PeriViewMR: Stereoscopic Continuous Real-World Perspective Sharing for Effective Collaboration

Yiyin Gu(✉) and Voicu Popescu

Purdue University, West Lafayette 47907, USA
{gu251,popescu}@purdue.edu

Abstract. In real-world co-located collaboration each collaborator sees the workspace from their own viewpoint. This can hinder collaboration when collaborator A refers to a part of the workspace that collaborator B cannot see due to occlusions. The usual solution is for B to move around to assume a viewpoint similar to that of A. This is inefficient, as it takes time and effort for B to move around, and it is also ineffective, as it disrupts the natural collaboration flow. We introduce *PeriViewMR*, a mixed reality (MR) approach that alleviates the viewpoint difference between two collaborators to enable efficient and effective real-world co-located collaboration. Both collaborators wear an MR headset in passthrough mode. Collaborator A continues to see the workspace as usual, from their own perspective. However, collaborator B sees a composite view that seamlessly integrates their own near-field perspective with A's far-field perspective. This allows B to see where A is pointing, without moving. At the same time, B can turn towards A for a natural conversation. *PeriViewMR* provides a multi-perspective stereoscopic visualization of a real-world scene in real time. An initial within-subjects user study (N = 27) showed that *PeriViewMR* significantly improved task completion time and correctness. A follow-up study (N = 14) with authentic peer collaboration showed a large accuracy improvement with good usability and no cybersickness concerns.

Keywords: Real-world collaboration · Co-located collaboration · Occlusion · MR interfaces · Stereoscopic continuous real-world · Perspective sharing

1 Introduction

Consider two collaborators standing beside each other in front of a real-world scene. This scenario is common in practical applications such as examining large machinery, where technicians must inspect multiple components at once—one

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-032-42847-9_17.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
M. Lorenz et al. (Eds.): EuroXR 2026, LNCS 17151, pp. 353–376, 2027.
https://doi.org/10.1007/978-3-032-42847-9_17

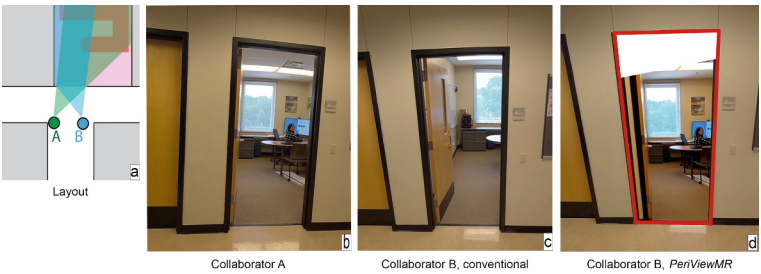


Fig. 1. Collaborators A and B are standing side by side, B right of A, facing the door of an office, as shown in (*a*). B cannot see the person in front of the desk in the room, but A can, as shown respectively in (*b*) and (*c*). *PeriViewMR* provides B with a visualization of the scene projected onto the door frame that integrates the perspective of A as shown in (*d*); the visualization shows B's perspective up to a vertical plane, where it switches from B to A with visual continuity at the bottom; using the *PeriViewMR* visualization, B can see the part of the scene hidden behind the walls. The images are actual frames from A and B's MR headsets, running in passthrough mode.

may monitor control panels while another observes internal mechanisms through an access port on the opposite side. Other use cases arise in safety-constrained environments, such as power plants, chemical facilities, or construction sites, where workers cannot freely reposition due to hazardous zones, moving equipment, or confined spaces [10]. Similarly, accessibility scenarios benefit from perspective sharing: a wheelchair user collaborating with a standing colleague may face occlusions from furniture or equipment at different heights, and physical repositioning may be impractical or impossible. Emergency response scenarios, where first responders must coordinate while maintaining safe positions behind barriers, represent another domain where eliminating the need to move for visual alignment provides substantial benefit. Collaborator A refers collaborator B to an element of interest of the scene, i.e., a target. Since A and B have different viewpoints, it can happen that B does not see the target due to occlusions. The typical solution is for B to take a few steps towards A to assume a viewpoint as close as possible to that of A, in the hope of seeing the target. The solution is inefficient, as it takes time and effort for B to assume a viewpoint that approximates that of A. Collaboration is also hindered by the fact that now A and B are not in the natural side by side collaboration position that allows them to alternate between looking at each other and looking at the scene. The solution is particularly ineffective when the scene is dynamic and the target is only visible for a short amount of time—B is highly unlikely to see such a transient target that disappears by the time B is able to assume A's viewpoint.

In this paper we introduce *PeriViewMR*, a mixed reality (MR) approach for efficient collaboration in real-world scenes. *PeriViewMR* allows two collaborators standing side by side to see each other as usual, by turning their head towards each other, but also to see the scene from the same perspective when looking forward, which avoids scene visualization mismatches due to occlusions. *PeriV-*

iewMR works by providing collaborator B with a composite visualization that shows the scene from their perspective up to a fixed virtual wall (the perspective switching plane, placed during setup at a chosen distance from the collaboration area), and then from the perspective of A beyond that. Collaborator A sees the scene as usual, from their own perspective. Figure 1 illustrates our method. Collaborator B does not have line of sight behind the office wall to see the person and the computer screen (*b*). A does have line of sight behind the wall through the open door (*c*), and if A refers to a target in this hidden region, B will not be able to see it. The *PeriViewMR* visualization (*d*) shows the scene up to the vertical plane from the perspective of B and then from the perspective of A beyond that. This way, B can see any workspace target that A sees, while still being able to turn to the left, for example to talk to A. The supplementary video shows the visualization in motion.

PeriViewMR is implemented by having both collaborators wear video passthrough MR headsets (e.g., Meta's Quest 3 [32]). The images in Fig. 1 are actual MR headset frames. The role of the headset of collaborator A is to acquire the scene from A's perspective and to send it to collaborator B. The headset of collaborator B acquires the scene from B's perspective and merges the two perspectives. Technically, the perspective integration is achieved through projective texture mapping: A's video frame is projected onto the fixed perspective switching plane, which serves as a geometric proxy for the far-field scene. Since A and B observe the same ground plane and share a common horizon, the perspectives align continuously along the ground, producing a seamless composite when the ground at the plane location is approximately planar—an assumption that holds in many scenes of practical interest. The composite visualization is also stereoscopic, as the left eye perspective of A is composited with the left eye perspective of B, and the right eye perspective of A is composited with the right eye perspective of B.

We evaluated *PeriViewMR* in two Institutional Review Board (IRB)-approved user studies. An initial study (N = 27) compared *PeriViewMR* to the conventional walk-around approach across four controlled tasks testing perspective sharing, gesture recognition, dynamic workspace monitoring, and visual continuity. Results show that *PeriViewMR* significantly outperforms the conventional method in task completion times and reduces physical effort and time pressure as quantified by NASA TLX [22]. A separate enhanced study (N = 14, 7 pairs) then evaluated *PeriViewMR* in authentic peer collaboration, where both collaborators were participants rather than researchers. This study showed a 30.5% point improvement in detection accuracy (49.5% vs 80.0%) with a large effect size (Cohen's d = 0.812), though the difference did not reach statistical significance with the available sample. Across both studies, *PeriViewMR* achieved "good" usability with average System Usability Scale (SUS) scores [9] of 75.3 and 72.7 [5], and did not raise cybersickness concerns, with average total SSQ scores [24] of 8.3 and 18.4 respectively, modest relative to the range reported across immersive visualization studies [7].

2 Prior Work

PeriViewMR is an MR approach for alleviating occlusions to make collaboration more efficient. As such, we review prior work on MR-enabled collaboration (Sect. 2.1), view and perspective sharing (Sect. 2.2), and MR occlusion management (Sect. 2.3).

