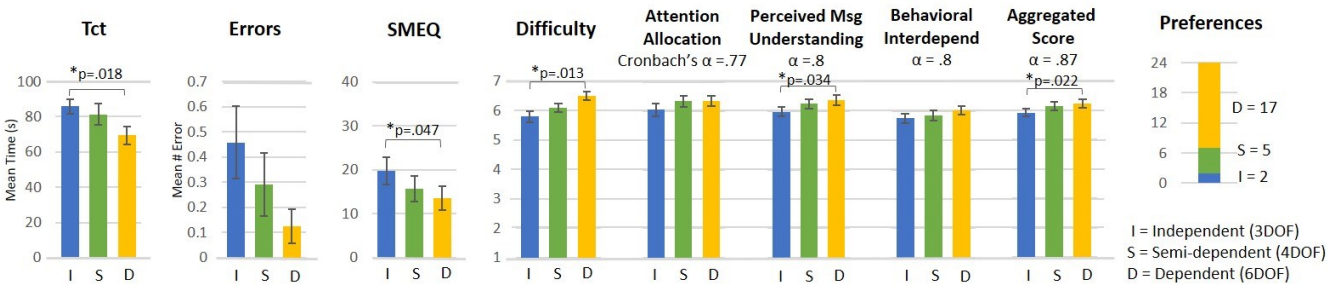

Figure 7: Results from Part B. Error bars represent standard errors.

participants used one hand to hold the 360-camera, which represented the *Miniature*. In addition to verbal communication, participants were allowed to use any method to complete the task such as holding the 360-camera right in front of the object or pointing at the object with their hand. Any wrong guess or incorrect binary question asked was counted as an error. The participants were given two training trials and six experimental trials, with 3 random letters, 1 number, 1 color, and 1 shape object. This part took approximately 20 minutes to complete.

4.5.3 Hypotheses. Our hypotheses for this part were:

B1. The higher degree of freedom of the Miniature control lowers subjective mental effort and task difficulty.
B2. The higher degree of freedom of the Miniature control improves user performance.
B3. The higher degree of freedom of the Miniature control increases the *Aggregated Social Presence score* (*ASoP*), in term of *Attention Allocation* (*Att*), *Perceived Message Understanding* (*Msg*), and *Perceived Behavioral Interdependence* (*Bhv*) subscales.
B4. Participants prefer full control of the Miniature's avatar (i.e. Dependent condition).

4.5.4 Results. Our dependent variables deviated from a normal distribution and therefore, we applied the Friedman and Wilcoxon signed-rank tests with Bonferroni correction (p-value adjusted) for post-hoc pairwise comparisons. Figure 7 presents our results.

4.5.4.1 Performance. Like Part A, we used the aggregated *Tct* and aggregated *Errors* resulting in 24 data points for each condition per variable. A Shapiro-Wilk test indicated that data from both variables were not normally distributed (*Tct* – W=.94, p=.0029, *Errors* – W=.55, p<.0001). A Friedman test yielded significant difference for task completion time (*Tct* - $\chi^2(2) = 7$, p=.03). Pairwise comparisons yielded a significant difference between *Dependent* and *Independent* condition (see Figure 7 - *Tct*). Note that the average *Errors* in all conditions were below one, indicating that there was almost no error in this part.

4.5.4.2 Subjective Mental Effort and Task Difficulty. We found significant differences for both variables (*SME* - $\chi^2(2) = 8.21$, p=.017, *Difficulty* - $\chi^2(2) = 10.09$, p=.0064). Pairwise comparisons gave a significant difference between *Dependent* and *Independent* condition for both variables (see Figure 7 – *SMEQ* and *Difficulty*).

4.5.4.3 Social Presence. Internal consistency yielded good results for all measured subscales of *Attention Allocation (Att)*, *Perceived Message Understanding (Msg)*, *Perceived Behavioral Interdependence (Bhv)*, and the combined *Aggregated Social Presence (ASoP)*. We found significant differences for *Msg* and *ASoP* (*Msg* – $\chi^2(2) = 10.45$, p=.0054, *ASoP* – $\chi^2(2) = 7.24$, p=.027). Pairwise comparisons yielded significant results between *Dependent* and *Independent* condition for both measures.

4.5.4.4 User Preferences. Most of the participants preferred the *Dependent* condition (70.8%) followed by the *Semi-dependent* condition (20.8%). A Chi-squared goodness of fit test yielded a significant difference against random choice, $\chi^2(2) = 15.75$, p<.001.

4.5.5 Discussion. Our results strongly support the hypothesis B4, where the majority of the participants preferred *Dependent* condition. Hypotheses B1, B2, and B3 were also partially supported as there were significant differences in favor of *Dependent* over *Independent* condition for lower *Tct*, *SMEQ*, and *Difficulty*, as well as higher *Msg*, and *ASoP*.

4.5.5.1 ***Better Control.*** Most participants preferred having full control of the *Miniature* avatar because it was easier to manipulate and required less verbal communication. Participants said:

"I could easily direct the remote user without having to give verbal instruction" – P2

Figure 8: Part C - a) 360-Video, b) original arrangement in the video, c) after rearranged. Part D - d) four camera placements.

