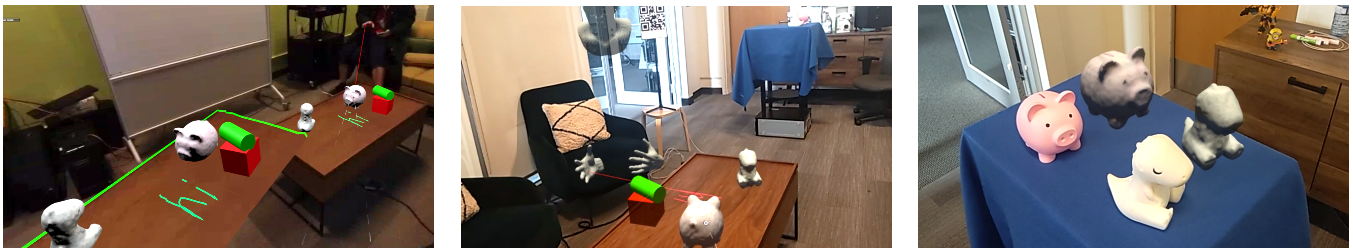
VirtualNexus: Enhancing 360° Video AR/VR Collaboration with Environment Cutout and Virtual Replicas

Anonymous Author(s)



(a) (b) (c)

Figure 1: *VirtualNexus* enhances 360° video AR/VR collaboration with environment cutouts and virtual replicas (e.g., the pig and the dinosaur). Starting from (a), the remote VR user is telepresent in the AR user's physical environment. They would like to collaborate on the desk. To have a close-up view and control on the desk, the VR user can create an *environment cutout* (the closer surface outlined in green in (a)) and pull it closer. Simultaneously in (b), the AR user sees the VR user's virtual avatar also moved closer to the desk from the camera's position (above the QRCode in (b)). The positions of the virtual annotations and objects are 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 [39]. In (c) we showcase virtual replicas of a pig and a dinosaur with their original physical copies.

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 physical environment to the remote VR user through 3D reconstructions (e.g. spatial meshes or point clouds) 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 allows the VR user to simultaneously interact with either or both high-fidelity 360° video and spatially-accurate 3D reconstructions of the physical environment. VR users define "cutouts" of the environment that they can interact with as a world in miniature, and their interactions are synchronized to the AR perspective. Furthermore, AR users can rapidly scan and share 3D virtual replicas of physical objects using neural rendering. We demonstrate our system's interaction capabilities and utility through three example applications and evaluate our system in dyadic user experiments with a collaborative storytelling task. We find that these novel features extend the interaction space of existing remote 360° collaboration systems, offering improved physical presence, versatility, and clarity in interactions.

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

CCS CONCEPTS

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

KEYWORDS

Virtual/Augmented Reality, Computer Mediated Communication

ACM Reference Format:
Anonymous Author(s). 2018. VirtualNexus: Enhancing 360° Video AR/VR Collaboration with Environment Cutout and Virtual Replicas. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 16 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Asymmetric remote AR/VR collaboration systems allow a remote VR user to be telepresent in a local AR's physical environment [18, 62, 65], 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 [61, 64] or point clouds [43, 62, 65]) or 360° videos. While 360° videos stream in higher quality, they are less interactable as they lack 3D spatial information.

Existing work has explored combining 360° videos with 3D reconstructions [61, 62]. However, they require users to switch between the 360° video and 3D reconstruction instead of harnessing their merits simultaneously. Contemporary systems [27, 30, 46] mobilize the 360° camera by attaching it to the local user's body or a robot. Although this extends the remote user's access to the physical environment (by changing their viewpoint), using a moving camera may cause simulator sickness, and robotic solutions can be less feasible for regular users. Prior research has also investigated enhancing interaction and collaboration in 360° video telepresence by additional reference cues such as gazes, gestures, ray pointers, and annotations [45, 61, 64]. However, the same enhancement has not been extended to the manipulation of objects. Lacking spatial information, virtual objects only float in front of the 360° video (instead of physically reacting with a 3D scene reconstruction), breaking the illusion of physical presence. Meanwhile, it is also challenging to incorporate physical objects to the 360° environment into the collaboration.

We present VirtualNexus, a system that enhances 360° video remote AR/VR collaboration with *environment cutouts* and virtual replicas. Our system allows the remote VR user to view a live 360° video with an embedded, invisible 3D reconstruction captured by the AR headset. The users perceive the virtual objects as a part of the video but can interact with a perceived real environment, preserving a sense of being physically present. Our system uses a stationary 360° camera but allows the 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 more precise control. In the AR environment, the VR user's avatar is virtually rendered relative to the cutout, giving the AR user awareness of the VR user's interactions and focus. All the changes a user makes in the environment cutout synchronize to the original 360° video and as overlays on the AR user's view of the physical environment. To further bridge the physical and virtual environments, we provide ad-hoc 3D virtual replica creation with Instant-NGP [39]. The local AR user can scan a physical object and obtain a sharable virtual replica within 1–3 minutes. Finally, for a coherent and complete AR/VR collaborative experience, VirtualNexus provides synchronized ray pointers, annotations, and shared virtual objects.

We demonstrate the utility of VirtualNexus through 3 application scenarios: 1) content authoring and prototyping, 2) collaborative learning and remote instruction, and 3) shared recreation activities. We evaluated our system in a user study with 14 participants, where they worked in dyads to design and create 4 3D scenes for a children's story.

We contribute a system that provides immersive features for 360° video-based AR/VR collaboration. Our system aligns a spatially-accurate 3D reconstruction of the physical environment, which enhances the physical presence of a remote VR user viewing the 360° video. Besides traditional collaborative techniques such as ray pointers, annotations, and shared virtual objects, we introduced two novel features: 1) a live environment cutout that allows the remote VR user better access to out-of-reach areas in the 360° scene, and 2) ad-hoc virtual replica creation that allows both users to incorporate physical objects in their collaboration. 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.

2 RELATED WORKS

Telepresence immersively brings a remote guest to a local user's physical environment [22, 47, 48, 60]. It has been a longstanding area of research, especially in the context of AR/VR remote interaction [23, 25, 62, 64]. To display the physical environment to the remote VR user, prior research has explored 3D reconstruction (i.e., textured spatial meshes [1, 45] or point clouds [43, 62]) and 360° videos [27, 34, 36, 46]. 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 [43, 65]. However, compared to 360° videos, real-time 3D reconstruction typically has lower quality, and it is difficult to cover a full scene without holes and occlusion. Holoportation [43] implements a pipeline that can stream high-quality full-scene reconstruction in real-time, but it requires high-end sensors, computing, and network infrastructure which are prohibitively difficult for use by regular users. In comparison, telepresence with 360° video provides higher quality (commodity 360° cameras are around $500 and can stream up to 6K videos) and thus better presence and immersion [57, 72]. 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 with 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. [62] proposed toggling between the modes of using 3D reconstruction or 360° video. However, the need to switch between two different media prevents seamlessly harnessing the merit of both simultaneously. The authors also reported that sudden changes in perspectives and interactive modalities when switching between modes may be hard to adjust to. Teo et al. proposed follow-up works [61, 64] 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 [64] or live 360° insertion [61] 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 Applying Interactive Modalities to 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 [30, 46, 61, 62, 64] by mounting a 360° camera to the local user's head or shoulder, 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 virtual reality [20, 46]. 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 [27] mounts a 360° camera on a locomotive robotic agent, allowing the remote user to explore the streamed space freely. However, besides also suffering from potential simulator sickness, using a robotic agent is too bulky and costly for regular users.