2.1 MR Collaboration

MR has the ability of integrating virtual 3D representations into the physical surroundings of its user. As such, one of the main applications of MR is in remote collaboration [40], where it integrates remote collaborators and their environments, hiding the distance between them. One challenge is the real-time, 3D acquisition of the collaborators and their environments. In order to allow the local participant to change the viewpoint freely, the remote collaborators and their environments have to be acquired not just in color, but also in depth [40]. A free viewpoint improves collaboration performance [45] and can increase the sense of co-presence between the remote collaborating parties [29]. However, depth acquisition is challenging, and so is reconstructing the scene from the acquired data [1]. As such, acquisition is often simplified to video acquisition, which is inexpensive, and which allows for good visualization of the 3D entities captured, as long as the viewing direction is more or less aligned with the acquisition direction [6]. Middle-ground solutions, between 3D and 2D acquisition, rely on homographies [8], correspondences [38], or warping [27,28].

MR is also beneficial in the case in which the collaborators share the same space. In such co-located collaboration [42], aligning the virtual spaces of individual collaborators consistently is an important challenge [31]. Co-located systems can exploit the fact that collaborators already share a physical environment, but they still have to address viewpoint-dependent occlusions and differences in what each collaborator can see.

2.2 View and Perspective Sharing

View sharing is central to collaboration because collaborators need to establish common ground about the objects, locations, and actions being discussed [11]. MR and AR systems often address this by inserting a collaborator's view into the user's view [51], but such heads-up or picture-in-picture approaches create a secondary frame that users must mentally register to the real workspace.

Prior work has explored perspective alignment, viewpoint sharing, and related co-located AR/MR collaboration contexts, including Decoupled Hands [21], viewpoint sharing in VR/AR collaboration [46], Vishnu for MR guidance [25], systems for merging environments and preserving workspace awareness [13,44], SPARC and ShiSha for shared perspectives in VR [23,43], co-located augmented presentations [20], and multi-view relation visualization in MR [30]. *PeriViewMR* differs by integrating a collaborator's live stereoscopic passthrough stream into the user's own passthrough view of the same physical workspace;

rather than presenting an inset or relying on known virtual geometry, it uses projective texture mapping [41] onto a world-locked planar proxy for a continuous near-field/far-field transition.

2.3 MR Occlusion Management

Improving the visualization of a real-world scene can sometimes require the *removal*, rather than the addition, of elements, through diminished reality approaches [35,52]. Removing the occluder reveals the background of interest, which is modeled with a collection of images [36] or with 3D geometry [3,37]. We do not take the approach of removing the occluder, but rather of integrating a perspective where the target is known to be visible.

A prior study compared cutaway, X-ray, picture-in-picture, and multiperspective visualization for integrating a secondary real-world perspective [26]. Cutaway and X-ray require detailed scene geometry, while picture-in-picture is familiar but discontinuous and low resolution. We therefore use a two-perspective multiperspective visualization that integrates the collaborator's secondary view with the user's primary view.

The *PeriViewMR* visualization is a simple multiperspective visualization. Multiperspective images enhance the visualization provided to a user with 3D scene samples to which the user does not have line of sight. The challenge is to integrate the samples of multiple perspectives with visual continuity. Constructing multiperspective visualizations of real-world 3D scenes is complicated by not knowing the scene geometry. One prior method leverages a high-level geometric proxy of an indoor scene to provide the user of an MR application with a multiperspective visualization that lets the user preview hidden parts of the scene, thereby increasing scene exploration efficiency [50]. Other approaches avoid full depth acquisition by using video, homographies, or view-dependent warping when the viewing and acquisition configurations are constrained [6,8,28]. We rely on the assumption that the ground is planar in the vicinity of the collaborators, which allows merging their perspectives with visual continuity.

One prior system is particularly relevant to *PeriViewMR*. **View Splicing** [48] addresses VR collaboration by splicing user and collaborator views using a multiperspective "switch camera" that transitions smoothly between viewpoints, with near geometry shown from the user's viewpoint and far geometry from the collaborator's viewpoint. Like *PeriViewMR*, View Splicing aims to let one user see the workspace from their collaborator's perspective without moving. However, View Splicing operates in purely virtual environments where scene geometry is fully known, enabling precise multiperspective ray construction. In contrast, *PeriViewMR* operates on *real-world* scenes captured through video passthrough, where 3D geometry is unavailable. We address this by using projective texture mapping onto a planar proxy, achieving stereoscopic visual continuity under the assumption of planar ground at the perspective switching distance.

3 Method

The overall goal of our method is to allow two collaborators standing side by side in a real-world scene to be able to collaborate efficiently, without the challenges posed by their different viewpoints. Specifically, when collaborator A refers to an element of the workspace, i.e., a target, for example verbally, by pointing to it with their hand, or by pointing to it with a physical laser pointer, collaborator B should be able to see the target, from B's current position, even if B doesn't have direct line of sight to the target. In other words, we want B to see around any occluder that might obstruct their view to the target.

3.1 Design Concerns

The following design concerns were derived from the collaboration breakdowns targeted by our system, from prior work on view sharing and occlusion management, and from practical constraints of portable MR deployment. Each names the study or observation it derives from:

- *DC1: in-place occlusion avoidance.* If collaborator A sees the target, then collaborator B should see it as well, from their current position, without navigating their viewpoint. This concern derives from the motivating breakdown: walking to see costs time, breaks the side-by-side arrangement, fails for transient targets (Sect. 1), and is unsafe in constrained environments [10].
- *DC2: natural interface for view selection.* Collaborators should alternate naturally between seeing each other and seeing the shared workspace through head and body orientation, not through keyboard, mouse, or menus. This concern derives from conversational grounding: to ground a reference, collaborators keep alternating attention between partner and workspace [11]; view selection must not hinder that motion.
- *DC3: natural communication between collaborators.* Collaborators should be able to talk to each other naturally, by turning their head to see each other, including gestures and body language. This concern derives from what co-location already gives: face-to-face speech, gesture, and body language precede any system, the setting surveyed in [42]; view sharing must leave them untouched.
- *DC4: visualization continuity.* Collaborator B should be provided a continuous visualization at the perspective boundary: physical points on and near the ground-plane intersection with the switching plane should project to nearby image locations. This concern derives from what an inset costs: a heads-up window [51] gives the user two pictures of one scene, and every glance must re-align them—the discontinuity the comparison of disocclusion interfaces found [26]; a continuous boundary leaves one picture and nothing to re-align.
- *DC5: stereoscopic.* The collaborators should see the scene in 3D with appropriate depth cues, not in 2D. This concern derives from a task observation: targets and occluders are separated in depth, and a monoscopic feed flattens exactly that ordering.

– *DC6: portability and low cost.* The collaboration approach should rely on low-cost, portable equipment to allow versatile and widespread deployment. This concern derives from a deployment observation: tripod cameras and projection enclosures anchor collaboration to instrumented spaces; consumer passthrough headsets do not.

3.2 *PeriViewMR* Design

Acquisition. To address the primary design concern *DC1* of in-place occlusion avoidance, collaborator B's view of the real-world has to be augmented with samples from a different viewpoint, i.e., including some samples to which A has line of sight but B does not. The simple method for acquiring a real-world 3D scene is with a video camera. If the camera is placed close to A's viewpoint, it will acquire many of the parts of the 3D scene A sees. The closer the camera to A's viewpoint, the smaller the chance that the camera misses a target seen by A. A tripod mounted camera placed close to A captures the samples needed, but portability is less than ideal. A head-mounted camera brings good proximity to A's viewpoint as well as portability, at the cost of physical encumbrance. MR headsets in passthrough mode strive to solve this exact problem of capturing their user's view of the scene, and they also track the user's view; furthermore MR headsets are portable and inexpensive, and therefore they are a good solution to our acquisition problem. Whereas the MR headset is needed for its tracked camera matching collaborator A's viewpoint, the display function of the MR headset is not strictly needed, unless the collaborators want the benefit of an AR application that augments the workspace with additional visual information.

Visual Interface Design. A natural interface for view selection through head motions (*DC2*) requires head tracking, and stereoscopic visualization (*DC5*) requires catering images with appropriate disparity to each of the user's eyes. The portability and low cost concerns (*DC6*) rule out room-scale projection enclosures in which images are projected on physical walls and users collaborate wearing stereo glasses. Again, like for collaborator A, having collaborator B wear an MR headset is a good solution that addresses these design concerns, at the cost of headset encumbrance.

