

VirtualNexus: Enhancing 360-Degree Video AR/VR Collaboration with Environment Cutouts and Virtual Replicas

Xincheng Huang*

University of British Columbia
Vancouver, BC, Canada
xchuang@cs.ubc.ca

Ziyi Xia

University of British Columbia
Vancouver, BC, Canada
zxia0101@cs.ubc.ca

Michael Yin*

University of British Columbia
Vancouver, BC, Canada
jiyin@cs.ubc.ca

Robert Xiao

University of British Columbia
Vancouver, BC, Canada
brx@cs.ubc.ca



Figure 1: *VirtualNexus* enhances 360° video AR/VR collaboration with environment cutouts and virtual replicas. In (a), the VR user is telepresent in the AR user's physical environment. To have a close-up view of the desk, the VR user can create a manipulable *environment cutout* that they have dragged closer. Simultaneously in (b), the AR user sees the VR user's avatar closer to the desk from the camera's position. Virtual annotations and objects are positionally synchronized across the cutout, the desk in the 360° scene, and the physical desk. *VirtualNexus* additionally implements ad-hoc 3D virtual replica creation from Instant-NGP [32]. In (c) we showcase virtual replicas of a pig and a dinosaur with their original physical copies. In (d) we showcase an example storyboard participants created with *VirtualNexus* in our study (from AR user's perspective).

ABSTRACT

Asymmetric AR/VR collaboration systems bring a remote VR user to a local AR user's physical environment, allowing them to communicate and work within a shared virtual/physical space. Such systems often display the remote environment through 3D reconstructions or 360° videos. While 360° cameras stream an environment in higher quality, they lack spatial information, making them less interactable. We present *VirtualNexus*, an AR/VR collaboration system that enhances 360° video AR/VR collaboration with *environment cutouts* and *virtual replicas*. VR users can define cutouts of the remote environment to interact with as a world-in-miniature, and their interactions are synchronized to the local AR perspective. Furthermore, AR users can rapidly scan and share 3D virtual replicas of physical objects using neural rendering. We demonstrated our system's utility through 3 example applications and evaluated our

system in a dyadic usability test. *VirtualNexus* extends the interaction space of 360° telepresence systems, offering improved physical presence, versatility, and clarity in interactions.

CCS CONCEPTS

- Human-centered computing → Mixed / augmented reality.

KEYWORDS

Virtual/Augmented Reality, Computer Mediated Communication

ACM Reference Format:

Xincheng Huang, Michael Yin, Ziyi Xia, and Robert Xiao. 2024. *VirtualNexus: Enhancing 360-Degree Video AR/VR Collaboration with Environment Cutouts and Virtual Replicas*. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*, October 13–16, 2024, Pittsburgh, PA, USA. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3654777.3676377>

*Both authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

UIST '24, October 13–16, 2024, Pittsburgh, PA, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 979-8-4007-0628-8/24/10
<https://doi.org/10.1145/3654777.3676377>

1 INTRODUCTION

Asymmetric remote AR/VR collaboration systems allow a remote VR user to be telepresent in a local AR user's physical environment [12, 52, 55], allowing them to communicate and work effectively within a shared virtual/physical space. Such systems usually display the physical environment to the remote VR user through 3D reconstructions (e.g., textured spatial meshes [51, 54], point clouds [37, 52, 55]), or 360° videos. Typically, the decision between

the two options induces a trade-off — while 360° videos stream in higher quality compared to 3D reconstructions, they hinder efficient bi-directional interaction as they lack 3D spatial (depth) information. Furthermore, virtual objects can only float in front of the 360° video (instead of physically reacting with a 3D scene reconstruction), breaking the illusion of being physically present.

To address the lack of spatial context, contemporary 360° systems [21, 24, 40] have explored mobile locomotion for the 360° camera. However, a locomotive camera may cause simulator sickness and is less feasible for regular users. In regards to the issue of virtual object reference and interaction, prior 360° systems have enhanced functionality in collaboration through additional non-verbal cues such as gazes, gestures, ray pointers, and annotations [28, 39, 51, 54]. However, similar enhancements have not been extended to object manipulation; it is also challenging to incorporate physical objects of the 360° environment into the collaboration. Thus, a clear gap emerges - how can we retain the high visual fidelity of a 360° display while extending the collaborative interaction with spatial manipulation, within the environment and with the virtual objects (more akin to 3D reconstruction)? Prior work has explored combining 360° videos and 3D reconstructions in telepresence and remote collaboration [10, 51, 61]; however, past research switches between these two views instead of harnessing their merits simultaneously.

We present *VirtualNexus*, a system that augments spatial interactivity in standard 360° video remote AR/VR collaboration using *environment cutouts* and *virtual replicas*. Environmental cutouts are a feature that allows a remote VR user to cut out a part of the 360° environment as a live textured mesh. The users can pull the environment cutout closer as a World in Miniature (WiM), bringing the environment within reach and offering precise control. Changes a user makes in the environment cutout synchronize to the original 360° video and overlay on the AR user's view of the physical environment. We further implemented ad-hoc 3D virtual replica creation with Instant-NGP [32], which allows the local AR user to scan a physical object and obtain a shared virtual replica within 1–3 minutes, further bridging the physical and virtual environments. We demonstrated the utility of these novel features through three application scenarios and evaluated our system in a user study. *VirtualNexus* is lightweight as we only require the use of an off-the-shelf 360° camera, AR and VR HMDs, and a consumer-grade computer to act as the server. We found that *VirtualNexus* extends the interaction space of 360° telepresence systems with enhanced physical presence, versatility, and clarity in interactions.

2 RELATED WORK

Telepresence immersively brings a remote guest to a local user's physical environment [16, 41, 42, 50]. It has been a longstanding area of research, especially in the context of AR/VR remote interaction [17, 19, 28, 52, 54]. To display the physical environment to the remote VR user, prior research has explored 3D reconstruction (i.e., textured spatial meshes [1, 39] or point clouds [37, 52]) and 360° videos [21, 27, 30, 40]. As 3D reconstructions are themselves virtual objects in VR, they have richer interactive potential than 360° videos. It is easier for users to move around and augment a 3D reconstruction in a virtual world [37, 55]. However, compared to 360° videos, real-time 3D reconstruction typically has lower quality, and

it suffers from holes and occlusions. Holoportion [37] implements a pipeline that can stream high-quality full-scene reconstruction in real-time, but it requires high-end sensors, computing, and network infrastructure. In comparison, telepresence with 360° video cost-effectively provides higher quality (commodity 6K 360° cameras are around \$500) and thus better presence and immersion [28, 48, 58]. Nevertheless, 360° videos are essentially a texture rendered on a spherical screen. Therefore, it is more challenging to incorporate common AR/VR interactive modalities in 360° telepresence.

2.1 Combining 360° Video and 3D reconstructions

Given the respective merits of 360° video and 3D reconstructions, prior work has explored combining the two in remote AR/VR collaboration. Teo et al. and Gao et al. proposed toggling between the modes of using 3D reconstruction or 360° video [10, 52]. However, the need to switch between two different media prevents simultaneously harnessing the merit of both. The authors also reported that frequently switching between perspectives and interactive modalities offered by different modes is challenging to adjust to. Young et al. extended this work, providing seamless transitions based on distance between users instead [61]. Teo et al. also proposed follow-up works [51, 54] that can insert 360° panorama as bubbles into 3D reconstructions. However, the 3D reconstruction in the proposed system has a static texture and is mostly used as context. Although users may update the 3D reconstruction's context with newly captured 360° images, they rely mostly on the live 360° video mode [54] or live 360° insertion [51] for real-time interaction. In our work, both content delivered through 360° and 3D reconstruction are live. We simultaneously provide a live 360° environment and live environment cutouts (spatial mesh textured with live video texture). We additionally provide enhanced interactivity with virtual objects and replicas. Thus, we now review common interactive requirements in AR/VR remote collaboration and how they apply to 360° video.

2.2 Interactivity in 360° Video Telepresence

To enhance the presence of the remote guest and the effectiveness of AR/VR remote collaboration, prior research has explored a variety of interactive modalities, and we review them as follows.

2.2.1 Access and Exploring a 360° Scene. It is straightforward to allow users to move and explore the remote environment in a 3D reconstruction. However, the same task is more challenging for 360° video telepresence as the remote users always take the perspective of the 360° camera. With a stationary camera, users can only access farther regions of the scene with far manipulation (e.g., far hand ray), reducing the precision of control. Prior research has proposed having the local user move the 360° camera in the physical space [24, 40, 51, 52, 54] by mounting a 360° camera to the local user, synchronizing the perspective of the local user and the remote guest. However, such an approach leads to an inconsistency between the remote user's physical and perceived motion, which could lead to simulator sickness in VR [14, 40]. More importantly, transferring the perspective control to the local user diminishes the remote user's freedom to explore the space, which could impair more comprehensive collaborative tasks (e.g., prototyping, gaming, and entertainment, tasks with divided labour). Alternatively,

VROOM [21] mounts a 360° camera on a locomotive robotic agent remotely controlled by the remote user. However, using a robotic agent is too bulky and costly for regular users.