2.2.2 Worlds in Miniature. Worlds in Miniature (WiM) [9] is a miniaturized representation of an entire or part of a physical or virtual world. The most common use of WiMs is navigation [28, 40], but prior research has extended their capability to manipulate virtual environments [6, 9, 58]. Similar to manipulating a Voodoo doll [44], synchronizing a user's inputs to a WiM with the larger world allows them to manipulate regions that are out of their reach. However, 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 [31, 45]. The ability to reference and augment the virtual world enriches the task and collaboration space of remote communication [3] and facilitates group awareness [17]. Prior research has enhanced 360° video collaboration with the use of gaze cue, ray pointers, and virtual annotations in 360° videos [61–64]. However, enhancing virtual object manipulation in 360° remote collaboration has not been well explored. 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 texture, instead of lying on a physical surface), hindering the sense of being physically present for the remote user. Rhee et al. [51] 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 a unique image-based lighting technique for 360° videos [50]. 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 [24] or deformable interfaces [14, 35]. 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 [12, 42]. However, most prior work requires virtual replicas to be created in advance with CAD tools [12, 42, 69, 70, 76] or only support creating from 2D contents or sketches [18, 21, 23]. Depth-based methods such as Kinect-Fusion [26] can quickly reconstruct an object or a scene. More recently, Neural Radiance Fields (NeRF) [2, 11, 37, 39] allow object and scene reconstruction with high quality. Notably, Instant-NGP [39] introduced a hashed encoding technique that 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, we propose concepts and designs for VirtualNexus as follows. We first focus on the rationale for the system design and detail the implementation in Section 4.

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 offered by the 360° video. As most state-of-art AR headsets (e.g., Microsoft HoloLens [1], Magic Leap [2]) always maintain a spatial map in the background, the process of creating and aligning a spatial reconstruction should be simple to the users to allow for low-friction setup experience.

3.2 Enhancing Access to Environments: Interactable Environment Cutouts

Prior work either has a side effect of simulator sickness or requires a robotic solution to enable a remote telepresent user to explore the streamed space in 360° videos. However, 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

[1]https://www.microsoft.com/en-ca/hololens
[2]https://www.magicleap.com/

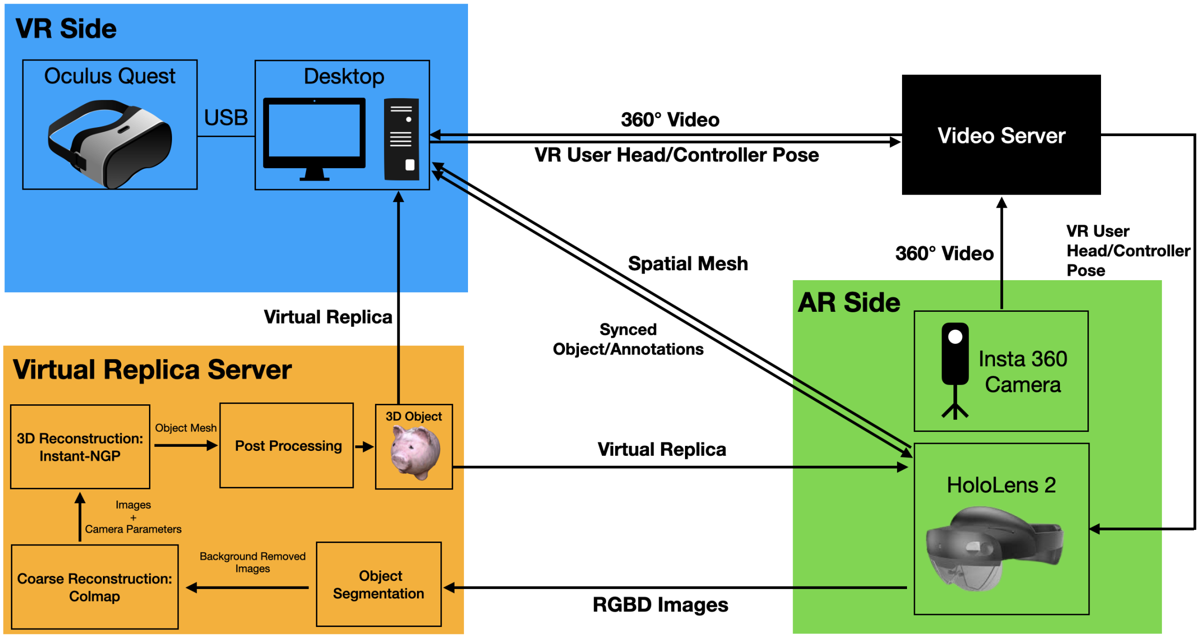


Figure 2: System Architecture: The system contains four major components: VR side, AR side, video server, and virtual replica server. VR side receives 360° video from the AR side via the video server. The AR side shares synced objects and annotations with the VR side and sends scanned RGBD images to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

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 creating prototypes rapidly [12, 76]. VirtualNexus incorporates ad-hoc creation of 3D virtual replicas in remote AR/VR collaboration. Our system allows the AR user to conveniently set an object on a platform, scan around it, and obtain a virtual replica shared between both users. VirtualNexus additionally stores the scanned virtual replica, enabling a "scan once, create many" experience.

3.4 Spatially Aligned and Synchronized Collaboration

Co-location in a spatially aligned and synchronized environment is a core premise for most 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, 17]. 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 when necessary, 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 SYSTEM IMPLEMENTATION

4.1 System Architecture and Apparatus

We implemented VirtualNexus mainly with Unity 2021.3.20f1. As our system has an asymmetric setting, it can be configured to build to either a VR or AR application. VirtualNexus uses Microsoft HoloLens 2 for AR and Meta Oculus Quest 2 for VR. In the physical space of the local user, we use an Insta360 X3[3] 360° camera that omnidirectionally streams the 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 [22] on a desktop machine with an Intel Core i7-9700K@3.6GHz 8-core CPU, 32GB of memory, and an NVIDIA GeForce RTX 2060 GPU. We attached QR codes to the front and back of 360° camera's tripod as the spatial anchor, aligning the VR world's origin with the 360° camera's lenses. Thus, the local and remote users share ray pointers, virtual annotations, and virtual objects in a spatially aligned manner. The local AR user also sees a virtual avatar overlay on the 360° camera with synchronized head and hand poses of the remote VR user. The virtual avatar moves around the local space once the remote VR user pulls out and interacts with an "environment cut-out". Finally, VirtualNexus runs virtual

[3] https://www.insta360.com/product/insta360-x3

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. We summarize the system architecture of VirtualNexus in Figure 2 and detail the implementation of the key system features in the following sections.

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 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 mapping [4]. We show the alignment between the spatial meshes and the 360° video in Figure 3. The HoloLens creates spatial reconstruction rapidly as the user moves around the space, allowing us to keep the spatial mesh synchronization process seamless for both AR and VR users.

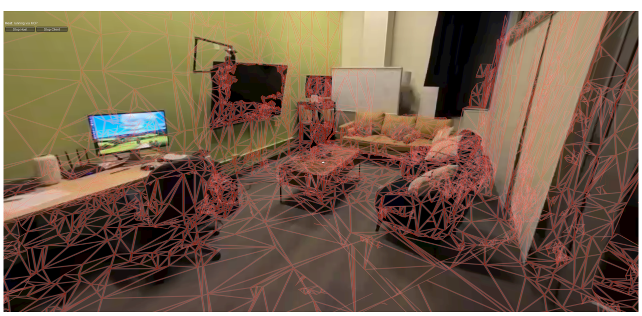


Figure 3: Spatial-accurate Alignment of 3D Reconstruction: we align the spatial mesh produced by HoloLens with the 360° video. Here we colour the edges of the spatial mesh in red for demonstrative purposes. Note that there is a hole in the TV as some black screens can be partially infrared absorbent.

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. While our work focuses on the interactive aspects of 360° remote collaboration, we also discuss how this monocular rendering impacted users and potential improvements in Section 6 and Section 8.

[4]https://docs.opencv.org/3.4/db/d58/group__calib3d__fisheye.html

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. Our system transforms the positions into the coordinate system of the opposite reality before they are sent, and then renders the ray pointers using Unity's line renders in red.

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 the controller button 'B' 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. All the annotations are implemented with Unity line renderers. 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) by pressing the left grip button. In the near annotation mode, a green sphere is rendered at the position of the poke pointer to indicate the status of annotation (see Figure 4).