Visualization Design. Collaborator B cannot rely exclusively on collaborator A's view of the scene. Imagine a solution where B's headset shows the frame of A's headset. First, providing B with A's first-person view instead of their own is unacceptable due to cybersickness concerns, as the scene visualization of B will not change based on their own head motions, but rather based on A's head motions. Even if A's view is stabilized to respond to B's head motions, B still has to see some of the scene from their own viewpoint, at least in their proximity, such that they can see A, to meet the natural communication requirement (*DC3*). It follows that B has to see the scene from both A's and B's perspective. We design the visualization such that collaborator B sees the scene from their own

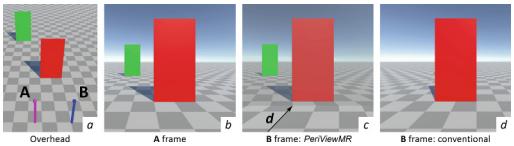


Fig. 2. *PeriViewMR* illustration. Collaborators A and B stand side by side (*a*). A sees both the red and green blocks (*b*). With a conventional frame *d*, B cannot see the green block which is hidden behind the red block. The *PeriViewMR* frame *c* is obtained by projecting the frame of A onto the perspective switching plane at distance ***d***. With *PeriViewMR*, B can see the green block, and, by turning their head to the left, their collaborator A. (Color figure online)

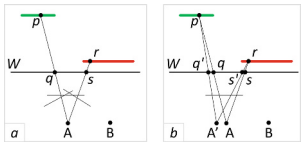


Fig. 3. *PeriViewMR* visualization stability under collaborator A view rotation (*a*) and translation (*b*). *a*: as the view direction of A rotates, *PeriViewMR* always projects on the perspective switching wall W the 3D point p at q and r at s, for a perfectly stable visualization. *b*: when A translates to A', the projection of p moves from q to q' and of r from s to s'.

perspective up to a distance d, which is sufficient to encompass collaborator A, and then from the perspective of A beyond d, which includes the workspace.

The A and B perspectives have to be integrated seamlessly in the visualization provided to B, to meet the continuity requirement *DC4*. Given two cameras with known intrinsic and extrinsic parameters, their images can be integrated with perfect continuity if the geometry of the 3D scene is known, using conventional projection texture mapping. However, to meet *DC6*, we want to bypass the real-time 3D acquisition of the real-world scene that could have intricate geometry, surface reflectance properties and animation. All we need is a cut line in collaborator B's frame where to switch perspectives with continuity. Consequently, we don't need to know the geometry of the entire scene, but rather only at the cut line. Finally, seamless integration of the two perspectives in the visualization provided to B has to be implemented for each eye, for the visualization to be stereoscopic (*DC5*). In other words, B should be provided appropriate depth cues for the part of the scene they see from their own perspective, and also for the part of the scene they see from A's perspective.

3.3 *PeriViewMR*: Stereoscopic Continuous Real-World Perspective-Sharing

Our *PeriViewMR* approach is illustrated in Fig. 2. The conventional frame of collaborator A is projected onto a finite vertical *switching plane rectangle* lying on a world-locked *perspective switching plane*. The plane is a virtual wall placed during setup, anchored to world coordinates, and fixed during the task. Its distance d is configurable for the environment, but collaborators must remain on their side of the plane. This fixed placement gives B a stable perspective boundary. Because the seam line lies on the ground where the vertical projection plane

meets the floor, A's projected view and B's direct view depict the same physical points at the same image locations, giving C^0 continuity at the boundary. *PeriViewMR* only requires planar ground at distance d; it does not require full scene geometry. Smaller values of d bring A's perspective closer but leave less unmodified near-field view for B, while larger values preserve more of B's natural view but may push the transition beyond the workspace. In the current implementation, d is selected during setup and remains fixed during a task; adapting d dynamically is left for future work.

Visualization Stability. Since collaborator A is standing, their view changes are primarily view rotations, as they pan and tilt their head. In the idealized case of a camera rotating about its optical center, yaw rotations (horizontal panning) preserve the projection of a fixed 3D point onto the perspective switching plane (Fig. 3, *a*). A real headset is not perfectly ideal: yaw, pitch, and roll also translate the left and right eye cameras slightly because the cameras are offset from the head's rotation center. These small translations change disparity and can produce small shifts in the projected stream. Pitch and roll can also make the shared stream feel tilted when A tilts their head, although the world-locked switching plane prevents the boundary itself from moving in B's space. A must therefore look toward the workspace (within approximately $\pm 45°$ of the forward direction shared with B) for meaningful perspective sharing; extreme rotations capture scene regions irrelevant to B. The *PeriViewMR* visualization is affected more substantially by translations of viewpoint A (Fig. 3, *b*). Figure 4 illustrates the small *PeriViewMR* frame changes when collaborator A translates their head 30 cm to the left, a generous upper bound on how much a human's eyes translate during ordinary panning. We conclude that the visualization is stable for collaborators who remain near their station and orient their head toward the shared workspace, while larger movements require resetting or adapting the switching plane.

Depth Cues. The same procedure is repeated for A's left and right eye frames, so the top region showing A's perspective is stereoscopic rather than a flat plane (Fig. 5). These depth cues correspond to A's binocular camera baseline rather than B's exact eye locations beyond the switching plane; they help B interpret A's visual evidence and pointing references, but they do not reconstruct full scene geometry from B's viewpoint. *PeriViewMR* projects the left and right eye images onto a large switching plane rectangle and only renders pixels covered by both streams, avoiding a stereo center with monoscopic edges. A monocular comparison remains future work.

4 Implementation Overview

We implemented our system, *PeriViewMR*, in the Unity 3D game engine [47]. The process begins with a setup phase where the experimenter places a virtual perspective switching plane into the workspace and defines a Meta spatial

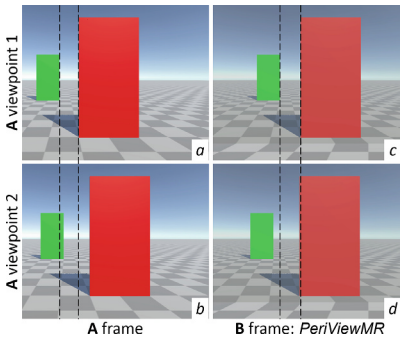


Fig. 4. *PeriViewMR* visualization stability under collaborator A viewpoint translation. Collaborator A translates their viewpoint 30 cm to the left (from top to bottom row). Vertical dashed lines were added to facilitate top/down comparison. In the A frame (*a* and *b*), the gap between the blocks increases due to motion parallax (*a* to *b*). In the *PeriViewMR* frame (*c* and *d*), the gap also increases slightly, with the red block moving little as it is close to the perspective switching vertical plane. The bottom part of the *PeriViewMR* frame is not affected by the view A change. (Color figure online)

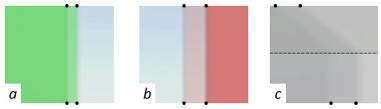


Fig. 5. Illustration of disparity between *PeriViewMR* left- and right-eye frames, corresponding to the scenario in Fig. 2, with a 7 cm interpupillary distance, and parallel eye optical axes (no convergence). The disparity is visualized with magnified fragments of the left and right frames blended together. The dots point to the ghosting region that indicates disparity. The farther green block (*a*) has less disparity than the closer red block (*b*), and disparity is continuous across the perspective switching line (illustrated with the dotted line in *c*). (Color figure online)

anchor [33]. Once placed, the plane remains stationary regardless of where the collaborators move within the workspace. The anchor's universally unique identifier (UUID) is uploaded to a cloud service for later retrieval. To initiate a session, both headsets download the UUID, localize the shared anchor, and use it as a common physical reference to align their coordinate systems. The Photon networking service [16] then transmits local coordinates between the headsets to continuously calculate their relative transform. In practice, the initialization handshake is: (1) place the perspective switching plane, (2) create and upload the spatial anchor UUID, (3) have both devices localize the same anchor, (4) exchange headset transforms through Photon, and (5) visually check that the projected switching boundary is stable in the expected workspace.

