VRoxy: Wide-Area Collaboration From an Office Using a VR-Driven Robotic Proxy

Mose Sakashita
Cornell University
Ithaca, NY, USA
ms3522@cornell.edu

Hyunju Kim
Cornell University
Ithaca, NY, USA
hk724@cornell.edu

Brandon Woodard
Brown University
Providence, RI, USA
brandon_woodard@brown.edu

Ruidong Zhang
Cornell University
Ithaca, NY, USA
rz379@cornell.edu

François Guimbretière
Cornell University
Ithaca, NY, USA
fvg3@cornell.edu

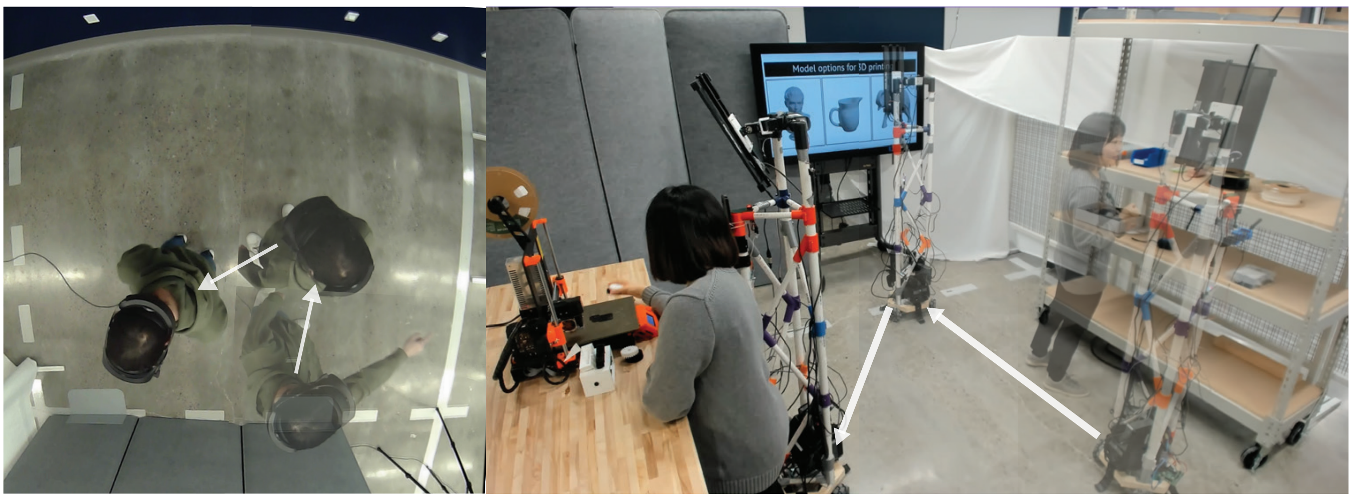

Figure 1: VRoxy is a robotic system intended to represent a VR user as a robotic proxy. While the VR user moves around the confined room (Left), the user can navigate through different task spaces (shelf, monitor, or desk) and interact with the collaborator on the other side (Right).

ABSTRACT

Recent research in robotic proxies has demonstrated that one can automatically reproduce many non-verbal cues important in co-located collaboration. However, they often require a symmetrical hardware setup in each location. We present the VRoxy system, designed to enable access to remote spaces through a robotic embodiment, using a VR headset in a much smaller space, such as a personal office. VRoxy maps small movements in VR space to larger movements in the physical space of the robot, allowing the user to navigate large physical spaces easily. Using VRoxy, the VR user can quickly explore and navigate in a low-fidelity rendering of the remote space. Upon the robot's arrival, the system uses the feed of a 360 camera to support real-time interactions. The system also facilitates various interaction modalities by rendering the micro-mobility around shared spaces, head and facial animations, and pointing gestures on the proxy. We demonstrate how our system can accommodate mapping multiple physical locations onto a unified virtual space. In a formative study, users could complete a design decision task where they navigated and collaborated in a complex 7.5m x 5m layout using a 3m x 2m VR space.

UIST '23, October 29-November 1, 2023, San Francisco, CA, USA

ACM ISBN 979-8-4007-0132-0/23/10…$15.00
https://doi.org/10.1145/3586183.3606743

CCS CONCEPTS

• **Human-centered computing → Collaborative and social computing devices**.

ACM Reference Format:
Mose Sakashita, Hyunju Kim, Brandon Woodard, Ruidong Zhang, and François Guimbretière. 2023. VRoxy: Wide-Area Collaboration From an Office Using a VR-Driven Robotic Proxy. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23), October 29-November 1, 2023, San Francisco, CA, USA*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3586183.3606743

1 INTRODUCTION

Collaborative design endeavors frequently necessitate intricate coordination of nonverbal signals and collaboration patterns among team members, particularly when engaging with physical artifacts. In co-located design scenarios, physical closeness and non-verbal communication facilitate rapid understanding of workspaces and colleagues' intentions [5, 9, 19].

To emulate this co-location experience within remote collaborative design, researchers have investigated embodied robotic proxies that convey a sense of physical co-presence by automatically mapping movements, such as head or body gestures [39, 48, 49]. By eschewing traditional control interfaces [7, 23] (e.g., joystick or keyboard), users can concentrate on primary tasks while maintaining peripheral awareness of each other's physical presence. These systems are often constrained by their reliance on symmetric configurations, necessitating similar hardware devices and a certain amount of space for both local and remote users to collaborate effectively. These constraints may present difficulties when, for example, a remote designer operates from home without adequate space or equipment to establish the same hardware arrangement.

Virtual reality (VR) can mitigate such limitations by offering flexible rendering options and facilitating immersive remote experiences. For instance, some systems for telepresence robots enable VR users to navigate between rooms in a spacious environment through a head-mounted display (HMD) [20, 25, 28]. However, these VR systems typically demand explicit robot control via manual inputs, such as controllers, which impose cognitive burdens and divert focus from primary tasks [42]. Moreover, many of these systems do not fully support important aspects of co-located interactions such as facial expressions or micro-mobility [32], the nuanced motions people use to signify attention or the collaborative patterns observed around shared spaces such as tables [55] and whiteboards [26].

In this paper, we present VRoxy, a controller-less, VR-controlled robotic proxy devised to enable navigation through expansive environments from a much more compact space (Fig. 1). VRoxy maps a VR user's body positioning to the movement of a mobile robot on the other side. The VR user can independently explore the 3D-rendered remote space, regardless of the robot's actions. When the exploration is completed and the robotic proxy catches up with the remote user's position, the VR user can see a real-time 360 camera feed. VRoxy can also deliver other interaction modalities during the real-time feed, such as micro-mobility around shared spaces, head movements, facial expressions, eye gaze, and pointing gestures. We demonstrate a scenario where a VR user can navigate between a lab space and an office through a hallway to cooperate with multiple collaborators on different tasks. Additionally, we investigate the potential of VRoxy for facilitating unique VR interface capabilities, like instantaneously transitioning between two remote locations. We demonstrate that VRoxy can adapt to both stationary and mobile robotic proxies using a consumer-grade headset (Quest Pro). We present preliminary user feedback results to demonstrate the potential of our approach.

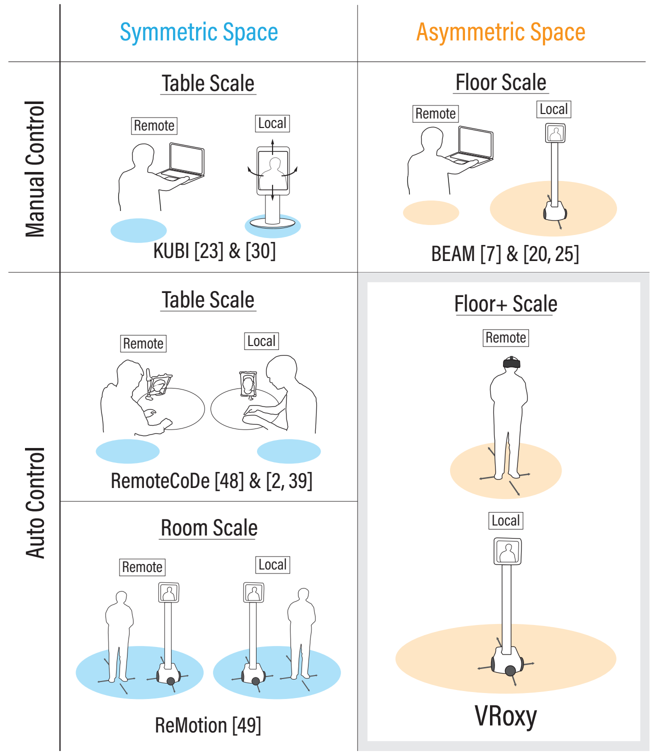

Figure 2: Prior work on telepresence robots within the outlined design space. VRoxy supports asymmetric settings and allow a wide spectrum of movement, from micro-mobility to interactions across buildings.

2 RELATED WORK

VRoxy is grounded on prior work in the field of remote collaboration, telepresence robots, and VR systems for wide-area navigation. Key references are mapped out in the design space, depicted in Fig. 2. We provide an overview of significant literature from past research and outline the unique aspects of our work.

