Vicarious: Context-aware Viewpoints Selection for Mixed Reality Collaboration

Faisal Zaman
Computational Media Innovation Centre, Victoria University of Wellington
New Zealand
faisal.zaman@vuw.ac.nz

Craig Anslow
School of Engineering and Computer Science, Victoria University of Wellington
New Zealand
craig.anslow@ecs.vuw.ac.nz

Taehyun Rhee
Computational Media Innovation Centre, Victoria University of Wellington
New Zealand
taehyun.rhee@ecs.vuw.ac.nz

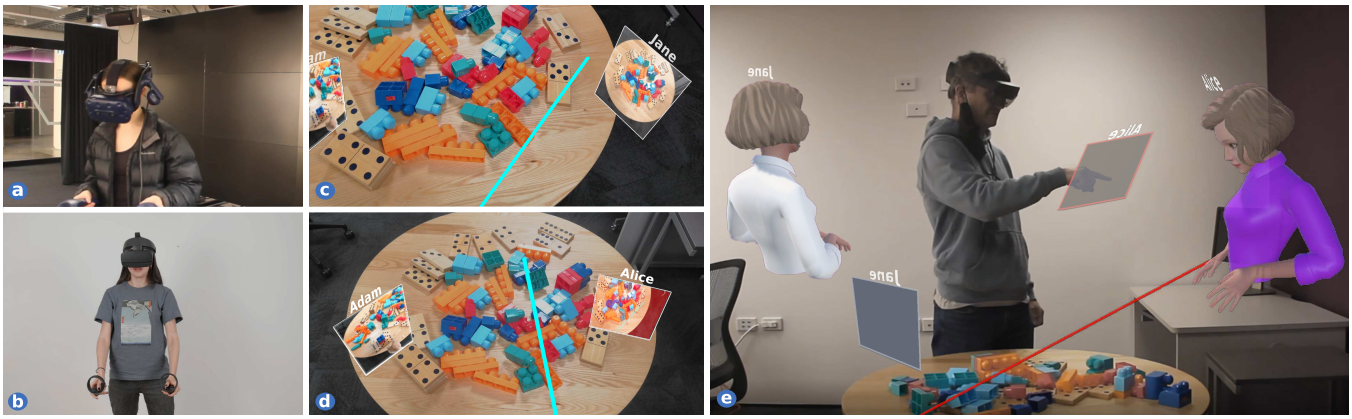


Figure 1: Two remote users (a and b), guide a local user (e) in a physical task. With *Vicarious*, all users have access to ego- and exocentric viewpoints, and the context-aware viewpoint selection method highlights the optimal viewpoint.

ABSTRACT

Mixed-perspective, combining egocentric (first-person) and exocentric (third-person) viewpoints, have been shown to improve the collaborative experience in remote settings. Such experiences allow remote users to switch between different viewpoints to gain alternative perspectives of the remote space. However, existing systems lack seamless selection and transition between multiple perspectives that better fit the task at hand. To address this, we present a new approach called *Vicarious*, which simplifies and automates the selection between egocentric and exocentric viewpoints. *Vicarious* employs a context-aware method for dynamically switching or highlighting the optimal viewpoint based on user actions and the current context. To evaluate the effectiveness of the viewpoint selection method, we conducted a user study ($n = 27$) using an asymmetric AR-VR setup where users performed remote collaboration tasks under four distinct conditions: *No-view, Manual, Guided*, and *Automatic* selection. The results showed that *Guided* and *Automatic* viewpoint selection improved users' understanding of the task space and task performance, and reduced cognitive load compared to *Manual* or *No-view* selection. The results also suggest that the asymmetric setup had minimal impact on spatial and social presence, except for differences in task load and preference. Based on these findings, we provide design implications for future research in mixed reality collaboration.

CCS CONCEPTS

• **Human-centered computing** → **Mixed / augmented reality**; **Collaborative interaction**.

KEYWORDS

Mixed Reality, Remote Collaboration, Telepresence, 360-degree Panoramic Video, Viewpoint Sharing, Perspective Sharing.

ACM Reference Format:
Faisal Zaman, Craig Anslow, and Taehyun Rhee. 2023. Vicarious: Context-aware Viewpoints Selection for Mixed Reality Collaboration. In *29th ACM Symposium on Virtual Reality Software and Technology (VRST 2023), October 9–11, 2023, Christchurch, New Zealand*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3611659.3615709

VRST '23, October 09–11, 2023, Christchurch, NZ

ACM ISBN 979-8-4007-0328-7/23/10…$15.00
https://doi.org/10.1145/3611659.3615709

1 INTRODUCTION

While traditional video conferencing tools have gained popularity for remote collaboration, they come with limitations, such as a lack of spatial presence and peripheral awareness, which hinder the seamless exchange of information and coordination among collaborators [47]. Immersive collaboration solutions such as VR/AR/MR

address these by creating shared spaces through 3D reconstruction using depth sensors and/or photogrammetry [61, 63], or 360° views captured by panorama cameras [48]. However, due to the limited field of view of head-mounted displays, the annotations or actions of remote users may not always be visible

Researchers have investigated various visual communication cues, including the pointer [25, 33], gaze [13, 18, 31], view frustum [10, 37, 59], and combinations of these cues [26, 27], to guide remote user attention [9, 42]. In complex tasks, a single viewpoint can be difficult to understand and explore the dynamic physical space. Moreover, it becomes difficult to perceive others' actions without constantly shifting one's focus [21, 22]. Therefore, prior work has explored multiple viewpoints by integrating egocentric and exocentric viewpoints and sharing them with remote users [8, 28]. Multiple viewpoints allow users to switch between different perspectives and share their viewpoints with others in real-time thus providing a more comprehensive view of the environment.

However, multiple viewpoints can be challenging for users to determine which views to focus on, potentially leading to overlooked information or user actions. While one possible solution is for users to physically move around or toggle between viewpoints to gain the desired view, this requires additional time and effort [17]. Moreover, relying on users to navigate or switch between views can be inefficient, as it adds to the task load of determining the most appropriate viewpoint to focus on at any given moment [1].

We propose a novel approach that simplifies and automates the selection of egocentric and exocentric viewpoints. We employ a context-aware method for selecting and dynamically switching or highlighting optimal viewpoints based on user actions and the current context. We conducted a user study (n=27) in which remote users guide a local user in a collaborative task in the local space under four different conditions: *No-view*, *Manual*, *Guided*, and *Automatic* selection. The results showed that *Guided* and *Automatic* viewpoint selection improved understanding of the local space and task performance and reduced cognitive load compared to *Manual* or *No-view* selection.

The contributions of this paper are as follows:

- A *context-aware viewpoint selection method* that simplifies and automates the selection of egocentric and exocentric viewpoints based on visual saliency, user actions, and speech patterns.
- A *user study (n = 27)* evaluate the impact of context-aware viewpoint selections method on collaboration performance and user experience under four distinct conditions (*No-view, Manual, Guided, and Automatic*) and two user roles (*local and remote*).
- The results provide *insights and recommendations* for design implications and directions for future research.

2 RELATED WORK

Our research builds upon earlier work on viewpoint sharing in telepresence and mixed-perspectives sharing. We review existing research within these areas and highlight the research gap that this paper addresses.

2.1 Viewpoint Sharing in Telepresence

Viewpoint sharing has been studied for decades, as it allows sharing remote user perspectives with other collaborators and experiencing the remote environment through others' eyes as if they were physically present at that location [53]. Telepresence systems have explored the concept of out-of-body view where a live 360° video of a local person's surroundings shared with remote collaborators [23, 28, 38]. It allows the user to seamlessly switch between a first-person perspective, and a third-person perspective to explore the remote workspace and improves the sense of presence for remote collaborators [28].

