

DualStream: Spatially Sharing Selves and Surroundings using Mobile Devices and Augmented Reality

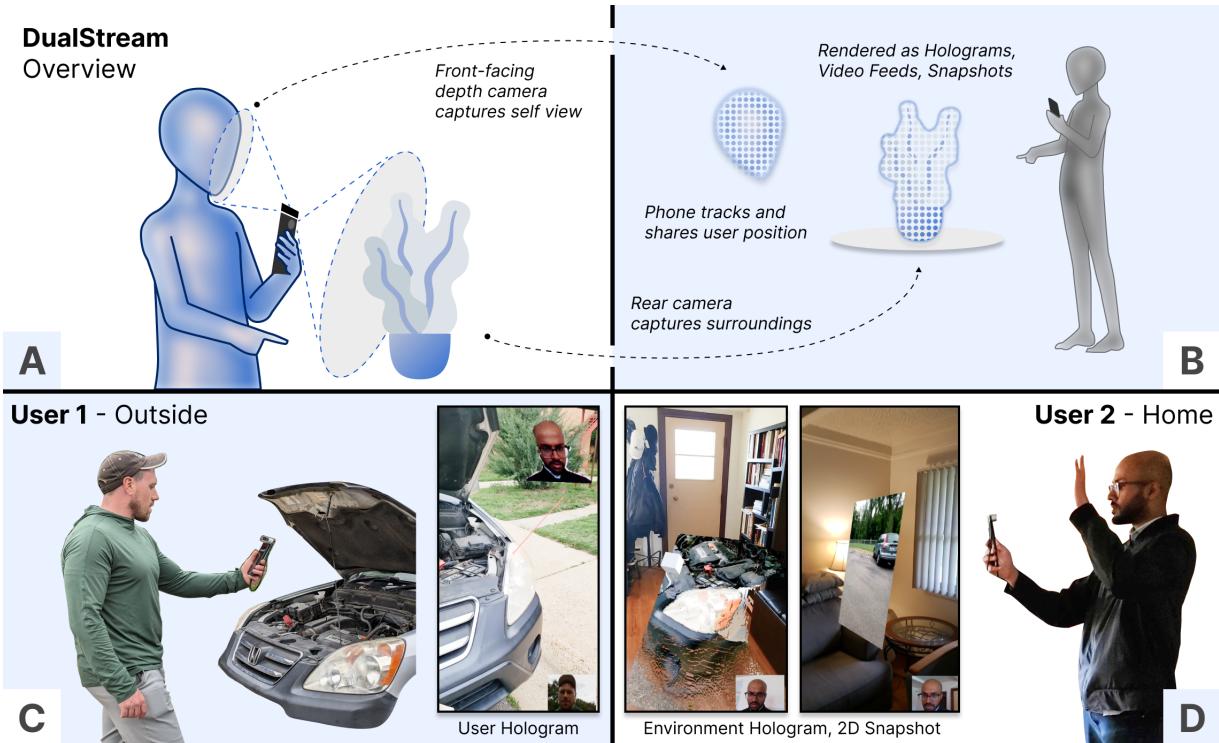
Rishi Vanukuru^{1*}Suibi Che-Chuan Weng¹
Mark D. Gross¹Krithik Ranjan¹
Ellen Yi-Luen Do^{1†}Torin Hopkins¹
Amy Banic²¹ ATLAS Institute, University of Colorado Boulder, USA² Interactive Realities Lab, University of Wyoming, USA

Figure 1: An overview of the DualStream system. (A & B) An illustration of how the front and rear cameras of a mobile device can be used to share information about self and surroundings with remote collaborators. The captured information is used to create 3D Holograms, Spatial Video Feeds, and Environment Snapshots. (C) DualStream being used to share information about a car's engine, while viewing a remote expert as if they were in the same location. (D) The expert viewing the shared 3D hologram of the car, and a 2D snapshot of the surrounding remote environment.

ABSTRACT

In-person human interaction relies on our spatial perception of each other and our surroundings. Current remote communication tools partially address each of these aspects. Video calls convey real user representations but without spatial interactions. Augmented and Virtual Reality (AR/VR) experiences are immersive and spatial but often use virtual environments and characters instead of real-life representations. Bridging these gaps, we introduce *DualStream*, a system for synchronous mobile AR remote communication that captures, streams, and displays spatial representations of users and their surroundings. DualStream supports transitions between user and environment representations with different levels of visuospatial fidelity, as well as the creation of persistent shared spaces using

environment snapshots. We demonstrate how DualStream can enable spatial communication in real-world contexts, and support the creation of blended spaces for collaboration. A formative evaluation of DualStream revealed that users valued the ability to interact spatially and move between representations, and could see DualStream fitting into their own remote communication practices in the near future. Drawing from these findings, we discuss new opportunities for designing more widely accessible spatial communication tools, centered around the mobile phone.

Index Terms: Human-centered computing—Mixed/augmented reality; Human-centered computing—Collaborative Interaction; Human-centered computing—Mobile computing

1 INTRODUCTION

Human communication is fundamentally tied to our perception of one another in a shared physical environment. We constantly process and add to a three-dimensional (3D) audio-visual canvas of gestures and conversations. Movement is also a key aspect, as embodied cognition [15] and proxemics shape the way we navigate space in

*Email: rishi.vanukuru@colorado.edu

†Email: ellen.do@colorado.edu

social contexts [24]. With remote communication becoming increasingly integral to work and personal life, a key challenge is to develop tools that can match the experience of in-person interactions. Video-conferencing is one of the most common forms of audio-visual remote communication today, and enables synchronous conversations using real-world representations of ourselves and the environments we inhabit. Presenting this information on two-dimensional (2D) screens, however, makes it difficult to establish a shared frame of reference and interact in a more spatial manner [34]. Recent advances in Augmented and Virtual Reality (AR/VR) devices, as well as camera technology that can simultaneously capture color and depth information, have resulted in systems where complete human representations and entire physical environments can be streamed and interacted with remotely in 3D [35, 44]. While such experiences may come close to replicating in-person communication, the systems involved are expensive, confined to specific locations, and rely on devices such as head-worn displays that are still relatively unfamiliar to wider audiences. In contrast, mobile AR applications have given millions of people around the world a glimpse of the power of spatially interacting with information [8]. By integrating the advanced video and depth capture technology present in modern mobile devices, and the spatial interaction capabilities of mobile AR applications, mobile devices are well-positioned to bridge the gap between spatial remote communication and wider availability.

To demonstrate this potential, we develop *DualStream*—a mobile remote communication platform where users share 3D views of each other and their surroundings, and can interact with these representations spatially in AR (Fig. 1). DualStream’s hardware comprises a mobile device with an externally-mounted front-facing depth camera. DualStream captures two “streams” of 3D information from the front and rear cameras, corresponding to the user and their immediate surroundings. Using these streams, we create remote 3D representations of (1) ourselves, which look and move as we do in real life, and (2) our surroundings, that are spatially consistent with their real-world locations. We develop a series of features such as seamlessly moving between fixed and spatial views, and using environment snapshots to create a persistent shared space for remote communication. Thus, DualStream enables users to simultaneously feel they are “being there” in a remote location (by spatially interacting with shared environment views), and that remote participants are “being here” in their local environment (as independently moving user representations in space). We demonstrate that unlike fixed volumetric capture setups tethered to single rooms and reliant on expensive hardware, DualStream can enable the sharing of selves and spaces, anywhere and anytime. DualStream leverages the familiarity of personal mobile computing, and provides a more spatial and immersive experience than the status quo of video conferencing.