"...more control makes sense. It also takes time for the remote user to rotate when it is easier to just point the avatar at the thing you want it to look at." – P5

5.5.5.2 ***More Intuitive.*** Some participants expressed having a full control being intuitive. Participants said:

"It feels as though the avatar is a part of myself, moves like I move, and generally feels more natural because of this. There is no disorientation, and minimal cognitive load experienced in this setup." – P6

4.5.5.2 ***Improved Perceived Message Understanding.*** Participants felt that their collaborator could see what they saw better with full control, although, it was a 360-video and the remote user could look anywhere in all conditions.

"I had full control of where the avatar could look... made it easier for the avatar to see what I was seeing. I could point it in much harder spots. Whereas I was doing more work to show the avatar what I was looking at in the Independent condition; as I knew I couldn't move him around." – P24

4.6 Study Part C – 360 Video Dependencies

Complementary to Part B, which examined the *Miniature* control by the *Giant*, Part C compared the corresponding 360-video viewing of those three levels of control. The participants, as the *Miniature*, were tasked with learning the surrounding environment from a playback of pre-recorded 360 videos and tracking data. In this study, we were interested in how the level of view dependencies of the 360-videos affects the resulting level of simulator sickness and spatial presence. View independence has been investigated in MR remote collaboration systems sharing live 2D video [12, 32] or 3D reconstructed scenes [14, 19, 68], yet not much with sharing live 360 panorama. The closest prior work is [30] which included an observational study on communication behavior but without direct comparison between dependent and independent views. Another prior work [45] investigated view dependency in a formal study, but the 360-camera was mounted on the user's head and did not include semi-dependent condition.

4.6.1 Study Design. Part C was a within-subject design where we investigated three levels of view dependencies of a 360-video (Figure 6): *Independent* (fully independent from camera's rotation), *Semi-dependent* (dependent on yaw rotation from the camera), *Dependent* (fully dependent on camera's rotation). These corresponded to the level of control that the *Giant* had over the *Miniature* avatar in Part B. In the *Independent* case, participants, the *Miniature*, could see a 360-video in a consistent orientation, viewing independently of the 360 camera's rotation. In the *Semi-dependent* condition, the *Miniature*'s view was affected by a sideway rotation around the ground vector. In the *Dependent* condition, the *Miniature* experienced the full impact of the 360 camera's rotation. The only objective dependent variable for this study was the measure of spatial understanding, which was taken as the score from solving spatial task. Our subjective measures were simulator sickness using the Kennedy Lane simulator sickness questionnaire (SSQ) [31]; Spatial Presence (MEC) [73] with 6 item scale on Spatial Situation model (SSM), and 6 item scale on Spatial Presence: Self Location (SPSL) subscales; SMEQ, task difficulty, and user preferences. SSQ were collected before and after each condition. Spatial Presence and SMEQ were collected after each condition and the user preference were collected at the end of Part C.

4.6.2 Tasks. To ensure consistent experience across all sessions, we pre-recorded the 360-video and the 360-camera's 6DOF tracking data. In this way, we could apply different level of dependencies to the same 360 video playback. To record the 360-video stream, we used OBS, Open Broadcaster Software [67]. We have created spatial tasks and recorded videos of three different tasks with an

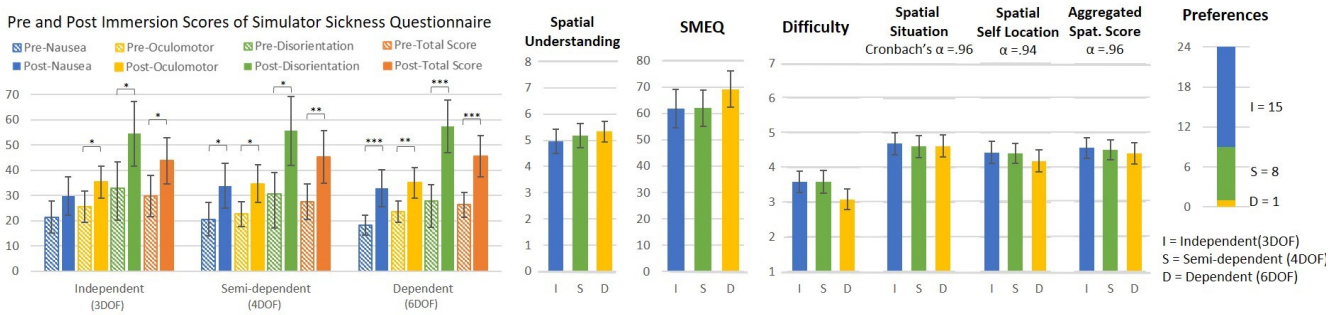

Figure 9: Results from Part C (*=p<.005, **=p<.01, and *<0.05). Error bars represent standard errors.**

actor holding a 360 camera, walking around the room, simulating a collaboration akin to Part A and B. Our focus was on the level of simulator sickness and spatial presence.

To evaluate spatial understanding, we created a spatial arrangement task where the participants had to arrange a set of objects in the room to match the original arrangement shown in the 360-video. While the participants were viewing the 360-video, the actor rearranged the room, making 8 changes. To score points, the participants needed to move the objects back to the original place. There was some flexibility allowed, e.g. numbered boxes could be in different order as far as they formed the correct layout. We created three unique arrangements and they were presented in the same order for every participant. However, the condition order was counter-balanced between participants, therefore each condition had the same number of these arrangements completed.