2.2.2 Worlds in Miniature. Worlds in Miniature (WiM) [6] is a miniaturized representation of an entire or part of a physical or virtual world. The most common use of WiMs is navigation [22, 33], but prior research has extended their capability to manipulate virtual environments [4, 6, 49]. Similar to manipulating a Voodoo doll [38], synchronizing a user’s inputs to a WiM with the larger world allows them to manipulate regions that are out of their reach. In an AR collaboration context, Yu et al. explore the idea of duplicated reality [62]. Their system creates a WiM (digital twin) that reconstructs a volume of the physical world, however, they rely on sensors in a small spatial region, limiting flexibility. Overall, using a WiM as an interactive technique in telepresence and remote collaboration has not been widely explored.

2.2.3 Reference and Augmentation. In remote mixed-reality collaboration, users often augment the shared space with pointers, virtual annotations, and virtual objects so they can better communicate ideas and collaborate [25, 39]. The ability to reference and augment the virtual world enriches the task and collaboration space of remote communication [3] and facilitates group awareness [11]. Prior research has enhanced 360° video collaboration with the use of gaze, ray pointers, and virtual annotations in 360° videos [51–54]. However, enhancing virtual object manipulation in 360° remote collaboration has not been well explored. While it is common to have virtual objects react to the physical environment with collision and physics in mixed reality, 360° videos lack spatial information to provide the same physicality (e.g., virtual objects float in front of the video, instead of lying on a physical surface), hindering the sense of being physically present for the remote user. Rhee et al. [44] incorporated synchronized ray pointers and virtual objects in remote collaboration. However, they took a graphical approach and focused on naturally blending virtual objects with the 360° video using an image-based lighting technique for 360° videos [43]. We take a physics approach: virtual objects are rendered on the 360° video, but physically react to an embedded 3D reconstruction.

2.3 Virtual Replicas and Neural Radiance Fields

It is challenging to provide remote users access to the physical environment they are telepresent in. Recent research has taken mechanical and robotic approaches, allowing remote users to move physical objects in the local user’s space with mini-robots [18] or deformable interfaces [9, 29]. However, such methods usually have a limited area of operation (e.g., a delegated platform like a desk) and introduce additional hardware overhead. An alternative approach is to provide indirect physical access through virtual replicas [8, 36]. However, most prior work requires virtual replicas to be created in advance with CAD tools [8, 36, 56, 57, 63] or only support creating from 2D contents or sketches [12, 15, 17]. While depth-based methods such as Kinect-Fusion [20] can quickly reconstruct an object or a scene, more recently, Neural Radiance Fields (NeRF) [2, 7, 31, 32] allow object and scene reconstruction with high quality. Notably, Instant-NGP [32] drastically reduces the training time of NeRF, making it feasible to reconstruct individual objects within seconds

or minutes. In our work, we incorporate virtual replica creation with Instant-NGP into our collaborative telepresence system.

3 INTERACTIVE DESIGN AND CONCEPTS

By distilling the requirements and gaps from related work, here we propose concepts and designs for *VirtualNexus*.

3.1 Preserving Spatial Physicality: Embedded 3D Reconstruction

360° VR telepresence allows a user to explore a remote space omnidirectionally with immersion. However, regular 360° videos lack spatial information to allow users to virtually interact with physics and collision (e.g., draw annotations on a wall, and bounce a virtual object on a desk), thus reducing the sense of being physically present. To solve this, we propose to align a spatial reconstruction with the 360° Video. While we render virtual objects with the 360° video, they behave like reacting with an actual physical environment when users manipulate them. The aligned spatial reconstruction should be transparent to preserve the higher reality of the 360° video. As most state-of-the-art AR headsets maintain a spatial map behind the scene, the process of creating and aligning a 3D reconstruction should be seamless and hidden from the users.

3.2 Enhancing Access to Environments: Interactable Environment Cutouts

In 360° telepresence, with a regular stationary 360° camera setup, users can only rely on far-hand manipulation (e.g., dragging an object with a long ray pointer) to access faraway regions in the scene, precluding precise interactions. Therefore, we introduce the concept of *environment cutouts*, allowing the remote VR user to create a “slice” of the 360° environment that can be interacted with at a different scale or position. For example, the VR user can select a part of the real world to make a copy, optionally scale it down (similar to a miniature diorama), pull it closer, and interact with this cutout (for example, placing virtual objects on this smaller world) while any such interactions are also reflected on the original world location. While the remote VR user can use the ray pointer to access farther objects, the ability to pull an environment cutout closer allows users to harness near interactions (e.g., grab, near draw) that have a higher precision. To convey the intention of the VR user to the AR user, the AR person will see the VR user’s avatar moving toward the physical counterpart of the cutout as they pull an environment cutout (e.g., the VR person pulling a whiteboard closer is rendered as them moving toward it).

3.3 From Reality to Virtual: Ad-hoc Creation of 3D Virtual Replicas

In immersive environments, virtual replicas are useful props for referring to objects, conveying ideas, and prototyping rapidly [8, 63]. *VirtualNexus* enables ad-hoc creation of 3D virtual replicas in remote AR/VR collaboration. The AR user can conveniently set an object on a platform, scan around it, and obtain a shared virtual replica. *VirtualNexus* additionally stores the scanned virtual replica, enabling a “scan once, create many” experience.

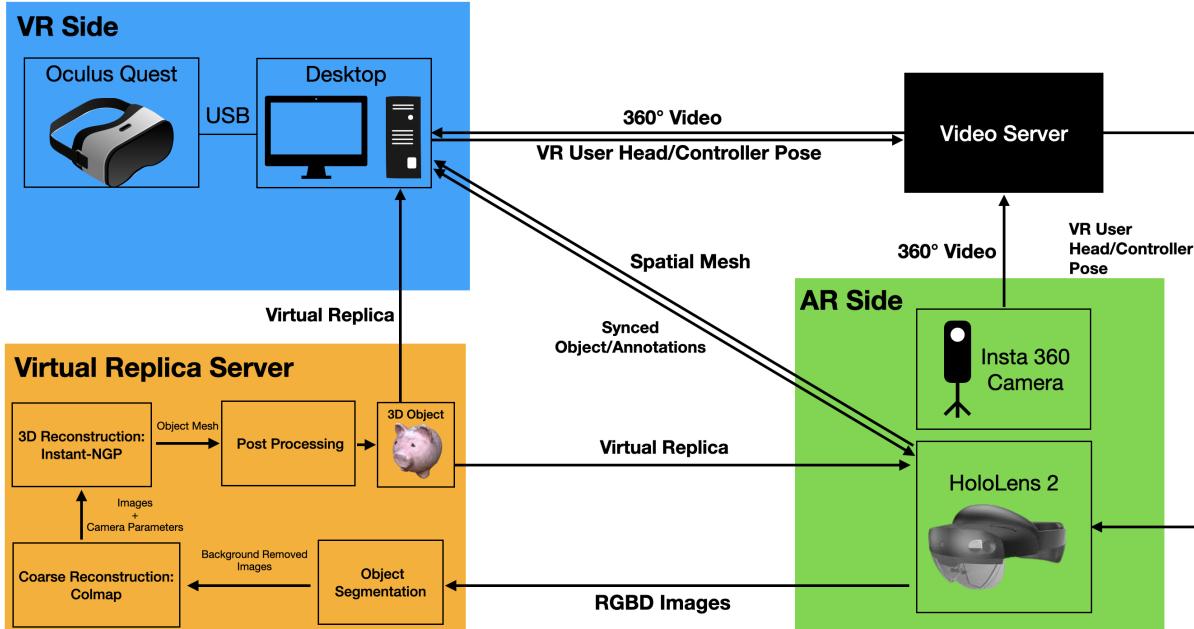


Figure 2: The system has four major components: VR side, AR side, video server, and virtual replica server. The VR side receives 360° video from the AR side via the video server. The AR side shares synced objects and annotations with the VR side and sends scanned RGBD images to the virtual replica server, which creates virtual replicas and sends it to both the AR and VR sides.

3.4 Spatially Aligned and Synced Collaboration

Co-location in a spatially aligned and synchronized environment is fundamental for remote AR/VR collaboration systems. Therefore, *VirtualNexus* offers synchronized ray pointers, annotations, and shared virtual objects, which are essential elements to maintain group awareness [3, 11]. For coherence, these features also adapt to the aforementioned system design: 1) annotations and virtual objects are able to collide and physically interact with the hidden spatial reconstructions, and 2) the environment cutout maintains a cloned copy of annotations and virtual objects that are synced with the original 360° environment and the AR physical environment.