2.1 Physical Task Collaboration

Buxton's framework highlights the importance of integrating person and task space [10] in facilitating remote collaboration. Reference space [9] where the person and task space overlap is essential to support through pointing gestures, helping with sharing of deictic reference [19]. Especially in physical tasks, rendering these gestural cues improve collaboration dynamics as they establish common ground and shorten verbal communication [27]. Some work in the field of supporting physical tasks attempts to render these missing gestures over task space through handheld devices [17, 52] or an HMD [15, 50]. Other systems render shared space in mixed reality. MiniMe flexibly presents a remote collaborator over physical space [41]. Loki supports live instruction through rendering shared space in point cloud [58]. They are designed for a more or less fixed task space and require every participant to wear a headset. In our work, due to the physical representation of a VR user, collaborators in physical space have peripheral awareness of the space and do not have to wear any device. The targeted shared workspace is also much larger.

2.2 Motion-controlled Robotic Proxies

Mobile robots such as ASTEROIDS [30] or Beam [7] offer mobility in a remote environment, but these telepresence robots require explicit control using a joystick or keyboard which can demand relatively high cognitive load [39, 42, 60]. The cognitive load introduced might distract from primary tasks [48, 49]. To afford automatic control, some robotic proxies have been proposed to offer implicit control, automatically mapping non-verbal cues to the robot. MeBot [2] is a robot with limbs that tracks a user's facial direction to control the robot's orientation but still requires users to use joysticks to manipulate the limbs. MMSpace [39] offers a fully motion-controlled kinetic display controlled with head rotation, supporting smooth face-to-face conversations around a table. RemoteCoDe [48] can remap head rotation to a kinectic robot for a hands-on design task for a more flexible workspace. These projects are designed to support collaborative tasks in a small setting, for example, around a table. ReMotion [48] extends the approach further to accommodate interactions in a more spacious room, reproducing body positioning in the space. However, the main limitation of these auto-mapping robot systems is their requirement of symmetrical environments in which similar hardware devices and amount of physical space are required in both locations. The design space on telepresece robots is presented in Fig. 2. Our work explores a means to only require a small area for collaboration in a much larger remote space by automatically mapping a VR user's movement to a robotic proxy.

2.3 VR Interface for Telepresence Robots

Existing work on VR interfaces for telepresence robots can be broadly categorized into two types: those that use more traditional input devices, such as joysticks or controllers, and those that map body movements to robot actions. For instance, the Telesar VI project enables users to control the posture of a slave robot using their own torso and upper limb movements [54]. Other works have explored adding social cues such as gaze direction or facial expressions to enhance face-to-face communication and collaboration [16, 47, 50]. However, these approaches are typically limited to scenarios where the robot is fixed in one location, resulting in a lack of mobility. In contrast, our work focuses on scenarios where remote users must navigate an ample space independently without being tethered to a fixed location, leveraging VR tracking.

Other work presents VR interfaces for mobile telepresence robots, affording mobility in a remote environment. For example, Sven et al. use a headset to provide an immersive view while controlling the robot through a desktop application [28]. Heshmat et al. explore the use of a VR HMD to share outdoor activities with a 360 view of the remote environment [20]. VROOMing system allows VR users to navigate a remote environment by controlling a Beam robot through joystick controllers while showing an overlay avatar on the robot [25]. Although they allow a remote user to explore a vast space, the user has to explicitly control the robot using input devices such as a joystick or controller, which can increase their mental load or divide their attention from their collaborative tasks.

Others have allowed walking movements to control the robot implicitly. 'Flying Head,' introduced by Higuchi et al., allows a user's head movement captured by a VR HMD to control the yaw, pitch, and altitude of a proxy UAV (quadcopter). Still, the design of the robot system does not focus on articulating complex social cues [21]. Controlling a walking humanoid using an omnidirectional treadmill has been explored [14], but it induces motion sickness due to the robot's walking motion. Our work attempts to mitigate motion sickness while mapping motions to the mobile robot's movements in a larger environment.

2.4 VR Interfaces for Wide-Area Navigation

Here we describe the approaches towards navigating wide-area virtual spaces enabled by VR interfaces upon which we design our VR interface for navigating a wide environment.

2.4.1 Natural Walking. We employ walking-based navigation, drawing from the research by Sayyad et al. that compares natural walking with a typical 'point and click' teleportation technique where the users' preferences favored that of the natural walking system by a significant margin [51]. When we explore new places, we subconsciously attempt to learn and remember the layout of our surroundings [35]. This mental map serves as a tool to help us understand our environment, and walking has been shown to enhance the formation of such cognitive maps [46]. Thus, we integrate walking into our navigation interface.

2.4.2 Locomotion Techniques. To circumvent VR users from reaching physical boundaries in their local environment (e.g., an office space), locomotion techniques are often used. Redirected walking, walking-in-place, and teleportation are common techniques used to provide virtual paths for users to remain within the physical space they occupy [6, 13, 24, 31, 51]. Unfortunately, the space redirected walking requires is still relatively large for personal use, and walking-in-place lacks spatial sensory feedback, therefore providing a less natural walking experience [53].

A potential compromise is the use of teleportation graphs users can navigate towards. Once triggered, the selected teleportation path can transport the user beyond the extent of their physical environment [24, 29]. In our work, we design a similar interface using some principles of teleportation graphs; except, our work focuses on mapping the VR user's movement to the physical robot without the use of a controller, rather than solely considering the navigation in a VR space.

3 VROXY DESIGN

In this section, we present the design of VRoxy, a VR system designed to allow a VR user to seamlessly control a remote embodiment robot in an expansive space much greater than the user's local space. In the following sections, we introduce the strategies and techniques we adapted to our VRoxy design to accommodate interactions on different scales, beginning with room-scale areas and gradually extending to substantially larger areas that surpass the VR user's available space. Additionally, we examine VRoxy's potential to facilitate seamless transition across multiple buildings, transcending the capabilities of traditional co-located interactions [22].

In co-located collaboration, participants can navigate expansive environments, seamlessly transitioning between task areas within a single room or across multiple rooms to initiate new tasks or engage with different individuals. For example, in a university setting, a professor may enter a spacious laboratory and see a student working on a prototype on a workbench. The professor can then approach the workbench to review the student's prototype and offer advice. Subsequently, the professor can take the student to a shelf, considering potential components that could enhance the project. They can move to a whiteboard to brainstorm ideas on integrating the selected component into the prototype. Once finished, the professor can exit the lab, traverse a connecting hallway, and enter another room to engage in a discussion with a fellow faculty member.

Although such interactions have been facilitated in room-scale settings with symmetric configurations (i.e., identical physical space and hardware setups [49]), it is important to note that remote collaborators may have limited space, such as a standard office, that is much smaller than the local environment at the other end of the connection. Furthermore, preparing an identical environment with multiple rooms and hallways connecting them would be impractical. As such, we aim to accommodate this asymmetric context, enabling the remote user to effectively use their limited space to explore a significantly larger area.

3.1 VR interface to control a Robotic Proxy

The new VR headsets such as the Quest Pro can capture rich information about real-time non-verbal cues from VR users, including spatial movement, head rotations, facial expressions, eye gaze, and hand gestures. This greatly simplifies the tracking of the remote user when compared to symmetric systems such as ReMotion [49]. Here we will present how this information could be used to control a robotic embodiment equipped with omnidirectional wheels similar to the one used in the ReMotion system.

3.1.1 Mapping Head and Face Animations through an Articulated Display. As illustrated in Fig. 3, mapping head rotations provided by the Quest Pro to physical movements can be achieved simply by using an articulated display. This approach fosters joint attention among collaborators [36, 49, 61]. VRoxy also uses face tracking information provided by Quest Pro to render the face of the VR user's avatar on the articulated display, which is generated through AvatarSDK [1]. We employ blend shape data to animate facial features in real-time, enabling remote collaborators to better comprehend emotions during collaborative tasks or discussions [25, 48, 49]. Finally, we use eye tracking data to support eye gaze animation, assisting users in establishing eye contact and discerning more subtle cues about attention.

3.1.2 Mapping VR Movement to a Mobile Robot. We began with the simplest setting: two similarly sized rooms in both the VR user's space and the remote environment. Our initial approach involved a one-to-one movement mapping, where the mobile robot moves in correspondence with the VR user's movement while displaying a live feed from a 360-degree camera (Ricoh Theta V) attached to the mobile robot. This camera provides high-fidelity, real-time rendering of the remote space [57]. This approach was unsuccessful due to motion sickness induced by the unstable video feed from the robot's continuous movement. Even with a gimbal stabilizer attached to the camera, the issue persisted, attributable to feed latency and delay in orientation adjustments by the stabilizer. Additionally, the user's movements were limited to the robot's speed, thereby slowing the navigation process. We investigated alternative navigation designs. For example, we showed image icons or text labels for each task area over the 360 video feed, which a VR user selects to move to the next position. However, this approach failed

Figure 3: Several examples of head rotation, facial expression, and eye gaze. The system replicates the VR user's head rotation through an articulated display and their facial expression & eye gaze through an avatar animation.