After filling out the pre-immersion SSQ, the participant watched the video, then filled out the post-immersion SSQ. They watched the same video again, and then tried to rearrange the objects to match the video. Finally, they filled out the per-condition questionnaire. The length of each 360 video was one minute and the resting period between watching a video from the last condition to the next one was around 10 minutes. They repeated this process for all three conditions, for 40 minutes.

4.6.3 Hypotheses. Our hypotheses for this part were:

C1. A lower degree of view dependency lowers subjective mental effort and task difficulty.
C2. A lower degree of view dependency causes a lower level of simulator sickness.
C3. A lower degree of view dependency increases *Spatial Presence score* (*ASpP*), in terms of *Spatial Situation model* (*SSM*), *Spatial Presence: Self Location* (*SPSL*) subscales.
C4. A lower degree of view dependency improves *spatial understanding* (*SpU*).
C5. Participants prefer the *Independent* viewing condition.

4.6.4 Results. We applied a Friedman test and Wilcoxon signed-rank test with Bonferroni correction (p-value adjusted) for post-hoc pairwise comparisons. Figure 9 presents our results.

4.6.4.1 Simulator Sickness. We did not find any significant difference between the conditions. However, significant differences were found between pre and post immersion, in the *Independent* condition except *Nausea* (*Oculomotor*, V=33.5, p=.044, *Disorientation*, V=23, p=.021, *Total Score*, V=41, p=.018), in the *Semi-dependent* condition (*Nausea*, V=8, p=.014, *Oculomotor*, V=11.5, p=.019, *Disorientation*, V=16, p=.024, *Total Score*, V=17.5, p=.0097), and in the *Dependent* condition (*Nausea*, V=9, p=.0039, *Oculomotor*, V=15, p=.0066, *Disorientation*, V=14, p=.0019, *Total Score*, V=20, p=.0016). We also tested the differential score (post - pre) between conditions but did not find any significant difference.

4.6.4.2 Subjective Mental Effort and Task Difficulty. We found no significant difference for both variables.

4.6.4.3 Spatial Presence. Internal consistency yielded excellent results (*Cronbach's* $\alpha > 0.96$) for all measured subscales, however, we did not find any significant difference in any of the measures.

4.6.4.4 Spatial Understanding. No significant difference was found for our spatial understanding measure.

4.6.4.5 User Preferences. Most of the participants preferred the Independent condition (63%) followed by the Semi-dependent condition (33%). A Chi-squared goodness of fit test yielded a significant difference from random choice, $\chi^2(2) = 12.25$, p=.002.

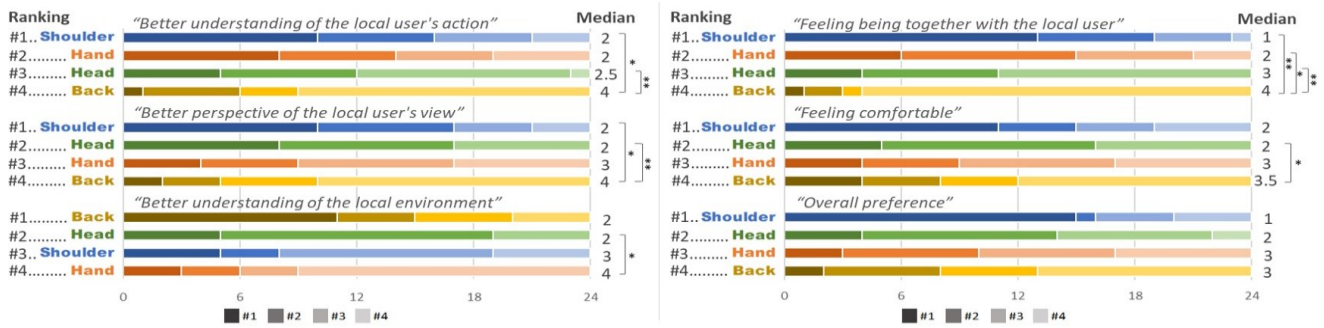

Figure 10: Results from Part D (=p<.01, and *<0.05). Error bars represent standard errors.**

4.6.5 Discussion. The results strongly support our hypothesis C5 that the participants preferred the *Independent* condition. We found significant differences between pre and post immersion for SSQ in each condition but not between conditions. Interestingly, the p-values were lower for the *Independent* condition, followed by *Semi-dependent*, and *Dependent* had the greatest p-value between pre and post immersion. There was no significant result to support hypotheses C1, C3 and C4.

4.6.5.1 ***Less Sickening?*** Although the SSQ results did not yield any significance, the translational movement remained, and this might be a possible cause for the increase of simulator sickness in all conditions. However, some participants reported feeling less sick in the independent viewing condition (3DOF). One said:

"It didn't make me feel sick and so I could concentrate on the task" – P2

We believe that further studies, with longer exposures to the 360-video, are required to better examine the effects of 360-video view dependencies. There are also opportunities for new techniques for reducing simulator sickness due to translational movement of the camera. The example known methods are reducing the FOV during the movement or freezing the video frame during a large translation. However, there are tradeoffs for each method needing further investigation along with other approaches.