4 VIRTUALNEXUS

4.1 System Architecture and Apparatus

We implemented *VirtualNexus* (Fig. 2) using Unity 2021.3.20f1, which can be configured as either a VR or AR application. *VirtualNexus* uses Microsoft HoloLens 2¹ for AR and Meta Oculus Quest 2² for VR. An Insta360 X3³ 360° camera omnidirectionally streams the local user's environment at 5.6K resolution and 30fps to the remote VR side. For efficient 360° video streaming, we re-implemented a foveated video compression pipeline introduced by prior work [16] on a desktop machine with an Intel Core i7-9700K 3.6GHz CPU, 32GB RAM, and an NVIDIA GeForce RTX 2060. QR codes on the front and back of 360° camera's tripod serve as the spatial anchors, aligning the VR world's origin with the 360° camera's lenses. The

local AR user sees a virtual avatar overlaid on the 360° camera with synchronized head and hand poses of the remote VR user. Finally, *VirtualNexus* runs virtual replica processing and VR-side rendering on the same machine, which has an Intel Core i9-12900KF 3.2GHz CPU, 64GB memory, and an NVIDIA GeForce RTX 4090 GPU. The AR user can scan a physical object and send the resulting photos to the virtual replica creation server. The server pre-processes the images, reconstructs a virtual replica with Instant-NGP, and sends it back to both AR and VR as shared virtual objects.

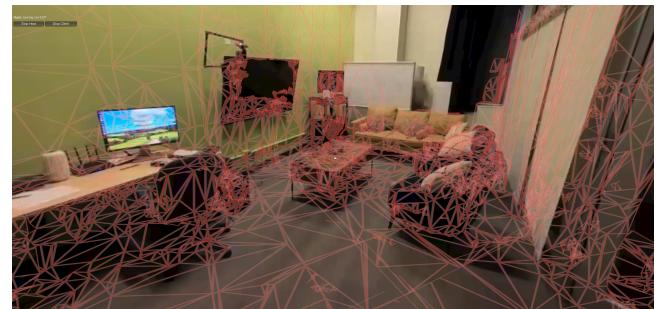


Figure 3: Spatial-accurate Alignment of 3D Reconstruction with the 360° Video. The edges of the spatial mesh are only coloured in red here for demonstrative purposes.

4.2 Combining 360° Video with Spatial Mesh

VirtualNexus embeds a spatially aligned 3D reconstruction of the physical environment with the 360° video, laying out the basis for

¹<https://www.microsoft.com/en-us/hololens>

²<https://www.meta.com/ca/quest/products/quest-2/>

³<https://www.insta360.com/product/insta360-x3>

spatial interaction with physicality. To achieve this, we utilized the spatial meshes created and maintained by Microsoft HoloLens 2, which is the foundation of a mixed-reality experience. As HoloLens creates or updates a spatial mesh, *VirtualNexus* extracts the vertices and triangles from the spatial meshes and transforms them into the VR world's coordinates. *VirtualNexus* then sends the mesh information through a TCP connection to the remote VR side and reconstructs the spatial meshes in real-time. To accurately align the 360° video with the reconstructed spatial mesh, we reverse-engineered the 360° camera's intrinsic parameters and projected the 360° video to the skybox with equidistant fisheye mapping⁴. We show the alignment between the spatial meshes and the 360° video in Fig. 3. We implement spatial mesh synchronization in a silent thread to keep it seamless for both AR and VR users.

4.3 Monocular-Binocular Trade-off

In virtual and augmented reality, virtual contents are rendered binocularly to generate a depth cue. However, as most 360° videos are monocular, users can only tell depth using their empirical knowledge of objects' sizes (i.e., closer objects look bigger and farther objects are smaller). We started with overlaying binocularly rendered objects in front of a monocular 360° video. However, we found that this causes an inconsistency regarding depth perception: a virtual object looks closer than a physical object in the 360° video even if they are placed in the same position. To mitigate this, in the VR build, we shifted the position of the right-eye camera leftwards by the VR headset's inter-pupillary distance (IPD), causing the VR headset to effectively render in monocular mode, thus rendering virtual objects as if they belonged to the 360° video. Such an adaptation may lead to difficulty in perceiving depth during object manipulation. In the future, we can opt to use binocular 360° cameras (already available as commodity products) creating 360° videos with binocular depth perception.

4.4 Spatially Synchronized Collaboration

As mentioned in 4.1, we aligned the AR and VR space using the QR Codes attached to the 360° camera as the spatial anchor. To facilitate synchronized remote collaboration, we implemented synchronized ray pointers, annotations, and virtual objects.

4.4.1 Synchronized Ray Pointers and Annotations. It is common to use ray pointers to convey ideas and intentions in virtual and augmented applications. In *VirtualNexus*, the AR and VR users can see each other's hand/controller ray pointers, which are implemented by constantly exchanging ray origin and direction information using UDP packets. We implemented shared annotations by adding two additional bytes to the same UDP packets exchanging ray pointer positions: a byte that indicates whether a user is drawing and a byte indicating the number of annotations a user has drawn. The drawing flag is set to 1 when a user is drawing annotations in their own world (the VR user presses a controller button and the AR user uses a pinch gesture), causing the user's synchronized pointer in the other user's world to draw annotations at the same time. We use the number of annotations as a sequence number to detect when a user starts a new annotation or deletes the latest annotation.

⁴https://docs.opencv.org/3.4/db/d58/group__calib3d_fisheye.html

We implemented the annotations with Unity line renderers. Users can create floating annotations or draw on the environment. In the latter case, we attach the annotations (as child objects) to the scene objects (e.g., depth mesh or environment cutouts) they are drawn on. To distinguish the ownership of annotations, the users see their own annotations in green and the collaborator's annotations in red. The VR user can switch between far and near annotations (i.e., between the ray pointer and the "poke" pointer). In the near annotation mode, a green sphere is rendered at the position of the poke pointer to indicate the status of annotation (see Fig. 4b).

4.4.2 Shared Virtual Objects. Both the AR and VR users can spawn shared virtual objects (see Fig. 6a). Virtual objects appear in the same location in the world for both the AR and VR user and can be freely controlled by either user via grab interactions. Users can also choose to edit their physics properties, their material, etc. Motions and edits are fully synchronized between sides by using a server-client setup built atop the Mirror⁵ library. Our implementation initially provides a default set of meshes representing some base shapes, such as a cube, sphere, etc. The local AR user can extend this set by scanning physical objects into shared virtual replicas. We detail the virtual replica creation in Sec. 4.6.

4.5 Environment Cutouts

To define an environment cutout, the VR user first makes a selection of 4 points with their ray pointer cast onto the depth mesh (Fig. 4a). These raycast points and the camera position define a selection frustum. We then select the triangles of the spatial mesh that lie in the selection frustum, which form new mesh objects that define the cutout. While users can create both 2D (Fig. 4b) and 3D cutouts (Fig. 4c), the latter may be subject to occlusions.

The cutout supports standard VR manipulations, such as grabbing, rotating, and scaling. Users can "select" a cutout to make it active, which causes the VR user's actions to be performed relative to the cutout, and causes their avatar to be rendered in AR relative to the cutout's physical counterpart (e.g. as shown in 7(b)). With no cutout, or if the cutout is deselected, the VR user's interactions will occur with respect to the 360° video, and they will be rendered at the location of the 360° camera (as in Fig. 7d).

We sync interactions across this copied cutout and the world-space 360° video. When the VR user creates a virtual object (annotation or mesh) while a cutout exists, they see two objects – one corresponding to the world space (the "original object"), and one relative to the cutout ("copy object"), for example, in Fig. 4b. Movements and edits are synchronized between the original, copy, and the virtual objects displayed to the AR user, but the copy itself is only visible to the VR user when using a cutout.

4.6 Virtual Replica Creation with Instant-NGP

VirtualNexus allows the AR user to create shared virtual replicas from physical objects in the environment. We chose Neural Radiance Fields (NeRF) to create virtual replicas from a futuristic standpoint: NeRFs have shown promise in producing photorealistic scans of objects and scenes with high-quality lighting and texture, and we feel that future object scanning pipelines may involve

⁵<https://mirror-networking.com/>

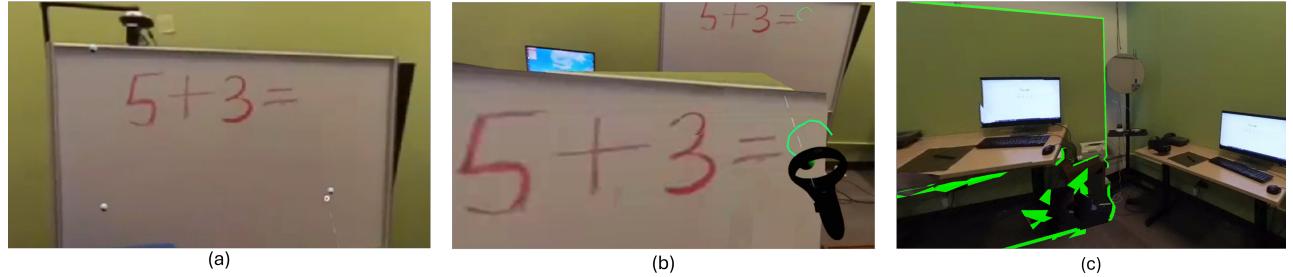


Figure 4: Environment Cutouts: In (a), the VR user defines a cutout of the whiteboard through 4 raycasted points. In (b), annotations on active cutouts are synced to the original location. In (c), users can cutout 3D space additional to 2D surfaces.

such technologies for fidelity, rather than traditional pipelines like KinectFusion [20]. However, for compatibility with existing mesh-rendering pipelines, we produce both a NeRF model and traditional vertex-coloured mesh from our object scanning pipeline.