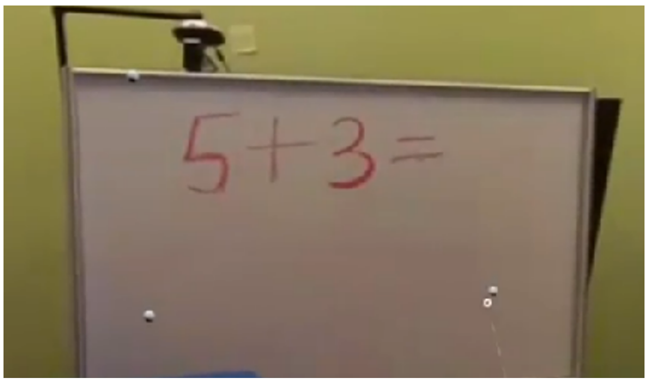
4.4.2 Shared Virtual Objects. Virtual objects appear in the same location in the world for both the AR and VR user and can be freely controlled by either user via physical grab interactions. Users can also choose to edit their physics properties, their material, etc. via virtual interactions. Motions and edits are fully synchronized between sides by using a server-client setup built atop the Mirror Networking library[5], using the VR application as the host and the AR application as the client.

Additionally, we provide a menu for virtual object creation in both AR and VR environments — users can select the specific object they want to spawn (see Figure 6(a)), its colour, and its physics

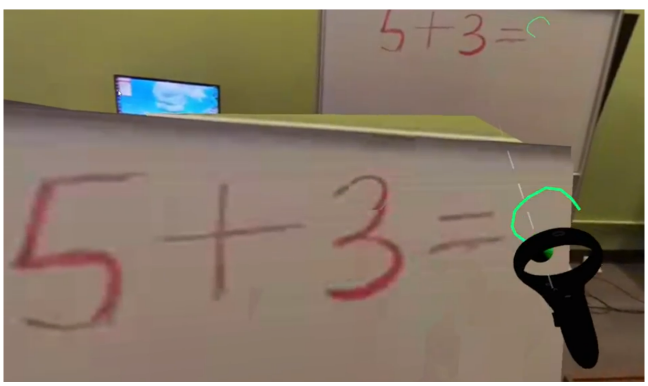
[5]https://mirror-networking.com/

properties, then press a button to spawn it. A similar menu is shown when editing or deleting objects (see Figure 8(a)).

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 virtual replicas. The AR user moves their head around the object to scan it; the RGB and depth data frames are then fed to our Instant-NGP [39] based mesh reconstruction pipeline discussed in Section 4.6, which outputs an object mesh and associated texture. This data is sent to the AR and VR environments, becoming new objects that can be spawned through the existing creation menus.



(a)



(b)

Figure 4: In (a), the VR user defines a cutout of the whiteboard through four raycasted points onto the mesh (indicated by the white dots). In (b), annotations done on the cutout when it is active are reflected back to the original location.

4.5 Environment Cutouts

To define an environment cutout, the remote VR user first makes a selection of 4 points through their ray pointer; each point is then raycast onto the depth mesh (Figure 4(a)). These raycast points along with the camera position define a pyramid-shaped selection volume which is slightly enlarged to capture all relevant vertices. Then, for each of the depth mesh objects within the scene (as the depth mesh is not a singular continuous object), we iterate through its vertices, checking whether they lie in the selection volume. We then subset each mesh object by selecting any triangle with a vertex in the selection volume, producing a new set of mesh objects that define the cutout.

The cutout supports standard VR manipulations, such as grabbing, translation, rotation, 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 7(d)). We employ a custom shader for the cutout's meshes to provide a selection highlight (gray for inactive, green for active) and to render the appropriate part of the 360° video texture on the mesh.

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 actually see two objects — one corresponding to the world space (the "original object"), and one relative to the cutout ("copy object"), for example, in Figure 4(b). 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 Rapid Virtual Replica Creation with Instant-NGP

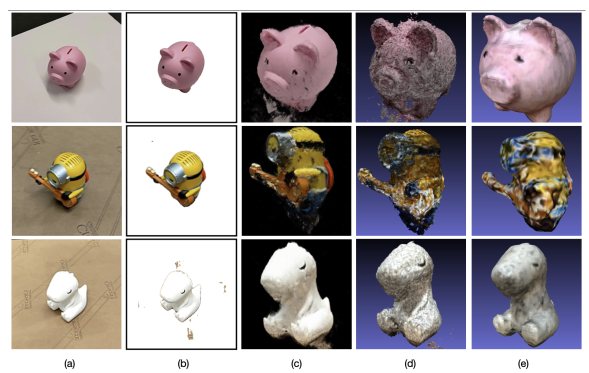


Figure 5: Intermediate results for rapid virtual replica creation: (a) one of the images from the multi-view 2D image set (b) background removed image (c) Instant-NGP created volumetric rendering object (d) Mesh object created from cube-marching (e) Voxelized and smoothed final object.

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 mostly from a futuristic standpoint: NeRFs have shown promise in producing highly photorealistic scans of objects and scenes with high-quality lighting and texture, and we feel that future object scanning pipelines may involve such technologies for fidelity, rather than traditional mesh-reconstruction pipelines like KinectFusion [26]. However, for compatibility with existing mesh-rendering pipelines, we produce both a NeRF model and traditional vertex-coloured mesh from our object scanning pipeline. Comparisons of the NeRF model and cube-marched mesh can be found in Figure 5.

To the best of our knowledge, our system is the first to adapt NeRF reconstruction for remote AR/VR collaboration. Among the variants of NeRF scene reconstruction techniques, Instant-NGP [39] 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. However, reconstructing full scenes with sufficient fidelity would take more than an hour, and so we scope VirtualNexus's reconstruction capability to individual objects. As we require a user to walk around the targeted object, our pipeline is best suitable for smaller objects that can be placed on a desk (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 Figure 2). The VR and AR applications connect to the server over TCP. The server is implemented in Python and incorporates Instant-NGP's Python API [6]. Examples of intermediate results at each stage of the pipeline can be seen in Figure 5.

4.6.1 *Colour and Depth Image Capturing.*

To scan an object, a user triggers the scanning function in the AR application and then walks around the targeted object. Our system renders a semi-transparent grey rectangle to help the user center the object in their field of view.

We capture colour and depth images of the object using the native Universal Windows Application API [7] and HoloLens 2's Research Mode API [8] at 5 FPS for about 15 seconds, accumulating around 75 images. For each frame, we reproject the depth image (built as a depth mesh) 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 mainly used for background segmentation and are also used to improve the efficiency and quality of the NeRF reconstruction [11].

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 [74]. We select all non-planar points as the initial 'coarse mask' of foreground pixels. Subsequently, we obtain a refined segmentation mask from Segment-Anything [32]. We average the coarse mask to obtain a point as the prompt to Segment-Anything [32] (i.e., indicating the targeted object's position). Segment-Anything [32] then outputs a hierarchy of masks, and we use the one that best overlaps with the coarse mask as the final segmenting mask applied to the colour image. Background-segmenting each frame takes about 0.3 seconds, and thus about 25 seconds for 75 frames.

[6]https://github.com/NVlabs/Instant-NGP
[7]https://learn.microsoft.com/en-us/windows/uwp/audio-video-camera/process-media-frames-with-mediaframereader
[8]https://github.com/microsoft/HoloLens2ForCV

4.6.3 *Colmap and Instant-NGP.*

Before providing the images to Instant-NGP[39], we need to obtain the camera poses for the images. Initially, we tried to use the HoloLens's reported camera poses for each frame directly, but found that the poses were not accurate enough for satisfactory reconstruction. Therefore, we used Colmap [53, 54], 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.

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 API[9] 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

VirtualNexus describes a general system that facilitates collaboration in mixed reality space. In this section, we outline applications of our system to practical scenarios in various domains.

5.1 Content Authoring and Prototyping

Collective content authoring is a key domain in which collaborative mixed reality research has been applied [23, 41]. VirtualNexus facilitates such real-time content authoring through the creation, sharing, and manipulation of virtual objects and annotations on both local and remote ends. 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 of the environment closer.