4.6.5.2 ***Easier for Spatial Understanding.*** Some participants expressed that *Independent* view gave them the freedom to learn the surrounding and circumvented distractions of the camera movements.

"It makes remembering things a lot easier when you don't have your focus taken from you all the time." – P5

"...it allowed me the most control out of all of the setups, with the least disorientation, and the most time afforded for accurately memorizing the positions and formations of the objects. I was the least aware of the partner's movements and position, and the most aware of my own movements and positions..." – P6

4.6.5.3 ***Semi-dependent, Best to Follow?*** Some participants stated that if the task was to understand the *Giant*'s attention or action, the *Semi-dependent* condition would be more preferred.

"The task was remembering layout of objects in whole environment, so I had to keep focus on each set of objects for a while. Semi-dependent and Dependent conditions made me to rotate my body or head a lot. If the task was to identify or to remember the specific object what local user intended, I'd prefer semi-dependent." – P22

4.7 Study Part D – Camera Placements

In Parts A to C, the 360-camera was always held in the *Giant*'s hand. In this last study Part D, we were interested in comparing different 360-camera placement positions on the *Giant*'s body. From past research, we picked two placement positions, Backpack (*Back*) [8, 69] and Head-worn (*Head*) [30, 44]. We also proposed two more positions, Shoulder-worn (*Shoulder*) and Handheld (*Hand*), see Figure 8 (d1, d2, d3, and d4).

4.7.1 Study Design. Part D was a within-subject design with one independent variable, the camera placement position with 4 conditions, *Back*, *Head*, *Hand*, *Shoulder*. The dependent variable was subjective ranking of conditions, as shown in Figure 10.

4.7.2 Task. We let the participants watch and explore the pre-recorded 360 videos in *Dependent* view recorded from the same location, simulating a remote house inspection scenario. In every video, the 360-camera was mounted on the *Giant*, played by the actor. The *Giant* performed actions and interacted with objects around the house, such as reading the power meter, examining the security alarm, etc. The duration of each video was one minute. We interviewed our participants after each condition to

learn what they liked or disliked about the placement. After the participants watched all four videos, they filled out a questionnaire to rank the placements under six categories. This last study took approximately 10 minutes to complete.

4.7.3 Results. The results are shown in Figure 10. A Friedman test yielded significant differences in every category (*Action* - $\chi^2(3) = 14.05$, p=.0028, *Perspective* - $\chi^2(3) = 17.75$, p=.0005, *Environment* - $\chi^2(3) = 14.6$, p=.0022, *Together* - $\chi^2(3) = 29.75$, p<.0001, *Comfort* - $\chi^2(3) = 8.55$, p=.036, *Overall* - $\chi^2(3) = 11.15$, p=.011). A Wilcoxon signed-rank test with Bonferroni correction (p-value adjusted) for post-hoc pairwise comparisons gave significant pairs.

4.7.4 Discussion. ***Shoulder*** was ranked best in five of six categories. From the interview, it was clear that mounting the 360-camera slightly in front of the shoulder allowed the remote user to see the environment in front of the local user as well as the face of the local user. The placement's height was almost the same as the local user's eye level. It was not affected by local user's head or hand's movement, while it was affected by the slight vertical displacement as the local user walked, which was also present in other conditions. ***Hand*** came second in terms of better understanding of user action and feeling being together. Participants found the camera's height too low as our actor held the camera in a natural pose, just below chest level. ***Head*** came second for better perspective, comfort, and the overall preference. It was found slightly too high compared to the expected eye level. ***Back*** came first for better understanding of environment. Some participants found this condition amusing as they could observe the *Giant* from a 3rd person perspective.

5 GENERAL DISCUSSION

The results of our four-part study have shown that our *GMC* with 360-video sharing and tangible interaction satisfied all three design requirements. Part A found having either an avatar or a frustum significantly improved the overall experience of the collaboration through a 360-video, and having no visualization hindered the local user's ability to notice the remote collaborator's attention. These findings support R3, which require the system to provide an appropriate visual representation of the remote VR user so that the local AR user is aware of the attention and to improve the communication fidelity.

The results of Part B and C showed that our *GMC* tangible interface offered mobility and flexibility for the local AR user to control the remote user's view, which upholds R2. However, it also revealed the conflict between the *Giant* and the *Miniature* on their opposing preferences for level of control and view dependency. Part D validated our belief that the shoulder-mounted position (with handheld supported) of the 360-camera would be the most favorable for the remote VR user to see the local AR user's face, action, and environment as required by R1. From these results, we give our design recommendations.

5.1 Design Recommendations

In this section, we share the insights for designing *GMC*. From part A, **we recommend using both, an avatar and a view frustum, as they are complementary**, but also offer an option to switch each of them on or off.

From Part B and C, we have learned that the *Giant* preferred having full control of the *Miniature*, but the *Miniature* preferred independent viewing. **This leads us to recommend a middle ground solution of *Semi-dependent (4DOF)* control and viewing as a default mode.** The *Giant* can still switch to *Dependent (6DOF)* mode, when full control is needed, and *Independent (3DOF)* mode, when the *Miniature* must take the lead. This view is similar to [40].