Veas et al. [65] showed that having multiple viewpoints improved spatial understanding and situational awareness during collaboration. Chellali et al. [6] found sharing viewpoints enhanced co-presence and awareness during remote object manipulation tasks. The lack of awareness of the remote user's view direction has been found to diminish users' sense of embodiment [29] while providing independent viewing directions improved the sense of presence for remote users [49, 68]. Integrating point clouds with 360-degree videos enhances the viewing experience by providing depth perception, 3D scene reconstruction, and improved interaction [61, 69]. This fusion offers a more immersive environment and a better understanding of spatial relationships. However, technical challenges like data processing and alignment need to be addressed for effective integration.

Flying or aerial telepresence explored alternative viewing positions based on unmanned aerial vehicles (UAVs) or drones emphasizing the significance of viewpoint in remote collaboration [19, 50], teleoperation [62], and telepresence [44]. Viewpoint manipulation in such applications had a large impact on user perception, control ease, collaboration, and overall system effectiveness [54].

While the use of viewpoint sharing in MR-based collaboration is not new, the novelty of this paper lies in combining visual cues with contextual information to automatically suggest which view users should be focusing on. This aspect has not been explored or evaluated in prior work.

2.2 Mixed-Perspectives Sharing

Mixed-perspective representations have been extensively studied in prior work. For example, in Dollhouse VR [20] multiple users are able to navigate an interior design using a combination of exocentric and egocentric perspectives were found useful for gaining an overall understanding of the interior design and spatial layout for users outside the VR. Similarly, ShareVR [12] combines floor projection, mobile displays, and positional tracking to render the virtual world to non-HMD users to enable multiple perspectives of the shared physical and virtual space. TransceiVR [64] leverages screen sharing and spatial annotation to enable external users to explore VR scenes, reducing communication barriers between users in VR and non-VR.

ARgus [8] explores the trade-offs of such mixed representations in an AR workspace, where they combine different view representations (Headset View, External View, and Virtual View). They found mixed representations improve efficiency and spatial understanding, and reduce reliance on verbal instructions.

Mixed perspectives are also explored in multi-scale interactions, such as Voodoo Dolls [41], which provide the user working at

Table 1: A high-level comparison of previous work outlined based on viewpoint sharing

	Selection Method	Input	Viewpoints	Annotation	Collaboration
Sasikumar et al. [51], Gao et al. [10]	×	Pointcloud	Ego	✓	One-to-One
Sodhi et al. [55]	×	2D, Pointcloud	Exo	✓	One-to-One
Le et al. [30]	×	2D, Pointcloud	Ego, Exo	✓	One-to-One
Muller et al. [36], Kratz et al. [29]	×	2D	Ego	×	One-to-One
Billinghurst et al. [4]	×	2D	Ego	✓	One-to-One
Piumsomboon et al. [45]	×	2D	Exo	✓	One-to-One
Young et al. [68]	×	2D,360	Ego	×	One-to-One
Ryskeldiev et al. [49]	×	2D,360	Ego	✓	One-to-One
Kasahara et al. [22], Komiyama et al. [28]	Manual	360,2D,3D	Ego,Exo	×	One-to-One
Le Chénéchal et al. [32]	Manual	360,2D,3D	Ego, Exo	✓	One-to-Many
Young et al. [69]	Manual	360, Pointcloud	Ego, Exo	×	One-to-One
Vicarious	Manual, Guided, and Automatic	360, 2D	Ego, Exo	✓	Many-to-Many

multiple scales the ability to manipulate both visible and occluded objects, along with an additional thumbnail view of the selected object. Snow Dome [43] introduces mixed-perspective representations by placing a remote VR user within a virtual 3D reconstruction of an AR user's space. This setup allows the VR user to experience the environment as a giant or miniature, offering them an overview of the AR user's workplace from different points of view and scales.

CollaVR [39], 360Anywhere [56], and SpaceTime [67] demonstrated viewpoint sharing increases the overall context in synchronous collaboration scenarios. Collaborative MR systems [34, 61] enable remote users to have an independent view of the overall task space through live 360 panoramas and reconstructed 3D models. Remote users can view and annotate the 3D scene from a different perspective, resulting in reduced task completion time.

While these existing systems offer manual view selection, there is potential for automatic or guided view selection to aid users in selecting viewpoints that enhance collaboration and provide valuable context that warrants further exploration.

2.3 Viewpoint Sharing and Transition

The earliest work on viewpoint sharing and transition techniques used a dot in the user's field of view (FoV) to indicate the gaze of the remote user to the local users and a picture-in-picture (PiP) mode to display remote users hand gesture [7]. Magic Book [3] presents a collaborative approach, where one person has an egocentric point of view of the inside of a book via VR, while another person can look from an exocentric perspective and allows seamlessly transport users between Reality and Virtuality. This approach has been seen as being particularly helpful for navigation tasks [11, 57].

Phillips and Piekarski [40] explore a possession metaphor in AR-to-VR transition that allows players to quickly switch viewpoints without physically traveling. While in Mobileportation [69], the user can switch between exo- and egocentric views simply by walking up to or away from their partner's avatar. Maintaining visual context is important during the transition as it helps the participant understand where they are in the global context. Fussell et al. [9] use a continuous transition to move from an egocentric perspective to an exocentric one. Pausch et al. [58] allow users of their immersive virtual reality system to place a camera icon on the world-in-miniature (WIM) map of their environment to seamlessly transition to different viewpoints. Lee et al. [32] use fade-in and fade-out effects when switching between a user's view to another's to prevent the expert from feeling sick. Influenced by these findings, *Vicarious* utilizes color glow around the viewpoint for highlighting and slide-in then zoom-in/out effects for smooth transitions.

Despite extensive research showing the benefits of providing multiple viewpoints of the collaboration space, we didn't find any prior work on automatic seamless selection and transitions between multiple viewpoints during collaboration. In Table 1, we have included some previous work that is more similar to ours and grouped them based on the input type, viewpoints selection method, annotation, and collaboration type. These categories were chosen based on the related work and more details are discussed in Section 3.

3 VICARIOUS

We provide a high-level overview of the design of *Vicarious*, highlighting the key components and outlining the main features and functionalities that enable context-aware viewpoint sharing and collaboration.

3.1 Design Overview

To facilitate efficient collaboration between local and remote users, the viewpoints of both users are captured and shared with each other. We utilize an asymmetric AR-VR collaboration setup where the local user's physical space is live-streamed using a 360° camera. This setup provides the remote user with a live panoramic view of the collaborative space. Remote users can offer assistance and support to the local user, even when they are not physically present. A local user uses an AR-HMD, while remote users use VR-HMDs. Local user's perspective is captured through the camera integrated into their AR-HMD and the remote users' perspective is captured based on what is rendered on their VR-HMD viewports.

Therefore, users have the option to choose from two different ways of viewing the environment at any given time:

- *Egocentric View*: Live video stream from the local user's AR camera is shared with the remote experts. The video stream is displayed in picture-in-picture (PiP) mode, which smoothly follows the user's field of view or can be anchored or pinned

to a fixed location in the world space. Users can interact with the PiP window by clicking on it (using either a VR controller or their hands in case of an AR user).

- *360° Camera View*: A live 360° camera, mounted in the local user's space, provides a panoramic view of the environment, which is then streamed to remote users. Remote users can view the video through a VR headset and have 3 degrees of freedom (3-DoF). Each VR user is represented as an avatar, and the movement of avatars is achieved by utilizing inverse kinematics from head and hand tracking.

Due to the limited field of view of the human peripheral vision and the constraints of AR-VR display technology, the remote user can only see a portion of the 360 spherical surface at any given time, depending on their viewport orientation. Therefore, the FoV of the remote user is also captured and broadcast to other users. This allows users to see where the remote user is currently looking. By combining these different environment representations, incorporating both ego- and exocentric perspectives, we employ a context-aware method to determine the optimal viewpoint.