For example, in Figure 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 scanned object 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 (which will be described later in Section 6), which involves collaboratively building a storyboard.

[9]https://docs.blender.org/manual/en/latest/modeling/modifiers/generate/remesh.html, https://docs.blender.org/api/current/index.html

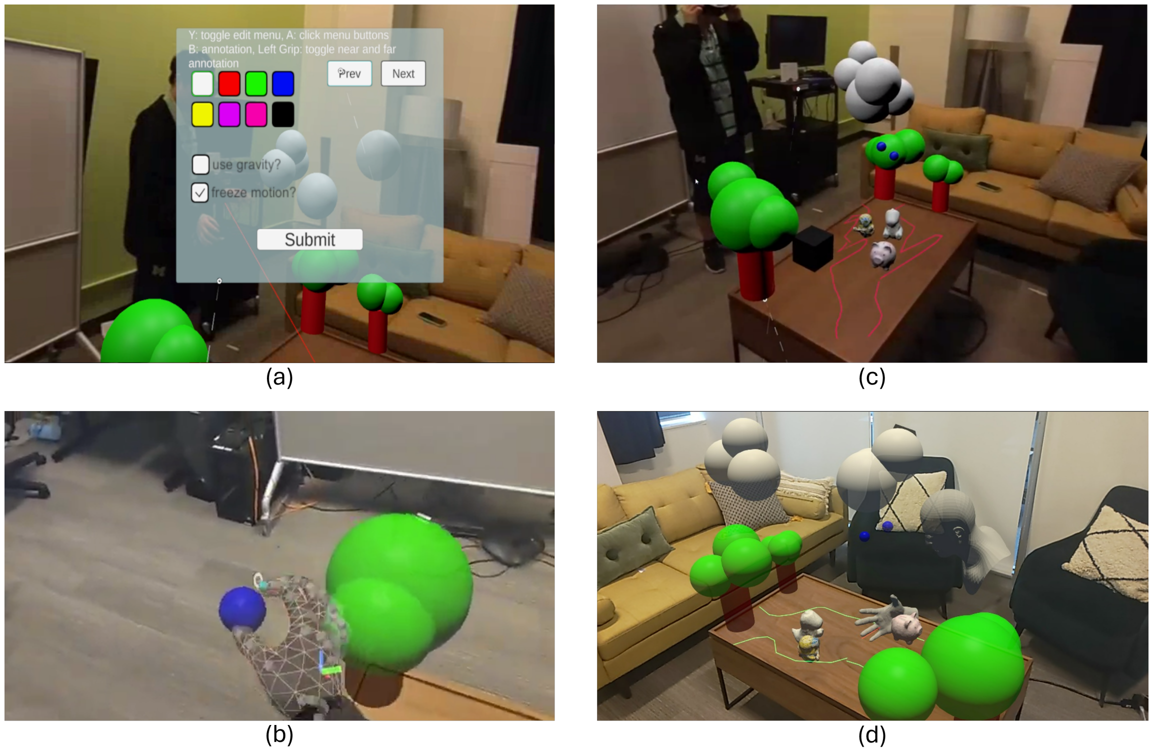


Figure 6: The remote VR and local AR users work collaboratively to create a story scene in which 3 characters walk through a forest. In (a), the VR user uses the create menu to generate new objects for the scene, which the AR user can also interact with, as shown in (b). Subimages (c) and (d) show the completed scene from the VR and AR perspectives respectively. The virtual characters are pre-scanned objects that have been added for use.

5.2 Collaborative Learning and Remote Instruction

Education and training are domains in which mixed reality systems have been heavily applied [29, 73]. Online learning platforms have become increasingly important in an increasingly digital world, and such platforms facilitate communication between educators and students even across long distances [7]. The spatial presence and immersion provided by mixed reality systems have lent further enticement towards its application in virtual classroom environments [55]. 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 (Figure 7(a) and 7(b)). The teacher first annotates a question on the whiteboard. The student might find the whiteboard to be too far to interact with suitably and precisely. In real life, the student may then walk up to the whiteboard to write their answer. However, using VirtualNexus, the student can do the opposite — bring the whiteboard closer by defining an environmental cutout of it. By doing so, the student then annotates their answer on this closer cutout, which is then reflected on the original whiteboard and seen by the teacher and audience.

To extend this education scenario, the teacher might ask students to mirror their interactions with virtual objects, e.g. in a virtual hands-on lesson (Figure 7(c) and 7(d)). To illustrate, we outline an arts-and-craft exercise in which the teacher is teaching about modelling and colouring, and wants the student to follow 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, there may be required objects necessary for the task that the student might not have (i.e. the model pig). To distribute this object digitally, the teacher can scan the object locally and create a replica that quickly becomes available to the remote students. A student can then use these virtual objects to replicate the teacher's instructions.

5.3 Shared Recreational Activities

Mixed reality mediums have seen heavy use in games and other recreational and entertainment-related domains. Using VirtualNexus, we can develop collaborative recreational activities that use virtual objects in co-located spaces. Such activities could take advantage of the spatial movement in AR/VR to encourage exercise, while also using the shared collaboration aspect of VirtualNexus to encourage socialization. We illustrate such an example using a bowling game, situated on a virtual alley overlaid on the observed local environment (Figure 8). To start, users can create virtual bowling pins, which both the remote and local users can move to the correct position on the floor of the environment. Then, either user can create a ball, which can be rolled at the pins. By taking turns

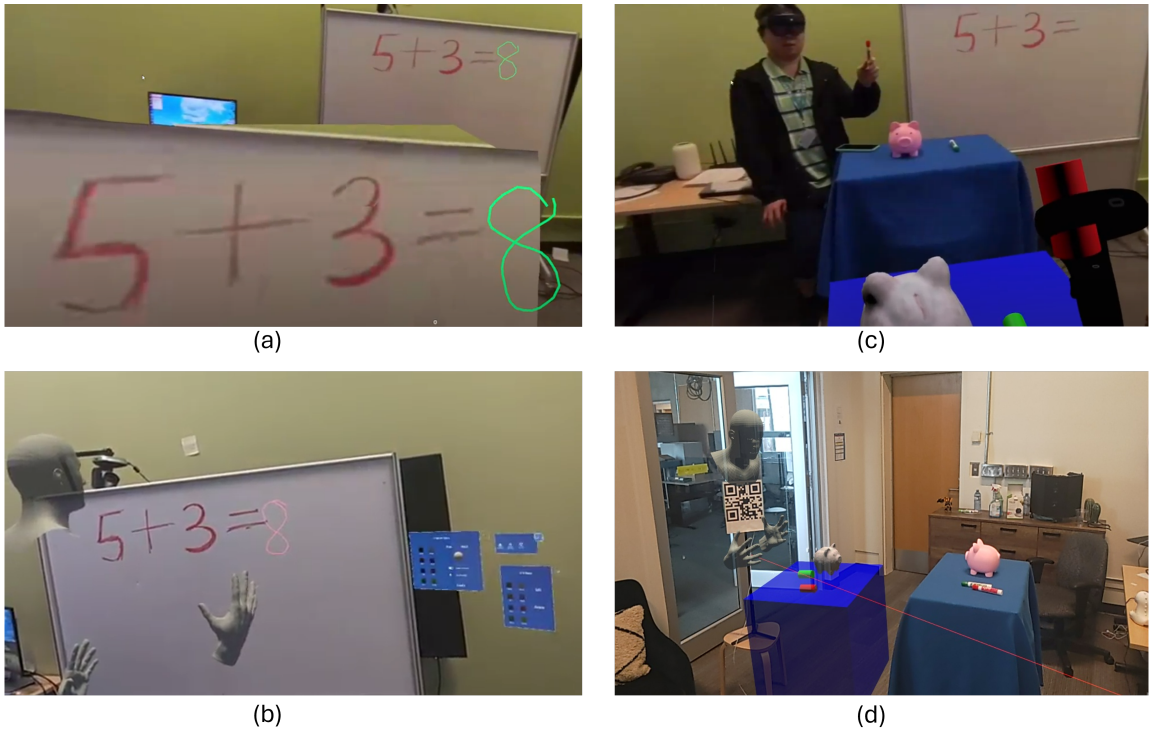


Figure 7: In (a), the VR student answers a question that the AR teacher has written on the whiteboard using the cutout. The AR teacher sees the answer written on the board in (b). The AR teacher also sees the VR student standing close to the board, as when the cutout is active, the avatar's position is shown relative to the cutout world space. 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'. Finally, (d) shows the perspective of this virtual classroom from the AR instructor's perspective.