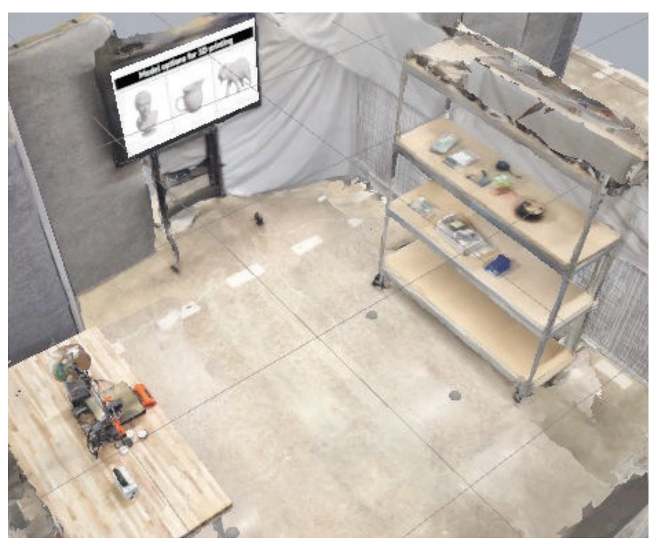

Figure 4: Example of a 3D mesh of the lab captured using Scaniverse, illustrating the environment for navigation.

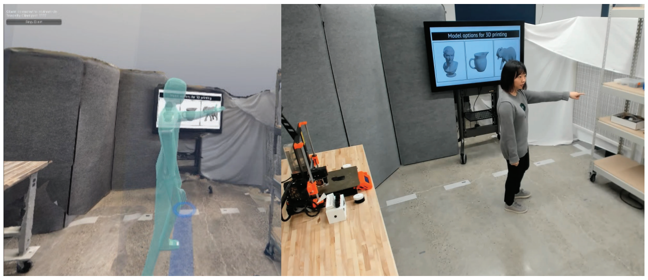

Figure 5: The virtual avatar is rendered in the navigation view to represent a collaborator on the other side.

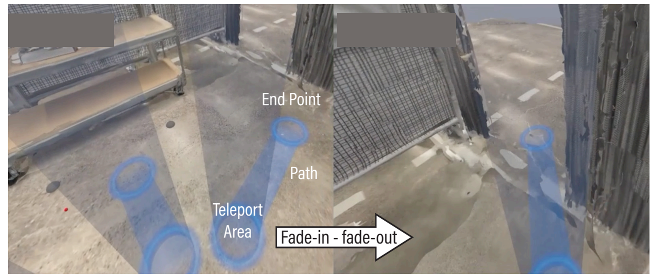

Figure 6: The teleport area connected with a path and endpoint. When a VR user enters the cylindrical zone, the user moves to the endpoint after a fade-in - fade-out effect.

to provide spatial awareness due to abrupt visual transitions. Our design exploration reinforced the need of quick and stable means of space exploration. We decided to accomplish this by displaying a 3D mesh of the space during large movement and switching to a live feed from the camera when the robotic embodiment matches the remote user's position in the space.

The VRoxy system's navigation view renders a pre-scanned 3D model of the remote space as shown in Fig. 4, which is obtained beforehand using a mesh captured by Scaniverse [38]. This approach ensures the maintenance of a consistent physical frame of reference between the physical and virtual environments, allowing collaborators to intuitively understand spatial relationships [9]. It also liberates the user from the constraint of the robot's speed. In this navigation view, a VR user can physically walk around in the shared space to explore the room. We selected a walking interface over joystick control for navigation, as the latter can cause motion sickness and hinder the use of hand gestures [8, 11, 12]. Moreover, physically walking around a virtual environment aids users in developing spatial orientation and awareness [44, 51].

The issue with the pre-captured model is that it does not provide real-time updates on activities happening in the environment, including information on the location of collaborators [26]. To enhance the VR user's awareness of their collaborators' location in a shared space, VRoxy generates full-body avatars representing the VR user within the navigation view as shown in Fig. 5. To track the movements of individuals in a room, our system uses Kinect Azure SDK and maps the movements from the physical space to the virtual environment in the navigation view. This feature enables the VR user to recognize the movements of the VR user within the shared space promptly.

Once a VR user wants to engage with collaborators at a certain task area, they can remain at the location in the navigation view. The system returns to the live view once the robot reaches the exact location. This enables an immersive perspective of the real-time camera feed from the 360-degree camera from a new position in the physical space. After completing live interactions at that location, the VR user can simply turn around to enter the navigation mode and begin navigating again to switch task areas, for instance, from a shelf to a table. VRoxy offers this automatic transition between the two modes to facilitate smooth navigation and real-time collaboration with the VR users.

3.1.3 Task Space through Camera Feed and Screen Sharing. Task space [10] here refers to objects of interest within the shared space, for example, around a table, shelf, or computer screen. To achieve smoother navigation, the system needs to ensure that real-time updates of the task space are visible within the navigation view, as the user waits for the robot to arrive at the targeted task areas. VRoxy incorporates a real-time camera feed or shared screen within the 3D models, such as on a virtual screen or table. This feature allows users to observe the updated task space even before the live view is activated.

3.2 Mapping between Area of Different Sizes

The design presented above cannot support a scenario in which a VR user needs to access a large space but only has a small space available to them. To accommodate the discrepancy in size, we decided to provide a simple teleportation mechanism between areas of interest in the target room. We begin by identifying an object of interest, such as a table, shelf, or monitor. Then we use the robotic embodiment tracking system to register the visual teleport path and orientation of the endpoint in front of each area. To facilitate navigation in such a wide space, the system must ensure that a VR user consistently returns to a similar position each time they visit a corresponding task area. The designer then manually adds teleportation markers linking the different areas of interest with expected paths in a manner that orients and positions the user, guiding them to maintain consistent placements. VRoxy renders teleportation areas on the ground, marked with a path and an endpoint, as shown in Fig. 6. To use a teleportation link, the user needs to step into the link before being teleported in VR space to the other side of the link. In addition to its simplicity, the primary advantage of this technique is its ability to sustain spatial orientation and relative positioning. This enables a VR user to be in a specific space in their room and simultaneously be in a matching task area within the navigation interface. This feature is advantageous when a VR user needs to add a physical requirement in their space (e.g., a chair for a sitting arrangement or tracked VR keyboard). Whenever the VR user moves to a new task area, the robot uses the registered robot pose to control its movements, simulating the VR user's intended body positioning. As soon as the robot replicates the registered position and orientation, the VR user can reaccess the immersive live view to initiate interactions with their collaborators. To switch to a different task area, a VR user can return to navigation mode by turning around to resume navigation.

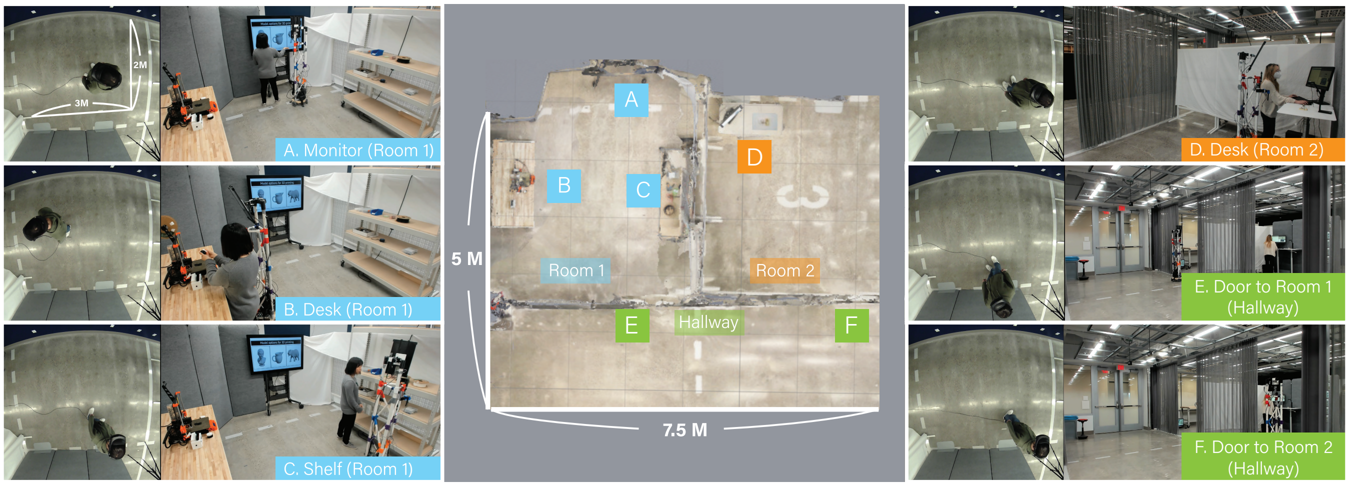

Figure 7: Demonstrations of the mapping of a wide environment in VR within limited space. The system enables a VR user to move around different task spaces in the room (A-C) and go to another room (D) via the hallway (E-F). While moving in VR, the user consistently occupies the same location in the physical space for each task area.