We conducted a formative evaluation of DualStream with users in the lab and outside, to understand how people compare mobile spatial communication with their current practices of remote collaboration. Findings from this study showcase the potential for DualStream to enhance everyday spatial communication, and provide insight into areas which require improvement before such experiences become widely adopted. Our key contributions are:

1. The development of a mobile-based remote communication platform—DualStream—that can simultaneously share spatial representations of users and their environments, with features that enable users to combine 2D and 3D representations of self and surroundings,
2. Real-world scenarios which showcase the applicability of mobile spatial communication to a range of contexts, and
3. Findings from a formative evaluation of DualStream in the lab and in the real world, showcasing its potential and highlighting challenges and opportunities for future work.

2 RELATED WORK

DualStream builds upon prior work on AR/VR collaboration, video conferencing, as well as the specific use of mobile devices to support both these forms of communication. We focus on systems that consider more realistic, real-time representations of users and their environments. We also draw on Harrison and Dourish’s distinction between interaction in “space”—using direct geometric arrangements or metaphors to make sense of interpersonal communication—and in a specific “place”—locations imbued with contextual and social meaning [11, 18].

2.1 Bringing Real-World Information into Immersive Collaboration

Immersive collaboration takes place in shared digital worlds via VR or by making use of each user’s local environment via AR. Interaction in space is a necessary characteristic of such immersive systems. However, questions of how best to represent one’s real self, and incorporate aspects of real-world places, still remain.

2.1.1 Real-World Selves

Early work by Bailenson et al. [2] demonstrated the value of using realistic representations of collaborators in virtual environments. Many studies have since further highlighted the beneficial impact of realism and fidelity [26, 47, 51] on presence and collaboration. Commercial immersive communication tools such as Spatial¹ also offer users the option to create realistic avatars. Another approach is to capture and stream real-time visual information of users. An early system by Billinghurst and Kato involving collaboration via AR headsets [4] overlaid real video feeds of users over fiducial markers. There have been further efforts to integrate video calls with immersive environments [43]. Recent improvements in 3D content capture have enabled systems where full-body information can be streamed and reconstructed remotely in real-time [16, 25, 35, 44, 55].

2.1.2 Real-World Places

Using 360-degree video capture, VR systems have been able to provide more complete pictures of local environments to remote users inhabiting a fixed perspective [56], moving based on local user control [39], or independently using telepresence robots [23]. With room-scale depth capture, projects such as XSpace [20] and Re-locations [14] enable the creation of blended locations for remote, multi-view collaboration. Systems such as Loki [44] and Holoporation [35] are also capable of streaming environmental depth information. Recent surveys have highlighted the importance of situated spatial collaboration in the context of remote assistance and collaboration on physical tasks [13, 49].

Taken together, these immersive systems have brought us close to replicating real-world environments and interactions in real-time. However, challenges to depth-based spatial communication such as scalability and transmission reliability are yet to be solved [38], and the devices required for capture and display are expensive, relatively unfamiliar, and far from widespread use.

2.1.3 Mobile Spatial Collaboration

Mobile devices have proven to be ideal platforms for making immersive experiences available to wider audiences today [8, 9]. Mobile AR applications are now commonplace, and have provided millions of people a glimpse of spatial computing via games and social apps like Snapchat² and Pokemon Go³. When considering collaborative mobile AR experiences, research has explored their use in co-located contexts, for games [3], social activities [9], and spatial problem solving [17, 50]. Mobile AR has also been used to support remote

¹Spatial: spatial.io

²Snapchat: snapchat.com

³Pokemon Go: pokemongolive.com

collaboration [10, 31] in the contexts of lab-based activities [46] and design critiques [28]. Work by Young et al. on mobile telepresence has considered the ways in which mobile devices can share user and environment representations in real-time, first via interaction in three degrees of freedom [52], and subsequently using 360 degree cameras mounted on mobile devices in the Mobileportation project [53].

2.2 Making Video Conferencing more Spatial

Similar to how immersive collaboration incorporates aspects of space by default, video conferencing tools rely on capturing and streaming real-world audio-visual information of remote participants in real-time. Prior research has studied the impact of screen-based spatial metaphors on the sense of space and co-presence [19]. These metaphors include viewing remote video feeds around conference tables [45], and enabling video feeds to move in and out of virtual rooms to indicate informal and formal meeting spaces [36]. While these systems offer a strong virtual frame of reference, the visual information is still limited to being viewed on fixed 2D screens, and the resulting sense of spatial interaction is inferred rather than direct. One way in which video conferencing has been made more spatial is by having displays and cameras that literally move in space. Through cameras controlled by remote users [40], robotic-arm-mounted displays that mirror head gestures [42], and telepresence robots [27, 32], studies have investigated how spatial cameras and displays can impart a greater sense of co-presence and a common shared reference frame. However, like the more advanced AR/VR collaboration systems discussed earlier, these spatial video conferencing setups are also cumbersome, expensive, and difficult to scale to more general communication. In contrast, mobile devices have brought video calling to the masses. Mobile video calls support informal “small talk” interactions, task-oriented functional conversations, as well as “show and talk” interactions where real-world objects are the focus [33]. The presentation of spatial information on fixed 2D screens, however, presents challenges around the asymmetries of control, participation, and awareness [22].

2.3 What DualStream does Differently

Through DualStream, we enable users to share representations of their selves and surroundings, and interact with remote collaborators in a spatial manner. Thus, DualStream replicates features provided by more elaborate volumetric setups [35, 44] and goes beyond existing mobile AR systems by focusing on the spatial sharing of both self and place-based information. A key difference between DualStream and prior systems such as Mobileportation and ARCritique [28, 52, 53] is the ability to share 3D information about oneself. DualStream offers more control over what is being shared, and enables users to freely combine 2D and 3D information depending on the context and level of detail required. DualStream also employs a fundamentally different scheme of capturing, compressing, streaming, and rendering depth information. We colorize and compress two streams of depth simultaneously (self and surroundings) and stream them using conventional video networks. The freedom to use DualStream anywhere and anytime presents possibilities not available in immersive AR/VR systems. Further, we lean into the familiarity of mobile devices and videoconferencing, by enabling users to seamlessly switch between spatial AR calls and screen-based video calls. By presenting information about self and shared places in AR, DualStream helps provide a strong frame of reference for inherently spatial media. The ability to visualize video feeds in 3D enables DualStream to simulate the notion of displays moving in space, as employed by related work on spatial video calls [42]. When taken together, DualStream enables spatial remote communication in personal spaces, truly remote assistance, and supports the creation of blended spaces for interaction.