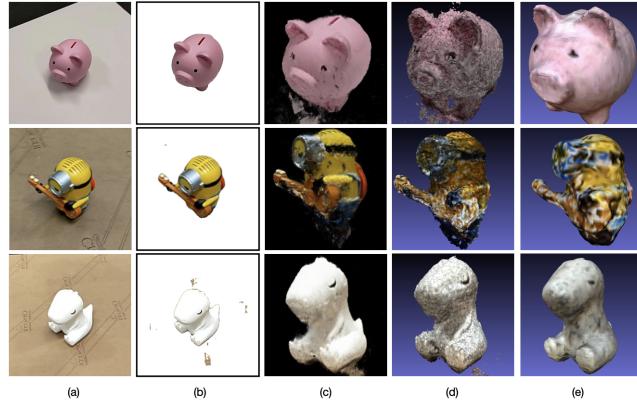


Figure 5: Intermediate results for rapid virtual replica creation: (a) original RGB image (b) background removed image (c) volume rendering by Instant-NGP (d) mesh object created from cube-marching (e) Voxelized and smoothed final object.

To the best of our knowledge, our system is the first to adopt NeRF reconstruction for remote AR/VR collaboration. Among the variants of NeRF scene reconstruction techniques, Instant-NGP [32] strikes a balance between training time and quality. In our preliminary exploration, reconstructing individual objects with Instant-NGP (i.e., as opposed to an entire scene) with sufficient quality only takes 1–3 minutes on our NVIDIA GeForce RTX 4090 GPU machine. As we require a user to walk around the targeted object, our pipeline is best suitable for smaller desktop objects (e.g., small appliances, toys, hand-held tools). *VirtualNexus*' end-to-end virtual replica creation pipeline has a server-client architecture (see bottom left of Fig. 2). The server is implemented in Python and incorporates Instant-NGP's Python API⁶. Examples of intermediate results at each stage of the pipeline can be seen in Fig. 5.

⁶<https://github.com/NVlabs/Instant-NGP>

4.6.1 Colour and Depth Image Capturing. To scan an object, a user triggers the function in the AR application and then walks around the target object. A semi-transparent grey rectangle is rendered to help the user center the object in their field of view (FoV). We capture colour and depth images of the object using the native Universal Windows API⁷ and HoloLens 2 Research Mode API⁸ at 5 FPS for about 15 seconds, accumulating around 75 images. For each frame, we reproject the depth image from the perspective of the colour camera to align the depth and colour images, then stream the resulting images to the reconstruction server. The depth images are then used for background segmentation and improving the efficiency and quality of the NeRF reconstruction [7].

4.6.2 Pre-processing: Background Segmentation. To reconstruct a clean virtual replica of an object, we first remove the background, which we assume is a planar surface (e.g. a table or platform). In each image, we start with the plane obtained from the HoloLens' built-in plane detection functionality, then use a RANSAC algorithm to refine the fit [60]. We select all non-planar points as the initial ‘coarse mask’ of foreground pixels. Subsequently, we obtain a refined segmentation mask from Segment-Anything [26] using the average of the coarse mask as a point prompt. Segment-Anything [26] outputs a hierarchy of masks, and we use the one that best overlaps with the coarse mask as the final segmenting mask. Our background segmenting process takes about 25 seconds.

4.6.3 Colmap and Instant-NGP. Before providing the images to Instant-NGP[32], we need to obtain the camera poses for the images. Initially, we tried to use the HoloLens' reported camera poses for each frame directly, but found that the poses were not accurate enough for satisfactory reconstruction. Therefore, we used Colmap [45, 46], a structure-from-motion technique that is used by most NeRF variants. We fed the images and the Colmap-determined camera poses to Instant-NGP, which reconstructs the object as a NeRF model and also outputs a vertex-coloured mesh with cube-marching. Running Colmap is the most expensive part of our pipeline, and can take anywhere from 20 seconds to 2 minutes. By contrast, Instant-NGP's training process takes about 15 seconds while cube-marching is practically instantaneous.

⁷<https://learn.microsoft.com/en-us/windows/uwp/audio-video-camera/process-media-frames-with-mediaframereader>

⁸<https://github.com/microsoft/HoloLens2ForCV>

4.6.4 Post-processing: Mesh Simplification and Smoothing. The initial mesh created by Instant-NGP contains too many vertices and often contains unsightly holes. Therefore, we apply a “Remesh Modifier” with voxelization and smooth shading using the Blender API⁹ on the initial mesh. This process both simplifies and smooths the mesh, and takes about 3 seconds. Our virtual replica creation server sends the mesh information as an obj file with per-vertex colours to both the AR and VR builds. We implemented a parser that can process colourized obj files at runtime, allowing either the AR or VR user to create them as shared virtual objects (Section 4.4.2)

5 APPLICATION SCENARIOS

Here we outline *VirtualNexus*’ application to various domains.

5.1 Content Authoring and Prototyping

Collective prototyping is a key application domain for collaborative mixed reality [17, 34]. *VirtualNexus* facilitates such real-time content authoring through the creation, sharing, and manipulation of virtual objects and annotations for remote users. Furthermore, the scanning and creation of instant replicas allow both users to integrate shapes beyond basic primitives, bringing in virtual objects that mimic real physical items. The cutout feature provides additional options for the remote user to interact and prototype, allowing increased precision by bringing further areas closer.

For example, in Fig. 6, we use *VirtualNexus* to create a collaborative virtual scene on a desk that the local user can walk around and view from different angles. In this scene, basic primitives such as cubes and spheres form the environment, and virtual replicas are used to create more detailed characters. Annotations are used to define areas in the environment (i.e. a path). The domain of content creation and prototyping also forms the basis of our user study (Section 6), which involves collaboratively building a storyboard.

5.2 Remote Education and Instruction

Remote mixed reality systems also apply to remote education and instruction [23, 47, 59]. Online learning platforms have become increasingly important in an increasingly digital world, and such platforms facilitate communication between educators and students despite distances [5]. We demonstrate a virtual classroom environment in which a teacher, in a classroom, can call upon a student, who may be joining remotely, to answer a question on a whiteboard (Fig. 7a and 7b). The teacher first annotates a question on the whiteboard. The student might find the whiteboard to be too far to interact with precisely. In real life, the student may walk up to the whiteboard. *VirtualNexus* allows the student to achieve the same by bringing the whiteboard closer using its environment cutout. The student can then annotate their answer on this closer cutout, which is then reflected on the original whiteboard.

To extend this education scenario, the teacher might ask students to mirror their interactions with virtual objects in a virtual hands-on lesson (Fig. 7c and 7d). To illustrate, we outline an arts-and-craft exercise in which the teacher is teaching about modelling and colouring while the student follows along. The teacher’s desk has equipment and objects found in the classroom (i.e. markers); the student can replicate it using virtual objects. However, some

required objects for the task may not be physically present for the student (i.e. the model pig). Thus, the teacher can scan the object locally and create a replica for the remote student. The student can then use these virtual objects to replicate the teacher’s instructions.

5.3 Shared Recreational Activities

Mixed reality mediums are often used in games and other recreational domains to encourage exercise and socialization. Using *VirtualNexus*, we can develop collaborative recreational activities that use virtual objects for remote users. We illustrate an example using a bowling game situated on a virtual alley overlaid on the observed local environment (Fig. 8). Users can create virtual bowling pins and lay them at the end of the virtual alley. Then, either user can create a ball, which can be rolled at the pins. By taking turns rolling the balls, the users can experience a fun virtual bowling session situated in a physical environment.

6 USER STUDY

We conducted a user study on *VirtualNexus*. With a collaborative storyboarding task, we assessed the usability of the novel interactive techniques (e.g., environment cutouts and virtual replicas).