From part D, **we recommend using a Protruded-shoulder-worn mounting** (Figure 1-a2) for mobility, stability (fixed to torso), convenience to install/remove the 360-camera, being within the FOV of the *Giant*, and providing a good perspective of collaborator's face and action. This mount can potentially be used with a gimbal for hardware stabilization and independent viewing. We believe that this is an improvement over a traditional on-the-shoulder mounting [38].

5.2 Limitations

Although the *Giant-Miniature* metaphor has a lot of promise there are shortcomings in the current work. First, Part B and C of our study could have been conducted together using dyads of participants instead of using an actor. We used an actor due to the concern for participants' well-being and for better control of random variables that could affect the quantitative results. In Part B, we wanted the participants to have the freedom to hold and use the 360-camera however they liked, without any constraint or potential complaints from the VR user if they were also participants. For example, some participants were curious about the avatar and swung the camera around that could have increased simulator sickness of the VR user. We took note of these behaviors so that we could design better detection and prevention

for potential causes of simulator sickness. Our study design decision traded some of our qualitative findings for better quantitative results. In Part C, we used pre-recorded videos and tracking data to control the camera movements across all participants for the same level of experience.

Second, our study did not include explicit cues, such as ray pointing. In fact, our system originally supported raycast pointing (explicit cue). However, as our focus was on the virtual representations (implicit cue) with minimal extra cues, we removed this feature from our study to minimize confounding factors. We believe that a raycast pointing feature would have improved performance, especially for the Study Part A.

Third, we measured the latency of the 360-video stream to be M=0.75 s (SD=0.25). This was noticeable on the Frustum's overlaid video displayed to the participants. However, the participants learned to quickly adapt to it. We believe that this latency will be reduced as the relevant technologies improve.

Lastly, we used a VST-HMD in our study for its wide FOV (96°) to ensure that the virtual representations were visible to the participants. The camera frame rate was at 90 Hz for comfort, but the output resolution was VGA (480p). The participants had no issue with the tasks given, however, for real world applications, an untethered optical see-through HMD would be more suitable.

6 CONCLUSION AND FUTURE WORK

In this paper, we presented a multi-scale MR collaborative system with two modes, *Conventional Scale Collaboration (CSC)* and *Giant-Miniature Collaboration (GMC)*. In *GMC*, a local AR user, as a Giant, collaborates with a remote VR user, as a *Miniature*, through a 6DOF tracked tangible 360-camera interface. We conducted a four-part study to evaluate *GMC*. Study Part A found both virtual representations, an avatar and a frustum, to be crucial for *GMC*. Part B showed that the *Giant* preferred full *Miniature* control. However, Part C found that the *Miniature* preferred independent viewing. Part D showed that the Protruded-shoulder-worn placement of the 360-camera was most preferred for *GMC*. Based on the results, we recommended using both avatar and frustum for virtual representation, *Semi-dependent (4DOF)* control and 360-video viewing, and a Protruded-shoulder-worn mounting as a default setting for *GMC*.

This paper reported on a well-controlled quantitative study that focused on each component of *GMC* design. In the future we plan to conduct a qualitative study that observes the differences in communication behavior between *CSC* and *GMC* with dyads of participants instead of an actor on one side. Furthermore, we would like to evaluate them in the real use case with more complex tasks that involve real object manipulation. We are also interested in autonomous switching of virtual representations or camera control, based on the context of collaboration. Beyond MR collaboration based on HMDs, the idea of *GMC* can be applied to collaboration with typical mobile devices as well. A 360-camera mounted mobile device can display the remote user's avatar and also serve as a tangible interface. The remote user can view the 360-video on a mobile device utilizing a built-in orientation sensor.

ACKNOWLEDGMENTS

This research was supported by South Australia Fellowship. The authors would like to thank the participants for their participation and their invaluable feedback.

REFERENCES

[1] Sigurdur O Adalgeirsson and Cynthia Breazeal. 2010. MeBot: a robotic platform for socially embodied presence. in *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, IEEE Press, 15-22.

[2] Judith Amores, Xavier Benavides and Pattie Maes. 2015. Showme: A remote collaboration system that supports immersive gestural communication. in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, ACM, 1343-1348.

[3] Sathya Barathan, Gun A. Lee, Mark Billinghurst and Robert W. Lindeman. 2017. Sharing Gaze for Remote Instruction *ICAT-EGVE 2017 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, The Eurographics Association.

[4] Stephan Beck, Andre Kunert, Alexander Kulik and Bernd Froehlich. 2013. Immersive group-to-group telepresence. *IEEE Transactions on Visualization and Computer Graphics*, 19 (4). 616-625.

[5] Mark Billinghurst, Hirokazu Kato and Ivan Poupyrev. 2001. The MagicBook: a transitional AR interface. *Computers & Graphics*, 25 (5). 745-753.

[6] Mark Billinghurst, Alaeddin Nassani and Carolin Reichherzer. 2014. Social panoramas: using wearable computers to share experiences. in *SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applications*, ACM, 25.

[7] Scott Brave, Hiroshi Ishii and Andrew Dahley. 1998. Tangible interfaces for remote collaboration and communication *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, ACM, Seattle, Washington, USA, 169-178.