3 THE DUALSTREAM SYSTEM

DualStream supports **spatial interaction** via mobile AR, and enables users to share various representations of **self** and **place** (their local surroundings) with remote collaborators. The implementation of this prototype consists of a mobile device with an externally mounted, front-facing depth camera (Intel RealSense D435) directly connected to the device via a USB-C cable. Although some modern mobile devices are already equipped with front and rear-facing depth cameras, it is not yet possible to access depth data while also capturing color video and running AR experiences. Therefore, we use the externally mounted camera to explore capabilities that will likely be achievable by on-device solutions in the near future. Even in this form, the phone and depth camera setup is not tethered to a specific location, and is more readily available than complex volumetric capture systems involving AR/VR headsets and multiple cameras.

3.1 Establishing a Shared Spatial Frame of Reference

When starting a call using DualStream, users must first scan horizontal surfaces in their local environment to help the phone track its own position and orientation. Users then touch an appropriate location on the screen to place the “anchor” object (represented as a small blue cube) for the shared room. The position and orientation of the user’s phone relative to the anchor is then streamed to any remote participants. This data helps to consistently position the shared user and environment representations.

3.2 Sharing Realistic Representations of Self

Once users have set up the shared reference frame, they can use the buttons on the interface to begin sharing representations of themselves. These representations are placed and moved based on user movement, thereby creating remote representations that look and move the way we do. The two main user representations in DualStream are (1) **3D Hologram**, where the local user sees a point-cloud representation of the remote user’s face, and (2) **Spatial Video**, where the front-facing camera feed of the remote user’s head and shoulders is displayed in real-time, with or without the remote background (Fig. 1C). As with other video-conferencing applications, users can see their self-view in the form of a small video in the bottom-right corner of the screen. Users have complete control over which representation of themselves is shared with remote participants, and can also turn off their representation, displaying a small white cube instead. These representations enable users to feel like the person they are talking with is in the same space as they are.

3.3 Sharing Real-world Environments

Users can also share their environments in real time. Similar to the self representations, users have full control over what environment representation they share remotely. DualStream offers two options for sharing environment information (Fig. 1D):

1. **Environment Hologram:** A 3D representation of the environment is projected as a point cloud from the position of the remote user, to appear to be situated in the local space. As the remote user moves around, this point cloud also moves to represent the changing surroundings.
2. **Environment Video Feed:** A user shares the view of their immediate surroundings as seen on the phone screen. The local user sees a rectangular frame with this video feed attached to the remote user at a fixed distance from their self-representation, mapped to the perspective and field of view of the remote user. As the remote user moves, this rectangular frame moves around with them, functioning as a portal into the remote space. This representation is useful for sharing parts of the environment that are primarily 2D in nature or distant objects whose depth cannot be accurately captured.

3.4 Key Features

DualStream consists of a basic user interface with a row of buttons in the bottom-left of the screen corresponding to the user and environment representations. By pressing these buttons, a secondary set of options appears, allowing users to switch between representations, take snapshots, activate the front-facing depth camera, and re-position the anchor object (Fig. 3L).

3.4.1 Environment Snapshots

Users can only share real-time information about the environment directly in front of themselves, as viewed by the phone's rear camera. DualStream enables users to take **snapshots** of the environment (as a video frame or hologram) in its current state. These snapshots then persist in the location where they were taken in the shared remote space. This can be used to freeze video frames or holograms in place for later discussion without the live feed (Fig. 1D). By taking multiple snapshots, users can share a greater amount of their local environment in a spatially persistent manner. Once a user places a snapshot of their environment in their collaborator's view, they see a spatially anchored annotation indicating the location from where the snapshot was taken (as a small cube), and the visual area that was covered by the snapshot (as a semi-transparent plane). This helps provide feedback about previously shared information and aids in the creation of multiple snapshots of continuous scenes. These annotations and snapshots can be selectively displayed or hidden, based on the context of the conversation.

3.4.2 Pointing

Users can point at remotely shared environments and snapshots by touching the phone display. This triggers a laser pointer that the remote user can view in their own local space. The pointer helps approximate a basic deictic gesture and aids with spatial referencing in the shared environment (Fig. 1C).

3.4.3 Switching from AR to Screen call

DualStream allows users to quickly and seamlessly switch from a spatial AR call to a conventional screen-based video call (Fig. 3B) using buttons located below their self-view in the bottom-right of the screen (Fig. 3L).

4 IMPLEMENTING DUALSTREAM

DualStream is implemented as an Android application, and can function on any Android device capable of supporting AR experiences via Google's ARCore API⁴. Across the development and formative evaluation, we used Samsung Galaxy S9 devices. We used Unity 2022.2⁵ to develop and build this application. In order to fully capture and stream both the “dual” streams, a front-facing depth camera (Intel RealSense D435⁶) is mounted on top of the Android device. DualStream does not need this external camera to function, and can be used to view representations of remote users and their environments, as well as share local environments remotely, even without the depth camera.

4.1 Local Data Capture

4.1.1 Self

We use the RealSense camera to simultaneously capture frames of Color (RGB24 format) and Depth (Z16 format). When the phone is held at a comfortable viewing distance (approximately 0.4 meters away from the face), this captures the user's head and upper torso. We use the RealSense SDK for Unity⁷ to access this data within the DualStream application. The large size of depth buffers makes

it bandwidth-intensive to transmit in real-time over conventional networks. Therefore, we convert the depth buffer into a color image, where the color value of each pixel corresponds to the depth at that point. We use the colorization process within the RealSense SDK, and tune it such that depth information of up to 0.8 meters away from the camera is encoded in the entire color range to ensure maximum depth resolution retention. We also align the resultant color and colorized depth images to represent the same visual area (accounting for differences in field of view).

4.1.2 Surroundings

We access the rear camera feed of the mobile device using the ARCore SDK for Unity. The ARCore Depth API can provide estimates of environment depth, which is inferred using image processing and sensor fusion [12]. Inspired by the colorization approach in the Intel RealSense SDK, we convert each depth frame obtained from ARCore into a single colorized RGB frame, where depth is encoded in the pixel hue using the Turbo color mapping scheme [30]. We process the colorized depth frames to encode distances of up to 2 meters away from the device.

DualStream System Overview

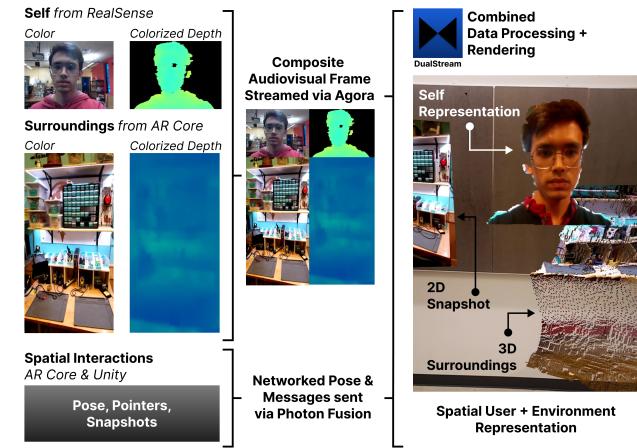


Figure 2: Implementation overview of DualStream: System diagram of data capture, processing, streaming, and rendering. Audio-visual information is streamed via Agora, while other messages relating to movement and snapshots are streamed via Photon Fusion.