6.1 Participants

We recruited 14 participants (8 females, 5 males, averaged 24.8 years old. 1 participant reported N/A for both demographic questions) through convenience sampling, forming 7 dyads. All participants had some prior experience with remote collaborative tools (e.g. Google Docs) and video communication tools (e.g. Zoom). Almost all participants had some experience with using VR in the past (13 out of the 14 participants); experience with AR headsets was rarer (8 out of the 14 participants). Our study was reviewed and approved by the institutional ethical board, and all participants reviewed and signed a consent form prior to the study.

6.2 Study Protocol

Upon arrival, the participant dyad was split into a local AR user and a remote VR user. Each participant underwent an individual short guided tutorial regarding interactions using *VirtualNexus* features. After the participants familiarized themselves with the system, they worked together to create a virtual storyboard for a researcher-provided narrative. This collaborative task took users through all features of the system - participants discussed and ideated the storyboard through cutouts and annotations, and then created it using virtual scanned objects. We instructed the participants to scan and create the first physical object they decided to use in the task, allowing them to experience the full virtual replica creation process. In the interest of time, we provide pre-scanned models for additional objects users may need afterwards. We continued the task until the users had finished creating enough scenes, or when we reached the allotted time. After completion, the researchers wrapped up the study through a final questionnaire. It involved 5-point Likert-scale questions that related to immersion, presence, ease of use, and task performance using the *VirtualNexus* system, as well as an optional open-ended field for each question (Fig. 9 and 10). The entire study took approximately 90 minutes, and participants were reimbursed \$24 CAD.

⁹<https://docs.blender.org/api/current/index.html>

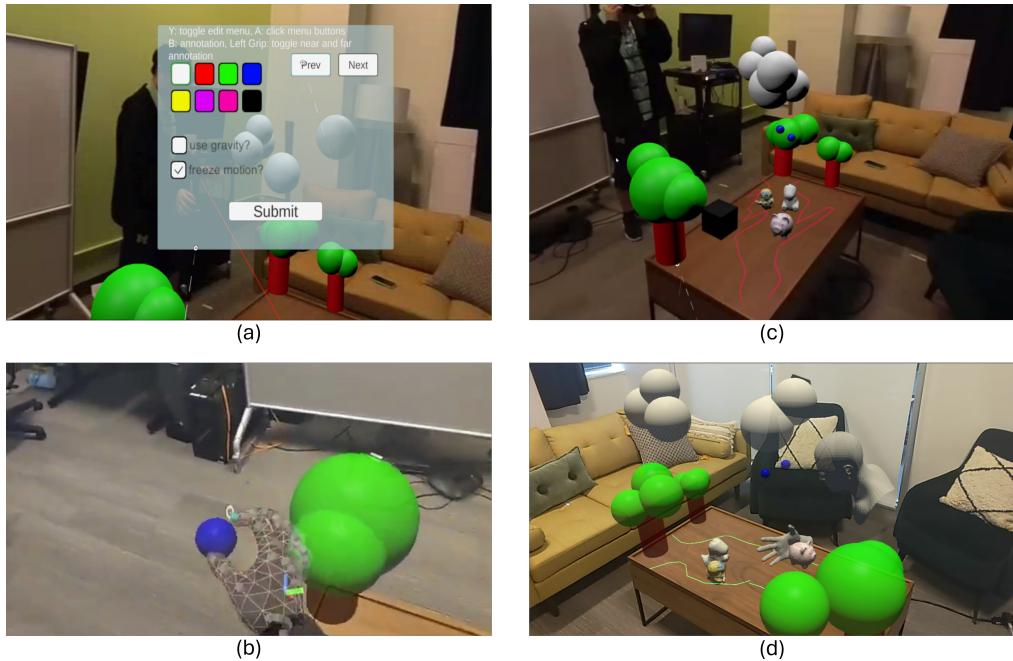


Figure 6: The collaborative prototyping scenario: The users collaboratively create a story scene. In (a) and (b), both VR and AR users use shared virtual objects for the scene. Subimages (c) and (d) show the completed scene from the VR and AR perspectives.

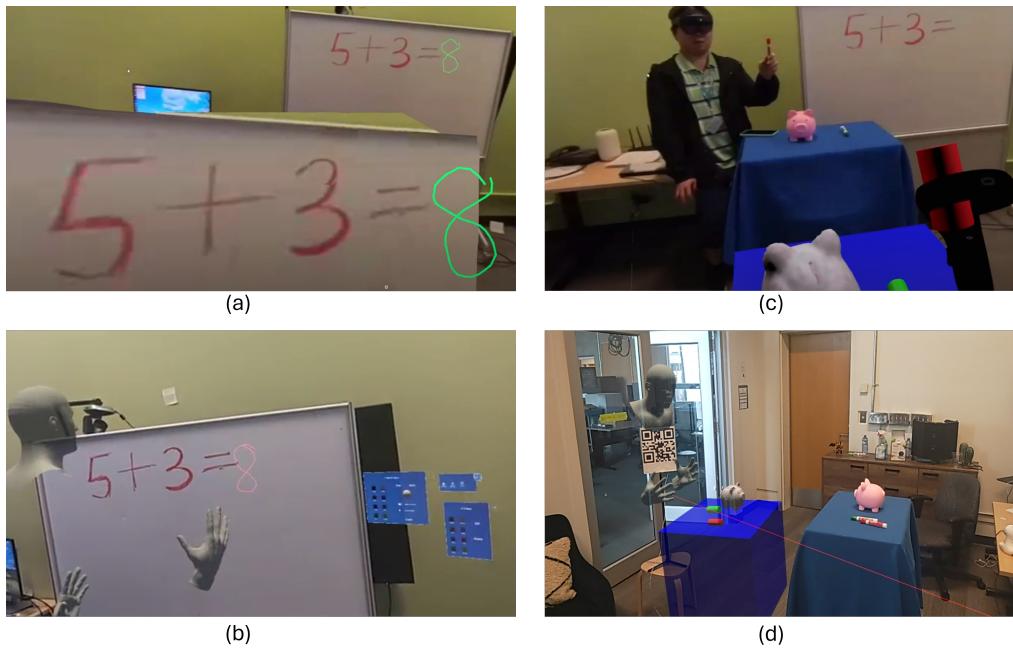


Figure 7: The remote education scenario: In (a), the VR student answers a question on the whiteboard using the pulled cutout. The AR teacher sees the answer written on the board in (b). The AR teacher also sees the VR student standing close to the board. In (c) and (d), we see an instructor and learner engaging in a crafts session. From the VR perspective in (c), virtual objects can mimic real-world items as both users hold up a red ‘marker’. (d) shows the AR instructor’s perspective.

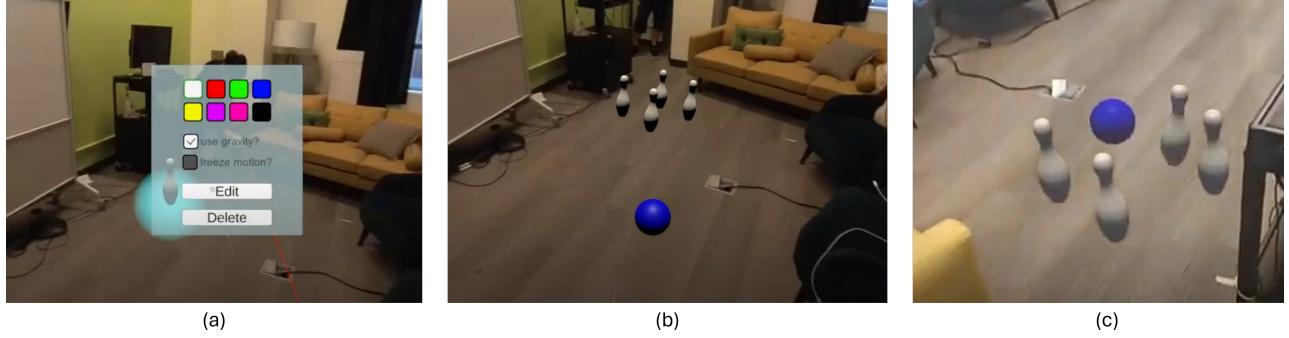


Figure 8: The shared recreational activity scenario: In (a), the remote and local users collaboratively create, edit, and arrange bowling pins with physics properties. (b) and (c) shows the users playing the remote bowling game in VR and AR.

6.3 Results and Findings

Participants generally were positive towards *VirtualNexus* (Fig. 9 and 10). Fig. 1d shows an example storyboard created in the study. We present the findings based on their responses to our questionnaires, especially those relevant to the two novel interactive techniques. A1-A7 and V1-V7 refer to the AR and VR users respectively.