[8] Minghao Cai, Soh Masuko and Jiro Tanaka. 2018. Gesture-based Mobile Communication System Providing Side-by-side Shopping Feeling. in *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*, ACM, 2.

[9] Katherine M. Everitt, Scott R. Klemmer, Robert Lee and James A. Landay. 2003. Two worlds apart: bridging the gap between physical and virtual media for distributed design collaboration *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Ft. Lauderdale, Florida, USA, 553-560.

[10] Faro. FOCUS-M 70. Retrieved January 1, 2019 from https://www.faro.com/en-sg/products/construction-bim/faro-laser-scanner-focus/.

[11] Cédric Fleury, Alain Chauffaut, Thierry Duval, Valérie Gouranton and Bruno Arnaldi. 2010. A Generic Model for Embedding Users' Physical Workspaces into Multi-Scale Collaborative Virtual Environments. in *ICAT 2010 (20th International Conference on Artificial Reality and Telexistence)*, Adelaide, Australia.

[12] Susan R Fussell, Leslie D Setlock and Robert E Kraut. 2003. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, 513-520.

[13] Susan R Fussell, Leslie D Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer and Adam DI Kramer. 2004. Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction*, 19 (3). 273-309.

[14] Lei Gao, Huidong Bai, Gun Lee and Mark Billinghurst. 2016. An oriented point-cloud view for MR remote collaboration. in *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications*, ACM, 8.

[15] Lei Gao, Huidong Bai, Thammathip Piumsomboon, G Lee, Robert W Lindeman and Mark Billinghurst. 2017. Real-time Visual Representations for Mixed Reality Remote Collaboration.

[16] Garmin. VIRB 360. Retrieved January 1, 2019 from https://buy.garmin.com/en-AU/AU/p/562010.

[17] Steffen Gauglitz, Cha Lee, Matthew Turk and Tobias Höllerer. 2012. Integrating the physical environment into mobile remote collaboration. in *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, ACM, 241-250.

[18] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk and Tobias Höllerer. 2014. In touch with the remote world: Remote collaboration with augmented reality drawings and virtual navigation. in *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, ACM, 197-205.

[19] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk and Tobias Höllerer. 2014. World-stabilized annotations and virtual scene navigation for remote collaboration. in *Proceedings of the 27th annual ACM symposium on User interface software and technology*, ACM, 449-459.

[20] Jan Gugenheimer, Evgeny Stemasov, Julian Frommel and Enrico Rukzio. 2017. ShareVR: Enabling Co-Located Experiences for Virtual Reality between HMD and Non-HMD Users *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ACM, Denver, Colorado, USA, 4021-4033.

[21] Kunal Gupta, Gun A Lee and Mark Billinghurst. 2016. Do You See What I See? The Effect of Gaze Tracking on Task Space Remote Collaboration. *IEEE Transactions on Visualization and Computer Graphics*, 22 (11). 2413-2422.

[22] Chad Harms and Frank Biocca. 2004. Internal consistency and reliability of the networked minds measure of social presence *In M. Alcaniz & B. Rey (Eds.), Seventh Annual International Workshop: Presence 2004*, Valencia: Universidad Politecnica de Valencia.

[23] Keita Higuchi, Katsuya Fujii and Jun Rekimoto. 2013. Flying head: A head-synchronization mechanism for flying telepresence. in *2013 23rd International Conference on Artificial Reality and Telexistence (ICAT)*, 28-34. 10.1109/ICAT.2013.6728902

[24] Keita Higuchi, Ryo Yonetani and Yoichi Sato. 2016. Can Eye Help You?: Effects of Visualizing Eye Fixations on Remote Collaboration Scenarios for Physical Tasks. in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, 5180-5190.

[25] HTC. Vive. Retrieved January 1, 2019 from https://www.vive.com/au/product/.

[26] HTC. Vive Pro, Retrieved January 1, 2019 from https://www.vive.com/au/product/vive-pro/.

[27] HTC. Vive SRWorks SDK. Retrieved January 1, 2019 from https://developer.vive.com/resources/knowledgebase/intro-vive-srworks-sdk/.

[28] HTC. Vive Tracker. Retrieved January 1, 2019 from https://www.vive.com/au/vive-tracker/.

[29] Hiroshi Ishii. 2008. The tangible user interface and its evolution. *Communications of the ACM*, 51 (6). 32-36.

[30] Shunichi Kasahara, Shohei Nagai and Jun Rekimoto. 2017. JackIn Head: Immersive Visual Telepresence System with Omnidirectional Wearable Camera. *IEEE Transactions on Visualization and Computer Graphics*, 23 (3). 1222-1234. 10.1109/TVCG.2016.2642947

[31] Robert S Kennedy, Norman E Lane, Kevin S Berbaum and Michael G Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3 (3). 203-220.

[32] Seungwon Kim, Mark Billinghurst and Gun Lee. 2018. The Effect of Collaboration Styles and View Independence on Video-Mediated Remote Collaboration. *Computer Supported Cooperative Work (CSCW)*. 1-39.

[33] Seungwon Kim, Gun A Lee, Nobuchika Sakata, Andreas Dunser, Elina Vartiainen and Mark Billinghurst. 2013. Study of augmented gesture communication cues and view sharing in remote collaboration. in *Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on*, IEEE, 261-262.