3.2 Context-aware Viewpoint Selection

With several viewpoints available to users at any given time, our goal is to identify the viewpoint that provides the most comprehensive information about the current state of the task. To determine the optimal viewpoint, we consider the following criteria:

3.2.1 Contextual Information. First, from the list of ego- and exocentric viewpoints, we extracted saliency information and used

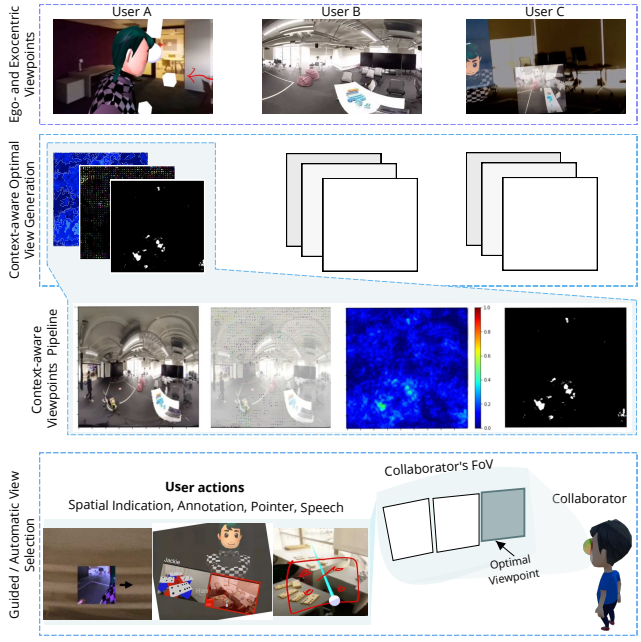


Figure 2: Pipeline of the pre-processing step. The pre-processing step first computes optical flow from an input 2D video, estimates saliency based on the optical flow, and finally, the binary threshold of the saliency map.

it as an indication of where the areas of interest are, this information is used as part of localizing attention. Optical flow and motion vector indicate the change of motion in the frame, so the value of optical and motion vector represents the motion saliency. We create saliency maps by combining three attention cues.

$$S(i) = w1 * K(i) + w2 * O(i) + w3 * G_\sigma(i) \quad (1)$$

where $w1$, $w2$, $w3$ map weights, $K(i)$ is the global contrast, and $O(i)$ optical flow. We use global contrast instead of local contrast for improved efficiency. We construct a histogram and calculate the difference for each color.

$$K(i) = \sum_{j=1}^{m} f(i)|C_i - C_j| \quad (2)$$

Lucas-Kanade method was used to track scene features points for changing, when local changes occur in the layers, motion vectors are recorded for later registration in the corresponding saliency map of the layer which resulted in $O = \{o_i \mid i = 1...T\}$ from the stacked video frames 2D video with T key frames $F = \{f_i \mid i = 1...T\}$. Next, the saliency maps are filled with a weight of 1 for the pixels corresponding to the registered feature points and facial regions for each layer. The probability of salient object appearance decreases with the distance from the center of an image and so we account for this by applying a Gaussian function $G_\sigma(i)$ to the saliency maps. This gives a saliency score for each viewpoint which highlights the most visually significant regions within the video.

3.2.2 Predefined Actions. While the saliency score is helpful in identifying visually prominent content, it is not the sole factor for determining the optimal viewpoint. We also consider the user's actions or features they are using at any given time. We provide a set of predefined functionalities or features, which users can access through user interfaces. This user-centric approach emphasizes incorporating the users' actions and tasks into the determination of the optimal viewpoint.

Verbal cues. All users can use voice chat to communicate among themselves. We used WebRTC's native AnalyserNode to identify voice activity within the audio stream and computed its intensity. First, we calculate the root mean square (RMS) value, which reflects the average amplitude of the audio signal and provides an indicator of voice intensity. Then, we periodically retrieve the time-domain data from the AnalyserNode, to calculate the voice intensity metric in real-time and analyze the audio stream for information about the voice intensity.

Visual cues. Users can point to any object using a 3D pointer, annotate, and perform gestures that are translated through avatars. We keep track of changes in these user activities on the scene and assign a weight value. Through a pilot study, we conducted experiments to fine-tune the weight values for each action performed using the user interface (UI). Specifically, we experimented with different combinations of weight values for three available actions: voice chat, annotating, and pointing (ordered by weight from high to low). Based on our experiments, we found that this ordering of weight values yielded better results.

A decision matrix is used to select the focus or automatically determine the viewpoint during collaboration based on the selected mode. Saliency scores and predefined action weights are assigned

to each viewpoint. Weighted scores are calculated by multiplying the ratings with their corresponding weights, and the viewpoint with the highest weighted score is selected.

3.2.3 Multiple-Viewpoint Awareness. Users can select their preferred view and also have the option to toggle between different views to see alternative perspectives. When a user selects their viewpoint, we visually highlight or indicate the presence of other viewpoints. This is done by displaying small thumbnail images or avatars of other users who have different viewpoints. Clicking on these thumbnails could provide a quick switch to that user's viewpoint. In Figure 2, we have illustrated lists of user actions, saliency scores, and active speakers that collectively influence the view selection process.

3.3 Visualization and Interaction

To visualize contextual information about important content outside of the user's current view, we overlay other users' points of view on the main thumbnail view, where the user can select them manually or automatically to bring them on top.

3.3.1 Interaction Interface. To ensure smooth transitions between viewpoints, the video gradually fades in when switching to a new viewpoint, avoiding sudden changes and providing a seamless transition. Similarly, when moving away from a viewpoint, the video fades out gradually to prevent abrupt visual changes that could cause disorientation.

In addition, the Picture-in-Picture (PiP) window adjusts dynamically to match the user's head movements. It smoothly follows the user's field of view, maintaining a consistent relative position within their visual perspective. This enables a natural and immersive viewing experience as the user looks around (Figure 3).

Alternatively, the PiP window can be anchored or pinned to a fixed location in the world space. It remains stationary relative to the surrounding environment, regardless of the user's head movements. This feature is useful when users need a consistent reference point or observe a specific area independent of the user's gaze.

3.3.2 Indication of Spatial Context. We track the direction and position of viewpoints thumbnails placed in world space. Since the user can position them anywhere, it leads to a disassociation between the avatar and their viewport. To address this issue, we used an arrow pointing toward the avatar of a remote user allowing local users to identify which remote user the viewport belongs to. By following the arrow, local users can quickly establish a connection between the viewpoint thumbnails they are observing and the corresponding remote user. The placement of egocentric and exocentric viewpoint thumbnails remains unchanged. In addition to the arrow indicator, we have also placed nametags on top of each viewpoint thumbnail, further assisting in identifying and associating specific viewports with the corresponding users.

3.4 Networking

Real-time audio and video communication are achieved using WebRTC within Unity3D, which acts as a client connecting to a Node.js web server through WebSocket. The web server acts as a central hub, facilitating signaling and negotiation between peers. To ensure direct connections between peers, TURN and STUN servers are utilized for NAT traversal. Each peer has dedicated video and audio streams for transmitting media, while non-media data exchange such as viewpoint synchronization, position, orientation, and annotations is handled through WebRTC Data Channels.

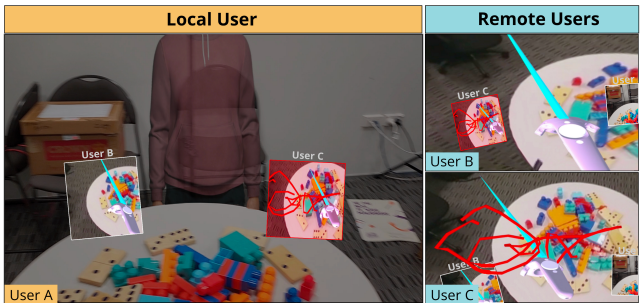


Figure 3: Viewpoint perspectives of local and remote users, highlighting remote user C's viewpoint in the GS condition.