6.3.1 Cutout Interactions. The VR users reported that the concept of cutting out partial environment and interacting with it was intuitive (Q6V: Mean = 4.57, STD = 0.53) – V5: “*It was a learning curve, but it became quite intuitive after a short exposure to the experience*”. V3 echoed V5 but recommended improvements in distinguishing between the cutout and the world space (as currently, the cutout space is not localized to a smaller volume). V2 indicated that the cutouts improved clarity: i.e. “*I would otherwise not be able to see what work they’re doing on the whiteboard and that would make things very difficult*”. V1 also supported this approach in terms of clarity and interaction precision: “*Pulling the miniature whiteboard closer was necessary for writing legibly with the annotation tool*.”

6.3.2 Scanned Physical Objects. Almost all users agreed (Q10A: Mean = 4.86, STD = 0.38; Q11V: Mean = 4.86, STD = 0.38) that having rapidly scanned physical objects enhanced the capability of collaboration compared to having only primitive shapes. From the AR perspective, both A2 and A3 thought it was more fun to interact with a virtual object compared to the same physical object. From the VR perspective, V5 mentioned that scanned objects better incorporate physical components from the scene when compared to only using the regular 360° video feed. Additionally, V3 praised the usefulness of having ad-hoc created replicas in the task, they stated: “*the basic shapes are not sufficient for modelling more complicated objects. (Without virtual replicas) in our task, the three characters would likely have had to be represented by geometric shapes rather than their scanned models..., potentially reducing our working efficiency*”.

7 DISCUSSION AND FUTURE WORK

Drawing from the user study results and comparing with representative prior research, here we discuss how *VirtualNexus* improves 360° telepresent collaboration and identify future work. We start with the implication of embedding a 3D reconstruction under the 360° video and reflect on the *VirtualNexus*’s interactive techniques.

7.1 Balancing 360° and 3D reconstruction

Past research has studied the trade-off between using 360° videos and 3D reconstruction for telepresence [51, 52, 54]. In 360° videos, despite the higher visual quality, remote VR users cannot move in the shared world without involving locomotive equipment such as robots [13, 21]. In contrast, telepresence systems with 3D reconstructions [51, 54] allows users to walk around the remote environment, but the reconstructed scene suffers from holes and artifacts due to imperfect scanning and occlusion.

Teo et al. [52] set out to balance this trade-off between 360° videos and 3D reconstructions in telepresence. This work allows the remote guest to switch between the 360° video and the point-cloud reconstruction of the same physical environment. However, to utilize the merits of both, the user needs to frequently context switch between two distinct sets of interactive modalities, potentially increasing the mental workload. *VirtualNexus* takes a slightly different approach: while the remote user always sees the physical environment in a high-quality 360° video, we align a transparent 3D reconstruction with the 360° video to enhance the interactivity. We believe such a design provides a more coherent interactive experience and better physical presence (Q1V from the questionnaire): In our study, we observed the remote users naturally annotating and placing virtual artifacts on remote physical surfaces, like they are physically in the remote environment wearing an AR headset.

7.2 Interactivity in 360° Video Telepresence

In the direction of improving the interactivity of 360° video telepresence, Rhee et al. [44] incorporated virtual annotations and artifacts. However, the main drawbacks of 360° telepresence remain: The inability to walk around and interact with the remote environment.

In co-located collaboration, users can freely walk around and utilize their surroundings. Such freedom is limited for remote users telepresent with 360° videos. Inspired by the concept of WiM[6], *VirtualNexus* implements the environment cutouts, which does the opposite by bringing parts of the environment toward the remote user. In our study, we found the cutouts also improved the clarity and precision: For further away areas, the remote user can pull environment cutouts closer for more precise annotation.

VirtualNexus additionally provides the remote user with better access to individual physical objects with ad-hoc creation of virtual

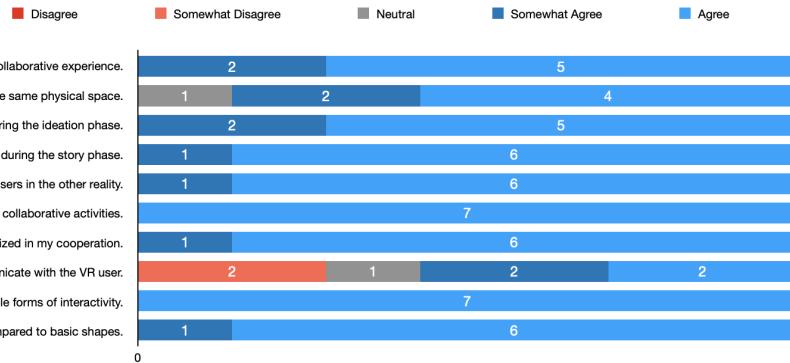


Figure 9: Results from the 5-point Likert scale questions presented to AR users.

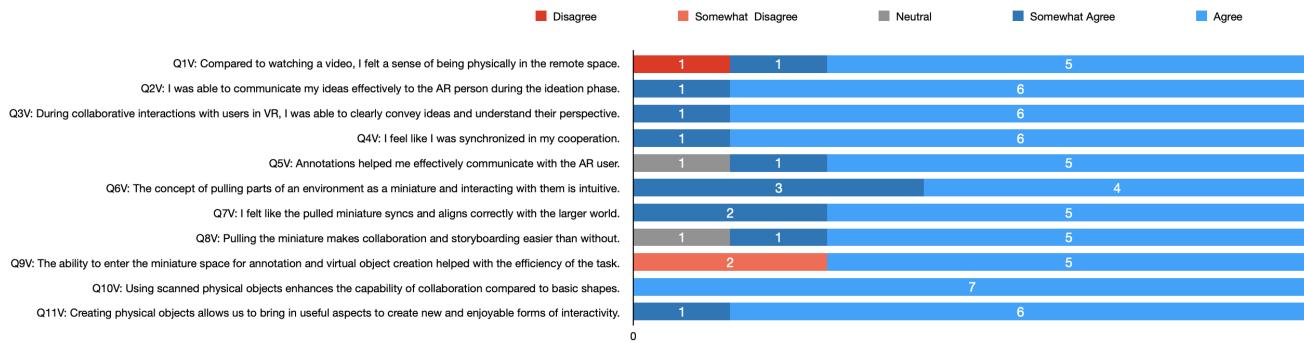


Figure 10: Results from the 5-point Likert scale questions presented to VR users.

replicas. These objects mimic reality, but take advantage of digital affordances — they can be replicated, scaled down or up, and can have different physics properties. Echoing past research on using virtual replicas for remote collaboration and instruction [8, 17, 36, 63], participants in our study found virtual replicas enhance interpretability. In contrast, representing objects as primitives adds an interpretation layer and hinders efficiency. Currently *VirtualNexus* takes 1–3 mins to create a unique virtual replica. Usability research by Nielsen [35] suggests that a response time over 10 seconds risks losing users’ attention, but can be alleviated by providing a progress bar. As we managed the virtual replica creation in a separate thread behind the scenes, we expect the users can work on other sub-tasks in parallel. For example, we observed some AR participants proceed to sketch on the whiteboard with the remote VR user. Future work can reduce the processing time for object reconstruction by replacing Colmap [45, 46] with refined HoloLens camera poses.

7.3 Awareness of the Virtual Other and Context Switching from Cutouts

Our study surfaced two future improvements for *VirtualNexus*: 1) the local AR user’s awareness of the remote VR user, and 2) context switching between the cutout space and the original environment.

AR users have more freedom to move and a smaller FoV, so they have a reduced awareness of the VR user. Such awareness is further reduced when the VR user relocates to the area they cut out

from the environment. Therefore, it is reasonable to smooth out the VR user’s relocation (e.g., by animating their motion) when they activate/deactivate the cutout. We can also provide additional cues to the local AR user when the VR user tries to communicate [39, 40].

Presently, the cutout space arbitrarily overlays the original world. The remote user needs to context switch as they redirect their focus from the cutout space to the environment and vice versa, making it hard to understand whether an object belongs to the cutout space or the original world. Future work can localize the cutout using a “snow globe” metaphor — the cutout acts as a localized miniature world: Objects outside this space are hidden, and the objects within would be more easily understood as belonging to the cutout.

8 CONCLUSION

We introduce *VirtualNexus*, a system for AR/VR collaboration that combines high-fidelity 360° video with accurate 3D reconstructions of physical environments. It supports traditional collaborative features like annotations and virtual object manipulation and adds novel features. In VR, users can create cutouts — miniature parts of the original world, while in AR, users can quickly scan and share physical items as virtual replicas. We outline applications of the *VirtualNexus* system and perform a user study with a collaborative storyboarding task. We find that the cutout system was intuitive and provided increased clarity and precision, and the scanning system enhanced the capabilities of the collaborative processes.

ACKNOWLEDGMENTS

This work was supported in part by the Natural Science and Engineering Research Council of Canada (NSERC) under Discovery Grant RGPIN-2019-05624 and by Rogers Communications Inc. under the Rogers-UBC Collaborative Research Grant: Augmented and Virtual Reality.