3.2.1 Navigating through Multiple Rooms. With the interface introduced above, VRoxy supports transitions between different task spaces within a single room. Here we aim to extend the design to accommodate multiple room settings. To demonstrate this, we prepare a 3x2m VR user's space, which is about a typical office size, which we use to map to a spacious environment where there are multiple rooms and a hallway connecting to these rooms. Using the navigation technique (Section 3.2), Fig. 7 demonstrates that VRoxy can allow a VR user to be at a certain location in the VR user's physical space for a corresponding task area.

3.3 Support for Micro-Mobility

In co-located collaboration, individuals employ subtle movements as cues to indicate their focus to others, aiding in the coordination of various collaborative coupling styles around a table [55] or a whiteboard [26]. These minor adjustments, known as "micro-mobility" [32], can enable the VR user to obtain a more detailed view or shift their attention within a task area. For instance, when examining a shelf, collaborators can move side-by-side to closely inspect different sections. The techniques discussed thus far present a limitation in their assumption that the robot will be at a fixed position during an interaction at each task area. However, in real-world situations, individuals often adjust their positions for optimal access, such as moving closer to a display or shelf.

To support these subtle movements, VRoxy allows the VR user to move around in the spherical rendering of the 'live view' at each task area. When the amount of movement reaches a certain threshold, the system uses this movement within the 'live view' to control the robot's movement (Fig. 8). As the robot applies small adjustments, the live view temporarily pauses and transitions to a grayscale image to assist the VR user in understanding that the view is not updated due to latency, and a new view will be provided shortly. Pausing the 360 videos while moving or slowing down navigation mitigates motion sickness. Once the robot aligns its position with the VR user's position, it stops moving and provides a stable, immersive live view of the local space. To avoid triggering teleportation to the next point by mistake, we disabled the teleport links in cases of the user not facing them (with adjustable sensitivity). This approach allows VRoxy to help the VR user examine task areas more closely or initiate discussions by altering their proximity to surrounding objects or collaborators.

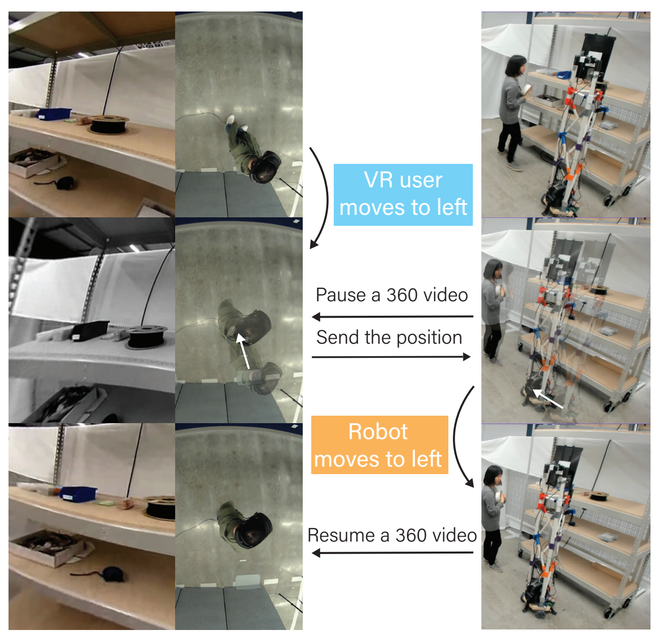

Figure 8: Demonstrations of the support for micro-mobility. When the VR user moves, the video feed is stopped in black and white while the robot updates its position. The video resumes once the robot is at the target position.

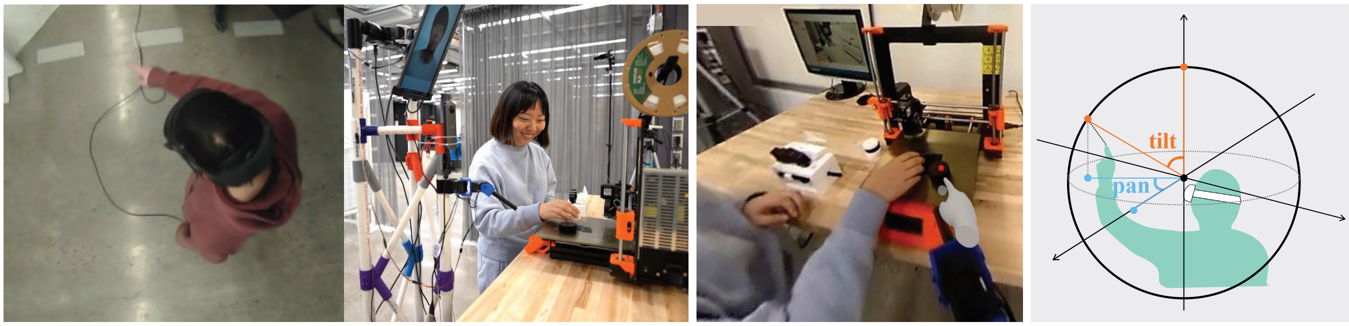

Figure 9: The pointing gesture is captured by the Quest Pro and rendered through the 2DoF point arm attached to the robotic proxy (Left). The pointing position is visualized as a red point on the 360 view for the VR user (Middle). Using the pointing position and the reference position of the pointing arm, we calculate the pan and tilt angles to control the pointing arm (Right).

3.4 Rendering Pointing Gesture

x In the live view mode, VRoxy physically renders the gesture through a 2DoF pointing arm as shown in Fig. 9. The position of the arm is computed using the information provided by the Oculus SDK's hand tracking API. After a pointing gesture is detected, we cast a ray from a pointing fingertip onto the 360 video sphere to obtain a pointing position. We compute the pan and tilt angles from the reference of our pointing device (see Fig. 9 Right). As the VR user points, the red point is visualized as the pointing spot on the panorama image to instantly inform the user where they are pointing at.

3.5 Beyond Being There: Switching between Multiple Buildings

To demonstrate the flexibility of our approach, we present how the system could be used to accommodate movements across buildings, going beyond simulating co-located interactions [22]. This exploration also leverages the concept of the re-embodiment [34], using VRoxy to enable VR users to move their social presence from one robotic proxy to another in distant locations.

3.5.1 Rendering of Complex Head Movements. As we implemented our system, we realized that the VR headset's tracking information presented the potential to render more complex head movement beyond what the pan-tilt display offered [48, 49]. Intrigued by this capability, we decided to design a new articulation system for head movement. Our system combined a 3-joints system to render leaning-over motions with a spherical parallel manipulator for more accurate head rotation. A potential design of a robotic proxy tailored to accommodate tasks around a table is shown in Fig. 10. Although we are presenting a desktop model, the same schematic can be adapted for our mobile robot.

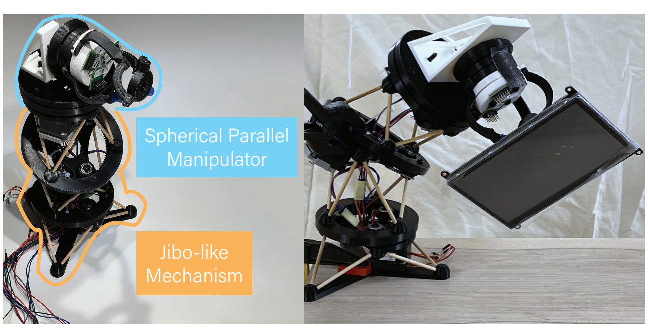

Figure 10: The robotic proxy for a stationary setting designed with a Jibo-like mechanism for leaning-over motion and a spherical parallel manipulator for realistic head rotation.

Leaning-Over Motion. Around a desk, people either look at each other or at artifacts. It is often the case that they lean over [18] to work closely on the artifact. To render the notion of leaning over, we implemented a kinematic chain similar to the Jibo system [43] as shown Fig. 10 as it can deliver smooth motion.

Precise Head Motion. In a traditional articulated display used for telepresence systems [2, 39, 48, 49], the points of actuation are typically distributed across multiple joints. This can introduce challenges in reproducing head-like animation. In contrast, a spherical parallel manipulator results in a more centralized point of actuation and smooth animation, as demonstrated by Pollen Robotics' Orbita [45]. Thus, we used the Open Source version of this design as our screen actuator to render a realistic animation of 3DoF head rotation. The display and its actuator are mounted on top of the leaning platform as shown Fig. 10.

3.5.2 Designing Movement Mapping Across Multiple Robots. To demonstrate the potential for a multiple-buildings scenario, we position the robot described above in an office situated in a different building 0.2 miles away from the building containing the lab space demonstrated in Section 3.2.1. In their VR space, the VR user sees a remote office in another building beyond a blue boundary, indicating that the office is separate from their current space. Utilizing the same navigation interface, the user can instantly move to the office using the teleport link. At that moment, to indicate that the user is leaving their social presence from the robot, the robot automatically returns to the hallway for parking. Upon the user's entry into the room, the other robot placed on a desk in the office indicates this re-embodiment by rotating the articulated display to face the collaborator in the office. The VR user can sit in a chair located in their local space (See Fig. 11). Owing to the fixed physical mapping of VRoxy for navigation, the user can consistently return to the same position and sit in the physical chair

Figure 11: Demonstrations of switching between robotic proxies across multiple buildings. Using VRoxy, the VR user can navigate from the lab space in Building B to the office in Building A where the fixed robot is located to meet another collaborator. On the other side, the mobile robot in Building B starts parking to indicate the leave, and the stationary robot in Building A turns the body to face the collaborator to signify the re-embodiment.

whenever they re-enter the office. The VR user can view the real-time feed from another 360-degree camera in the office and lean over to closely examine a breadboard, while the robot recreates this action through articulated animations. This demonstrates the VR user's ability to switch between environments, as the system automatically alternates between the two embodiment systems on the other end.