4 USER STUDY

We conducted a user study to investigate the effectiveness of context-aware viewpoint selection as described in **Section 3**. The study employed a 4×2 mixed factorial design. The within-subjects variable, *Viewpoint Condition*, consisted of four levels (No-view, Manual, Guided, and Auto), while the between-subjects variable, *Participants Role*, consisted of two levels (Local and Remote user).

Hypotheses: We investigate two research questions:

RQ1 Would context-aware viewpoint selection increase the sense of presence in remote collaboration?

RQ2 Would context-aware viewpoint selection improve the task performance compared to using no-view selection?

Based on the research questions we formulated the following hypotheses:

H1 Guided viewpoint selection will result in a higher level of spatial presence and social presence compared to other conditions.

H2 Both guided and automatic viewpoint selection will lead to reduced task completion time compared to the no-view and manual selection conditions.

H3 Usability will be significantly higher in the guided and automatic viewpoint selection conditions.

H4 Participants will prefer guided viewpoint selection more than other conditions.

Participants: We recruited 27 participants (21 identified as male, 6 as female) aged 18 to 55 years (M = 28.13, SD = 7.32) through advertisements and local university and community center flyers. All participants had normal or corrected normal vision using glasses or contact lenses. The sample was diverse in terms of ethnicity, with 12 participants identifying as European, 6 as Asian, 5 as mixed race, 3 as Latino/Hispanic, and 1 participant identifying as Pacific Islander. The majority of participants (n = 18) reported having no prior experience with AR/VR technology, while a few (n = 9) had used it for an average of 10 hours (SD = 5.35) in the past year.

Experimental Conditions: Altogether, we had the four experimental conditions as follows:

(1) *No-view Selection (NS)*: is a stripped down version of *Vicarious* with all the view selection tools removed. Instead, users have full control over their viewing perspective without interference.
(2) *Manual Selection (MS)*: allows users to manually choose between ego- or exocentric viewpoints by clicking or moving PiP windows within their field of view. The selected view gradually fills the user's field of view and can be reverted by clicking anywhere on the window.
(3) *Guided Selection (GS)*: refers to the viewpoint selection technique that visually highlights the optimal view from the FoV list to prompt the local user to manually select the optimal viewpoint for the remote user's actions.
(4) *Automatic Selection (AS)*: the ego- or exocentric view is automatically selected as the main view, which the system considers to be the optimal viewpoint representing user actions.

Experimental Tasks and Setup: A collaborative task was designed, involving multiple remote users (using VR-HMD) guiding a local user (using AR-HMD) in building an assigned model using Legos and dominoes. The local user is in the task space where Legos and dominoes are randomly placed on a table. A 360° camera livestreams the task space to the remote users, who are located in a separate room. Both remote users have access to a visual representation of the desired final model and can provide instructions or guidance to the local user in locating specific Legos and dominoes in the task space. They can communicate steps to the local user through speech description, gesture pointing, and/or 3D annotation.

The local user finds the objects and builds the model using the physical Legos or dominoes based on the guidance instructions received from the remote users. Different model structures were used for each condition, and the order and combinations of tasks and conditions are counterbalanced to reduce bias. Depending on the experiment condition, participants had access to either ego- and/or exocentric viewpoints.

Procedure: At the start of the study, participants signed a consent form and provided demographic information, including any prior experience with AR/VR. After a brief overview of the experiment, participants were divided into groups of three, with two participants acting as remote experts and one as the local user constructing the model. Then they were given about 5 minutes to acquaint themselves with the system. Participants then proceeded to complete four conditions sequentially, with a 5-minute break in-between. Each condition lasts approximately 10-12 minutes. After completing each condition, the participants were given 5 minutes to complete subjective questionnaires. The task concluded when the local user completed building the model. After completing the study, participants were given a post-study questionnaire and asked to rate their preferences across four conditions based on different criteria. On average, the study lasted slightly over an hour.

Measures: We logged users' activities, including voice chat and completion time. To measure the spatial presence, we used the *iGroup Presence Questionnaire* (IPQ) [52], which has a 7-point Likert scale with four subscales: general presence (GP), realism (RL), involvement (INV), and spatial presence (SP). For social presence, we compiled a questionnaire with a 7-point Likert scale, including questions from various subscales: co-presence (CP) (from Bail [2] and Hauber [16]), as well as mutual attention (MA), mutual understanding (MU), and behavioral engagement (BE) (from the *NMM Social Presence Questionnaire* [14]). The workload was measured using the *NASA-TLX* [15], and system usability with the System Usability Scale (SUS) [5]. Motion sickness was evaluated using the *Simulator Sickness Questionnaire* (SSQ) [24]. The post-study questionnaire included measuring participants' preferences across various categories, which also included qualitative feedback through open-ended questions.

5 RESULTS

The Shapiro-Wilk test was conducted on the residuals to check for normality. Two-way mixed ANOVA ($\alpha = 0.05$) with Tukey's HSD post hoc test was performed for normally distributed data, and two-way mixed ANOVA with the Aligned Rank Transform (ART) [66] was used otherwise. Holm-Bonferroni correction was used for all post hoc tests.

Social Presence: Figure 4 shows the average Social Presence (SP) score of view selection conditions. Participants in the GS condition gave a significantly higher rating on the SP scale overall (Local: M = 5.78, SD = .21; Remote: M = 5.26, SD = .14) than those in the NS, MS, and AS. There was a significant main effect of the conditions (BE: $F_{3,183} = 3.11$, $p = .027$; CP: $F_{3,399} = 7.52$, $p < .001$; MA: $F_{3,291} = 14.06$, $p < .001$; MU: $F_{3,183} = 10.47$, $p < .001$), indicating that conditions had a statistically significant impact on the dependent variable. However, the main effect of the role was not statistically significant (BE: $F_{1,25} = 1.52$, $p = .228$; CP: $F_{1,25} = 2.16$, $p = .153$; MA: $F_{1,25} = 0.02$, $p = .885$; MU: $F_{1,25} = 0.17$, $p = .680$). Additionally, the interaction between conditions and role was not statistically significant (BE: $F_{3,183} = 0.46$, $p = .709$; MA: $F_{3,291} = 1.08$, $p = .356$; MU: $F_{3,183} = 0.86$, $p = .465$) except for CP: $F_{3,399} = 3.62$, $p = .013$. Post hoc pairwise comparisons suggest that participants in the GS condition felt significantly higher social presence compared to the NS (BE: $p = .017$; CP: $p < .001$; MA: $p < .001$; MU: $p < .001$) and MS (CP: $p = .018$; MA: $p = .0002$, MU: $p = .001$).