4.2 Networking & Rendering

When users launch the DualStream application, they automatically join a shared room and can begin streaming information to other remote participants. During testing, we used DualStream to support communication between up to four simultaneous users.

4.2.1 Position and Interaction Data

To stream the mobile device pose and other messages that handle user interactions, we use the Photon Fusion SDK for Unity⁸. A custom script synchronizes position and orientation relative to the anchor object in each user's local frame of reference. Interaction messages (such as changing the type of user and environment representations, pointing, taking snapshots) are also handled via Fusion. Photon is optimized for sharing position and interactions with low latency. However, it is not suited for real-time video, and so we use a different mechanism for audio-visual streaming.

⁴Google AR Core: developers.google.com/ar

⁵Unity: unity.com

⁶Intel RealSense D435: intelrealsense.com/depth-camera-d435/

⁷RealSense SDK: intelrealsense.com/sdk-2

⁸Photon Fusion: photonengine.com

4.2.2 Audio-Visual Data

We use the Agora Real-time communication SDK for Unity⁹, which provides the ability to stream audio and video information over managed cloud servers. Audio is captured from the device microphone and streamed in a single channel. To efficiently share color and (colorized) depth information from both front and rear cameras, we combine the four individual color frames into a single composite frame, and stream this composite frame via a custom camera capture software device (see Fig. 2). This also ensures that all streams of individual color and depth information are synchronized across the network, with an average latency of under one second.

4.2.3 Remote Depth Rendering

Once a remote composite frame is received, we extract the individual color and colorized depth frames, and pass them through custom shaders that generate various representations of the user and their environment. The shaders take into account the camera field of view and maximum depth captured in order to render accurate representations of real-world information. They are also optimized to render representations that appear like point clouds but are less computationally intensive, allowing for a smooth experience on mobile devices. The quality of the 3D environment is lower than that of the 3D user hologram. This is because although the RealSense captures real depth information, the ARCore Depth API only provides a best estimate at the time. However, we assessed that the quality of both 3D renderings was sufficient for a proof-of-concept, in order to get initial feedback from users. The user representations could convey essential facial contours, and the environment representations provided enough depth to distinguish between horizontal, vertical, and curved surfaces. Despite the many parallel computational processes involved, DualStream runs consistently at 30 frames per second on the Samsung Galaxy S9.

5 DUALSTREAM IN ACTION

The core value of DualStream comes not from replicating cutting-edge volumetric collaboration systems (e.g., [35, 44]), but rather from its ability to enable interaction scenarios that are difficult to achieve via immersive AR/VR systems today. The freedom of using a system based on widely available devices and not tethered to a specific physical location provides unique opportunities. We now demonstrate some examples of the various remote communication contexts that DualStream can support, using scenarios where two friends — Elijah and Daneel — interact with each other. Each example in this section is a result of the research team brainstorming, implementing, and walking through scenarios that highlight the breadth of possible configurations of use. The figures representing each scenario are all taken from screen captures of DualStream, with further footage available in the supplementary video figure.

5.1 Sharing Personal Spaces

While mobile video conferencing helps show places to remote viewers, DualStream enables users to bring friends and family into their homes from afar, and take them along during travels near and far. For example, Elijah can call Daneel via DualStream from an art gallery, and share views of a painting with Daneel at home. They can also revert to a screen-based video call at anytime, whenever the spatial component of the conversation is completed (Fig. 3 A, B). Both Elijah and Daneel could be hiking outdoors, and can use DualStream to have a quick chat to share views of monuments, trees, and the creek (Fig. 3 C, D). DualStream can also support larger spatial gatherings. For example, instead of conducting a daily check-in meeting over desktop video-conferencing, Elijah and their teammates could join a quick spatial call via DualStream, and appear to all be present around each other’s workspaces (Fig. 3 E, F, G). Once

everyone has shared their updates, the members can return to mobile or desktop-based video calls for longer discussions.

5.2 Truly Remote Assistance

DualStream can potentially enable people to seek spatial remote assistance from anywhere given a mobile device and internet connection. For example, Elijah could ask Daneel for help in fixing the battery of their broken-down car by sharing a 3D hologram that Daneel can view and point at from their home (Fig. 1 C and D). Elijah can also see Daneel spatially move around them. Similarly, Elijah could share a 3D hologram of a laser cutter that they are setting up, to give a remote walk-through to Daneel who will work on the machine soon (Fig. 3 H, I). The option to capture and stream the environment video feed enables DualStream to function outdoors as well, and share video information of distant objects that depth cameras cannot capture. For example, Elijah can share 3D close-up views of the soil, and 2D views of the tree when seeking gardening advice from Daneel (Fig. 3 J, K).

5.3 Blended Spaces for Interaction

Unlike existing work in which environment sharing is determined by the system, DualStream gives users agency to construct their own blended, shared place. This enables users to combine and overlap remote environments. For example, Elijah and Daneel can share 2D snapshots of objects in each others’ rooms such as paintings and a TV, while also sharing a 3D hologram of a piece of furniture they plan to swap. Elijah can then see the remote couch overlaid on the real couch in their own space (Fig. 3 O, P). This agency enables users to capture not just space, but moments of time in space. Elijah can take a snapshot of a painting, move it to another wall, and take a second snapshot. In Daneel’s space, there now exist two copies of the painting, which they can use to advise Elijah on which arrangement looks better (Fig. 3 M, N).

6 FORMATIVE EVALUATION

We conducted a formative evaluation of DualStream to (1) gain feedback on the experience of interacting with remote collaborators via DualStream, and (2) identify key areas of improvement for the prototype. The approach of this formative evaluation was inspired by the notion of Experience Prototyping [6]. DualStream in its current form is therefore designed as a probe to enable “... others to engage directly in a proposed new experience”, and “... provide inspiration, confirmation or rejection of ideas based upon the quality of experience”. This evaluation took place in the lab and remotely, and each study session lasted between 45 minutes and 1 hour. The study procedures were approved by our university’s institutional review board (CU IRB Protocol 23-0035).