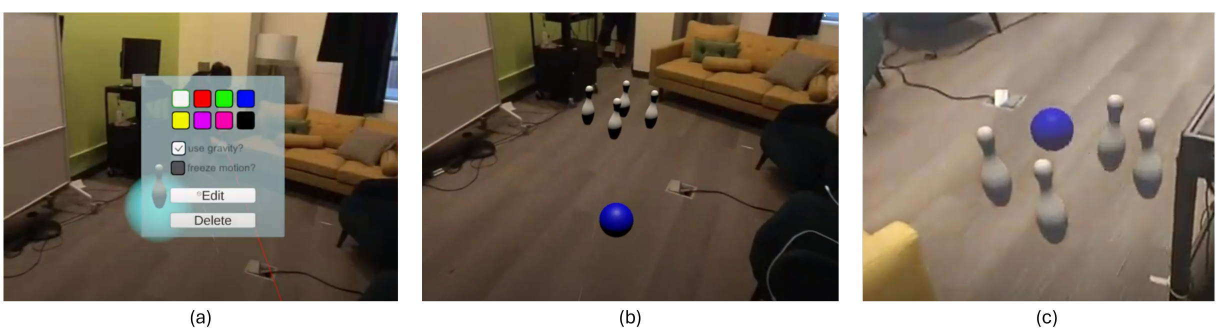


Figure 8: In (a), the remote and local users collaboratively create, edit, and manipulate bowling pins into the correct positions with the correct physics properties. After the pins have been set up, the VR user can create a ball and roll it at the pins to knock them down. Subimage (b) shows this from the VR user's perspective; subimage (c) is from the AR user's perspective.

rolling the balls, the users can experience a fun virtual bowling session situated in a physical environment.

6 USER STUDY

To understand how VirtualNexus affects task performance, immersion, and presence during collaboration between a local and remote user, we performed a dyadic user study. Users were introduced to the system through a tutorial; we then gave them more free rein in performing the collaborative task of developing a virtual storyboard for a provided narrative.

6.1 Participants

Our recruitment methods were primarily based on convenience sampling. We recruited a total of 14 participants from the local institute, forming 7 dyads. The average age of the participants was 24.8 and the gender distribution was 8 females, and 5 males (1 participant reported N/A for both questions). All participants had at least 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 headsets in the past (13 out of the 14 participants); experience with AR headsets was more rare (8 out of the 14 participants indicated some prior experience). Our study was reviewed and approved by the institutional ethical board.

6.2 Study Protocol

Upon arriving at the lab, the pair of participants were first introduced to the VirtualNexus system — the lead researcher provided a high-level briefing on the goals of the system, the various features, and the novel interactions. Participants also reviewed and signed a consent form pertaining to the study. Then, we moved on to the formal user study, which consisted of several components. The entire study took approximately 90 minutes in total, and participants were reimbursed $24 CAD.

6.2.1 Tutorial. The local AR user underwent a tutorial regarding interactions on the Microsoft HoloLens, before being introduced to the specific VirtualNexus system. At this point, the participants were separated; the local AR user was escorted to a prepared space, and the remote VR user was kept in the lab. Within the VirtualNexus system, the AR user, with the help and guidance of one of the researchers, engaged in a walkthrough of the Instant-NGP scanning process, the virtual object creation process, the annotation features, etc. Simultaneously, on the remote VR user's side, another researcher provided a tutorial on the features of interacting with a cutout, near and far annotations, virtual objects, etc.

6.2.2 Task: Create a Virtual Storyboard for a Narrative. After the participants had been familiarized with the system, they moved on to the main task of creating a virtual storyboard for a given narrative, which described the journey of several characters in an epic adventure. This collaborative task would take the users through all the features of the VirtualNexus system that they had learned in the tutorial. When introducing the task, we encouraged the participants to first discuss and ideate the scenes through annotations, and then create them using virtual objects (including scanned objects that were more representative of the characters of the story). In the interest of time, the rapid scanning process was illustrated through a single object in the local environment; other required objects were pre-scanned and sent to the system through a script. After this introduction, the task was performed largely independently by the dyadic pair. Researchers were present at both sites to provide guidance and assistance when necessary. We continued the task until the users had finished creating enough scenes, or when we reached the allotted time. Some examples of the final created virtual scenes can be seen in Figure 9(a) for the remote VR user, and in Figure 9(b) for the AR user.

6.2.3 Questionnaire. After completing the prior task, 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,

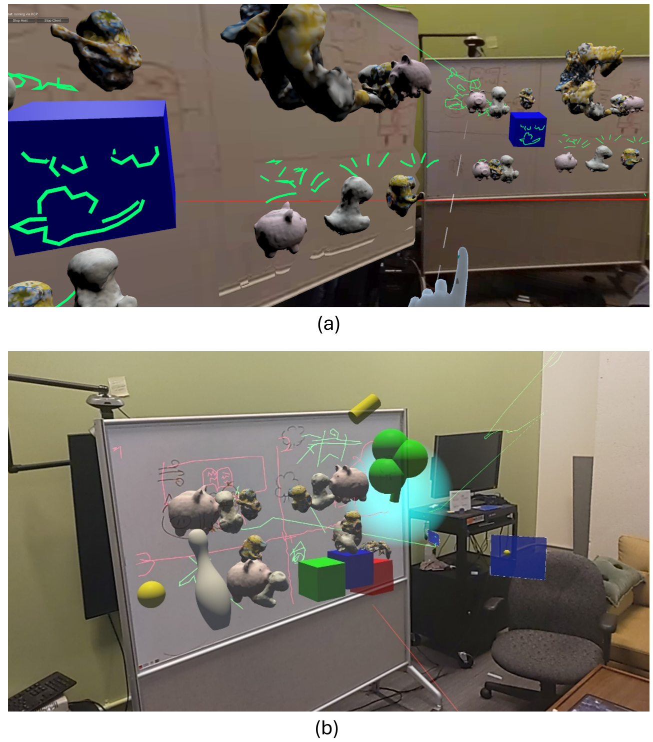


Figure 9: The final storyboards from two different study sessions using the various features of VirtualNexus from the (a) VR and (b) AR perspectives. Both storyboards split the whiteboard into 4 quadrants to create a 3D mixed-reality comic.

as well as an optional open-ended field for each question for participants to expand on their selection (see Figures 10 and 11). These questions were different for the AR and VR users to reflect the varying features.

6.3 Results and Findings

Participants generally were positive towards the VirtualNexus experience. We present the following findings based on their quantitative and open-ended responses to our questionnaires. We use A1-A7 to represent the AR users and V1-V7 to represent the VR users.