3.6 Implementation

We used the Unity engine for the VR interface and robotic embodiment control and rendering. We distributed processing and connected multiple client applications using Mirror Networking for Unity. To achieve low latency and high framerate in video streaming, we implemented a WebRTC sender/receiver using Unity Render Streaming [56]. Both blend shape data for facial expressions and eye gaze direction are provided by Quest Pro through the Movement SDK. We used Luos [33] as an interface between a Unity client and the Orbita [45] parallel spherical actuator used in the articulated head demonstrated in Section 3.1.1. Fig. 12 shows an overview of the software architecture of the VRoxy system.

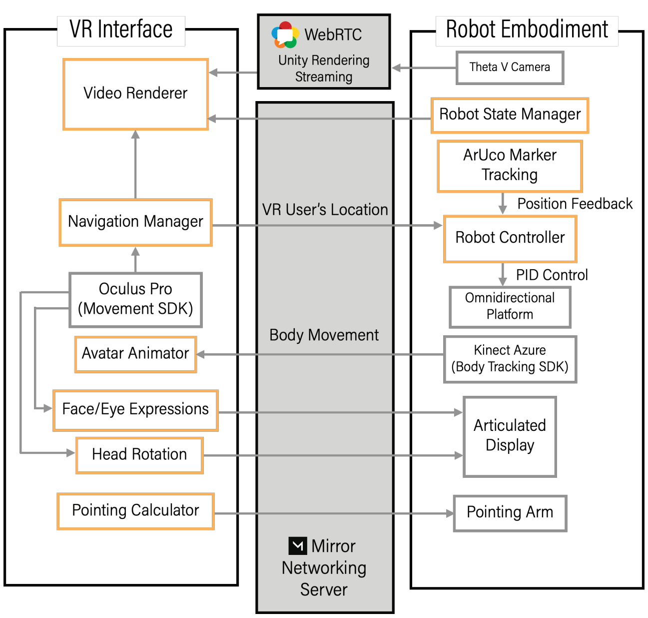

Figure 12: The overview of VRoxy software architecture

3.6.1 *Hardware.*

We used a similar set of hardware devices to the mobile robotic embodiment used in ReMotion [49]. To track the robot, we placed multiple ArUco boards on the ceiling of multiple rooms and (simulated) corridors. Note that there could be significant gaps in coverage as we used a wide-angle lens for tracking and the robot could tolerate a small amount of dead reckoning to move from one tracked area to the next. We placed a Kinect Azure camera in a corner of the large lab space for tracking a remote collaborator's body in the shared environment.

4 VISITING A LAB USING VROXY

To showcase the capabilities of VRoxy, we simulated a collaborative scenario where a collaborator remotely visits two adjacent labs using the system.

Fig. 13 illustrates a typical sequence of interactions employing VRoxy for remote collaboration in an extensive environment. Initially, a VR user greets a remote collaborator in the hallway (1). The remote VR user is able to engage in face-to-face communication through a live view. The VR user then enters a lab space, with the VR user following the navigation interface using a few simple steps. On the other end, the robot replicates these movements, entering the room to signify the VR user's position. The VR user recognizes the remote collaborator's location near a shelf via a rendered animated avatar in a 3D virtual space and approaches that spot for a discussion (2). The local participant can see the robot approaching to interact with them. Once the robot arrives, the VR user is presented with an immersive, real-time video feed (3). The VR user can

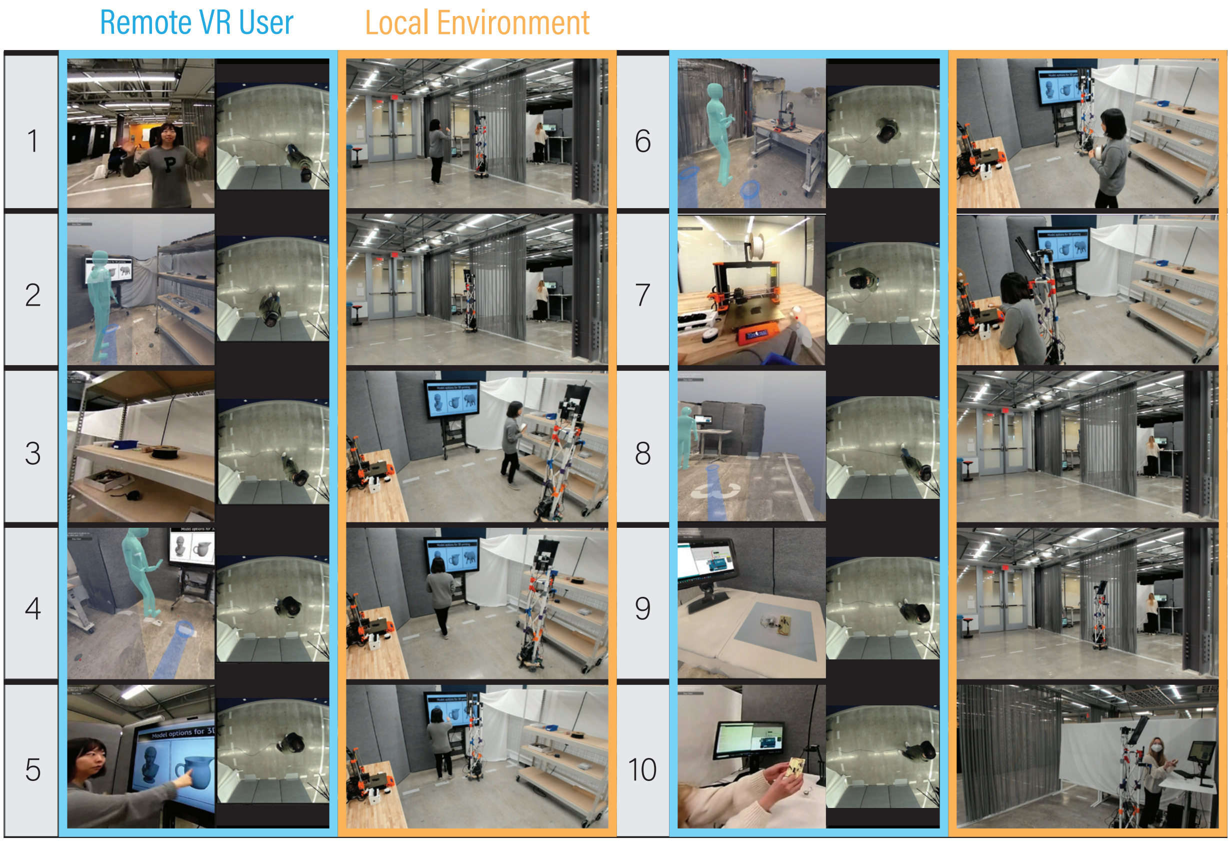

Figure 13: Demonstrations of navigating and completing collaborative design tasks using VRoxy. The VR user employs the blue teleport graphs on the ground to traverse a greater distance in the remote environment than in the local physical space. The user switches to a magnified view of objects or areas of interest (3, 5, 7, 10). The robot traverses the hallway to a different room in the same laboratory (9).

then point to one of the filaments on the shelf to indicate his preference within the live view, and the remote collaborator can pick it up, discerning the intended reference through the arm's pointing gesture. The VR user can make a small side-by-side movement to have a new view of the shelf from a different angle that becomes available after the robot adjusts its position. The collaborators now move to the monitor to discuss which 3D model to print on a 3D printer (4, 5). Then they go to the 3D printer placed at a table to do some printing tasks using non-verbal cues and gestures rendered through the articulated display and the pointing arm (6, 7). As the VR user completes the task with the collaborator, the VR user says goodbye to the collaborator and exits the lab space, then visits a smaller office through the hallway to initiate a discussion with the other collaborator on a prototype project (8). The robot executes these transitions for the VR user. While the robot moves into the room, the VR user can observe the real-time rendering of a task camera showing a breadboard and a shared screen with a circuit diagram and firmware programming to prepare for a discussion with the collaborator (9). Upon the robot's arrival, the VR user can point to a specific part for discussion while being able to have a face-to-face discussion.

5 FORMATIVE EVALUATION

We conducted a formative study to evaluate VRoxy's capability of enabling users to remotely navigate spacious environments within limited space and interact with remote collaborators, using the similar scenario demonstrated in Section 4. To investigate whether participants were aware of the spatial relationships of the remote space through VRoxy, we did not show the actual space until the end of the survey.

We recruited 6 participants (2 male, 4 female), ranging in age from 22 to 32 years old, from our institution, all of whom were compensated with $15. All participants had prior experience using videoconferencing tools such as Zoom, and only two had used VR headsets for remote communication before.