Spatial Presence: Figure 5 shows the average Spatial Presence (SP) score of view selection conditions. There was a significant main effect of the conditions for GP: $F_{3,75} = 0.97$, $p = .414$; INV: $F_{3,183} =$

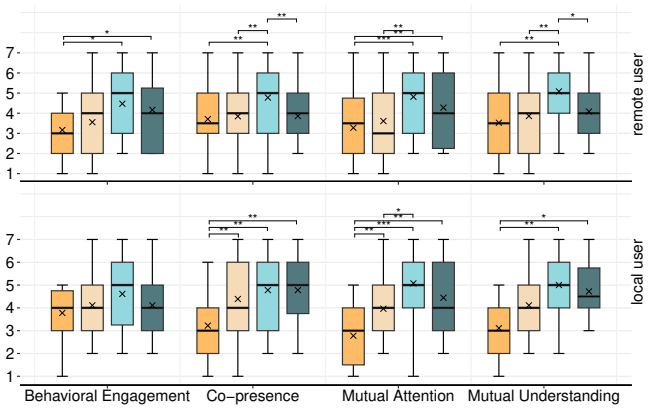


Figure 4: Social Presence results in each subscale (*=p<.05, **=p<.01, ***=p<.001). NS MS GS AS

5.99, $p = .001$; RL: $F_{3,183} = 5.89$, $p < .001$), with marginal main effect for SP: $F_{3,507} = 2.57$, $p = .053$ indicating that conditions had a statistically significant impact on the dependent variable. However, the main effect of the role was not statistically significant (GP: $F_{1,25} = 0.98$, $p = .331$; INV: $F_{1,25} = 0.83$, $p = .370$; RL: $F_{1,25} = 1.45$, $p = .239$; SP: $F_{1,25} = 0.0009$, $p = .999$). Additionally, the interaction between conditions and role was not statistically significant (GP: $F_{3,75} = 0.58$, $p = .629$; INV: $F_{3,183} = 0.18$, $p = .904$; RL: $F_{3,183} = 1.28$, $p = .282$; SP: $F_{3,507} = 1.56$, $p = .198$). Participants in the GS condition gave a significantly higher rating on the SP scale overall (Local: M = 6.21, SD = 1.49; Remote: M = 6.08, SD = 1.38) than those in the NS, MS, and AS. Post hoc pairwise comparisons suggest that participants in the GS condition felt significantly higher spatial presence compared to the NS (SP: $p = .053$), MS (INV: $p < .001$, RL: $p < .001$), and AS (RL: $p = .043$).

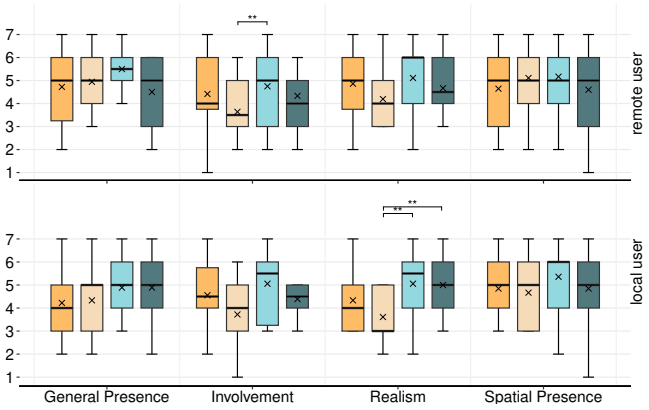


Figure 5: Spatial Presence results in each subscale (*=p<.05, **=p<.01, ***=p<.001). NS MS GS AS

System Usability: SUS scores for both VR and AR fell within an acceptable range (see Figure 6). The VR role received an average rating ($M = 68.71$, $SD = 22.17$), while the AR role received an "ok" rating ($M = 60.13$, $SD = 23.63$). These findings suggest that there were no significant differences in usability scores between VR and AR and that the different conditions did not significantly affect the overall SUS scores. A two-way ANOVA results indicated that there was no significant main effect of role ($F_{1,25} = 4.04$, $p = .053$), suggesting that the SUS scores in VR (M = 68.71, SD = 22.17) did not differ significantly from AR (M = 60.14, SD = 23.63). Similarly, the main effect of conditions was not significant ($F_{3,75} = .98$, $p = .404$), and thus failed to show a significant difference between conditions. The interaction between role and conditions was also not significant ($F_{3,75} = .79$, $p = .497$), suggesting that the relationship between role and conditions did not significantly influence the SUS scores.

Completion Time and Task Load: The completion time for searching Lego blocks shown in Figure 7, did not differ significantly among the conditions ($F_{3,140} = .07$, $p = .973$). Bartlett's test of sphericity was not statistically significant ($\chi^2 = 2.97$, $p = .395$). For assembling Lego blocks, there was a significant difference in completion time among the conditions ($F_{3,140} = 8.96$, $p < .001$). Pairwise comparisons showed significant differences between the *GS* and both *MS* ($p = .001$) and *NS* ($p < .001$) conditions. In total

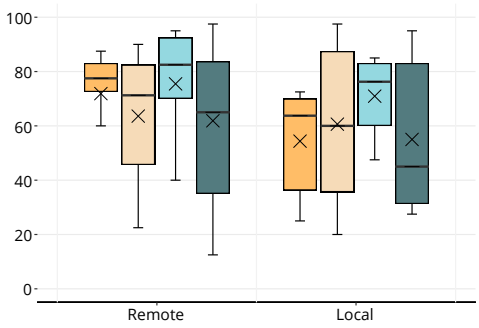


Figure 6: System usability results (80.3 or higher is considered good, 68 and above classified as average, and below 51 considered poor). NS MS GS AS

time for all Lego blocks, there was a significant difference among the conditions ($F_{3,140} = 14.63$, $p < .001$). The assumption of equal variances was reasonable (Bartlett's $\chi^2 = 7.55$, $p = .056$). Pairwise comparisons showed significant differences between the *GS* condition and the other conditions (*AS*, $p = .002$; *MS* and *NS*, $p < .001$).

Furthermore, we summed up the six subscales of the NASA-TLX with their weights to obtain the overall NASA-TLX score (see Figure 8). The overall NASA-TLX score met the assumption of homogeneity of variances and indicate a significant effect on the overall NASA-TLX score ($F_{2,18} = 5.70$, $p = .010$). Results of the posthoc test indicate that the overall NASA-TLX score in the GS condition (M = 35.6, SD = 17.94) was significantly lower than that in the other conditions.

Simulator Sickness: Figure 9 shows the average score of SSQ questionnaire [24], with 16 items rated from 0: none - 3: severe, then calculated the three subscales (nausea, oculomotor, and disorientation) and the total score. The SSQ was administered pre-experiment and post-experiment for each task in each condition. The results indicate very low simulator sickness scores for MS conditions. SSQ values for the NS_{post} (M = 4.71, SD = 4.41) and AS_{post} (M = 3.76, SD = 3.84) do not show significant differences between SSQ values for the NS_{pre} (M = 4.57, SD = 6.62) and AS_{pre} (M = 4.22, SD = 5.72), indicating that the different viewpoint selection conditions may not have induced simulator sickness.

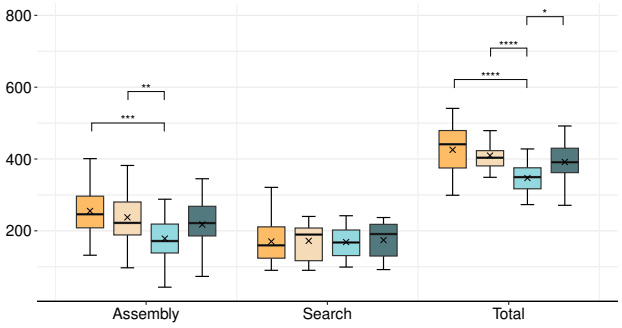


Figure 7: Average task completion time for each condition (*=p<.05, **=p<.01, ***=p<.001). NS MS GS AS

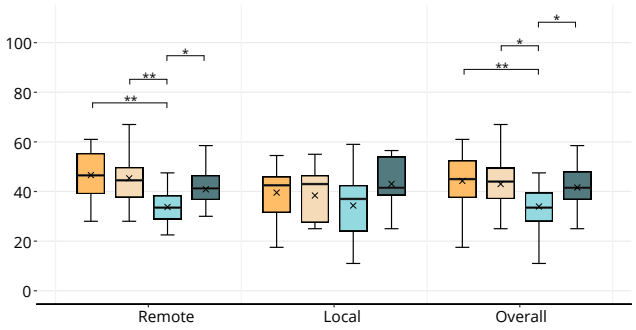


Figure 8: NASA-TLX score (0: very low to 100: very high) (*=p<.05, **=p<.01, ***=p<.001). NS MS GS AS