6.1 Participants

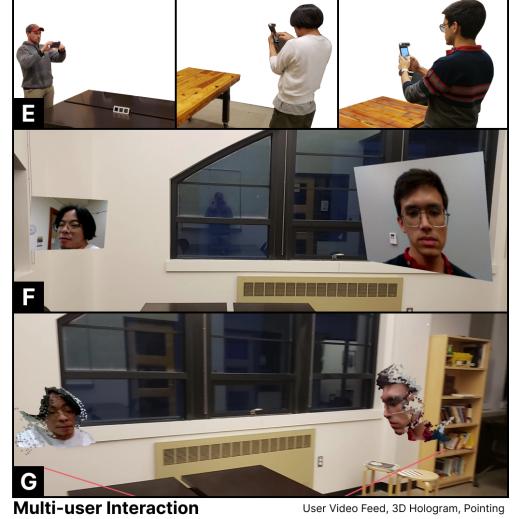
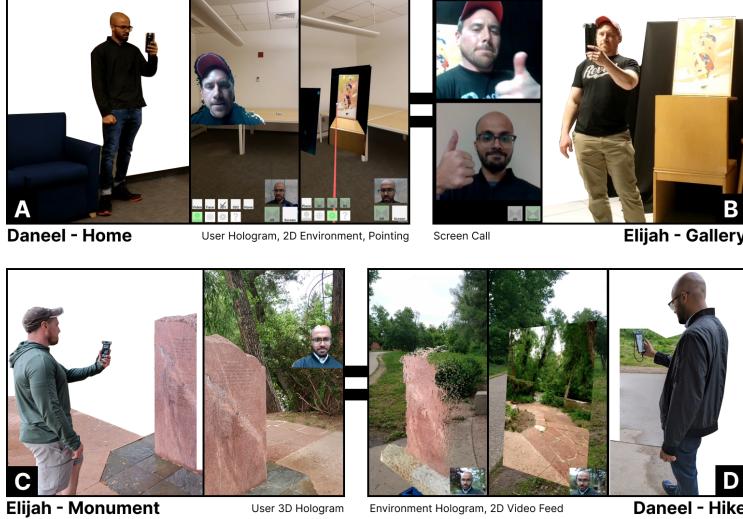
To evaluate DualStream in a controlled environment, we put out an open call for participants in our university community. Of those who responded, we excluded members who were already aware of the prototype through demonstrations of earlier versions, and thus we recruited five in-person participants. We also invited five participants from outside our city to use DualStream from their locations and engage in a remote call with a researcher. These participants were known to the researchers, but were unaware of this project and had not used the prototype before. We used the remote evaluation as a means of validating DualStream in real contexts where such a prototype is likely to be used. Details about the participants, their familiarity with AR/VR technology, and usage of mobile remote communication tools can be found in Table 1.

6.2 In-lab Evaluation

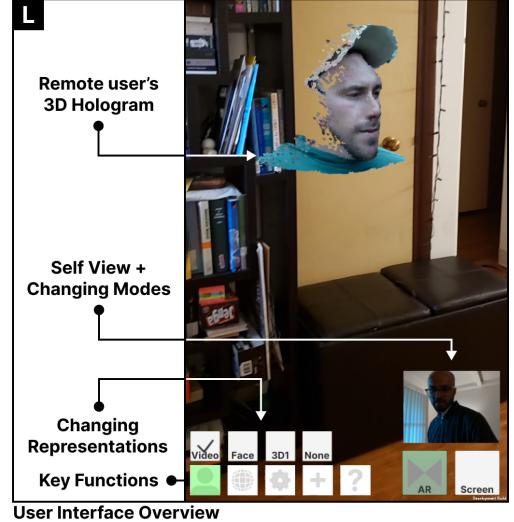
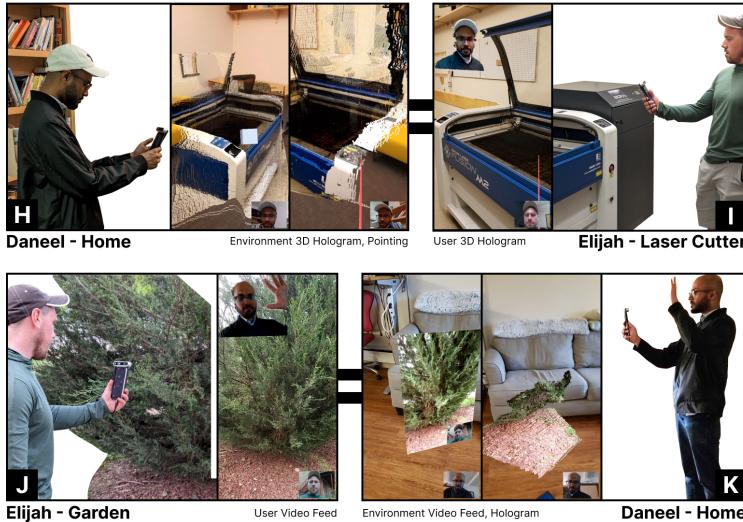
The session began with an introductory discussion where the participant and researcher were in the same physical lab space. The

⁹Agora Real-time Communication: agora.io

Sharing Personal Spaces



Truly Remote Assistance



Blended Spaces for Interaction

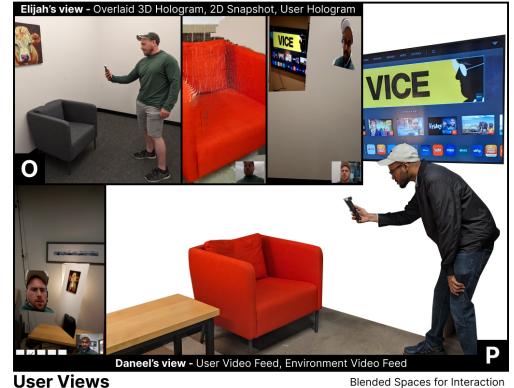


Figure 3: DualStream System Scenarios: (Top) Examples of Elijah and Daneel using DualStream to share personal spaces, while one user is in an art gallery (A, B), on a hike (C, D), or discussing with multiple members in a team (E, F, G). (Middle) DualStream being used for remote assistance in a range of contexts - to learn how to work with a laser cutter (H, I), or seeking advice on gardening (J, K). Sub-figure L consists of an overview of the user interface. (Bottom) Demonstrations of how DualStream creates blended spaces for interaction by capturing moments of space in time (M, N): working with different configurations of the same physical painting, and enabling both users to dynamically share and blend information about screens and furniture from local and remote spaces (O, P).

participants answered questions about their usage of remote communication tools and experience with AR/VR technologies. The participants were then directed to a separate lab space, where the researcher helped them start using DualStream by placing the anchor object and briefing them about the various features of the application. The researcher then returned to the original lab space and joined the call. Both lab spaces were controlled and had a variety of 2D and 3D media to enable a rich range of possible shared content (Fig. 4 Top). Both the participant and the researcher had a mobile phone with an external front-facing depth camera, enabling the bi-directional sharing of self and surroundings. Once in the DualStream call, the researcher demonstrated the various user representations, environment-sharing capabilities, and interactions. The participants were then asked to try the same features. This demonstration took between 15 and 20 minutes. The participant and researcher returned to the first lab space and engaged in a semi-structured interview regarding the participant's experience with the prototype, seeking feedback for future development. We recorded audio and took notes during this interview, which lasted between 15 and 30 minutes.

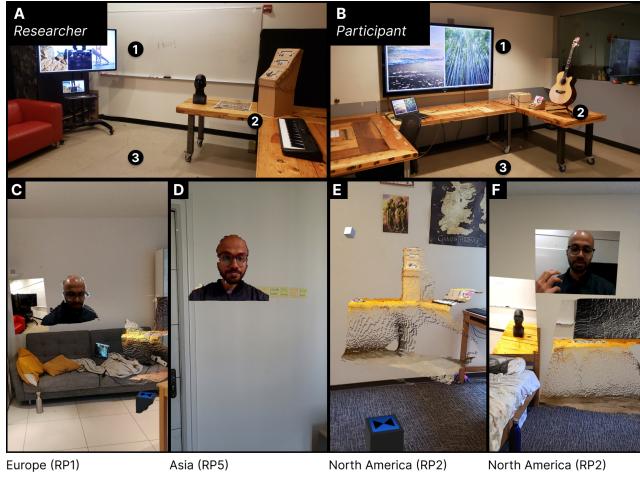


Figure 4: (Top) An overview of the in-lab study setup. Participants were in location B, while the researcher was in location A. Locations were arranged to provide sufficient room to move and view remote content (3). Both spaces had a range of media such as (1) whiteboards and displays, and (2) 3D objects. (Bottom) A collection of screen views from the remote study (C, D, E, F).