6.3.1 Cutout Interactions. The VR users reported that the concept of cutting out parts of the environment and interacting with it was intuitive (Q6V: Mean = 4.57, STD = 0.53) — V5 commented that: *"It was a learning curve, but it became quite intuitive after a short exposure to the experience"*. V3 agreed that the creation of the cutout was intuitive and easily understood but recommended improvements in distinguishing between the cutout space and the world space (as currently, the cutout space is not localized to a smaller volume). V2 indicated that the cutouts were a creative solution towards having increased 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*

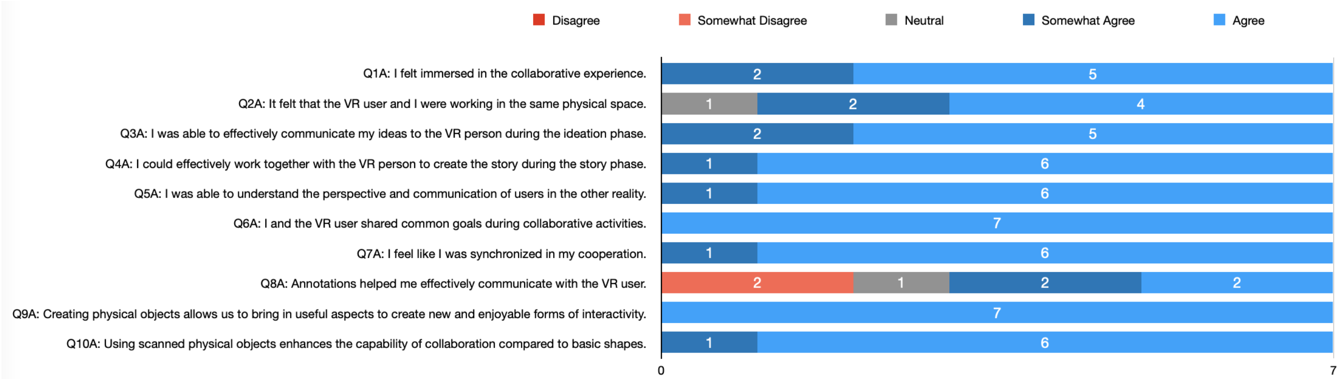


Figure 10: Results from the 5-point Likert scale questions presented to AR users.

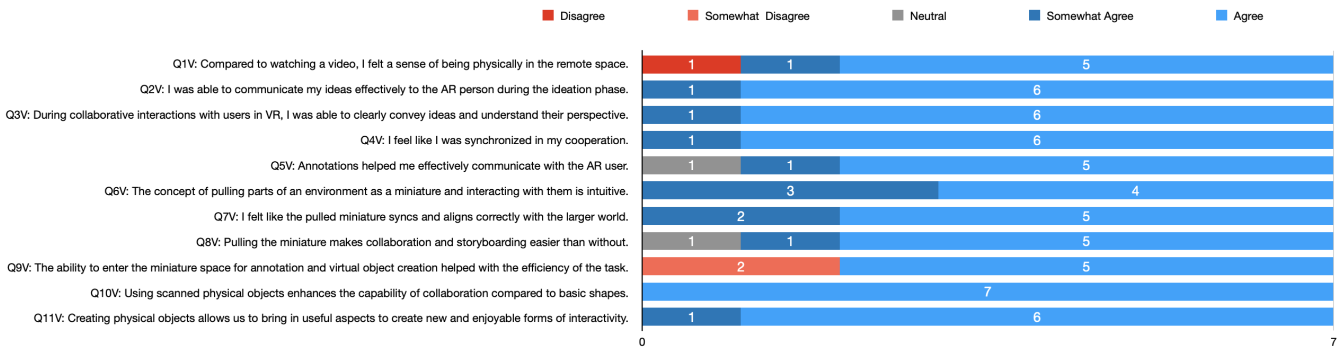


Figure 11: Results from the 5-point Likert scale questions presented to VR users.

was necessary for writing legibly with the annotation tool. Moving it further away allowed me to see the entire picture".

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 when 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. Commenting on the study task itself, V2 stated that they enjoyed storyboarding a narrative in 3D as it felt more realistic. 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. This would have created an extra interpretation layer, potentially reducing our working efficiency".*

6.3.3 Annotations. Though both annotation using the physical marker and virtual annotation were provided to the AR users, participants had middling reactions towards using the virtual annotations (Q8A: Mean = 3.75, STD = 1.27). The AR users reported that, while virtual annotation was a useful feature, it was sometimes difficult to achieve stable drawing while holding their hands in the air (possibly relating to arm fatigue for mid-air interactions [4, 49]), especially when compared to simply using a physical marker. On the other hand, remote VR users (who needed to use virtual annotation to draw in the virtual world) were more positive and indicated that the annotation feature helped them effectively communicate their ideas (Q5V: Mean = 4.57, STD = 0.79). In particular, VR users appreciated the annotation function when used in tandem with cutouts, allowing them to utilize near-annotation which provides more precise control. V3 mentioned that: *"The annotations were very helpful for drawing complex shapes and writing text... while the long-range annotation tool had a rather low sampling rate resulting in jagged drawings, the close-range tool was smooth and allowed me to write intelligibly and draw acceptable drawings."*

6.3.4 Presence and Co-Location. Our study simulated a scenario in which a remote VR user was teleported to the physical space of a local AR user. Overlaying the virtual content with the physical scene enriched the AR users' immersion in the collaborative experience (Q1A: Mean = 4.71, STD = 0.49). Our findings revealed that presence can be achieved through interactive methods, instead of solely through direct attention and context cues as in prior research. Under this paradigm, A1 and A3 mentioned that synchronous manipulation of the shared objects afforded the feeling that they were collaborating in the same space. On the other side, VR users generally agreed they felt physically present in the remote space (Q1V: Mean = 4.28, STD = 1.50). Echoing how interactive methods can afford presence, V3 felt that the ability to drag 3D objects around the room enhanced the feeling of being present, but they wished that they could also physically affect the objects in the local environment.

6.3.5 Effect on Collaboration. Overall, both AR users and VR users reported relative ease in mutually sharing and comprehending ideas during the ideation (Q3A: Mean = 4.71, STD = 0.49; Q2V: Mean = 4.86, STD = 0.38) phase. Furthermore, we found that both AR and VR users were able to understand the perspectives and ideas of the user in the other reality (Q5A: Mean = 4.86, STD = 0.38; Q3V: Mean = 4.86, STD = 0.38) and both sides felt synchronized during cooperation (Q7A: Mean = 4.86, STD = 0.38; Q4V: Mean = 4.86, STD = 0.38). We found that our system offered different strengths for VR and AR users during collaboration. VR users focused well on tasks that benefited from having a broader, more macroscopic perspective. For example, V7 appreciated the fact that they could see what the AR side was doing so they could provide real-time feedback during collaboration. However, we did observe occasional challenges regarding depth perception — V2 reported challenges in gauging distance between virtual objects and the 360° environment from the VR side. Additionally, we observed some users (V3, V7) were able to utilize collision between virtual objects and the hidden spatial reconstruction (e.g., when placing an object on the desk or whiteboard) as a secondary depth cue.

AR users had better spatial awareness as they could move freely in the physical room and view the scene from different perspectives. We observed that in some dyads, the VR users started to rely on the AR users for microscopic manipulation. They would use the synchronized ray pointers to refer to objects for the AR user to fine-tune the pose and scale of the objects.

7 DISCUSSION AND DESIGN INSIGHTS

We contextualize the insights drawn from our results within the findings of prior related works. We discuss how our work fits into prior research, and how this leads to possible design insights and open questions for the future.

7.1 Tying Together the Real and the Virtual

Milgram et al. first introduced the seminal concept of the reality-virtuality continuum, categorizing mixed-reality display systems based on a three-dimensional taxonomy of 1) the extent of world knowledge (the extent to which the real world is modelled), 2) reproduction fidelity (level of reproduction of the real world), and 3) the extent of presence metaphor (the extent to which the display is similar to viewing reality) [38, 56]. Zhang et al. related this concept to cross-reality systems, which allow multiple users to interact in co-located space with different degrees of virtuality (as in VirtualNexus) [75]. In particular, their work examined how users perform object manipulation tasks differently depending on the level of virtuality. Thus, task performance is one metric affected by the reality-virtuality continuum; Davis et al. also considered how factors of immersion and sense-being are affected by it [10].

VirtualNexus presents two novel interactions that create a shift on the reality-virtuality continuum when compared to prior mixed-reality interaction works. The first is the scanning of physical objects to create virtual replicas. These objects form mimicked representations of reality, yet take advantage of the affordances of the digital environment — they can be replicated many times, they can be scaled down or up, and they can have different physics properties. Our results showed that the ability to create such manipulable objects extended the range of interactive possibilities to both users. In particular, it helped with virtual interpretability, as representing objects as primitives would add an interpretation layer which can be confusing and less efficient. This contrasts past works such as ColabAR [67], which rely on physical substitutes for representation.