Once co-location is established, the view-sharing pipeline is activated. The Meta Passthrough API [34] is used to acquire the raw left and right eye video frames, along with camera intrinsics and extrinsics, from both headsets. These camera streams are then transmitted bidirectionally between collaborators via the Agora video streaming service [2]. The core of the technique is a stereoscopic projective texture mapping algorithm that executes locally on each device. This process is fully symmetric: collaborator B's headset uses the stream and trans-

form data from A to render A's perspective beyond the perspective switching plane, while collaborator A's headset simultaneously performs the same computation using the stream from B. This allows both users to seamlessly transition to the other's point of view.

The system's video-streaming latency was measured using the Agora internal statistics API, yielding an average delay of 486.21 ms over WiFi with download speeds of 452 Mbps and upload speeds of 503 Mbps. Prior research has shown that delayed visual feedback affects collaborative physical-object tasks but can remain usable for referential communication within sub-second delays [18]; recent Social XR work similarly reports a 900 ms delay threshold for user acceptance in volumetric communication [14]. Our measured latency was sufficient for the verbal and finger-count referencing used in our study. However, fast hand gestures or rapid manual actions could feel delayed or detached from the physical person, and lower-latency streaming is an important engineering target for future versions.

5 Evaluation

We conducted two IRB-approved studies. Study 1 evaluates the core visualization in controlled researcher-participant tasks. Study 2 evaluates the approach in an authentic peer-collaboration task where both collaborators are participants.

5.1 Study 1: Controlled Researcher-Participant Tasks

Methods

Participants. We recruited 27 participants from our department's student population for the initial study (Tasks 1–4): 20 male, 7 female, average age 20.4 years. Most had prior VR experience (25/27 had used VR at least once; 19/27 had used MR headsets).

Study Design. We compared *PeriViewMR* (experimental condition, EC) to walking to establish line of sight (control condition, CC) using a within-subjects design. Participants served as collaborator B, following instructions from collaborator A (a researcher). Condition order was counterbalanced, with half of participants experiencing EC first and half experiencing CC first. Task order was balanced using a Latin-square design. The within-subjects design allows finding differences with fewer participants while controlling for individual differences.

Procedure. Each participant was involved for up to 60 min and compensated with a gift card equivalent to 30 USD. A participant: (1) read, discussed with the researcher, and signed a consent form; (2) filled out demographics via online Qualtrics form accessed with QR code; (3) performed tasks in the first condition; (4) answered condition-specific questionnaires; (5) performed tasks in the second condition; and (6) answered questionnaires for the second condition. Collaborator and participant stood side by side with 1.5 m between them. In the study, the

perspective switching plane was fixed for all trials to keep the evaluation geometry consistent. The plane was parallel to the line connecting the two participants' starting positions and offset 2 m in front of them. This placed collaborator A and B in the unmodified near-field while positioning the task workspace beyond the perspective boundary. In EC, both wore Quest 3 headsets [32] in passthrough mode, with answers entered via floating panel using a handheld controller. In CC, neither wore headsets, simulating real-world collaboration; answers were entered on a handheld tablet.

T1: Verbal Reference to a 3D Workspace. The first task tests *PeriViewMR*'s ability to let participants see the workspace from their collaborator's viewpoint and identify workspace elements their collaborator references. A board with five (digit, letter) pairs was positioned so participants could only see the left half from their initial position (Fig. 6). Without moving, participants could see the digit but not its letter. The collaborator, located to the participant's left, had line of sight to entire pairs. The collaborator calls out a number and participants must identify the corresponding letter. In CC, participants took steps to see the letter pairing the called digit, then entered it on the tablet. In EC, participants used *PeriViewMR* to see the letter pairing the digit and entered it via the headset panel. After entering a letter, the collaborator called another number until all five pairs were used.

T2: Pointing Reference to a 3D Workspace. This task investigated *PeriViewMR*'s ability to let participants see the workspace from their collaborator's perspective and see their collaborator naturally by turning their head. The task resembles *T1*, but instead of saying numbers aloud, the collaborator shows numbers 1–5 using left hand fingers. This requires participants to see their collaborator to know which number's pair to find.

T3: Monitoring a Dynamic 3D Workspace. This task investigates *PeriViewMR*'s ability to let participants monitor two workspace parts T_A and T_B simultaneously. Participants have line of sight to T_B but only collaborators see T_A. Two tablets were positioned at T_A and T_B (Fig. 6). Participants view T_B directly but not T_A, which only collaborators see. T_A sits beyond the fixed perspective switching plane. Each tablet plays a 40 s video showing random two-digit numbers (11–19). Each number displays for two seconds. Usually tablets show different numbers; occasionally they show the same number. Participants must indicate numbers shown simultaneously by both tablets. Participants can miss the first k matching pairs, then report the $k+1^{th}$ pair. In CC, participants were not required to return to their starting position after moving; they could choose any physical position during the trial as long as T_B remained visible.

T4: Connecting Perspectives Continuously. This task investigates *PeriViewMR*'s ability to merge the participant's perspective on the real-world scene up to the fixed perspective switching plane with their collaborator's perspective beyond it, with good visual continuity. For this task, five lines were installed on

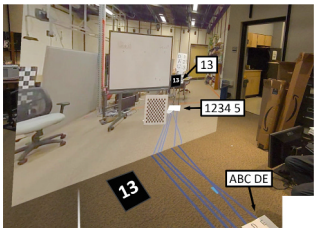


Fig. 6. Illustration of the setup for Tasks 1, 2, 3, and 4 in the *PeriViewMR* condition. The tablet positions for Task 3 are labeled as T_A and T_B.

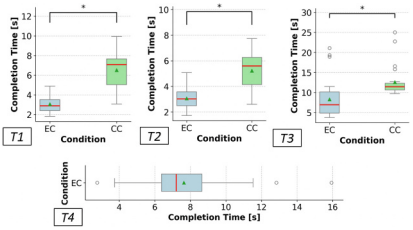


Fig. 7. Task T1, T2, T3 completion times comparison and T4 completion times.

the floor using colored adhesive tape, starting from the participant and extending beyond the plane (Fig. 6). Each line has a letter label at the near end and a number label at the far end. Participants can see the near end but not the far end, including far-end labels. However, collaborators can see the far end, including labels. The collaborator calls out a number from a line's far end, and participants must trace the line to identify the letter label at the near end. Participants perform this task only in EC to stay within one hour total time and avoid fatigue.

Data Collection. Objective metrics included task completion time (from stimulus visibility to verbal response) and correctness. Task completion time used identical start triggers across conditions. Subjective metrics included NASA TLX [22] for both conditions, and SUS [9] and SSQ [24] for EC only.

Data Analysis. We analyze the data using descriptive statistics (average, median, standard deviation) shown graphically through box plots and numerically in tables. Inferential statistics compare conditions using a significance level $\alpha = 0.05$. For normally distributed continuous data, we use paired samples t-tests appropriate for our within-subjects design. When normality assumptions are not sustained, we use the non-parametric Wilcoxon signed-rank test [49]. Effect sizes are calculated using Cohen's d for paired-samples t-tests [15], and $r = |z|/\sqrt{N}$ for Wilcoxon tests [39]. Following Cohen's guidelines [12], effect sizes of $|r| \approx 0.1$, 0.3, and 0.5 indicate small, medium, and large effects respectively. The TLX, SUS, and SSQ data are analyzed by computing overall and sub-scale scores using procedures specific to each standard instrument.

Research Hypotheses. We formulated hypotheses covering four groups of *PeriViewMR*'s potential benefits:

– **RH1–RH2, perspective access and collaborator awareness**: *PeriViewMR* allows B to see the workspace from A's perspective and to see A naturally by turning their head.

- **RH3–RH6, task performance**: *PeriViewMR* enables simultaneous monitoring, visual continuity, faster completion, and more accurate task completion for dynamic workspaces.
- **RH7–RH9, workload and user experience**: *PeriViewMR* reduces physical and temporal task load, has good usability as measured by SUS, and does not cause cybersickness as measured by SSQ.