6.3 Remote Evaluation

The DualStream application was first sent to each participant, along with instructions on how to install it on their phones. The researcher and participant then joined a common video call where the researcher conducted a similar initial interview and briefing as the in-lab sessions. The researcher walked the remote participant through the process of starting DualStream and placing the anchor object. After this, the researcher joined the shared experience using DualStream running on a mobile phone with an attached depth camera. The participants did not have an external depth camera and were thus unable to share self-representations. The rest of the experience was functionally identical to the in-lab study, with the concluding interview taking place over the video call. Fig. 4 (bottom) shows examples of the screen view of remote participants during the study session (included here with permission). Despite the large distances between study sites, the DualStream application was able to function smoothly without much latency. The screen views also demonstrate that the quality of rendering at remote sites was comparable to what was achieved in controlled settings.

Participant	Gender	Age	AR Exp.	Mobile Comm.
P1	F	30	Limited	Daily
P2	F	32	Limited	Daily
P3	M	28	Familiar	2-3x / week
P4	M	24	Limited	4-5x / week
P5	F	28	Limited	2x / week
RP1 (Europe)	M	26	Familiar	1x / week
RP2 (N. America)	M	27	Familiar	Rarely
RP3 (N. America)	F	22	None	Daily
RP4 (N. America)	M	22	None	Rarely
RP5 (Asia)	F	26	Limited	2-3x / week

Table 1: Participants' (In lab - P and Remote - RP) experiences with Mobile AR (AR Exp.) and mobile video communication applications (Mobile Comm.). Mobile AR experience is categorized as *none* (rarely used), *limited* (used a few times at showcases), *familiar* (used before for personal use).

6.4 Results and Findings

After completing all sessions, two members of the research team qualitatively analyzed the notes and interview transcripts. After independently coding the data into various categories, the researchers discussed their findings and compiled a final set of observations based on overarching themes. These themes relate to the perception of users and environments, interactions in shared spaces, areas for improvement, and potential contexts of future use.

6.4.1 The Perception of User Representations

Across users, the spatial movement of the researcher's self-representation increased engagement, presence, and strengthened the feeling of sharing space. Reactions to the user representations themselves were mixed. On the positive extreme, RP5 mentioned that the moving 3D hologram almost made them believe that the researcher was visiting their home for the first time. The negative impact of the "uncanniness" of the 3D representation was discussed by several participants (P1, P2, P3, P5). However, P5 mentioned that the point cloud aesthetic helped make the 3D representation feel less uncanny, and suggested that such abstractions could be used to ease this feeling. P1, P2, P4, and P5 indicated that a more full-body representation would be desirable. P1 and RP4 mentioned that the spatial video without background was the most useful for them, as it retained a high enough resolution while also removing overlapping information about the researcher's background. In contrast, RP2 said they would prefer to use the full user video feed, as it was familiar and reminded them of traditional video calls.

6.4.2 On the Persistence of Environment Objects

Many participants felt the persistence of objects (as captured by snapshots) helped anchor conversations in space. P1, P2, P3, and RP2 mentioned that capturing 3D objects in snapshots could enable users to continue referring to shared objects throughout conversations. RP2 spoke of how "*snapshots enable you to move on with the conversation and still quickly refer back to a point*", while P2 mentioned that "*you can take a snapshot, continue with conversations and not forget about the object that is still there*". P1, P4, and RP2 noted how the ability to build scenes using multiple snapshots helped in constructing a mental map of the remote space. P1 also mentioned that they could see people using this feature to take snapshots of other members of the family as they moved around the house, bringing in the potential for multi-user interactions around a single device. RP2 stated that the act of following the researcher while they were sharing their environment significantly added to their level of engagement and spatial comprehension.

6.4.3 Comparing 2D and 3D Surroundings

While the utility of 3D environment representations was clear to most users, in its current form, many preferred the spatial environment video feed (P1, RP1, RP2), and mentioned that it was sufficient to gain a sense of the relative positions of objects. P2 mentioned

practical reasons for using 2D content, stating “*if I’m on a beach or large landscape, 2D is better than 3D because I can’t share the ocean*”. P5 discussed using 2D video for sharing a large environment if the remote user is in a small space, or perhaps “*scaling down the environment*”. Regarding 3D representations, P3 mentioned that “*though rendering could use improvement, it really felt like the objects were in my space*”, and this sentiment was expressed by many other participants. P2 also frequently navigated “out” of the shared space by switching to the screen-based call. They mentioned that viewing 3D content all the time was overwhelming, and they appreciated the option to move to the screen-based call because it was familiar.

6.4.4 Areas for Improvement

Concerning the self representations, users mentioned the need for more context about the body (P1, P2, P5), but also cautioned that a combination of a virtual body and a real face might make the experience feel even more uncanny (P5). Spatial constraints and sharing of vastly different environments posed issues, with P5 suggesting that the application should be able to scale the shared environment down to fit smaller spaces. All participants mentioned that the rendering of 3D information required improvements to be truly usable. The user interface was minimal, and participants often required instructions from the researcher to use certain functions. Several participants (P1, P2, P3, P5) suggested that the AR content should be editable. The amount of persistent AR content was overwhelming in some cases, and the need to delete unused snapshots became apparent.

6.4.5 Future Applications: Practical and Playful

Many users mentioned that they could see themselves using DualStream to share new spaces and view them remotely (P1, P3, RP1, RP2, RP5), in the context of looking for apartments to move into for example. Architectural planning was also indicated as a potential context, with P2 and P5 discussing how sharing scaled snapshots of environments between indoor and on-site locations would be particularly helpful. Some users considered contexts of play, by proposing shared escape rooms (P5), bringing in interactive AR content (P1, P4), and adding playful interactions to snapshots, such as walking to trigger actions or moving between persistent scenes (P1).

7 DISCUSSION

7.1 The Promise of Mobile Spatial Communication

In developing DualStream, we were primarily concerned (like many contemporary projects in the aftermath of the COVID-19 pandemic) with adding to the experience of “being there” [5]. This objective naturally led to incorporating more realistic spatial representations of each other. In doing so, we uncovered interactions that empower users to not only share, but blend their surroundings in ways that go “beyond” in-person interaction [21]. These interactions have analogues in the space of immersive collaboration. Projects have discussed ways to merge [20, 48], duplicate [54], and remix [29] immersive environments. Some of these ideas also emerged from participant reactions to the DualStream prototype. Participants wondered if shared environments could be scaled, edited, and invited to merge in ways that would foster personal connections. Future work can explore how such concepts and ideas from immersive AR/VR can be adapted in mobile contexts. Mobile Spatial Communication is also uniquely positioned to act as a bridge modality for cross-reality systems. Applications such as Mozilla Hubs already allow users to join shared virtual spaces using personal computers and AR/VR headsets. By incorporating features from DualStream into these systems, mobile devices can help users to better navigate the transition between the desktop computing paradigms of today, and the immersive computing environments of the near-future.