5.1 Procedure

In the task, participants put on the Quest Pro and explored the remote space (7.5m x 5m) in VR, while physically moving around their room (3m x 2m). The remote space featured two rooms of varying sizes and several task spaces. The task required participants to communicate with the instructor to observe how to operate a 3D printer. The instructor also moved around spaces and guided them in visiting multiple task areas by interacting with the robotic proxy. Prior to the task, participants had a training session of 15 minutes, which included watching demo videos of VRoxy and practicing navigation. During the training session, we presented pre-recorded mobile robot footage, avoiding any exposure to the room to prevent prior familiarity and evaluate the ability to create the proper conceptual model of the space after a task. A fisheye camera was attached to the ceiling and captured the position of participants as they put on the VR headset and moved to task spaces in VR.

After completing the task, participants were asked to fill out opened-ended questionnaires to provide their opinion about the usefulness of the various features of VRoxy. Afterward, the instructor requested to draw a simple map of the space where they had explored through VRoxy approximately 10 minutes prior. Finally, we conducted semi-structured interviews with participants to gain more insight into the effectiveness of the system.

5.2 Results

During the user study, the physical positions of VR users in different task spaces were recorded to show how the navigation system worked. We report the drawing results of the participants to show how users perceive the remote space in VR. Lastly, we summarize the answers to the open-ended questions and interviews.

5.2.1 The physical position of VR user. Based on the recorded videos, we found the consistent physical location of VR users at each task area, demonstrating the technique of mapping between asymmetric spaces.

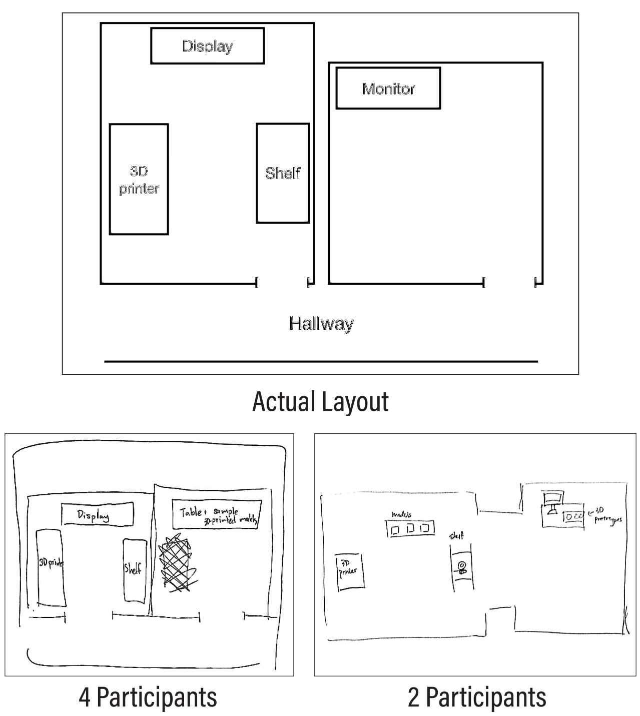

Figure 14: Selected drawings from participants reporting how they perceive the remote space after the study. They were clustered according to the consistency of the participants' depictions with the actual layout.

5.2.2 Drawings of VR user. Overall, all participants demonstrated an understanding of the spatial relationships of several task areas and accurately remembered the room configurations (See Fig. 14). However, a few participants (P2, P5) could not understand the spatial relationship between the two rooms and excluded the hallway. One potential reason could be that the hallway was long but the user had to teleport from one end of the hallway to another end immediately so they might have not been able to reliably gauge the physical distance in the actual environment.

5.2.3 Qualitative Feedback. With respect to the navigation view, some participants noted that the VR environment in the navigation view was similar to the live view, allowing for easy navigation through the space (P4) and not distracting them in transition (P6). Additionally, one participant (P3) argued that the presence of the avatar was enough to help them understand where the person was in the remote space. Yet P1 mentioned that it decreased their feelings of co-presence because *"my collaborator will turn into a blue avatar in navigation mode"*. It suggests that a full-body avatar that looks like an actual user would increase their feeling of being together. We tested a higher fidelity avatar created with AvatarSDK [1] to substitute the abstract blue avatar, but we noted that it can be socially awkward without facial expressions and requires creating an avatar for every participant in advance.

With respect to the live view, several participants mentioned that using the live feed allowed them to easily understand where the collaborator was (P4) and to see the whole body including the nonverbal cues (P6). Yet P5 reported experiencing discontinuity of communication due to the paused video during the micro-mobility. Some participants reported feeling distracted when transiting from the live view to the navigation view unintentionally, for example, by turning their head back (P1, P2). We offer a few suggestions on improving VRoxy's transition feature in Section 6.1.

As with our space compression feature, participants (P3, P4, P6) commented they could easily move in the remote space due to the walking interfaces. However, P1 stated the joystick would be more convenient in navigating a virtual environment. Nevertheless, the same participant (P1) also noted, *"it definitely reduces the possibility of motion sickness."*, which was the issue in the previous work using joysticks [25]. P2 commented that they found that the subtle movement provided the ability to easily move around each task area; *"I was able to make small adjustments to the robot position easily."*

It is noteworthy that even though our participants knew that the robot would represent themselves in the remote space, they were not much mindful of a robot representing them on the other side. P3 stated: *"I did not even remember it"*; and P6 commented *"I was so focused on myself barely thought of being a robot on the other side."* Nevertheless the waiting time for the robot in the navigation view was still a problem: P4 told us *"I kept thinking of whether my robot reached the spot that I'm at in the VR space."* and P6 said that *"(the waiting time) might be a little long."*

This is of course encouraging but it is important to note this was a formative study and that we did not directly evaluate the remote collaborators interacting with robots. Our focus was on the VR user's experience. Further research is needed to fully understand the impact of this system of representation on remote collaboration.

6 DISCUSSION AND FUTURE WORK

6.1 Blending Navigation and Live View

VRoxy combines two types of view modes for navigation and real-time interactions, as previously explored [57]. However, as commented by participants, this led to some inconsistency between these two view modes. While we implemented a fade-in-fade-out transition for blending views, there is room to explore other blending techniques, such as panoramic stitching and blending employed by Google Street View [4].

To increase spatial understanding in viewing a 360 image, we explored an approach of adding depth of field to these 360 images (Fig. 15). For this, we used a deep learning model, called Pano3D [3], that predicts a depth map image for a 360 image. We used this depth for displacing the sphere mesh to have geometry representing the 360 scenes. However, we found inconsistency between different 360 pictures in their depth maps, and it did not work well when a human was in the picture. These design iterations led us to create the separate navigation view mode described in Section 3.1.2.

Another approach to making a uniform interface would be to let the VR user interact with remote collaborators in the 3D model view, even after a robot arrives. For this, one could render real-time task and person spaces directly in the 3D model view. For example, RGB-D cameras, such as Kinect Azure, can be attached to the robot and present a point cloud to the VR user, although this approach may not provide a live view at as high fidelity as a 360 video does [40, 58].

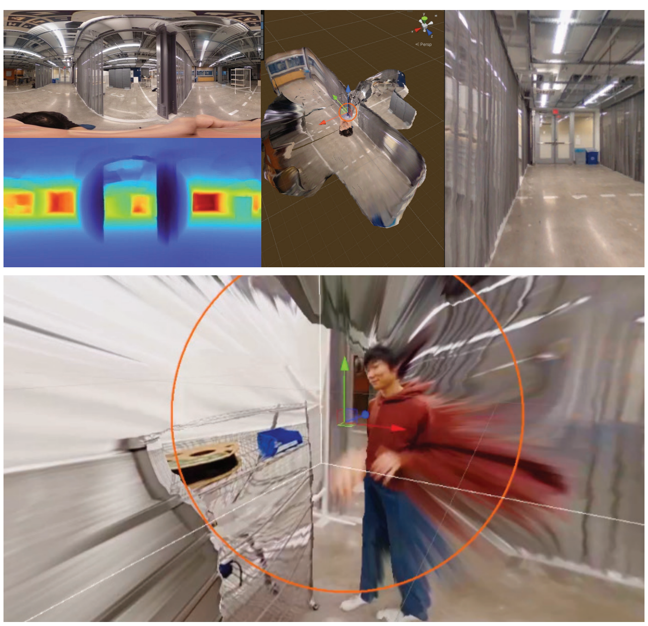

Figure 15: A test applying displacement to a 360 rendering using a depth map created with Pano3D (Top), a 360 scene including a complex object and a person that does not work well with a generated depth map (Bottom).

6.2 Support for Impromptu Social Interactions

One limitation of our navigation approach is the lack of awareness of the robot's location while it is in motion. This means that the VR user is unable to encounter people or have small chats in the same way as face-to-face communication. Certainly, rendering an avatar during navigation could help, but of course, the VR user will need to wait for the proxy to catch up to be able to engage with remote collaborators. This will be more problematic as the compression factor increases. One solution might be to design a faster robot by finding the shortest path or by increasing speed. Another might be to implement a way for the robot to send signals to the VR users, indicating when it is in a location where impromptu social interactions are possible.

6.3 Flexible Robot Control for New Environment and Obstacles

One of the main drawbacks of our setup is that it takes some time to configure a new site: AR Tags need to be installed in the ceiling, and the different rooms and corridors need to be scanned. Additionally, if there are any changes to the environment, such as the addition of furniture, the robot is unaware of them.