Results. We present the results for each metric, task, and condition. Broader interpretation is consolidated in Sect. 6.

Objective Metrics: Time and Correctness. *Task completion times* for the four tasks and the two conditions are presented numerically in Table 1, and graphically in Fig. 7. For tasks *T1*, *T2*, and *T4*, where participants completed multiple repeated measures, we first calculated the mean task completion time for each participant. EC has shorter task completion times than CC for *T1–T3*, and the differences are significant; *T4* was EC-only. For *T1*, *T2* and *T3* the data is normally distributed; the Cohen's d effect sizes of 2.22, 1.84 and 0.99 correspond to the qualitative labels of "very large" ($d > 1.2$), "very large" and "large" ($d > 0.8$), indicating practically meaningful differences between conditions.

Interpretation. The consistent time advantage across the directly compared tasks indicates that eliminating the need to walk for line of sight provides a substantial efficiency gain. These results support **RH5**.

Task completion correctness percentages showed that EC condition achieved 100%, 98.5%, 69.2%, 96.2%, and 71.1% for Tasks 1, 2, 3 (first pair), 3 (second pair), and 4 respectively, while CC condition achieved 99.3%, 99.2%, 0%, 92.3%, and was not applicable for the same tasks. For *T1* and *T2*, CC also achieved high accuracy, as the static workspace allows time to walk and find targets.

Interpretation. The high EC accuracy in *T1* can only be achieved by seeing the workspace from collaborator A's perspective, supporting **RH1**. Similarly, high EC accuracy in *T2* requires switching naturally between seeing the collaborator and seeing their perspective on the workspace, supporting **RH2**. The high CC accuracy for these static tasks is expected: participants have time to walk to the needed viewpoint.

Task *T3* involved a dynamic workspace with transient targets—pairs of identical numbers displayed briefly before disappearing. In CC, no participant (0%) was able to find the first pair of matching numbers, as the pair disappeared by the time they assumed a viewpoint showing both tablets. In EC, 69.2% of participants found the first pair. Both conditions achieved higher accuracy on the second pair (96.2% EC, 92.3% CC) as participants learned the task timing.

Interpretation. The 0% vs 69.2% first-pair detection gap demonstrates a qualitative capability difference: CC cannot support simultaneous monitoring of

Table 1. Task performance time in seconds.

Task	Condition	Descriptive Statistics			Inferential Statistics		
		Avg	Median	Std.Dev	Test	p	EffectSize
T1	CC	6.51	7.08	1.89	Pairedt-test	$< 0.001^{*}$	2.22
	EC	3.04	2.88	0.86			
T2	CC	5.22	5.59	1.39	Pairedt-test	$< 0.001^{*}$	1.84
	EC	3.06	3.02	0.82			
T3	CC	12.59	11.42	3.68	Pairedt-test	$< 0.001^{*}$	0.99
	EC	8.30	6.94	4.86			
T4	EC	7.64	7.21	2.94	-	-	-

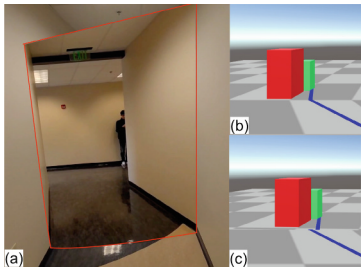


Fig. 8. (a) Perfect continuity on the ground across the 'wall'. (b) VR simulator normal view. (c) VR simulator analysis: 2° error.

spatially separated, transient targets, while EC can. These results support **RH3** (simultaneous monitoring) and **RH6** (accuracy advantages for dynamic workspaces).

Task *T4* achieved 71.1% accuracy (EC only). Participants took on average 7.64 s to trace tapes from far to near end (Table 1), with some taking up to 16 s. Theoretically, two perspectives should have perfect continuity along the ground line where the switching plane intersects the floor (Fig. 8a), but even a 2° rotational error causes noticeable misalignment for thin, closely spaced tracing lines (Fig. 8c). We observed that this task was sensitive to one-time initialization error from Meta's spatial anchor system: some sessions achieved near-perfect alignment while others exhibited small offsets. The error did not appear to accumulate during use; it remained constant once the session began.

Interpretation. These results provide partial support for **RH4**. The 71.1% accuracy reflects a demanding practical case for the method: the tapes are thin, close together, and intersecting, and the task is affected by passthrough resolution, spatial anchor alignment, and the projection assumptions of the switching plane [33]. We therefore interpret T4 as evidence that the continuity mechanism is useful but sensitive to initialization precision and fine visual detail, rather than as evidence of perfect continuity in all setups (Fig. 8).

Subjective Metrics: TLX, SUS, and SSQ. *NASA TLX* scores (Fig. 9; Table 2) show *PeriViewMR* has low task load levels on all sub-scales, with an overall score of 1.8/10.0 for EC. EC showed lower average task load than CC for physical demand, temporal demand, effort, frustration, and overall workload. Significant EC advantages were found for physical demand Ph (EC: 1.3, CC: 3.1, $p < 0.001$, $r = 0.86$) and time demand T (EC: 2.0, CC: 2.7, $p = 0.012$, $r = 0.78$). Mental demand was slightly higher in EC (2.4 vs 1.7), though not statistically significant ($p = 0.169$). The performance sub-scale P was inverted in Fig. 9 and Table 2 so that lower plotted values indicate better perceived performance; this sub-scale favored CC (EC: 1.9, CC: 1.0, $p = 0.025$, $r = 0.77$), so we do not

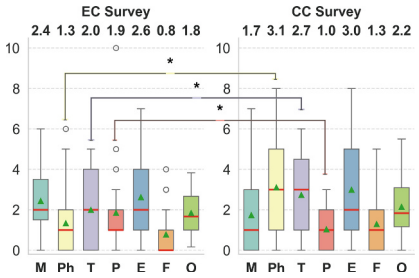


Fig. 9. Box-plots of the NASA TLX sub-scales and overall score: M–mental demand, Ph–physical demand, T–time demand, P–performance, E–effort, F–frustration, O–overall. Performance sub-scale (P) was inverted using the formula $10-P$ to ensure that all sub-scales follow the interpretation that lower scores indicate better outcomes. Asterisks (*) indicate statistically significant differences ($p < 0.05$).

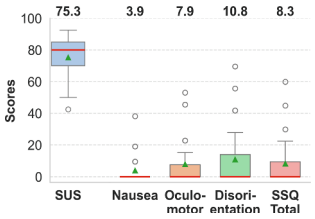


Fig. 10. Task 1–4 box-plots and average values of participants' answers to the SUS and SSQ questionnaires.

interpret it as support for reduced workload. The frustration sub-scale showed a large effect size ($r = 0.77$) despite non-significant $p = 0.087$.

Interpretation. The large effects on physical demand and time demand support the physical and temporal components of **RH7**. The slightly higher mental demand in EC likely reflects the learning curve of a new visualization modality. The performance sub-scale result suggests that participants felt especially confident in CC for the controlled static tasks, even though EC was faster and reduced physical movement.

SUS scores averaged 75.3, placing *PeriViewMR* between "Excellent" and "Good" according to Bangor et al.'s adjective ratings [5], supporting **RH8**. Based on experimenter observation during the sessions and the Task 4 error analysis above, the main usability challenges include spatial anchor misalignment and the requirement that collaborator A must look at the target for B to see it in the shared perspective. If A pans significantly away from the workspace, their frame will not project useful content onto the perspective switching plane.

SSQ scores averaged 8.3 (Fig. 10), considered minimal for immersive visualization applications [7]. At the study's start, participants were instructed not to fight through motion sickness symptoms and to remove the headset immediately if needed, with assurance of compensation regardless of completion. All participants completed the study without incident, supporting **RH9**.

5.2 Study 2: Authentic Peer Collaboration

While our initial study demonstrated *PeriViewMR*'s effectiveness in controlled researcher-participant tasks, we conducted an enhanced study (N = 14) to eval-

Table 2. Main study (Tasks 1–4) NASA TLX sub-scales and overall score. P was inverted ($10 - P$) so lower scores indicate better outcomes. M = Mental, Ph = Physical, T = Temporal, P = Performance, E = Effort, F = Frustration, O = Overall.