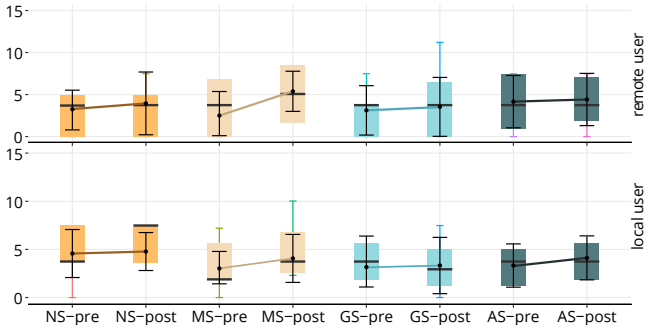
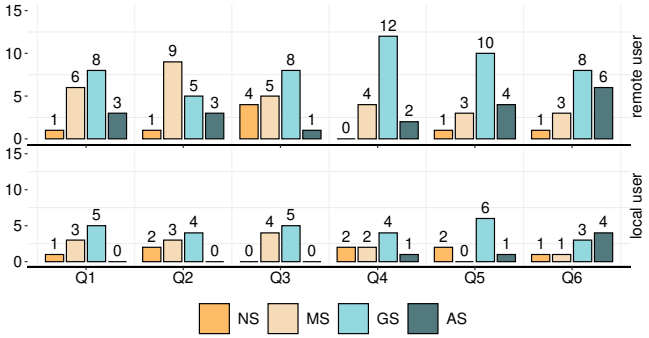


Figure 9: Increase in the SSQ score (pre- and post-exposure)



Q1	Which condition did you find easiest to use?
Q2	Which condition did you find most helpful for understanding the task environment?
Q3	Which condition did you find most helpful for understanding your partner's instructions?
Q4	Which condition did you find most effective for communicating with your partner?
Q5	Which condition did you find most helpful for completing the task?
Q6	Which condition did you prefer the most?

Figure 10: User preference among the four conditions.

Preferences: Figure 10 shows the responses to the user preference questionnaire regarding six different aspects of collaboration for both local and remote users across all four conditions. In general, the majority of participants showed a preference for the GS condition (48%), followed by the MS condition (27%). A chi-square goodness of fit test was used to determine whether the four conditions (NS, MS, GS, AS) were equally preferred and the results show preference was not equally distributed. For local users, the MS condition and for remote users, the GS condition was reported as the most user-friendly in terms of ease of use (Q1) $\chi^2(3) = 11.96$, $p = .007$. Regarding task environment understanding (Q2) $\chi^2(3) = 9$, $p = .029$ and task completion (Q5) $\chi^2(3) = 17.29$, $p < .001$, participants found both the GS and MS conditions to be most helpful. Additionally, the GS conditions were perceived as the most effective for understanding the partner's instructions (Q3) $\chi^2(3) = 12.55$, $p = .005$ and communicating effectively with partners (Q4) $\chi^2(3) = 18.18$, $p < .001$. Finally, both local and remote users reported GS as the most preferred condition (Q6) $\chi^2(3) = 8.70$, $p < .003$, while remote users specifically preferred both AS and GS. Chi-Square Test of Independence to determine any significant association between the participants' roles (VR vs. AR) and their preferences for specific conditions shows no significance.

6 DISCUSSION

The user study results indicate that Vicarious had a positive effect on improving collaboration in the MR remote collaboration compared to no-view selection.

RQ1 focused on assessing whether Vicarious enhances the sense of presence in remote collaboration.

As **H1** hypothesized that Guided Selection (GS) would lead to a higher level of presence compared to other conditions. The experimental results supported this hypothesis, as participants in the GS condition had higher IPQ scores in all subscales than in the other conditions. Participants found GS's highlighted feature effective understanding instructions, while NS lacked viewpoints and MS/AS had in the way, resulting in lower presence scores. [*"It felt like we were in the same room. By looking at the viewport, I could easily see what they were drawing and looking at, which made it easier."*]

RQ2 aimed to assess whether Vicarious improve the task performance compared to using the no-view selection (NS) condition.

As **H2** hypothesized GS and AS would reduce task completion time compared to NS and MS but the result didn't fully support this. Although GS had a faster task completion time AS took longer for assembly and searching tasks. This increase may be due to AS mode distracting the user's attention while automatically bringing the viewpoint into focus, resulting in an increased learning curve and task load (TLX: M=41.61, SD=13.63).

Similarly, the study partially supported **H3**, which hypothesized that usability would be highest in the GS and AS viewpoint selection conditions. However, while GS had a slightly higher usability score than AS, the difference was not statistically significant.

Lastly, **H4**, hypothesized that participants would have a strong preference for guided viewpoint selection. While this hypothesis was supported, the study also revealed a slightly similar preference for automatic selection. This suggests that both GS and AS helped participants navigate the remote environment effectively, and made them aware of other user activities. [*"The tool (GS) was spot-on, and the mini viewport window helped me understand exactly what my partner was seeing."*]

Communication Patterns: The remote users within the same group took turns speaking, alternating between the remote guide providing instructions and the physical builder seeking clarification or requesting assistance. The local user provided feedback to

the remote users regarding the progress, challenges, and questions related to the Lego blocks, and the remote guide responded with appropriate instructions and guidance. The remote users extensively used spatial deictic references to indicate specific locations on the Lego blocks. For example, they employed terms like "over there," "to the left," or "next to the blue brick" to provide precise instructions or indicate a point of reference. During this time, the viewpoint window was particularly observed as being used most. Per session, the system switched to the optimal viewpoints an average of 11.89 times during the AS condition ($SD = 2.55$). Remote users used their 3D pointer extensively, while the local user focused on the avatar and the viewing window of the remote user to find a reference. Additionally, during those instances, local users adjusted the PiP window with their hands. On the other hand, the local user utilized demonstrative pronouns (e.g., "this," "that") to refer to specific Lego blocks, picking them up to show the remote user. By employing these pronouns, they established a shared understanding of the objects being discussed. For the remote user's viewpoint, it was most used during that time, which we believe helped the remote user understand which piece the local user was referring to from an egocentric view rather than an exocentric view, although this wasn't uniformly the case. In situations requiring time-based instructions, partners used temporal deictic language. For instance, they might say, "Wait for a moment," "Place the brick after the blue one," or "Build this section first."

Design Implications: We learned several design implications from our user study which we think can be significant for future MR remote collaboration systems:

(1) *Flexibility in Viewpoint Selection:* Providing users with the flexibility to choose between different viewpoint selection conditions can enhance user satisfaction and adaptability to their specific preferences and needs. By offering a range of options, such as manual selection, guided selection, automatic selection, and no-view selection, systems cater to different user preferences and task requirements.
(2) *Guided Selection for Optimal Viewpoints:* The results indicate that the Guided Selection (GS) condition was preferred and performed well in various aspects. This suggests that incorporating audio and visual cues and highlighting the optimal viewpoints can assist users in understanding the task environment, communicating effectively with partners, and completing tasks. Designing intuitive and informative cues for guiding users toward optimal viewpoints can enhance the overall experience and task performance.
(3) *Consideration of User Context:* The study involved participants who were both remotely located and physically present, representing different contexts of collaboration. These findings highlight the importance of considering user context when designing viewpoint selection mechanisms. Also, user-specified preferences for predefined actions and adding more dynamic user action recognition would be helpful.
(4) *Usability and Workload Considerations:* The usability scores and NASA-TLX workload results provide insights into the user experience and cognitive load associated with each condition. Manually selecting the viewpoint adds additional workload and requires learning or getting used to, while automatic selection may require an initial understanding of its functionality.