To address this challenge, a navigation method combined with algorithms such as YOLO [59] and SLAM [37] could be adapted for the robot to be aware of surrounding objects and humans while navigating the space. With the use of a camera or a depth camera like the Kinect Azure, we can create a map of the environment that can later be used to control the robot in the mapped environment. This approach can eliminate the need for ArUCo markers and provide more flexible navigation around obstacles and humans.

6.4 Versatile Camera Position

The mobile robot currently implemented in VRoxy is equipped with a camera fixed at a certain height. However, introducing an adjustable height system for both the camera and the robot itself could significantly enhance the user experience. This feature would empower remote users to more closely examine objects or switch interaction modalities–like transitioning from a standing to a seated position at a table. Furthermore, it would provide the flexibility needed to accommodate users of varying heights.

6.5 Virtual versus Physical Embodiment

Our formative evaluation explored VRoxy's efficacy in enabling VR users to navigate and collaborate through a mobile robot in a larger remote space while understanding the structure of the space. Moving forward, bidirectional studies will be needed to comprehend how robotic movement, as a mediator, influences both interactions and remote users' perceptions of the robot as their VR collaborator's physical representation. These insights can inform a more dynamic and responsive robotic design. For instance, enhancing the VRoxy's robot with capabilities such as robotic arms could allow remote users to manipulate real-world objects and convey more intricate gestural signals in physical space.

In our VRoxy system, we aimed to accommodate an asymmetric setup so that a VR user can use a small space to explore a wide space. However, this also means that the VR user and remote collaborator see a different representation of each other. This asymmetry in the rendering may raise some research questions. One central question arising from this asymmetry is whether physical robots or virtual avatars are the better representation of the VR user for the collaborator on the other side. For example, a physical robot may provide a more realistic experience of co-presence for the collaborator with the robot but may limit the range of movements and interactions. In contrast, a virtual avatar may provide more flexibility in terms of movements and interactions but may not provide the same level of presence as a physical robot. Better understanding the impact of these different representations on collaboration and communication via in-depth study can allow one to optimize the VRoxy system to provide the most effective collaborative experience possible for asymmetric scenarios.

7 CONCLUSIONS

We introduce the VRoxy system, designed to facilitate dynamic remote collaboration by allowing a VR user to explore and interact with collaborators through an automatic robotic proxy. VRoxy employs a novel approach to map the VR user's body positioning in a small space to the robot's movement in a larger environment. The system blends a quick navigation through a pre-scanned view with 360-degree live views during direct collaboration. Additionally, VRoxy captures multiple modalities of non-verbal cues from the VR user through Quest Pro, including head rotation, facial expressions, eye gaze, and pointing gestures. These cues are then represented through a robotic embodiment at the remote location. With VRoxy, a collaborator can not only navigate and collaborate within a vast remote environment from a smaller physical space but also instantly switch between separate workspaces that are physically apart.

ACKNOWLEDGMENTS

This research was supported by NSF Award IIS-1925100, and the Nakajima Foundation. We would like to thank itSeez3D for providing a free license to use AvatarSDK [1] for research purposes. Our gratitude goes out to Wan Qing Chua and Tao Long for helping with an early prototype, Andy Wilson for his insight on using multiple Kinect devices, and to Rachel Lynne Witzig and Nels Numan for providing feedback to refine the paper.

REFERENCES

[1] 2022. Lifelike avatars for the metaverse. https://avatarsdk.com/

[2] S. O. Adalgeirsson and C. Breazeal. 2010. MeBot: A robotic platform for socially embodied telepresence. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 15–22. https://doi.org/10.1109/HRI.2010.5453272

[3] Georgios Albanis, Nikolaos Zioulis, Petros Drakoulis, Vasileios Gkitsas, Vladimiros Sterzentsenko, Federico Alvarez, Dimitrios Zarpalas, and Petros Daras. 2021. Pano3D: A Holistic Benchmark and a Solid Baseline for 360° Depth Estimation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 3722–3732. https://doi.org/10.1109/CVPRW53098.2021.00413

[4] Dragomir Anguelov, Carole Dulong, Daniel Filip, Christian Frueh, Stéphane Lafon, Richard Lyon, Abhijit Ogale, Luc Vincent, and Josh Weaver. 2010. Google Street View: Capturing the World at Street Level. *Computer* 43 (2010). http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5481932&tag=1

[5] Ignacio Avellino, Cédric Fleury, Wendy E. Mackay, and Michel Beaudouin-Lafon. 2017. CamRay: Camera Arrays Support Remote Collaboration on Wall-Sized Displays. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6718–6729. https://doi.org/10.1145/3025453.3025604

[6] Mahdi Azmandian, Timofey Grechkin, Mark T Bolas, and Evan A Suma. 2015. Physical Space Requirements for Redirected Walking: How Size and Shape Affect Performance.. In *ICAT-EGVE*. 93–100.

[7] Beam. 2023. Suitable Technologies Inc. Retrieved Feb 6, 2023 from https://suitabletech.com

[8] Costas Boletsis and Jarl Erik Cedergren. 2019. VR locomotion in the new era of virtual reality: an empirical comparison of prevalent techniques. *Advances in Human-Computer Interaction* 2019 (2019).

[9] Bill Buxton. 2009. *Mediaspace – Meaningspace – Meetingspace*. Springer London, London, 217–231. https://doi.org/10.1007/978-1-84882-483-6_13

[10] William A. S. Buxton. 1992. Telepresence: Integrating Shared Task and Person Spaces. In *Proceedings of the Conference on Graphics Interface '92* (Vancouver, British Columbia, Canada). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 123–129.

[11] Zekun Cao, Jason Jerald, and Regis Kopper. 2018. Visually-induced motion sickness reduction via static and dynamic rest frames. In *2018 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE, 105–112.

[12] Weiya Chen, Anthony Plancoulaine, Nicolas Férey, Damien Touraine, Julien Nelson, and Patrick Bourdot. 2013. 6DoF navigation in virtual worlds: comparison of joystick-based and head-controlled paradigms. In *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology*. 111–114.

[13] Kayla Davis, Taylor Hayase, Irene Humer, Brandon Woodard, and Christian Eckhardt. 2022. A Quantitative Analysis of Redirected Walking in Virtual Reality Using Saccadic Eye Movements. In *Advances in Visual Computing*, George Bebis, Bo Li, Angela Yao, Yang Liu, Ye Duan, Manfred Lau, Rajiv Khadka, Ana Crisan, and Remco Chang (Eds.). Springer Nature Switzerland, Cham, 205–216.

[14] Mohamed Elobaid, Yue Hu, Giulio Romualdi, Stefano Dafarra, Jan Babic, and Daniele Pucci. 2020. Telexistence and Teleoperation for Walking Humanoid Robots. In *Intelligent Systems and Applications*, Yaxin Bi, Rahul Bhatia, and Supriya Kapoor (Eds.). Springer International Publishing, Cham, 1106–1121.

[15] Mehrad Faridan, Bheesha Kumari, and Ryo Suzuki. 2023. ChameleonControl: Teleoperating Real Human Surrogates through Mixed Reality Gestural Guidance for Remote Hands-on Classrooms.. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3544548.3581449

[16] Lars Fritsche, Felix Unverzag, Jan Peters, and Roberto Calandra. 2015. First-person tele-operation of a humanoid robot. In *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*. 997–1002. https://doi.org/10.1109/HUMANOIDS.2015.7363475

[17] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. In Touch with the Remote World: Remote Collaboration with Augmented Reality Drawings and Virtual Navigation. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology* (Edinburgh, Scotland) *(VRST '14)*. Association for Computing Machinery, New York, NY, USA, 197–205. https://doi.org/10.1145/2671015.2671016

[18] William W. Gaver, John Bowers, Andrew Boucher, Hans Gellerson, Sarah Pennington, Albrecht Schmidt, Anthony Steed, Nicholas Villars, and Brendan Walker. 2004. The Drift Table: Designing for Ludic Engagement. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems* (Vienna, Austria) *(CHI EA '04)*. Association for Computing Machinery, New York, NY, USA, 885–900. https://doi.org/10.1145/985921.985947

[19] C. Gutwin and S. Greenberg. 2004. A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Computer Supported Cooperative Work (CSCW)* 11 (2004), 411–446.

[20] 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* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173933

[21] Keita Higuchi and Jun Rekimoto. 2013. Flying head: a head motion synchronization mechanism for unmanned aerial vehicle control. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 2029–2038.