REFERENCES

- [1] Julie Artois, Glenn Van Wallendael, and Peter Lambert. 2023. 360DIV: 360-degree Video Plus Depth for Fully Immersive VR Experiences. In *2023 IEEE International Conference on Consumer Electronics (ICCE)*. 1–2. <https://doi.org/10.1109/ICCE56470.2023.10043369>
- [2] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. 2021. Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. *CoRR* abs/2111.12077 (2021). arXiv:2111.12077 <https://arxiv.org/abs/2111.12077>
- [3] Bill Buxton. 2009. Mediaspace – Meaningspace – Meetingspace. In *Computer Supported Cooperative Work*. Springer London, London, 217–231.
- [4] Dane Coffey, Nicholas Malbraaten, Trung Le, Iman Borazjani, Fotis Sotiropoulos, and Daniel F. Keefe. 2011. Slice WIM: a multi-surface, multi-touch interface for overview+detail exploration of volume datasets in virtual reality. In *Symposium on Interactive 3D Graphics and Games* (San Francisco, California) (*I3D ’11*). Association for Computing Machinery, New York, NY, USA, 191–198. <https://doi.org/10.1145/1944745.1944777>
- [5] Noah Q. Cowit and Lecia Barker. 2023. How do Teaching Practices and Use of Software Features Relate to Computer Science Student Belonging in Synchronous Remote Learning Environments?. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1* (Toronto, ON, Canada) (*SIGCSE 2023*). Association for Computing Machinery, New York, NY, USA, 771–777. <https://doi.org/10.1145/3545945.3569876>
- [6] Kurtis Danyluk, Barrett Ens, Bernhard Jenny, and Wesley Willett. 2021. A Design Space Exploration of Worlds in Miniature. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (*CHI ’21*). Association for Computing Machinery, New York, NY, USA, Article 122, 15 pages. <https://doi.org/10.1145/3411764.3445098>
- [7] Kangle Deng, Andrew Liu, Jun-Yan Zhu, and Deva Ramanan. 2022. Depth-supervised NeRF: Fewer Views and Faster Training for Free. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [8] Carmine Elvezio, Mengu Sukan, Ohan Oda, Steven Feiner, and Barbara Tversky. 2017. Remote collaboration in AR and VR using virtual replicas. In *ACM SIGGRAPH 2017 VR Village* (Los Angeles, California) (*SIGGRAPH ’17, Article 13*). Association for Computing Machinery, New York, NY, USA, 1–2.
- [9] Sean Follmer, Daniel Leithinger, Alex Olwal, Akimitsu Hogue, and Hiroshi Ishii. 2013. inFORM: dynamic physical affordances and constraints through shape and object actuation. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (St. Andrews, Scotland, United Kingdom) (*UIST ’13*). Association for Computing Machinery, New York, NY, USA, 417–426. <https://doi.org/10.1145/2501988.2502032>
- [10] Lei Gao, Huidong Bai, Mark Billinghurst, and Robert W. Lindeman. 2021. User Behaviour Analysis of Mixed Reality Remote Collaboration with a Hybrid View Interface. In *Proceedings of the 32nd Australian Conference on Human-Computer Interaction* (Sydney, NSW, Australia) (*OzCHI ’20*). Association for Computing Machinery, New York, NY, USA, 629–638. <https://doi.org/10.1145/3441000.3441038>
- [11] Carl Gutwin and Saul Greenberg. 2002. A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Comput. Support. Coop. Work* 11, 3 (Sept. 2002), 411–446.
- [12] Zhenyi He, Ruofei Du, and Ken Perlin. 2020. CollaboVR: A Reconfigurable Framework for Creative Collaboration in Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 542–554. <https://doi.org/10.1109/ISMAR50242.2020.00082>
- [13] Yasamin Heshmat, Brennan Jones, Xiaoxuan Xiong, Carman Neustaedter, Anthony Tang, Bernhard E. Riecke, and Lillian Yang. 2018. Geocaching with a Beam: Shared Outdoor Activities through a Telepresence Robot with 360 Degree Viewing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI ’18)*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3173574.3173933>
- [14] Teresa Hirzle, Maurice Cordts, Enrico Rukzio, Jan Gugenheimer, and Andreas Bulling. 2021. A Critical Assessment of the Use of SSQ as a Measure of General Discomfort in VR Head-Mounted Displays. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3411764.3445361>
- [15] Erzhen Hu, Jens Emil Sloth Grønbæk, Wen Ying, Ruofei Du, and Seongkook Heo. 2023. ThingShare: Ad-Hoc Digital Copies of Physical Objects for Sharing Things in Video Meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI ’23*). Association for Computing Machinery, New York, NY, USA, Article 365, 22 pages. <https://doi.org/10.1145/3544548.3581148>
- [16] Xincheng Huang, James Riddell, and Robert Xiao. 2023. Virtual Reality Telepresence: 360-Degree Video Streaming with Edge-Compute Assisted Static Foveated Compression. *IEEE Transactions on Visualization and Computer Graphics* 29, 11 (2023), 4525–4534. <https://doi.org/10.1109/TVCG.2023.3320255>
- [17] Xincheng Huang and Robert Xiao. 2024. SurfShare: Lightweight Spatially Consistent Physical Surface and Virtual Replica Sharing with Head-mounted Mixed-Reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 4, Article 162 (jan 2024), 24 pages. <https://doi.org/10.1145/3631418>
- [18] Keiichi Ihara, Mehrad Faridan, Ayumi Ichikawa, Ikkaku Kawaguchi, and Ryo Suzuki. 2023. HoloBots: Augmenting Holographic Telepresence with Mobile Robots for Tangible Remote Collaboration in Mixed Reality. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) (*UIST ’23*). Association for Computing Machinery, New York, NY, USA, Article 119, 12 pages. <https://doi.org/10.1145/3586183.3606727>
- [19] Andrew Irillitti, Mesut Latifoglu, Qiushi Zhou, Martin N Reinoso, Thuong Hoang, Eduardo Veloso, and Frank Vetere. 2023. Volumetric Mixed Reality Telepresence for Real-time Cross Modality Collaboration. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI ’23, Article 101*). Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3544548.3581277>
- [20] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, and Andrew Fitzgibbon. 2011. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (Santa Barbara, California, USA) (*UIST ’11*). Association for Computing Machinery, New York, NY, USA, 559–568. <https://doi.org/10.1145/2047196.2047270>
- [21] Brennan Jones, Yaying Zhang, Priscilla N. Y. Wong, and Sean Rintel. 2021. Belonging There: VROOM-ing into the Uncanny Valley of XR Telepresence. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 59 (apr 2021), 31 pages. <https://doi.org/10.1145/3449133>
- [22] M. Kalkusch, T. Lidy, N. Knapp, G. Reitmayer, H. Kaufmann, and D. Schmalstieg. 2002. Structured visual markers for indoor pathfinding. In *The First IEEE International Workshop Agumented Reality Toolkit*, 8 pp.–. <https://doi.org/10.1109/ART.2002.1107018>
- [23] Dorota Kamińska, Tomasz Sapiński, Sławomir Wiak, Toomas Tikk, Rain Eric Haamer, Egils Avots, Ahmed Helmi, Cagri Ozcinar, and Gholamreza Anbarjafari. 2019. Virtual Reality and Its Applications in Education: Survey. *Information* 10, 10 (Oct. 2019), 318. <https://doi.org/10.3390/info10100318>
- [24] 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* (Beijing, China) (*VRST ’15*). Association for Computing Machinery, New York, NY, USA, 217–225.
- [25] 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* (Glasgow, Scotland UK) (*CHI ’19*). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300403>
- [26] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. 2023. Segment Anything. arXiv:2304.02643 [cs.CV]
- [27] 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* (Bangkok, Thailand) (*SA ’17*). Association for Computing Machinery, New York, NY, USA, Article 14, 4 pages. <https://doi.org/10.1145/3132787.3139203>
- [28] Gun A. Lee, Theophilus Teo, Seungwon Kim, and Mark Billinghurst. 2018. A User Study on MR Remote Collaboration Using Live 360 Video. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 153–164. <https://doi.org/10.1109/ISMAR.2018.00051>
- [29] Daniel Leithinger, Sean Follmer, Alex Olwal, and Hiroshi Ishii. 2014. Physical telepresence: shape capture and display for embodied, computer-mediated remote collaboration. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) (*UIST ’14*). Association for Computing Machinery, New York, NY, USA, 461–470. <https://doi.org/10.1145/2642918.2647377>
- [30] Zhengqing Li, Liwei Chan, Theophilus Teo, and Hideki Koike. 2020. OmniGlobeVR: A Collaborative 360° Communication System for VR. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA ’20)*. Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/3334480.3382869>
- [31] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2021. NeRF: representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 1 (dec 2021), 99–106. <https://doi.org/10.1145/3444548.3581277>