Limitation and Future Work: In the study, a 360° camera was used, which is essentially a 2D representation. However, future research may consider incorporating depth information or utilizing teleoperating robots to introduce additional viewpoints. Future work should also refine the user actions triggering guided and automatic selection by incorporating cues like gaze tracking and view frustum analysis. Additionally, advanced machine learning algorithms can be explored for scene feature extraction and gesture recognition. It is worth noting that the current implementation prioritized real-time operation, which was crucial for collaboration, further experiments are needed to evaluate the effectiveness of the pipeline and identify potential enhancements. Decoupling the head-tracked viewpoint could offer advantages for external users, allowing more natural navigation and aligning visual perspectives with camera angles [60]. Instructors can lead learners through different viewpoints to facilitate remote mentoring and ensure precise guidance [35, 46]. Future user studies can also aim for larger and more diverse participant samples. Despite the small group size, this study provided valuable initial insights into the usability and preferences of viewpoint selection conditions for remote collaboration, indicating how these conditions may manifest for larger teams.

7 CONCLUSION

We presented *Vicarious*, a context-aware viewpoint selection method that simplifies and automates the selection between egocentric and exocentric viewpoints. Our user study ($n = 27$) evaluated four experimental conditions (*No-view*, *Manual*, *Guided*, and *Automatic*) in an asymmetric multiuser AR-VR setup. Results indicate that *Vicarious* improved users' understanding of the task space and task performance while reducing cognitive load. The Guided Selection (GS) was the most preferred, performing well across multiple metrics such as user preference, system usability, understanding of task space, and task completion. Automatic Selection (AS), was particularly favored in the AR context (local users) whereas VR users had a split preference between the GS and AS conditions. Lower scores for the No-view Selection (NS) across all metrics highlight the need for viewpoint selection methods presented in this paper.

Future research should compare various attributes to enhance our understanding of viewpoint selection mechanisms and their influence on collaboration.

Overall, the findings of our study underscore the effectiveness of the *Vicarious* method in simplifying and automating viewpoint selection for remote collaboration tasks.

ACKNOWLEDGMENTS

This project was supported by the Entrepreneurial University Program funded by TEC, New Zealand. Special thanks to Dr. Jacob Young for his valuable feedback and suggestions.

REFERENCES

[1] Parastoo Abtahi, Mar Gonzalez-Franco, Eyal Ofek, and Anthony Steed. 2019. I'm a giant: Walking in large virtual environments at high speed gains. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
[2] Jeremy N Bailenson, Kim Swinth, Crystal Hoyt, Susan Persky, Alex Dimov, and Jim Blascovich. 2005. The independent and interactive effects of embodied-agent

appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence* 14 (2005), 379–393.
[3] Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. 2001. The MagicBook: a transitional AR interface. *Computers & Graphics* 25, 5 (2001), 745–753.
[4] 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*. 1–1.
[5] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189 (1996), 4–7.
[6] Amine Chellali, Isabelle Milleville-Pennel, and Cédric Dumas. 2013. Influence of contextual objects on spatial interactions and viewpoints sharing in virtual environments. *Virtual Reality* 17, 1 (2013), 1–15.
[7] Sung Ho Choi, Minseok Kim, and Jae Yeol Lee. 2018. Situation-dependent remote AR collaborations: Image-based collaboration using a 3D perspective map and live video-based collaboration with a synchronized VR mode. *Computers in Industry* 101 (2018), 51–66.
[8] Arthur Fages, Cédric Fleury, and Theophanis Tsandilas. 2022. Understanding Multi-View Collaboration between Augmented Reality and Remote Desktop Users. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–27.
[9] Susan R Fussell, Robert E Kraut, and Jane Siegel. 2000. Coordination of communication: Effects of shared visual context on collaborative work. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. 21–30.
[10] Lei Gao, Huidong Bai, Rob Lindeman, and Mark Billinghurst. 2017. Static local environment capturing and sharing for MR remote collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*. 1–6.
[11] Raphaël Grasset, Philip Lamb, and Mark Billinghurst. 2005. Evaluation of mixed-space collaboration. In *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)*. IEEE, 90–99.
[12] Jan Gugenheimer, Evgeny Stemasov, Julian Frommel, and Enrico Rukzio. 2017. Sharevr: Enabling co-located experiences for virtual reality between hmd and non-hmd users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 4021–4033.
[13] 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 (2016), 2413–2422.
[14] Chad Harms and Frank Biocca. 2004. Internal consistency and reliability of the networked minds measure of social presence. In *Seventh annual international workshop: Presence*, Vol. 2004. Universidad Politecnica de Valencia Valencia, Spain.
[15] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.
[16] Jörg Hauber, Holger Regenbrecht, Mark Billinghurst, and Andy Cockburn. 2006. Spatiality in videoconferencing: trade-offs between efficiency and social presence. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*. 413–422.
[17] Carrie Heeter. 1992. Being there: The subjective experience of presence. *Presence: Teleoperators & Virtual Environments* 1 (1992), 262–271.
[18] Keita Higuch, 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*. 5180–5190.
[19] Aaron Hitchcock and Kelvin Sung. 2018. Multi-view augmented reality with a drone. In *Proceedings of the 24th ACM symposium on virtual reality software and technology*. 1–2.
[20] Hikaru Ibayashi, Yuta Sugiura, Daisuke Sakamoto, Natsuki Miyata, Mitsunori Tada, Takashi Okuma, Takeshi Kurata, Masaaki Mochimaru, and Takeo Igarashi. 2015. Dollhouse vr: a multi-view, multi-user collaborative design workspace with vr technology. In *SIGGRAPH Asia 2015 Emerging Technologies*. 1–2.
[21] Hyungeun Jo and Sungjae Hwang. 2013. Chili: viewpoint control and on-video drawing for mobile video calls. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 1425–1430.
[22] Shunichi Kasahara and Jun Rekimoto. 2014. JackIn: integrating first-person view with out-of-body vision generation for human-human augmentation. In *Proceedings of the 5th augmented human international conference*. 1–8.
[23] Shunichi Kasahara and Jun Rekimoto. 2015. JackIn head: immersive visual telepresence system with omnidirectional wearable camera for remote collaboration. In *Proceedings of the 21st ACM symposium on virtual reality software and technology*. 217–225.
[24] 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 (1993), 203–220.
[25] Seungwon Kim, Allison Jing, Hanhoon Park, Soo-hyung Kim, Gun Lee, and Mark Billinghurst. 2020. Use of Gaze and Hand Pointers in Mixed Reality Remote Collaboration. In *The 9th International Conference on Smart Media and Applications. SMA, Jeju, Republic of Korea*. 1–6.
[26] Seungwon Kim, Gun Lee, Mark Billinghurst, and Weidong Huang. 2020. The combination of visual communication cues in mixed reality remote collaboration. *Journal on Multimodal User Interfaces* 14 (2020), 321–335.
[27] Seungwon Kim, Gun Lee, Weidong Huang, Hayun Kim, Woontack Woo, and Mark Billinghurst. 2019. Evaluating the combination of visual communication cues for HMD-based mixed reality remote collaboration. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.
[28] Ryohei Komiyama, Takashi Miyaki, and Jun Rekimoto. 2017. JackIn space: designing a seamless transition between first and third person view for effective telepresence collaborations. In *Proceedings of the 8th Augmented Human International Conference*. 1–9.
[29] Sven Kratz, Don Kimber, Weiqing Su, Gwen Gordon, and Don Severns. 2014. Polly: " being there" through the parrot and a guide. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. 625–630.
[30] 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 *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE)*. IEEE, 9–12.
[31] Morgan Le Chénéchal, Thierry Duval, Valérie Gouranton, Jérôme Royan, and Bruno Arnaldi. 2019. Help! i need a remote guide in my mixed reality collaborative environment. *Frontiers in Robotics and AI* 6 (2019), 106.
[32] Geonsun Lee, HyeongYeop Kang, JongMin Lee, and JungHyun Han. 2020. A User Study on View-sharing Techniques for One-to-Many Mixed Reality Collaborations. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 343–352.
[33] Gun A Lee, Seungjun Ahn, William Hoff, and Mark Billinghurst. 2020. Enhancing first-person view task instruction videos with augmented reality cues. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 498–508.
[34] 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*. 1–4.
[35] Chengyuan Lin, Edgar Rojas-Munoz, Maria Eugenia Cabrera, Natalia Sanchez-Tamayo, Daniel Andersen, Voicu Popescu, Juan Antonio Barragan Noguera, Ben Zarzaur, Pat Murphy, Kathryn Anderson, et al. 2020. How about the mentor? effective workspace visualization in ar telementoring. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 212–220.
[36] Jörg Müller, Tobias Langlotz, and Holger Regenbrecht. 2016. PanoVC: Pervasive telepresence using mobile phones. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 1–10.
[37] Jens Müller, Roman Rädle, and Harald Reiterer. 2017. Remote collaboration with mixed reality displays: How shared virtual landmarks facilitate spatial referencing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 6481–6486.
[38] Shohei Nagai, Shunichi Kasahara, and Jun Rekimoto. 2015. Livesphere: Sharing the surrounding visual environment for immersive experience in remote collaboration. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*. 113–116.
[39] Cuong Nguyen, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. 2017. CollaVR: collaborative in-headset review for VR video. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. 267–277.
[40] Keith Phillips and Wayne Piekarski. 2005. Possession techniques for interaction in real-time strategy augmented reality games. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*. 2–es.
[41] Jeffrey S Pierce, Brian C Stearns, and Randy Pausch. 1999. Voodoo dolls: seamless interaction at multiple scales in virtual environments. In *Proceedings of the 1999 symposium on Interactive 3D graphics*. 141–145.
[42] Thammathip Piumsomboon, Arindam Dey, Barrett Ens, Gun Lee, and Mark Billinghurst. 2019. The effects of sharing awareness cues in collaborative mixed reality. *Frontiers in Robotics and AI* 6 (2019), 5.
[43] 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*. 1–4.
[44] 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* 24, 11 (2018), 2974–2982.
[45] Thammathip Piumsomboon, Gun A Lee, Andrew Irlitti, Barrett Ens, Bruce H Thomas, and Mark Billinghurst. 2019. 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*. 1–17.
[46] Kevin Ponto, Hyun Joon Shin, Joe Kohlmann, and Michael Gleicher. 2012. Online real-time presentation of virtual experiences forexternal viewers. In *Proceedings*