[34] Seungwon Kim, Gun Lee, Nobuchika Sakata and Mark Billinghurst. 2014. Improving co-presence with augmented visual communication cues for sharing experience through video conference. in *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*, IEEE, 83-92.

[35] Kiyoshi Kiyokawa, Haruo Takemura and Naokazu Yokoya. 1999. A collaboration support technique by integrating a shared virtual reality and a shared augmented reality. in *Systems, Man, and Cybernetics, 1999. IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on*, IEEE, 48-53.

[36] Scott R. Klemmer, Mark W. Newman, Ryan Farrell, Mark Bilezikjian and James A. Landay. 2001. The designers' outpost: a tangible interface for collaborative web site *Proceedings of the 14th annual ACM symposium on User interface software and technology*, ACM, Orlando, Florida, 1-10.

[37] Regis Kopper, Tao Ni, Doug A Bowman and Marcio Pinho. 2006. Design and evaluation of navigation techniques for multiscale virtual environments. in *Virtual Reality Conference, 2006*, Ieee, 175-182.

[38] Takeshi Kurata, Nobuchika Sakata, Masakatsu Kourogi, Hideaki Kuzuoka and Mark Billinghurst. 2004. Remote collaboration using a shoulder-worn active camera/laser. in *Eighth International Symposium on Wearable Computers*, 62-69. 10.1109/ISWC.2004.37

[39] Hideaki Kuzuoka, Shinya Oyama, Keiichi Yamazaki, Kenji Suzuki and Mamoru Mitsuishi. 2000. GestureMan: a mobile robot that embodies a remote instructor's actions *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, ACM, Philadelphia, Pennsylvania, USA, 155-162.

[40] Joel Lanir, Ran Stone, Benjamin Cohen and Pavel Gurevich. 2013. Ownership and control of point of view in remote assistance *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Paris, France, 2243-2252.

[41] Morgan Le Chénéchal, Thierry Duval, Valérie Gouranton, Jérôme Royan and Bruno Arnaldi. 2016. Vishnu: virtual immersive support for HelpiNg users an interaction paradigm for collaborative remote guiding in mixed reality. in *Collaborative Virtual Environments (3DCVE), 2016 IEEE Third VR International Workshop on*, IEEE, 9-12.

[42] Morgan Le Chénéchal, Jérémy Lacoche, Jérôme Royan, Thierry Duval, Valérie Gouranton and Bruno Arnaldi. 2016. When the giant meets the ant an asymmetric approach for collaborative and concurrent object manipulation in a multi-scale environment. in *Collaborative Virtual Environments (3DCVE), 2016 IEEE Third VR International Workshop on*, IEEE, 18-22.

[43] Gun A Lee, Seungwon Kim, Youngho Lee, Arindam Dey, Thammathip Piumsomboon, Mitchell Norman and Mark Billinghurst. 2017. Improving Collaboration in Augmented Video Conference using Mutually Shared Gaze.

[44] Gun A Lee, Theophilus Teo, Seungwon Kim and Mark Billinghurst. 2017. Mixed reality collaboration through sharing a live panorama. in *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, ACM, 14.

[45] Gun A Lee, Theophilus Teo, Seungwon Kim and Mark Billinghurst. 2018. A User Study on MR Remote Collaboration using Live 360 Video *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR '18)*. , 153–164

[46] Daniel Leithinger, Sean Follmer, Alex Olwal and Hiroshi Ishii. 2014. Physical telepresence: shape capture and display for embodied, computer-mediated remote collaboration *Proceedings of the 27th annual ACM symposium on User interface software and technology*, ACM, Honolulu, Hawaii, USA, 461-470.

[47] Tamotsu Machino, Satoshi Iwaki, Hiroaki Kawata, Yoshimasa Yanagihara, Yoshito Nanjo and Kenichiro Shimokura. 2006. Remote-collaboration system using mobile robot with camera and projector. in *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, 4063-4068. 10.1109/ROBOT.2006.1642326

[48] Akira Matsuda, Takashi Miyaki and Jun Rekimoto. 2017. ScalableBody: a telepresence robot that supports face position matching using a vertical actuator *Proceedings of the 8th Augmented Human International Conference*, ACM, Silicon Valley, California, USA, 1-9.

[49] Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77 (12). 1321-1329.

[50] Jörg Müller, Tobias Langlotz and Holger Regenbrecht. 2016. PanoVC: Pervasive telepresence using mobile phones. in *Pervasive Computing and Communications (PerCom), 2016 IEEE International Conference on*, IEEE, 1-10.

[51] Benjamin Nuernberger, Kuo-Chin Lien, Tobias Höllerer and Matthew Turk. 2016. Interpreting 2d gesture annotations in 3d augmented reality. in *3D User Interfaces (3DUI), 2016 IEEE Symposium on*, IEEE, 149-158.

[52] Benjamin Nuernberger, Matthew Turk and Tobias Höllerer. 2017. Evaluating snapping-to-photos virtual travel interfaces for 3D reconstructed visual reality. in *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, ACM, 22.