- org/10.1145/3503250
- [32] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Trans. Graph.* 41, 4, Article 102 (July 2022), 15 pages. <https://doi.org/10.1145/3528223.3530127>
 - [33] Alessandro Mulloni, Hartmut Seichter, and Dieter Schmalstieg. 2012. Indoor navigation with mixed reality world-in-miniature views and sparse localization on mobile devices. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (Capri Island, Italy) (AVI '12). Association for Computing Machinery, New York, NY, USA, 212–215. <https://doi.org/10.1145/2254556.2254595>
 - [34] Michael Nebeling, Katy Lewis, Yu-Cheng Chang, Lihai Zhu, Michelle Chung, Piaoyang Wang, and Janet Nebeling. 2020. XRDirector: A Role-Based Collaborative Immersive Authoring System. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3313831.3376637>
 - [35] Jakob Nielsen. 1994. *Usability Engineering*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
 - [36] Ohan Oda, Carmine Elvezio, Mengu Sukan, Steven Feiner, and Barbara Tversky. 2015. Virtual Replicas for Remote Assistance in Virtual and Augmented Reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) (UIST '15). Association for Computing Machinery, New York, NY, USA, 405–415. <https://doi.org/10.1145/2807442.2807497>
 - [37] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Kamhis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchny, Cem Keskin, and Shahram Izadi. 2016. Holoportation: Virtual 3D Teleportation in Real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) (UIST '16). Association for Computing Machinery, New York, NY, USA, 741–754. <https://doi.org/10.1145/2984511.2984517>
 - [38] 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* (Atlanta, Georgia, USA) (I3D '99). Association for Computing Machinery, New York, NY, USA, 141–145. <https://doi.org/10.1145/300523.300540>
 - [39] Thammatip Piumsomboon, Arindam Day, Barrett Ens, Youngho Lee, Gun Lee, and Mark Billinghurst. 2017. Exploring enhancements for remote mixed reality collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications* (Bangkok, Thailand) (SA '17). Association for Computing Machinery, New York, NY, USA, Article 16, 5 pages. <https://doi.org/10.1145/3132787.3139200>
 - [40] Thammatip 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* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–17. <https://doi.org/10.1145/3290605.3300458>
 - [41] Feng Qian, Bo Han, Qingyang Xiao, and Vijay Gopalakrishnan. 2018. Flare: Practical Viewport-Adaptive 360-Degree Video Streaming for Mobile Devices. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking* (New Delhi, India) (MobiCom '18). Association for Computing Machinery, New York, NY, USA, 99–114.
 - [42] Feng Qian, Lusheng Ji, Bo Han, and Vijay Gopalakrishnan. 2016. Optimizing 360 video delivery over cellular networks. In *Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges* (New York City, New York) (ATC '16). Association for Computing Machinery, New York, NY, USA, 1–6.
 - [43] Taehyun Rhee, Lohit Petikam, Benjamin Allen, and Andrew Chalmers. 2017. MR360: Mixed Reality Rendering for 360° Panoramic Videos. *IEEE Transactions on Visualization and Computer Graphics* 23, 4 (2017), 1379–1388. <https://doi.org/10.1109/TVCG.2017.2657178>
 - [44] Taehyun Rhee, Stephen Thompson, Daniel Medeiros, Rafael Dos Anjos, and Andrew Chalmers. 2020. Augmented Virtual Teleportation for High-Fidelity Telecollaboration. *IEEE Trans. Vis. Comput. Graph.* 26, 5 (May 2020), 1923–1933. <https://doi.org/10.1109/TVCG.2020.2973065>
 - [45] Johannes Lutz Schönberger and Jan-Michael Frahm. 2016. Structure-from-Motion Revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.
 - [46] Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. 2016. Pixelwise View Selection for Unstructured Multi-View Stereo. In *European Conference on Computer Vision (ECCV)*.
 - [47] Sharad Sharma, Ruth Agada, and Jeff Ruffin. 2013. Virtual reality classroom as an constructivist approach. In *2013 Proceedings of IEEE Southeastcon*. 1–5. <https://doi.org/10.1109/SECON.2013.6567441>
 - [48] Mel Slater and Sylvia Wilbur. 1997. A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments* 6, 6 (1997), 603–616. <https://direct.mit.edu/pvar/article-abstract/6/6/603/18157>
 - [49] 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* (Denver, Colorado, USA) (CHI '95). ACM Press/Addison-Wesley Publishing Co., USA, 265–272. <https://doi.org/10.1145/223904.223938>
 - [50] Anthony Tang, Omid Fakourfar, Carman Neustaedter, and Scott Bateman. 2017. Collaboration with 360 Videochat: Challenges and Opportunities. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (Edinburgh, United Kingdom) (DIS '17). Association for Computing Machinery, New York, NY, USA, 1327–1339.
 - [51] Theophilus Teo, Ashkan F. Hayati, Gun A. Lee, Mark Billinghurst, and Matt Adcock. 2019. A Technique for Mixed Reality Remote Collaboration Using 360 Panoramas in 3D Reconstructed Scenes. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology (VRST '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. <https://doi.org/10.1145/335996.3364238>
 - [52] 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 (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3290605.3300431>
 - [53] Theophilus Teo, Gun A. Lee, Mark Billinghurst, and Matt Adcock. 2018. Hand gestures and visual annotation in live 360 panorama-based mixed reality remote collaboration. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction* (Melbourne, Australia) (OzCHI '18). Association for Computing Machinery, New York, NY, USA, 406–410. <https://doi.org/10.1145/3292147.3292200>
 - [54] Theophilus Teo, Mitchell Norman, Gun A. Lee, Mark Billinghurst, and Matt Adcock. 2020. Exploring Interaction Techniques for 360 Panoramas inside a 3D Reconstructed Scene for Mixed Reality Remote Collaboration. *Journal on Multimodal User Interfaces* 14, 4 (Dec. 2020), 373–385. <https://doi.org/10.1007/s12193-020-00343-x>
 - [55] Balasaranavan Thorai Kumaravel, Fraser Anderson, George Fitzmaurice, Björn 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* (New Orleans, LA, USA) (UIST '19). Association for Computing Machinery, New York, NY, USA, 161–174.
 - [56] Peng Wang, Xiaoliang Bai, Mark Billinghurst, Shusheng Zhang, Sili Wei, Guangyao Xu, Weiping He, Xiangyu Zhang, and Jie Zhang. 2021. 3DGAM: using 3D gesture and CAD models for training on mixedreality remote collaboration. *Multimed. Tools Appl.* 80, 20 (Aug. 2021), 31059–31084. <https://doi.org/10.1007/s11042-020-09731-7>
 - [57] Peng Wang, Yue Wang, Mark Billinghurst, Huizhen Yang, Peng Xu, and Yanhong Li. 2023. BeHere: a VR/SAR remote collaboration system based on virtual replicas sharing gesture and avatar in a procedural task. *Virtual Real.* (Jan. 2023), 1–22. <https://doi.org/10.1007/s10055-023-00748-5>
 - [58] Bob G Witmer and Michael J Singer. 1998. Measuring presence in Virtual Environments: A presence questionnaire. *Presence* 7, 3 (June 1998), 225–240. <https://doi.org/10.1162/105474698565686>
 - [59] Hsin-Kai Wu, Silvia Wen-Yu Lee, Hsin-Yi Chang, and Jyh-Chong Liang. 2013. Current Status, Opportunities and Challenges of Augmented Reality in Education. *Computers & Education* 62 (March 2013), 41–49. <https://doi.org/10.1016/j.compedu.2012.10.024>
 - [60] Robert Xiao, Julia Schwarz, Nick Throm, Andrew D. Wilson, and Hrvoje Benko. 2018. MRTouch: Adding Touch Input to Head-Mounted Mixed Reality. *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (2018), 1653–1660. <https://doi.org/10.1109/TVCG.2018.2794222>
 - [61] Jacob Young, Tobias Langlotz, Steven Mills, and Holger Regenbrecht. 2020. Mobileportation: Nomadic Telepresence for Mobile Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 2, Article 65 (jun 2020), 16 pages. <https://doi.org/10.1145/3397331>
 - [62] Kevin Yu, Ulrich Eck, Frieder Pankratz, Marc Lazarovici, Dirk Wilhelm, and Nassir Navab. 2022. Duplicated Reality for Co-located Augmented Reality Collaboration. *IEEE Transactions on Visualization and Computer Graphics* 28, 5 (2022), 2190–2200. <https://doi.org/10.1109/TVCG.2022.3150520>
 - [63] Xiangyu Zhang, Xiaoliang Bai, Shusheng Zhang, Weiping He, Peng Wang, Zhuo Wang, Yuxiang Yan, and Quan Yu. 2022. Real-time 3D video-based MR remote collaboration using gesture cues and virtual replicas. *Int. J. Adv. Manuf. Technol.* 121, 11 (Aug. 2022), 7697–7719. <https://doi.org/10.1007/s00170-022-09654-7>