Secondly, the cutout functionality of VirtualNexus allows remote users to create virtual miniatures of the real environment. Once again, this took advantage of the reproduced real world from the remote user's perspective combined with the manipulation affordances of a virtual system. The VR user can see and interact with what appears to be a mimicry of the real world, but can take advantage of virtual features such as moving this "world" closer, scaling it, annotating on it, etc; such interactions that would not be possible in a solely physical environment. Ultimately, VirtualNexus supports the immersion and presence of the remote user of being in a real environment but takes advantage of digital interactions to augment the experience in terms of interpretability and clarity, offering new options regarding collaborative interaction.

7.2 Balancing 360° and 3D reconstruction

The present implementation places the remote VR user into a singular, immovable (but rotatable) camera perspective, dependent on the location of the 360° camera. There has been research into reconstruction from 360° video, e.g. by Da Silveira et al. [8]. In VirtualNexus, we use environmental depth data collected by the HoloLens as the input for digital reconstruction, overlaid on top of the 360° video skybox. The difference in interactions between the remote user being in a 3D reconstructed environment of the local environment versus being in an overlaid 360° video has been studied in past research, especially in the work of Teo et al. [62, 64]. Their research as well as our own observations reveal advantages and disadvantages for each. For 360°, there are limitations in precise interaction and inflexibility for spatial interactions, which are made up with through stability in video quality. In short, remote VR users generally cannot move in the shared world, unless the system involves locomotive equipment such as robots, e.g. in Jones et al. [27] or Heshmat et al.'s work [19], which can be costly. On the other hand, reconstruction using the HoloLens spatial mesh produces holes and artifacts due to imperfect scanning and occlusion, which negatively affects quality. Teo et al. discuss how 360° video is better suited for search tasks and presence, but 3D reconstructed scenes can also serve as a complement [62].

VirtualNexus addresses the issues with remote 360° interaction precision and inflexibility through a different approach — the cutout feature. In 3D reconstructed virtual environments, the user ideally can walk around the environment and interact with virtual objects precisely; our system performs the exact opposite by allowing users to bring parts of the environment towards them. In addressing the issues associated with solely 360°, our findings revealed that the cutout system was able to improve the clarity and precision of remote interactions. In addition, users generally found the system intuitive to understand and easy to use.

7.3 Awareness of the Virtual Other

Considering the collaborative nature of VirtualNexus, users must understand each other's intentions and references with a high level of awareness and presence [71]. Salimian discussed how communication cues for group awareness are split into responsive cues (affected by actions in physical space), communication-based cues (directly addressing the other), environmental cues (to help understand the environment), and event-based cues (semi-scripted events) [52]. In VirtualNexus, two primary non-verbal ways exist in which we tried to establish mutual awareness of the virtual other in our system — avatarization of the local user from the remote user's perspective, and pointer cues of the virtual other for both users.

In regards to 'seeing' the virtual other, the VR user almost always has the AR user in their view, due to the larger field of view and their stationary perspective. Furthermore, they directly see the real-life representation of the AR user as the 360° camera captures it. However, the reverse is not true — AR users have more freedom in moving their camera and a smaller field of view, so they may not always be looking at the VR user, and they do not see a real-life representation; in our system, they see an avatar with only face and hands. We can take inspiration from Piumsomboon et al.'s work on awareness cues to alleviate this issue, giving the local AR user notice that the VR user is attempting to communicate something [45, 46]. Pointer cues were another way in which we explicitly provided awareness cues of the virtual others to the users. They were observed to be helpful, as users in one reality could see what the other user was pointing at while conversing. Further spatial cues such as gaze could be future extensions towards extending spatial awareness [45].

8 LIMITATIONS AND FUTURE EXTENSIONS

Finally, we highlight some of the technical and design limitations of our present work, and provide suggestions for areas of future work based on prior research.

8.1 Depth Cues

Our system trades off binocular depth cues for a coherent depth perception of the virtual objects and the 360° video (detailed in 4.3). However, participants did report occasional difficulty in perceiving depth while manipulating the objects. In the future, we can opt to use binocular 360° cameras which are already available as commodity products, creating 360° videos with binocular depth perception.

8.2 User Privacy and Personal Spaces

The present implementation of VirtualNexus provides both remote and local users with a shared collaborative space for interaction. However, the concept of private space on both ends requires future consideration, as discussed in Wang et al.'s *Slice of Light* work [68], and implemented in He et al.'s *CollaboVR* system [18]. Drawing inspiration from these works, privacy on both the remote VR and local AR side can occur through having an option or space for personal interaction, object creation, and interactions. Outside of the shared space, the user can use this individual space to more privately interact with the environment, which can then be shared and synced when they are ready.

8.3 Localized Cutout Spaces

Presently, the cutout space is not separated from the rest of the environment. This can introduce visual clutter which makes it hard to understand which objects belong to the cutout space and which objects belong to the original world. Furthermore, because it has no localized volume, objects belonging to the cutout can be very far away, yet still show up — we found this can at times be confusing to the remote user during interactions. In the future, we plan to more clearly localize the cutout using a 'snow globe' metaphor — the cutout acts as a localized miniature world that takes up a portion of visible space. Objects outside this space would no longer show up, and the objects within would be more easily understood as belonging to the cutout (e.g. with a radial outline). These improvements will reduce visual clutter and improve the user experience of VirtualNexus.

Furthermore, our present system only allows the remote user to have a single cutout in the scene at a time. Having the possibility for multiple cutout spaces would be a possible extension to the system. This could potentially increase the efficiency in engaging in a task, as it provides users with the option to more quickly interact and affect different areas of the scene quickly. However, the potential interactions and the features to enable context switching would require future study. For instance, one possibility could be to allow virtual objects to 'switch' cutouts, enabling an object to rapidly move between two distant environmental locations.

8.4 Remote Users and Non-HMD Guests

The current prototype of the VirtualNexus system supports (and was tested with) a single local AR user and a single remote VR user, but given the server-client nature of our system, it could theoretically expand to allow for several remote VR users to connect to the AR system. This could potentially allow for collaboration with multiple users in a single space. However, this would likely need to encode a different set of interactions, especially at scale [33, 59]. For instance, multiple remote users could use multiple cutouts with avatars for each.

We also considered how VirtualNexus could expand to include non-HMD guests in the collaboration process. Prior research has looked into add-on or projected displays for VR headsets [5, 15, 16], which allow them to communicate with non-HMD people. We could take advantage of such displays to communicate the interactions of the remote VR user to non-HMD guests in a simple manner. Furthermore, non-HMD guests could play a more direct role in the collaboration process by joining as remote clients using non-HMD media, such as through mobile devices (such as phones or tablets) [66], desktop displays [13], or other hardware apparatuses [36].

8.5 Rapidly Scanned Objects

We demonstrated the usefulness of having physical objects scanned as virtual replicas to enhance interactivity. However, to enable a sense of truly free manipulation of the physical scene, there are two directions we could further improve on — space and time. Our system still requires users to walk around the objects to capture multiple 2D images to achieve 3D reconstruction. A better approach to try in the future is to hand-hold the object to perform the scanning process. This process would involve dynamically removing occlusions, which remains a challenging task in computer vision. Additionally, our scanning time varies from 1–3 mins due to the random initialization in Colmap. The Colmap process could be replaced by refining camera poses captured from HoloLens with point cloud registration. With appropriate optimizations, our reconstruction pipeline could be as fast as 30 seconds end-to-end.

9 CONCLUSION

In this paper, we introduce VirtualNexus, a system to facilitate AR/VR collaboration combining both high-fidelity 360° video with spatially-accurate 3D reconstructions of a physical environment. Our system supports traditional interactive collaborative features such as annotations and virtual object manipulation but also introduces novel features unique to the environment. On the remote VR side, a user can define cutouts — copied 3D meshes that replicate the original world, providing a miniature slice of the original environment. On the local AR side, users can rapidly scan and share physical items, allowing them to become interactable virtual objects. We outline applications of the VirtualNexus system and deploy it in user studies centred around a collaborative storyboarding task. Our findings suggest that VirtualNexus improves collaboration and the sharing of ideas between remote and local users. The cutout system was intuitive and provided increased clarity and precision, and the scanning system enhanced the capabilities of the collaborative processes. Finally, we situate the implications of our findings within broader mixed-reality collaboration research and outline future extensions of our work.