of the 18th ACM symposium on Virtual reality software and technology. 45–52.

[47] Jason Rambach, Gergana Lilligreen, Alexander Schäfer, Ramya Bankanal, Alexander Wiebel, and Didier Stricker. 2021. A survey on applications of augmented, mixed and virtual reality for nature and environment. In *Virtual, Augmented and Mixed Reality: 13th International Conference, VAMR 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings*. Springer, 653–675.

[48] Taehyun Rhee, Stephen Thompson, Daniel Medeiros, Rafael Dos Anjos, and Andrew Chalmers. 2020. Augmented virtual teleportation for high-fidelity telecollaboration. *IEEE transactions on visualization and computer graphics* 26 (2020), 1923–1933.

[49] Bektur Ryskeldiev, Michael Cohen, and Jens Herder. 2018. Streamspace: Pervasive mixed reality telepresence for remote collaboration on mobile devices. *Journal of Information Processing* 26 (2018), 177–185.

[50] Mehrnaz Sabet, Mania Orand, and David W. McDonald. 2021. Designing telepresence drones to support synchronous, mid-air remote collaboration: An exploratory study. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.

[51] Prasanth Sasikumar, Lei Gao, Huidong Bai, and Mark Billinghurst. 2019. Wearable RemoteFusion: A Mixed Reality Remote Collaboration System with Local Eye Gaze and Remote Hand Gesture Sharing. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 393–394.

[52] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments* 10 (2001), 266–281.

[53] Mickael Sereno, Xiyao Wang, Lonni Besançon, Michael J Mcguffin, and Tobias Isenberg. 2020. Collaborative work in augmented reality: A survey. *IEEE Transactions on Visualization and Computer Graphics* 28, 6 (2020), 2530–2549.

[54] Hanieh Shakeri and Carman Neustaedter. 2019. Teledrone: Shared outdoor exploration using telepresence drones. In *Conference companion publication of the 2019 on computer supported cooperative work and social computing*. 367–371.

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

[56] Maximilian Speicher, Jingchen Cao, Ao Yu, Haihua Zhang, and Michael Nebeling. 2018. 360anywhere: Mobile ad-hoc collaboration in any environment using 360 video and augmented reality. *Proceedings of the ACM on Human-Computer Interaction* 2 (2018), 1–20.

[57] Aaron Stafford, Bruce H Thomas, and Wayne Piekarski. 2008. Efficiency of techniques for mixed-space collaborative navigation. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE, 181–182.

[58] Richard Stoakley, Matthew J Conway, and Randy Pausch. 1995. Virtual reality on a WIM: interactive worlds in miniature. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 265–272.

[59] Matthew Tait and Mark Billinghurst. 2015. The effect of view independence in a collaborative AR system. *Computer Supported Cooperative Work (CSCW)* 24 (2015), 563–589.

[60] Masanari Tanase and Yasuyuki Yanagida. 2015. Video stabilization for HMD-based telexistence-Concept and prototype configuration. In *2015 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 106–111.

[61] Theophilus Teo, Louise Lawrence, Gun A Lee, Mark Billinghurst, and Matt Adcock. 2019. Mixed reality remote collaboration combining 360 video and 3d reconstruction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–14.

[62] John Thomason, Photchara Ratsamee, Kiyoshi Kiyokawa, Pakpoom Kriangkomol, Jason Orlosky, Tomohiro Mashita, Yuki Uranishi, and Haruo Takemura. 2017. Adaptive view management for drone teleoperation in complex 3D structures. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. 419–426.

[63] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. 2019. Loki: Facilitating remote instruction of physical tasks using bi-directional mixed-reality telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 161–174.

[64] Balasaravanan Thoravi Kumaravel, Cuong Nguyen, Stephen DiVerdi, and Bjoern Hartmann. 2020. TransceiVR: Bridging Asymmetrical Communication Between VR Users and External Collaborators. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 182–195.

[65] Eduardo Veas, Alessandro Mulloni, Ernst Kruijff, Holger Regenbrecht, and Dieter Schmalstieg. 2010. Techniques for view transition in multi-camera outdoor environments. In *Proceedings of Graphics Interface 2010*. Citeseer, 193–200.

[66] 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*. 143–146.

[67] Haijun Xia, Sebastian Herscher, Ken Perlin, and Daniel Wigdor. 2018. Spacetime: Enabling fluid individual and collaborative editing in virtual reality. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. 853–866.

[68] Jacob Young, Tobias Langlotz, Matthew Cook, Steven Mills, and Holger Regenbrecht. 2019. Immersive telepresence and remote collaboration using mobile and wearable devices. *IEEE transactions on visualization and computer graphics* 25 (2019), 1908–1918.

[69] Jacob Young, Tobias Langlotz, Steven Mills, and Holger Regenbrecht. 2020. Mobileportation: Nomadic Telepresence for Mobile Devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4 (2020), 1–16.