[53] Ohan Oda, Carmine Elvezio, Mengu Sukan, Steven Feiner and Barbara Tversky. 2015. Virtual replicas for remote assistance in virtual and augmented reality. in *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, ACM, 405-415.

[54] Corey Pittman and Joseph J LaViola Jr. 2014. Exploring head tracked head mounted displays for first person robot teleoperation. in *Proceedings of the 19th international conference on Intelligent User Interfaces*, ACM, 323-328.

[55] Thammathip Piumsomboon, Gun A Lee, Barrett Ens, Bruce H Thomas and Mark Billinghurst. 2018. Superman vs Giant: A Study on Spatial Perception for a Multi-Scale Mixed Reality Flying Telepresence Interface. *IEEE Transactions on Visualization and Computer Graphics*. 1-1. 10.1109/TVCG.2018.2868594

[56] Thammathip Piumsomboon, Arindam Day, Barrett Ens, Youngho Lee, Gun Lee and Mark Billinghurst. 2017. Exploring enhancements for remote mixed reality collaboration *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, ACM, Bangkok, Thailand, 1-5.

[57] Thammathip Piumsomboon, Gun A Lee and Mark Billinghurst. 2018. Snow Dome: A Multi-Scale Interaction in Mixed Reality Remote Collaboration. in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, ACM, D115.

[58] Thammathip Piumsomboon, Gun A Lee, Jonathon D Hart, Barrett Ens, Robert W Lindeman, Bruce H Thomas and Mark Billinghurst. 2018. Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration. in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ACM, 46.

[59] VR QuickTime. 1995. An Image-Based Approach to Virtual Environment Navigation, Shenchang Eric Chen, Apple Computer, Inc. in *Siggraph, Computer Graphics Proceedings, Annual Conference Series*, 29-38.

[60] Abhishek Ranjan, Jeremy P. Birnholtz and Ravin Balakrishnan. 2007. Dynamic shared visual spaces: experimenting with automatic camera control in a remote repair task *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, San Jose, California, USA, 1177-1186.

[61] Peter Robinson and Philip Tuddenham. 2007. Distributed Tabletops: Supporting Remote and Mixed-Presence Tabletop Collaboration. in *Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'07)*, 19-26. 10.1109/TABLETOP.2007.15

[62] Bektur Ryskeldiev, Michael Cohen and Jens Herder. 2017. Applying rotational tracking and photospherical imagery to immersive mobile telepresence and live video streaming groupware. in *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, ACM, 5.

[63] Nobuchika Sakata, Takeshi Kurata, Takekazu Kato, Masakatsu Kourogi and Hideaki Kuzuoka. 2003. WACL: Supporting telecommunications using wearable active camera with laser pointer. in *null*, IEEE, 53.

[64] MHD Yamen Saraiji, Tomoya Sasaki, Reo Matsumura, Kouta Minamizawa and Masahiko Inami. 2018. Fusion: full body surrogacy for collaborative communication *ACM SIGGRAPH 2018 Emerging Technologies*, ACM, Vancouver, British Columbia, Canada, 1-2.

[65] Jeff Sauro and Joseph S Dumas. 2009. Comparison of three one-question, post-task usability questionnaires. in *Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, 1599-1608.

[66] Rajinder S. Sodhi, Brett R. Jones, David Forsyth, Brian P. Bailey and Giuliano Maciocci. 2013. BeThere: 3D mobile collaboration with spatial input *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Paris, France, 179-188.

[67] Open Broadcaster Software. OBS Studio. Retrieved January 1, 2019 from https://obsproject.com/.

[68] Matthew Tait and Mark Billinghurst. 2015. The Effect of View Independence in a Collaborative AR System. *Comput. Supported Coop. Work*, 24 (6). 563-589. 10.1007/s10606-015-9231-8

[69] Anthony Tang, Omid Fakourfar, Carman Neustaedter and Scott Bateman. 2017. Collaboration in 360 Videochat: Challenges and Opportunities, University of Calgary.

[70] Franco Tecchia, Leila Alem and Weidong Huang. 2012. 3D helping hands: a gesture based MR system for remote collaboration. in *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry*, ACM, 323-328.

[71] Unity3D. Unity3D Game Engine. Retrieved January 1, 2019 from https://unity3d.com.

[72] Valve Corp. SteamVR Plugin for Unity. Retrieved January 1, 2019 from https://assetstore.unity.com/packages/templates/systems/steamvr-plugin-32647.

[73] Peter Vorderer, Werner Wirth, Feliz Ribeiro Gouveia, Frank Biocca, Timo Saari, Lutz Jäncke, Saskia Böcking, Holger Schramm, Andre Gysbers and Tilo Hartmann. 2004. MEC Spatial Presence Questionnaire. *Retrieved Sept*, 18. 2015.

[74] Jacob O Wobbrock, Leah Findlater, Darren Gergle and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. in *Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, 143-146.

[75] Catherine Zanbaka, Paula Goolkasian and Larry Hodges. 2006. Can a virtual cat persuade you?: the role of gender and realism in speaker persuasiveness. in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM, 1153-1162.

[76] Xiaolong Zhang and George W Furnas. 2005. mCVEs: Using cross-scale collaboration to support user interaction with multiscale structures. *Presence: Teleoperators & Virtual Environments*, 14 (1). 31-46.