[22] Jim Hollan and Scott Stornetta. 1992. Beyond Being There. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, USA) *(CHI '92)*. Association for Computing Machinery, New York, NY, USA, 119–125. https://doi.org/10.1145/142750.142769

[23] Xandex Inc. 2023. KUBI Telepresence Robot. Retrieved Feb 6, 2023 from https://kubiconnect.com/

[24] Victoria Interrante, Brian Ries, and Lee Anderson. 2007. Seven League Boots: A New Metaphor for Augmented Locomotion through Moderately Large Scale Immersive Virtual Environments. In *2007 IEEE Symposium on 3D User Interfaces*. https://doi.org/10.1109/3DUI.2007.340791

[25] Brennan Jones, Yaying Zhang (yaying zhang), Priscilla N. Y. Wong, and Sean Rintel. 2021. Belonging There: VROOM-ing into the Uncanny Valley of XR Telepresence. In *CSCW 2021*. ACM. https://www.microsoft.com/en-us/research/publication/belonging-there-vroom-ing-into-the-uncanny-valley-of-xr-telepresence/

[26] Wendy Ju, Lawrence Neeley, Terry Winograd, and Larry Leifer. 2006. *Thinking with Erasable Ink: Ad-hoc Whiteboard Use in Collaborative Design*. Technical

Report.

[27] David Kirk and Danae Fraser. 2006. Comparing remote gesture technologies for supporting collaborative physical tasks, Vol. 2. 1191–1200. https://doi.org/10.1145/1124772.1124951

[28] Sven Kratz and Fred Rabelo Ferriera. 2016. Immersed remotely: Evaluating the use of Head Mounted Devices for remote collaboration in robotic telepresence. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 638–645. https://doi.org/10.1109/ROMAN.2016.7745185

[29] Changyang Li, Haikun Huang, Jyh-Ming Lien, and Lap-Fai Yu. 2021. Synthesizing Scene-Aware Virtual Reality Teleport Graphs. *ACM Trans. Graph.* 40, 6, Article 229 (dec 2021), 15 pages. https://doi.org/10.1145/3478513.3480478

[30] Jiannan Li, Maurício Sousa, Chu Li, Jessie Liu, Yan Chen, Ravin Balakrishnan, and Tovi Grossman. 2022. ASTEROIDS: Exploring Swarms of Mini-Telepresence Robots for Physical Skill Demonstration. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 111, 14 pages. https://doi.org/10.1145/3491102.3501927

[31] Yi-Jun Li, Frank Steinicke, and Miao Wang. 2022. A comprehensive review of redirected walking techniques: Taxonomy, methods, and future directions. *Journal of Computer Science and Technology* 37, 3 (2022), 561–583.

[32] Paul Luff and Christian Heath. 1998. Mobility in Collaboration. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work* (Seattle, Washington, USA) *(CSCW '98)*. Association for Computing Machinery, New York, NY, USA, 305–314. https://doi.org/10.1145/289444.289505

[33] Luos. 2023. *Luos: Open-source and real-time orchestrator for cyber-physical-systems, to easily design, test and deploy embedded applications and digital twins*. https://www.luos.io

[34] Michal Luria, Samantha Reig, Xiang Zhi Tan, Aaron Steinfeld, Jodi Forlizzi, and John Zimmerman. 2019. Re-Embodiment and Co-Embodiment: Exploration of social presence for robots and conversational agents. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 633–644.

[35] Daniel R. Montello. 1998. A New Framework for Understanding the Acquisition of Spatial Knowledge in Large-Scale Environments. https://api.semanticscholar.org/CorpusID:40864171

[36] Pieter Moors, Filip Germeys, Iwona Pomianowska, and Karl Verfaillie. 2015. Perceiving where another person is looking: the integration of head and body information in estimating another personâs gaze. *Frontiers in Psychology* 6 (2015). https://doi.org/10.3389/fpsyg.2015.00909

[37] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. 2015. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE transactions on robotics* 31, 5 (2015), 1147–1163.

[38] Inc. Niantic. 2023. Scaniverse 3D Scanner with LiDAR for iPhone and iPad. Retrieved Feb 6, 2023 from https://scaniverse.com/

[39] K. Otsuka. 2016. MMSpace: Kinetically-augmented telepresence for small group-to-group conversations. In *2016 IEEE Virtual Reality (VR)*. 19–28. https://doi.org/10.1109/VR.2016.7504684

[40] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew Wilson. 2016. Room2Room: Enabling Life-Size Telepresence in a Projected Augmented Reality Environment. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. ACM, New York, NY, USA, 1716–1725. https://doi.org/10.1145/2818048.2819965

[41] Thammathip Piumsomboon, Gun A. Lee, Jonathon D. Hart, Barrett Ens, Robert W. Lindeman, Bruce H. Thomas, and Mark Billinghurst. 2018. Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 46, 13 pages. https://doi.org/10.1145/3173574.3173620

[42] Irene Rae, Bilge Mutlu, and Leila Takayama. 2014. Bodies in Motion: Mobility, Presence, and Task Awareness in Telepresence. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 2153–2162. https://doi.org/10.1145/2556288.2557047

[43] Pranav Rane, Varun Mhatre, and Lakshmi Kurup. 2014. Study of a home robot: Jibo. *International journal of engineering research and technology* 3, 10 (2014), 490–493.

[44] Bernhard E Riecke, Bobby Bodenheimer, Timothy P McNamara, Betsy Williams, Peng Peng, and Daniel Feuereissen. 2010. Do we need to walk for effective virtual reality navigation? physical rotations alone may suffice. In *Spatial Cognition VII: International Conference, Spatial Cognition 2010, Mt. Hood/Portland, OR, USA, August 15-19, 2010. Proceedings 7*. Springer, 234–247.

[45] Pollen Robotics. 2023. *Orbita: A 3D joint for robotic head motion realism*. https://medium.com/pollen-robotics/orbita-is-turning-heads-literally-d10d378550e2

[46] Roy A. Ruddle, Ekaterina Volkova, and Heinrich H. Bülthoff. 2011. Walking Improves Your Cognitive Map in Environments That Are Large-Scale and Large in Extent. *ACM Trans. Comput.-Hum. Interact.* 18, 2, Article 10 (jul 2011), 20 pages. https://doi.org/10.1145/1970378.1970384

[47] Mose Sakashita, Tatsuya Minagawa, Amy Koike, Ippei Suzuki, Keisuke Kawahara, and Yoichi Ochiai. 2017. You as a Puppet: Evaluation of Telepresence User Interface for Puppetry. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 217–228. https://doi.org/10.1145/3126594.3126608

[48] Mose Sakashita, E. Andy Ricci, Jatin Arora, and François Guimbretière. 2022. RemoteCoDe: Robotic Embodiment for Enhancing Peripheral Awareness in Remote Collaboration Tasks. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 63 (apr 2022), 22 pages. https://doi.org/10.1145/3512910

[49] Mose Sakashita, Ruidong Zhang, Xiaoyi Li, Hyunju Kim, Michael Russo, Cheng Zhang, Malte F. Jung, and François Guimbretière. 2023. ReMotion: Supporting Remote Collaboration in Open Space with Automatic Robotic Embodiment.. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3544548.3580699

[50] MHD Yamen Saraiji, Tomoya Sasaki, Reo Matsumura, Kouta Minamizawa, and Masahiko Inami. 2018. Fusion: full body surrogacy for collaborative communication. In *ACM SIGGRAPH 2018 Emerging Technologies*. 1–2.

[51] Ehsan Sayyad, Misha Sra, and Tobias Höllerer. 2020. Walking and teleportation in wide-area virtual reality experiences. In *2020 IEEE international symposium on mixed and augmented reality (ISMAR)*. IEEE, 608–617.

[52] Rajinder S. Sodhi, Brett R. Jones, David Forsyth, Brian P. Bailey, and Giuliano Maciocci. 2013. BeThere: 3D Mobile Collaboration with Spatial Input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. ACM, New York, NY, USA, 179–188. https://doi.org/10.1145/2470654.2470679

[53] Frank Steinicke, Gerd Bruder, Jason Jerald, Harald Frenz, and Markus Lappe. 2010. Estimation of Detection Thresholds for Redirected Walking Techniques. *IEEE Transactions on Visualization and Computer Graphics* 16, 1 (2010), 17–27. https://doi.org/10.1109/TVCG.2009.62

[54] Susumu Tachi, Yasuyuki Inoue, and Fumihiro Kato. 2020. TELESAR VI: Telexistence Surrogate Anthropomorphic Robot VI. *International Journal of Humanoid Robotics* 17, 5 (Oct 2020), 2050019(1–33).

[55] Anthony Tang, Melanie Tory, Barry Po, Petra Neumann, and Sheelagh Carpendale. 2006. Collaborative Coupling over Tabletop Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI '06)*. Association for Computing Machinery, New York, NY, USA, 1181–1190. https://doi.org/10.1145/1124772.1124950

[56] Unity Technologies. 2023. *Unity Render Streaming*. https://github.com/Unity-Technologies/UnityRenderStreaming

[57] 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* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300431

[58] 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. https://doi.org/10.1145/3332165.3347872

[59] Yunong Tian, Guodong Yang, Zhe Wang, Hao Wang, En Li, and Zize Liang. 2019. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Computers and electronics in agriculture* 157 (2019), 417–426.

[60] Katherine M. Tsui and Holly A. Yanco. 2013. Design Challenges and Guidelines for Social Interaction Using Mobile Telepresence Robots. *Reviews of Human Factors and Ergonomics* 9, 1 (2013), 227–301. https://doi.org/10.1177/1557234X13502462 arXiv:https://doi.org/10.1177/1557234X13502462

[61] Roel Vertegaal. 1999. The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 294–301. https://doi.org/10.1145/302979.303065