Subscale	Condition	Descriptive Statistics			Inferential Statistics		
		Avg	Median	Std. Dev	Test	p	Effect Size
M	CC	1.74	1.00	1.87	Paired t-test	0.169	-0.39
	EC	2.44	2.00	1.74			
Ph	CC	3.11	3.00	2.59	Wilcoxon's	$< 0.001^{*}$	0.86
	EC	1.33	1.00	1.57			
T	CC	2.74	3.00	2.16	Wilcoxon's	0.012^{*}	0.78
	EC	2.00	2.00	1.90			
P	CC	1.04	1.00	1.16	Wilcoxon's	0.025^{*}	0.77
	EC	1.85	1.00	1.99			
E	CC	3.00	2.00	2.39	Wilcoxon's	0.638	0.49
	EC	2.63	2.00	2.15			
F	CC	1.30	1.00	1.54	Wilcoxon's	0.087	0.77
	EC	0.78	0.00	1.12			
O	CC	2.15	1.83	1.49	Wilcoxon's	0.409	0.21
	EC	1.84	1.67	1.03			

Table 3. Task 5 (Enhanced Study) NASA-TLX subscale comparison. EC = headset, CC = tablet. M = Mental, Ph = Physical, T = Temporal, P = Performance (inverted), E = Effort, F = Frustration.

Subscale	Condition	Descriptive Statistics			Inferential Statistics		
		Avg	Median	Std. Dev.	Test	p	Effect Size
M	CC	4.9	5.0	3.0	Paired t-test	0.023^{*}	−0.69
	EC	2.6	2.0	2.5			
Ph	CC	1.4	0.5	2.2	Paired t-test	0.149	0.41
	EC	3.0	1.5	3.1			
T	CC	7.2	7.5	1.9	Paired t-test	0.002^{*}	−1.03
	EC	3.7	4.0	3.5			
P	CC	5.5	5.5	2.7	Paired t-test	0.001^{*}	−1.15
	EC	2.6	2.5	2.3			
E	CC	5.9	7.0	2.7	Paired t-test	0.008^{*}	−0.84
	EC	2.6	1.5	3.0			
F	CC	2.2	1.5	2.5	Wilcoxon	0.168	0.62
	EC	1.3	1.0	1.9			

uate it in authentic peer collaboration scenarios. This follow-up addresses an important question: do users benefit from perspective-sharing in real-world collaboration with actual peers rather than researchers?

Methods. We recruited 7 additional pairs (N = 14): 9 male, 5 female, average age 22.9 years (SD = 4.1), with similar VR/MR experience levels. Each pair completed one EC round and one CC round, with condition order counterbalanced as in Study 1. Both collaborators were participants; roles were assigned by rock-paper-scissors and did not switch during the task. Participant A was the primary responder, participant B was the support collaborator, and participants could talk freely in both conditions.

The task was active collaboration through synchronized color-shape detection. Two target displays were positioned with spatial occlusion so that a single person could not easily observe both simultaneously. Each display presented changing shapes and colors at 2-second intervals. In EC, both participants wore Quest 3 headsets because the view-sharing cameras and displays were headset-based; B's headset stream was shared to A through *PeriViewMR* and projected onto a virtual room-scale rectangular wall/screen. In CC, participants used verbal communication only, with B describing the colors and shapes on one display so that A could identify matches. The tablet in CC was used only for entering answers and did not provide view sharing.

A detection was recorded when the primary responder pressed the response button, either in the headset interface in EC or on the tablet in CC. A response was scored as correct only if both displays were showing the same shape and

color at the time of the button press. Button presses made when the displays did not match were counted as false positives. Subjective measures included SUS, SSQ, NASA TLX, and a 13-item custom questionnaire assessing Collaboration Quality (items 1–6) and Technology Acceptance (items 7–13) on 7-point Likert scales.

Results. Detection accuracy was higher in EC (80.0%, SD = 11.6%) versus CC (49.5%, SD = 29.3%), with a 30.5% point mean difference, 95% CI [−4.2, 65.2] pp (Fig. 11). A paired samples t-test showed this trend approached but did not reach significance, t(6) = 2.148, p = 0.075, Cohen's d = 0.812 (large effect). With only 7 paired groups, the study was underpowered to detect this effect; a post-hoc power analysis suggests approximately 12 pairs would be needed. Response latency for correct detections showed no significant difference between conditions (EC: M = 2.02 s, SD = 0.25 s vs CC: M = 1.92 s, SD = 0.89 s, paired t(6) = 0.317, p = 0.762), indicating comparable response speeds across conditions. SUS scores were significantly higher in EC (M = 72.7, SD = 20.5) versus CC (M = 55.5, SD = 16.5), paired t-test p = 0.0285, Cohen's d = 0.658 (medium effect). NASA-TLX showed significant EC advantages in mental demand, temporal demand, performance, and effort (Table 3; Fig. 12). Physical demand was higher in EC than CC (3.0 vs. 1.4), but this difference was not statistically significant (p = 0.149), likely reflecting the weight and encumbrance of wearing and operating current headsets during active collaboration.

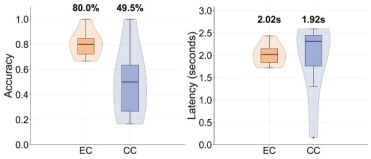


Fig. 11. Task 5 objective performance: detection accuracy and response latency for EC (headset) vs CC (tablet). Error bars show standard deviation.

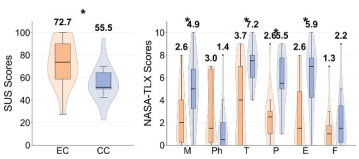


Fig. 12. Task 5 SUS usability and NASA-TLX workload scores. The Performance subscale (P) was inverted so lower scores indicate better outcomes.

Subjective Measurements. *Custom Questionnaire Results.* Participants rated access to the partner's viewpoint highest (Q2, M = 6.1/7, SD = 1.2), followed by coordination effectiveness (Q1, M = 5.6), ease of learning (Q9, M = 5.6), preference over traditional methods (Q7, M = 5.5), performance improvement and work-usage intention (Q8/Q10, M = 5.4), confidence in accuracy (Q11, M = 5.2), and real-work usage intention (Q6, M = 5.1). Focus maintenance, integration smoothness, and natural collaboration feel were moderate (Q3/Q5/Q12, M = 5.0; Q4, M = 4.9), while interference with real-world interaction was low (Q13, M = 3.9, reverse-scored), indicating generally favorable collaboration quality and technology acceptance.

Motion Sickness. SSQ scores for the headset condition showed generally low motion sickness symptoms: Nausea M = 7.5, Oculomotor M = 16.8, Disorientation M = 26.8, and Total M = 18.4 (SD = 25.6, Median = 5.6).

Interpretation. The enhanced study provides converging evidence that *PeriViewMR* offers meaningful benefits in authentic collaborative scenarios. The 30.5% point accuracy improvement (80.0% vs 49.5%) with a large effect size (Cohen's d = 0.812) supports **RH6**; the non-significant $p = 0.075$ reflects limited statistical power with seven pairs rather than a negligible effect. Reduced NASA-TLX mental demand, temporal demand, effort, and inverted performance support **RH7** for this collaborative detection task. This contrasts with Study 1, where inverted performance favored CC, indicating that perceived-performance effects depend on task context. The physical-demand direction shows that headset-based collaboration can still introduce bodily cost because current headsets add weight and encumbrance. Good usability (M = 72.7) supports **RH8**, and low SSQ scores support **RH9**. The system appears most useful when spatial constraints create genuine collaboration bottlenecks: real-time observation across separated locations, conflicting optimal viewing positions, or rapid switching between information sources.

6 Discussion

Perspective Sharing and Common Ground. The controlled tasks show that *PeriViewMR* can reduce the cost of establishing visual common ground [11]: B can understand A's references without walking to approximate A's viewpoint. Unlike prior work aligning hands, avatars, or virtual viewpoints [21,23,43], *PeriViewMR* shares a live stereoscopic view of a real co-located workspace and integrates it with B's own near-field view.