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] Adrien Chaffangeon Caillet, Alix Goguey, and Laurence Nigay. 2023. 3D Selection in Mixed Reality: Designing a Two-Phase Technique To Reduce Fatigue. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 800–809. https://doi.org/10.1109/ISMAR59233.2023.00095

[5] Liwei Chan and Kouta Minamizawa. 2017. FrontFace: Facilitating Communication between HMD Users and Outsiders Using Front-Facing-Screen HMDs. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '17)*. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/3098279.3098548

[6] 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

[7] 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

[8] Thiago L. T. da Silveira, Paulo G. L. Pinto, Jeffri Murrugarra-Llerena, and Cláudio R. Jung. 2022. 3D Scene Geometry Estimation from 360° Imagery: A Survey. *Comput. Surveys* 55, 4 (Nov. 2022), 68:1–68:39. https://doi.org/10.1145/3519021

[9] 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

[10] L. Davis, J. Rolland, F. Hamza-Lup, Yonggang Ha, J. Norfleet, and C. Imielinska. 2003. Enabling a Continuum of Virtual Environment Experiences. *IEEE Computer Graphics and Applications* 23, 2 (March 2003), 10–12. https://doi.org/10.1109/MCG.2003.1185574

[11] 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)*.

[12] 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.

[13] Arthur Fages, Cédric Fleury, and Theophanis Tsandilas. 2022. Understanding Multi-View Collaboration between Augmented Reality and Remote Desktop Users. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 549:1–549:27. https://doi.org/10.1145/3555607

[14] Sean Follmer, Daniel Leithinger, Alex Olwal, Akimitsu Hogge, 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

[15] Jan Gugenheimer, Evgeny Stemasov, Julian Frommel, and Enrico Rukzio. 2017. ShareVR: Enabling Co-Located Experiences for Virtual Reality between HMD and Non-HMD Users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 4021–4033. https://doi.org/10.1145/3025453.3025683

[16] Jan Gugenheimer, Evgeny Stemasov, Harpreet Sareen, and Enrico Rukzio. 2018. FaceDisplay: Towards Asymmetric Multi-User Interaction for Nomadic Virtual Reality. 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/10.1145/3173574.3173628

[17] 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.

[18] 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

[19] 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

[20] 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

[21] 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

[22] 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

[23] 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

[24] 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

[25] Andrew Irlitti, Mesut Latifoglu, Qiushi Zhou, Martin N Reinoso, Thuong Hoang, Eduardo Velloso, 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

[26] 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

[27] 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

[28] M. Kalkusch, T. Lidy, N. Knapp, G. Reitmayr, 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

[29] 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

[30] 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.

[31] 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

[32] 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]

[33] Geonsun Lee, HyeongYeop Kang, JongMin Lee, and JungHyun Han. 2020. A User Study on View-sharing Techniques for One-to-Many Mixed Reality Collaborations. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 343–352. https://doi.org/10.1109/VR46266.2020.00054

[34] Gun A. Lee, Theophilus Teo, Seungwon Kim, and Mark Billinghurst. 2017. Mixed reality collaboration through sharing a live panorama. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications* (Bangkok, Thailand) *(SA '17)*. Association for Computing Machinery, New York, NY, USA, Article 14, 4 pages. https://doi.org/10.1145/3132787.3139203

[35] 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

[36] 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

[37] 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/10.1145/3503250

[38] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1995. Augmented Reality: A Class of Displays on the Reality-Virtuality Continuum. In *Telemanipulator and Telepresence Technologies*, Vol. 2351. SPIE, 282–292. https://doi.org/10.1117/12.197321

[39] 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

[40] 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

[41] Michael Nebeling, Katy Lewis, Yu-Cheng Chang, Lihan 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

[42] 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

[43] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, 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 Lutchyn, 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

[44] 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

[45] Thammathip 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

[46] Thammathip Piumsomboon, Gun A. Lee, Andrew Irlitti, Barrett Ens, Bruce H. Thomas, and Mark Billinghurst. 2019. On the Shoulder of the Giant: A Multi-Scale Mixed Reality Collaboration with 360 Video Sharing and Tangible Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3290605.3300458

[47] 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 Networki ng* (New Delhi, India) *(MobiCom '18)*. Association for Computing Machinery, New York, NY, USA, 99–114.

[48] 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.

[49] Vincent Reynaert, Yosra Rekik, Florent Berthaut, and Laurent Grisoni. 2023. The Effect of Hands Synchronicity on Users Perceived Arms Fatigue in Virtual Reality Environment. *International Journal of Human-Computer Studies* 178 (Oct. 2023), 103092. https://doi.org/10.1016/j.ijhcs.2023.103092

[50] 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

[51] 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

[52] Mohamad H. Salimian. 2015. Exploring Group Awareness in a Mixed Reality Collaborative Environment. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (UbiComp/ISWC'15 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 495–501. https://doi.org/10.1145/2800835.2801654

[53] Johannes Lutz Schönberger and Jan-Michael Frahm. 2016. Structure-from-Motion Revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

[54] 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)*.

[55] 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

[56] Richard Skarbez, Missie Smith, and Mary C. Whitton. 2021. Revisiting Milgram and Kishino's Reality-Virtuality Continuum. *Frontiers in Virtual Reality* 2 (March 2021). https://doi.org/10.3389/frvir.2021.647997

[57] 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

[58] Richard Stoakley, Matthew J. Conway, and Randy Pausch. 1995. Virtual reality on a WIM: interactive worlds in miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '95)*. ACM Press/Addison-Wesley Publishing Co., USA, 265–272. https://doi.org/10.1145/223904.223938

[59] Anthony Tang, Michael Boyle, and Saul Greenberg. 2005. Understanding and mitigating display and presence disparity in mixed presence groupware. *Journal of research and practice in information technology* 37, 2 (2005), 193–210.

[60] 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.

[61] 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/10.1145/3359996.3364238

[62] 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

[63] 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

[64] 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

[65] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. 2019. Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 161–174.

[66] Balasaravanan Thoravi Kumaravel, Cuong Nguyen, Stephen DiVerdi, and Bjoern Hartmann. 2020. TransceiVR: Bridging Asymmetrical Communication Between VR Users and External Collaborators. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (UIST '20)*. Association for Computing Machinery, New York, NY, USA, 182–195. https://doi.org/10.1145/3379337.3415827

[67] Ana Villanueva, Zhengzhe Zhu, Ziyi Liu, Feiyang Wang, Subramanian Chidambaram, and Karthik Ramani. 2022. ColabAR: A Toolkit for Remote Collaboration in Tangible Augmented Reality Laboratories. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (April 2022), 81:1–81:22. https://doi.org/10.1145/3512928

[68] Chiu-Hsuan Wang, Chia-En Tsai, Seraphina Yong, and Liwei Chan. 2020. Slice of Light: Transparent and Integrative Transition Among Realities in a Multi-HMD-User Environment. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 805–817. https://doi.org/10.1145/3379337.3415868

[69] 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

[70] 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

[71] Xiangyu Wang, Peter E. D. Love, Mi Jeong Kim, and Wei Wang. 2014. Mutual Awareness in Collaborative Design: An Augmented Reality Integrated Telepresence System. *Computers in Industry* 65, 2 (Feb. 2014), 314–324. https://doi.org/10.1016/j.compind.2013.11.012

[72] 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

[73] 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

[74] 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

[75] Shuhao Zhang, Yue Li, Ka Lok Man, Yong Yue, and Jeremy Smith. 2023. Towards Cross-Reality Interaction and Collaboration: A Comparative Study of Object Selection and Manipulation in Reality and Virtuality. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 330–337. https://doi.org/10.1109/VRW58643.2023.00075

[76] 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