7.2 The Potential for Machine Learning

Ongoing work in the areas of machine learning and computer vision can potentially help improve the fidelity and completeness of representations created by DualStream. Systems such as Pose-on-the-Go [1] already leverage mobile sensors to predict human pose and gestures. Combining this full-body information with real-time facial capture would be a step towards more complete user representations. Recent work on creating photo-realistic avatars using smartphone cameras [7] demonstrates how mobile devices might soon be capable of computing and rendering high-quality faces for applications in AR/VR communication. Integrating methods for 3D reconstruction such as Neural Radiance Fields is another promising direction, with steps taken towards mobile device-based capture and processing of high-quality environment renderings [37, 41]. While these approaches generate realistic faces and spaces, there needs to be a balance between information that is streamed as-is (like DualStream) or computed, based on the context at hand.

7.3 Limitations

DualStream is a proof-of-concept of capabilities that are likely to be available on mobile devices in the near future. Throughout the design and development process, we made decisions that prioritized wider access, such as choosing cloud-based networking services to reach a broader audience outside the lab. There are potentially other combinations of present-day hardware (using iOS devices with depth cameras, or local networking setups) that might yield more optimized systems for in-lab studies. While the quality of 3D representations can be improved, testing DualStream at this stage revealed how even 2D information, when presented spatially in AR, might suffice for specific tasks. Our formative evaluation provides initial insights into how users interact with DualStream and the contexts in which they might want to use it. The study helped us understand which of the many features we should prioritize, while also uncovering some unexpected and interesting findings. The open-ended nature of the session was instrumental in achieving this. However, the lack of a concrete task makes it difficult to comment on the more immediate utility of DualStream. We intend to conduct follow-up evaluations focused on specific aspects of the prototype (e.g., user representations, pointing and gestures) in applied contexts, in order to gain a deeper understanding of how such a system can be designed to best support remote collaboration.

8 CONCLUSION

DualStream explores how mobile devices can better support spatial remote collaboration. By leveraging the combination of cameras and sensors on mobile phones, as well as their ubiquity relative to AR/VR headsets, DualStream enables the sharing of spatial representations of both self and place, anywhere and anytime. We discuss the benefits of DualStream through scenarios and examples showcasing how it can support the sharing of personal spaces, truly remote assistance, and interaction in blended spaces. Findings from a formative evaluation conducted in the lab and in the real world indicate that users found value in real representations, spatial interactions via mobile AR, and the ability to construct persistent shared spaces. Our work lays out new directions for enabling more widely accessible spatial communication tools with mobile devices as the focus.

ACKNOWLEDGMENTS

The authors would like to thank Greg Phillips, Alvin Jude, Gunilla Berndtsson, Per-Erik Brodin, Amir Gomroki, and Per Karlsson for their support and advice throughout this project. We also wish to thank Sandra Bae and Ruojia Sun for their assistance and many helpful discussions. This project was supported by a grant from Ericsson Research, and by the National Science Foundation under Grant No. IIS-2040489.