When the Technique Helps Most. The strongest objective benefit appears when the collaboration problem has a real viewpoint bottleneck. Static tasks give B enough time to walk around, so the main benefit is speed and reduced physical demand. Dynamic monitoring is more demanding: by the time B physically reaches a better viewpoint, the target can disappear. This explains the qualitative gap in T3 first-pair detection and the large Task 5 accuracy effect, even though the latter did not reach statistical significance with seven pairs.

Stereoscopic Sharing and Its Trade-Offs. The stereoscopic stream gives B depth cues for A's far-field view, supporting interpretation of pointing and spatial references. However, the depth is inherited from A's captured eye streams rather than reconstructed as full scene geometry for B. As a result, large differences in A/B distance to the target, head roll, or large yaw-induced camera translations can change the perceived depth and orientation of the shared region. These effects are acceptable for the tasks we evaluated, but they define important bounds for faster manual collaboration and close-range manipulation.

System Bounds. PeriViewMR is best suited to side-by-side collaborators facing the same general workspace, with a switching plane between the near-field communication space and the far-field workspace that needs sharing. The fixed plane gives B a stable spatial boundary, but it assumes collaborators stay on their side of the plane and that A keeps the task-relevant region in view.

7 Limitations and Future Work

From an implementation standpoint, difficult access to the video passthrough feed of current MR headsets contributes to latency and alignment error. Headset co-registration could also be improved. The Task 4 tracing result shows that fine visual continuity tasks are sensitive to anchor initialization, passthrough resolution, thin visual features, and the projection assumptions of the switching plane.

PeriViewMR's visual continuity relies on the assumption of planar ground at the fixed perspective switching plane. On irregular terrain (slopes, stairs, grates) or when obstacles protrude through the plane (furniture, machinery, a target straddling the boundary), visible seams may appear. The implementation uses a fixed plane per task; adapting the distance d to the workspace, task, or collaborators is future work, and dynamic movement must preserve visual stability.

Beyond the fixed plane, B's original far-field view is *replaced* by A's perspective rather than augmented, a trade-off that may matter in safety-critical environments. *PeriViewMR* is also optimized for collaborators positioned side-by-side facing the same general direction. Performance may degrade at substantially different heights, on opposite sides of an object, or under divergent facing directions, as the integration assumes significant viewpoint overlap.

Our empirical evaluation also has limitations. Study 2 had only seven pairs and was therefore underpowered for the observed accuracy effect. Our participants were primarily university-affiliated, which may not represent end-users in industrial, emergency-response, or accessibility settings. Finally, we did not compare against a monocular version, so the contribution of stereo depth cues remains future work.

PeriViewMR augments the user's real-world view with scene color samples acquired from a different perspective, and can be integrated with MR applications whose users already wear headsets. Headset encumbrance remains a barrier to natural communication, but hardware face displays [4] and software headset inpainting [17,19] are improving rapidly.

8 Conclusion

We have presented *PeriViewMR*, a mixed reality approach for alleviating viewpoint differences between two side-by-side collaborators working on a real-world scene. *PeriViewMR* lets a collaborator see the scene from their own perspective up to a fixed distance and from their collaborator's beyond it, with stereoscopic

continuity at the boundary. A controlled study (N = 27) showed significant time benefits and reduced physical and temporal workload versus walking to establish line of sight. A second peer-collaboration study (N = 14, 7 pairs) showed a large accuracy improvement in real-time collaborative monitoring, with good usability and low cybersickness scores. These results suggest that live stereoscopic perspective sharing helps when co-located collaborators need common visual access and repositioning is slow, disruptive, or impossible.

Acknowledgments. This work was supported by the National Science Foundation under Grant Numbers 2212200, 2219842, 2309564, 2318657 and 2417510. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Disclosure of Interests. The authors have no competing interests relevant to the content of this article.

References

1. Adcock, M., Anderson, S., Thomas, B.: RemoteFusion: real time depth camera fusion for remote collaboration on physical tasks. In: Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, pp. 235–242 (2013). https://doi.org/10.1145/2534329.2534331
2. Agora: Agora Real-Time Voice and Video Engagement. https://www.agora.io/en/. Accessed 6 Sept 2024
3. Andre, E., Hlavacs, H.: Diminished reality based on 3D-scanning. In: van der Spek, E., Göbel, S., Do, E.Y.-L., Clua, E., Baalsrud Hauge, J. (eds.) ICEC-JCSG 2019. LNCS, vol. 11863, pp. 3–14. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-34644-7_1
4. Apple: Apple Vision Pro. https://www.apple.com/apple-vision-pro/. Accessed 6 Sept 2024
5. Bangor, A., Kortum, P., Miller, J.: Determining what individual sus scores mean: adding an adjective rating scale. J. Usability Stud. **4**(3), 114–123 (2009)
6. Barakonyi, I., Fahmy, T., Schmalstieg, D.: Remote collaboration using augmented reality videoconferencing. In: Proceedings of Graphics Interface 2004, GI '04, pp. 89–96. Canadian Human-Computer Communications Society, Waterloo, CAN (2004). https://dl.acm.org/doi/10.5555/1006058.1006070
7. Bimberg, P., Weissker, T., Kulik, A.: On the usage of the simulator sickness questionnaire for virtual reality research. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp. 464–467. IEEE (2020). https://doi.org/10.1109/VRW50115.2020.00098
8. Bradley, A., Klivington, J., Triscari, J., van der Merwe, R.: Cinematic-L1 video stabilization with a log-homography model. In: 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1040–1048. IEEE (2021). https://doi.org/10.1109/WACV48630.2021.00108
9. Brooke, J.: SUS: A quick and dirty usability scale. In: Usability Evaluation in Industry, pp. 207–212. CRC Press (1996). https://doi.org/10.1201/9781498710411-35

10. Camacho-Fidalgo, G., et al.: Analyzing the impact of augmented reality head-mounted displays on workers' safety and situational awareness in hazardous industrial settings. arXiv preprint arXiv:2503.04075 (2025). https://doi.org/10.48550/arXiv.2503.04075
11. Clark, H.H., Brennan, S.E.: Grounding in communication. In: Perspectives on Socially Shared Cognition, pp. 127–149. American Psychological Association (1991). https://doi.org/10.1037/10096-006
12. Cohen, J.: Statistical Power Analysis for the Behavioral Sciences. L. Erlbaum Associates (1988). https://books.google.com/books?id=gA04ngAACAAJ
13. Congdon, B.J., Wang, T., Steed, A.: Merging environments for shared spaces in mixed reality. In: Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology, VRST '18, pp. 1–8. ACM (2018). https://doi.org/10.1145/3281505.3281544
14. Cortés, C., et al.: Delay threshold for social interaction in volumetric extended reality communication. ACM Trans. Multimed. Comput. Commun. Appl. **20**(7), 1–22 (2024). https://doi.org/10.1145/3651164
15. Dunlap, W.P., Cortina, J.M., Vaslow, J.B., Burke, M.J.: Meta-analysis of experiments with matched groups or repeated measures designs. Psychol. Meth. **1**(2), 170–177 (1996). https://doi.org/10.1037/1082-989X.1.2.170
16. Exit Games: Photon Engine, Photon Unity Networking. https://www.photonengine.com/pun. Accessed 6 Sept 2024
17. Frueh, C., Sud, A., Kwatra, V.: Headset removal for virtual and mixed reality. In: ACM SIGGRAPH 2017 Talks, pp. 1–2. Association for Computing Machinery (2017). https://doi.org/10.1145/3084363.3085083
18. Gergle, D., Kraut, R.E., Fussell, S.R.: The impact of delayed visual feedback on collaborative performance. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '06, pp. 1303–1312. ACM, New York, NY, USA (2006). https://doi.org/10.1145/1124772.1124968
19. Ghorbani Lohesara, F., Egiazarian, K., Knorr, S.: Expression-aware video inpainting for HMD removal in XR applications. In: Proceedings of the 20th ACM SIGGRAPH European Conference on Visual Media Production, pp. 1–9. Association for Computing Machinery (2023). https://doi.org/10.1145/3626495.3626497
20. Gottsacker, M., Chen, M., Saffo, D., Lu, F., Lee, B., MacIntyre, B.: Examining the effects of immersive and non-immersive presenter modalities on engagement and social interaction in co-located augmented presentations. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, CHI '25, pp. 1–19. ACM (2025). https://doi.org/10.1145/3706598.3713346
21. Gottsacker, M., Numan, N., Steed, A., Bruder, G., Welch, G.F., Feiner, S.: Decoupled hands: an approach for aligning perspectives in collaborative mixed reality. In: Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA '25, pp. 1–7. ACM (2025). https://doi.org/10.1145/3706599.3720219
22. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Human Mental Workload, pp. 139–183. Advances in Psychology, Elsevier (1988). https://doi.org/10.1016/S0166-4115(08)62386-9
23. Hoppe, A.H., van de Camp, F., Stiefelhagen, R.: ShiSha: enabling shared perspective with face-to-face collaboration using redirected avatars in virtual reality. Proc. ACM Hum. Comput. Interact. **4**(CSCW3), 1–22 (2021). https://doi.org/10.1145/3432950

24. Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G.: Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int. J. Aviat. Psychol. **3**(3), 203–220 (1993). https://doi.org/10.1207/s15327108ijap0303_3
25. Le Chénéchal, M., Duval, T., Gouranton, V., Royan, J., Arnaldi, B.: 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), pp. 9–12. IEEE (2016). https://doi.org/10.1109/3DCVE.2016.7563559
26. Liao, S., Zhou, Y., Popescu, V.: AR interfaces for disocclusion—a comparative study. In: 2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR), pp. 530–540 (2023). https://doi.org/10.1109/VR55154.2023.00068
27. Liu, F., Gleicher, M., Jin, H., Agarwala, A.: Content-preserving warps for 3D video stabilization. ACM Trans. Graph. **28**(3), 1–9 (2009). https://doi.org/10.1145/1531326.1531350
28. Liu, Y.L., Lai, W.S., Yang, M.H., Chuang, Y.Y., Huang, J.B.: Hybrid neural fusion for full-frame video stabilization. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 2279–2288. IEEE (2021). https://doi.org/10.1109/ICCV48922.2021.00230
29. Luo, L., Weng, D., Hao, J., Tu, Z., Jiang, H.: Viewpoint-controllable telepresence: a robotic-arm-based mixed-reality telecollaboration system. Sensors **23**(8), 4113 (2023). https://doi.org/10.3390/s23084113
30. Luo, W., et al.: Beyond links: exploring visual representations of multi-view relations in mixed reality. In: Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems, CHI '26, pp. 1–19. ACM (2026). https://doi.org/10.1145/3772318.3791398
31. McGill, M., Gugenheimer, J., Freeman, E.: A quest for co-located mixed reality: aligning and assessing SLAM tracking for same-space multi-user experiences. In: Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology, pp. 1–10. Association for Computing Machinery (2020). https://doi.org/10.1145/3385956.3418968
32. Meta: Quest 3: New Mixed Reality Headset. https://www.meta.com/quest/quest-3/. Accessed 6 Sept 2024
33. Meta: Meta Spatial Anchor Overview (2024). https://developers.meta.com/horizon/documentation/unity/unity-spatial-anchors-overview/. Accessed 16 Sept 2024
34. Meta: Passthrough Camera API Overview (2025). https://developers.meta.com/horizon/documentation/spatial-sdk/spatial-sdk-pca-overview/. Accessed 6 Sept 2025
35. Mori, S., Ikeda, S., Saito, H.: A survey of diminished reality: techniques for visually concealing, eliminating, and seeing through real objects. IPSJ Trans. Comput. Vis. Appl. **9**(1), 1–14 (2017). https://doi.org/10.1186/s41074-017-0028-1
36. Mori, S., Shibata, F., Kimura, A., Tamura, H.: Efficient use of textured 3D model for pre-observation-based diminished reality. In: 2015 IEEE International Symposium on Mixed and Augmented Reality Workshops, pp. 32–39. IEEE (2015). https://doi.org/10.1109/ISMARW.2015.16
37. Queguiner, G., Fradet, M., Rouhani, M.: Towards mobile diminished reality. In: 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 226–231. IEEE (2018). https://doi.org/10.1109/ISMAR-Adjunct.2018.00073

38. Raj, R., et al.: Feature based video stabilization based on boosted HAAR cascade and representative point matching algorithm. Image Vis. Comput. **101**, 103957 (2020). https://doi.org/10.1016/j.imavis.2020.103957
39. Rosenthal, R.: Meta-Analytic Procedures for Social Research. Sage Publications (1991)
40. Schäfer, A., Reis, G., Stricker, D.: A survey on synchronous augmented, virtual, and mixed reality remote collaboration systems. ACM Comput. Surv. **55**(6), 1–27 (2022). https://doi.org/10.1145/3533376
41. Segal, M., Korobkin, C., van Widenfelt, R., Foran, J., Haeberli, P.: Fast shadows and lighting effects using texture mapping. In: Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '92, pp. 249–252. ACM (1992). https://doi.org/10.1145/133994.134071
42. Sereno, M., Wang, X., Besançon, L., McGuffin, M.J., Isenberg, T.: Collaborative work in augmented reality: a survey. IEEE Trans. Vis. Comput. Graph. **28**(6), 2530–2549 (2022). https://doi.org/10.1109/TVCG.2020.3032761
43. Simões, J., Maciel, A., Moreira, C., Sousa, M., Jorge, J.: SPARC: shared perspective with avatar distortion for remote collaboration in VR. In: Magnenat-Thalmann, N., Kim, J., Sheng, B., Deng, Z., Thalmann, D., Li, P. (eds.) Advances in Computer Graphics, pp. 99–112. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-82021-2_7
44. Sousa, M., Mendes, D., dos Anjos, R.K., Lopes, D.S., Jorge, J.: Negative space: workspace awareness in 3D face-to-face remote collaboration. In: Proceedings of the 17th International Conference on Virtual-Reality Continuum and its Applications in Industry, VRCAI '19, pp. 1–2. ACM (2019). https://doi.org/10.1145/3359997.3365744
45. Tait, M., Billinghurst, M.: The effect of view independence in a collaborative AR system. Comput. Support. Coop. Work (CSCW) **24**(6), 563–589 (2015). https://doi.org/10.1007/s10606-015-9231-8
46. Tserenchimed, T., Kim, H.: Viewpoint-sharing method with reduced motion sickness in object-based VR/AR collaborative virtual environment. Virtual Reality **28**(3), 122 (2024). https://doi.org/10.1007/s10055-024-01005-z
47. Unity Technologies: Unity 3D - Game Engine. Unity Technologies (2023). https://unity.com/
48. Wang, L., Wu, W., Zhou, Z., Popescu, V.: View splicing for effective VR collaboration. In: 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 509–519. IEEE (2020). https://doi.org/10.1109/ISMAR50242.2020.00079
49. Woolson, R.F.: Wilcoxon signed-rank test. In: Wiley Encyclopedia of Clinical Trials, pp. 1–3. Wiley (2008). https://doi.org/10.1002/9780471462422.eoct979
50. Wu, M.L., Popescu, V.: Efficient VR and AR navigation through multiperspective occlusion management. IEEE Trans. Vis. Comput. Graph. **24**(12), 3069–3080 (2018). https://doi.org/10.1109/TVCG.2017.2778249
51. Yoo, G.S., Ji, Y.G.: A study on view sharing AR interface for improving situation awareness during military operations. Int. J. Hum. Comput. Interact. **41**(4), 2211–2226 (2025). https://doi.org/10.1080/10447318.2024.2316949
52. Zokai, S., Esteve, J., Genc, Y., Navab, N.: Multiview paraperspective projection model for diminished reality. In: 2003 Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 217–226. IEEE (2003). https://doi.org/10.1109/ISMAR.2003.1240705