REFERENCES

- [1] K. Ahuja, S. Mayer, M. Goel, and C. Harrison. Pose-on-the-Go: Approximating User Pose with Smartphone Sensor Fusion and Inverse Kinematics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, pp. 1–12. Association for Computing Machinery, New York, NY, USA, May 2021. doi: 10.1145/3411764.3445582
- [2] J. N. Bailenson, N. Yee, D. Merget, and R. Schroeder. The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction. *Presence: Teleoperators and Virtual Environments*, 15(4):359–372, Aug. 2006. doi: 10.1162/pres.15.4.359
- [3] P. Bhattacharyya, R. Nath, Y. Jo, K. Jadhav, and J. Hammer. Brick: Toward A Model for Designing Synchronous Colocated Augmented Reality Games. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pp. 1–9. Association for Computing Machinery, New York, NY, USA, May 2019. doi: 10.1145/3290605.3300553
- [4] M. Billinghurst and H. Kato. Collaborative Augmented Reality. *Communications of the ACM*, 45(7):64–70, July 2002. doi: 10.1145/514236.514265
- [5] S. A. Bly, S. R. Harrison, and S. Irwin. Media spaces: bringing people together in a video, audio, and computing environment. *Communications of the ACM*, 36(1):28–46, Jan. 1993. doi: 10.1145/151233.151235
- [6] M. Buchenau and J. F. Suri. Experience Prototyping. In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, DIS '00, p. 424–433. Association for Computing Machinery, New York, NY, USA, 2000. doi: 10.1145/347642.347802
- [7] C. Cao, T. Simon, J. K. Kim, G. Schwartz, M. Zollhoefer, S.-S. Saito, S. Lombardi, S.-E. Wei, D. Belko, S.-I. Yu, et al. Authentic Volumetric Avatars from a Phone Scan. *ACM Transactions on Graphics (TOG)*, 41(4):1–19, 2022.
- [8] D. Chatzopoulos, C. Bermejo, Z. Huang, and P. Hui. Mobile Augmented Reality Survey: From Where We Are to Where We Go. *IEEE Access*, 5:6917–6950, 2017.
- [9] E. Dagan, A. M. Cárdenas Gasca, A. Robinson, A. Noriega, Y. J. Tham, R. Vaish, and A. Monroy-Hernández. Project IRL: Playful Co-Located Interactions with Mobile Augmented Reality. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1):62:1–62:27, Apr. 2022. doi: 10.1145/3512909
- [10] D. Dateu, S. G. Lukosch, and H. K. Lukosch. Handheld Augmented Reality for Distributed Collaborative Crime Scene Investigation. In *Proceedings of the 2016 ACM International Conference on Supporting Group Work*, GROUP '16, pp. 267–276. Association for Computing Machinery, New York, NY, USA, Nov. 2016. doi: 10.1145/2957276.2957302
- [11] P. Dourish. Re-space-ing place: "place" and "space" ten years on. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, CSCW '06, pp. 299–308. Association for Computing Machinery, New York, NY, USA, Nov. 2006. doi: 10.1145/1180875.1180921
- [12] R. Du, E. Turner, M. Dzitsiuk, L. Prasso, I. Duarte, J. Dourgarian, J. Afonso, J. Pascoal, J. Gladstone, N. Cruces, S. Izadi, A. Kowdle, K. Tsotsos, and D. Kim. DepthLab: Real-time 3D Interaction with Depth Maps for Mobile Augmented Reality. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, UIST '20, pp. 829–843. Association for Computing Machinery, New York, NY, USA, Oct. 2020. doi: 10.1145/3379337.3415881
- [13] C. G. Fidalgo, Y. Yan, H. Cho, M. Sousa, D. Lindlbauer, and J. Jorge. A Survey on Remote Assistance and Training in Mixed Reality Environments. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2291–2303, 2023. doi: 10.1109/TVCG.2023.3247081
- [14] D. I. Fink, J. Zagermann, H. Reiterer, and H.-C. Jetter. Re-locations: Augmenting Personal and Shared Workspaces to Support Remote Collaboration in Incongruent Spaces. *Proceedings of the ACM on Human-Computer Interaction*, 6(ISS):556:1–556:30, Nov. 2022. doi: 10.1145/3567709
- [15] L. Foglia and R. A. Wilson. Embodied cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(3):319–325, 2013.
- [16] G. Gamelin, A. Chellali, S. Cheikh, A. Ricca, C. Dumas, and S. Otmane. Point-cloud avatars to improve spatial communication in immersive collaborative virtual environments. *Personal and Ubiquitous Computing*, 25(3):467–484, June 2021. doi: 10.1007/s00779-020-01431-1
- [17] J. G. Grandi, H. G. Debarba, I. Bemdt, L. Nedel, and A. Maciel. Design and Assessment of a Collaborative 3D Interaction Technique for Handheld Augmented Reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 49–56. IEEEVR, Reutlingen, Germany, Mar. 2018. doi: 10.1109/VR.2018.8446295
- [18] S. Harrison and P. Dourish. Re-place-ing space: the roles of place and space in collaborative systems. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, CSCW '96, pp. 67–76. Association for Computing Machinery, New York, NY, USA, Nov. 1996. doi: 10.1145/240080.240193
- [19] J. Hauber, H. Regenbrecht, M. Billinghurst, and A. Cockburn. Spatiality in videoconferencing: trade-offs between efficiency and social presence. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, CSCW '06, pp. 413–422. Association for Computing Machinery, New York, NY, USA, Nov. 2006. doi: 10.1145/1180875.1180937
- [20] J. Herskovitz, Y. F. Cheng, A. Guo, A. P. Sample, and M. Nebeling. XSpace: An Augmented Reality Toolkit for Enabling Spatially-Aware Distributed Collaboration. *Proceedings of the ACM on Human-Computer Interaction*, 6(ISS):568:277–568:302, Nov. 2022. doi: 10.1145/3567721
- [21] J. Hollan and S. Stornetta. Beyond being there. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, pp. 119–125. Association for Computing Machinery, New York, NY, USA, June 1992. doi: 10.1145/142750.142769
- [22] B. Jones, A. Witcraft, S. Bateman, C. Neustaedter, and A. Tang. Mechanics of Camera Work in Mobile Video Collaboration. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pp. 957–966. Association for Computing Machinery, New York, NY, USA, Apr. 2015. doi: 10.1145/2702123.2702345
- [23] B. Jones, Y. Zhang, P. N. Y. Wong, and S. Rintel. Belonging There: VRoom-ing into the Uncanny Valley of XR Telepresence. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):59:1–59:31, Apr. 2021. doi: 10.1145/3449133
- [24] A. Kendon. *Conducting interaction: Patterns of Behavior in Focused Encounters*, vol. 7. Cambridge University Press, UK, 1990.
- [25] S. F. Langa, M. Montagud, G. Cerniglio, and D. R. Rivera. Multiparty Holomeetings: Toward a New Era of Low-Cost Volumetric Holographic Meetings in Virtual Reality. *IEEE Access*, 10:81856–81876, 2022. Conference Name: IEEE Access. doi: 10.1109/ACCESS.2022.3196285
- [26] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, VRST '17, pp. 1–10. Association for Computing Machinery, New York, NY, USA, Nov. 2017. doi: 10.1145/3139131.3139156
- [27] J. Li, M. Sousa, C. Li, J. Liu, Y. Chen, R. Balakrishnan, and T. Grossman. ASTEROIDS: Exploring Swarms of Mini-Telepresence Robots for Physical Skill Demonstration. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, pp. 1–14. Association for Computing Machinery, New York, NY, USA, Apr. 2022. doi: 10.1145/3491102.3501927
- [28] Y. Li, S. W. Lee, D. A. Bowman, D. Hicks, W. s. Lages, and A. Sharma. ARCritique: Supporting Remote Design Critique of Physical Artifacts through Collaborative Augmented Reality. In *Proceedings of the 2022 ACM Symposium on Spatial User Interaction*, SUI '22, pp. 1–12. Association for Computing Machinery, New York, NY, USA, Dec. 2022. doi: 10.1145/3565970.3567700
- [29] D. Lindlbauer and A. D. Wilson. Remixed Reality: Manipulating Space and Time in Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 1–13. Association for Computing Machinery, New York, NY, USA,

- Apr. 2018. doi: 10.1145/3173574.3173703
- [30] A. Mikhailov. Turbo, An Improved Rainbow Colormap for Visualization. *Google AI Blog*, 10(15-16):8, 2019.
- [31] J. Müller, R. Rädle, and H. Reiterer. Remote Collaboration With Mixed Reality Displays: How Shared Virtual Landmarks Facilitate Spatial Referencing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pp. 6481–6486. Association for Computing Machinery, New York, NY, USA, May 2017. doi: 10.1145/3025453.3025717
- [32] C. Neustaedter, S. Singhal, R. Pan, Y. Heshmat, A. Forghani, and J. Tang. From Being There to Watching: Shared and Dedicated Telepresence Robot Usage at Academic Conferences. *ACM Transactions on Computer-Human Interaction*, 25(6):33:1–33:39, Dec. 2018. doi: 10.1145/3243213
- [33] K. O'Hara, A. Black, and M. Lipson. Everyday practices with mobile video telephony. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pp. 871–880. Association for Computing Machinery, New York, NY, USA, Apr. 2006. doi: 10.1145/1124772.1124900
- [34] K. O'hara, J. Kjeldskov, and J. Paay. Blended interaction spaces for distributed team collaboration. *ACM Transactions on Computer-Human Interaction*, 18(1):3:1–3:28, May 2011. doi: 10.1145/1959022.1959025
- [35] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou, V. Tankovich, C. Loop, Q. Cai, P. A. Chou, S. Mennicken, J. Valentin, V. Pradeep, S. Wang, S. B. Kang, P. Kohli, Y. Lutchyn, C. Keskin, and S. Izadi. Holoportation: Virtual 3D Teleportation in Real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, pp. 741–754. Association for Computing Machinery, New York, NY, USA, Oct. 2016. doi: 10.1145/2984511.2984517
- [36] P. Panda, M. J. Nicholas, M. Gonzalez-Franco, K. Inkpen, E. Ofek, R. Cutler, K. Hinckley, and J. Lanier. AllTogether: Effect of Avatars in Mixed-Modality Conferencing Environments. In *2022 Symposium on Human-Computer Interaction for Work*, CHIWORK 2022, pp. 1–10. Association for Computing Machinery, New York, NY, USA, June 2022. doi: 10.1145/3533406.3539658
- [37] M. Park, B. Yoo, J. Y. Moon, and J. H. Seo. InstantXR: Instant XR Environment on the Web Using Hybrid Rendering of Cloud-based NeRF with 3D Assets. In *Proceedings of the 27th International Conference on 3D Web Technology*, Web3D '22, pp. 1–9. Association for Computing Machinery, New York, NY, USA, Nov. 2022. doi: 10.1145/3564533.3564565
- [38] R. Petkova, V. Poulikov, A. Manolova, and K. Tonchev. Challenges in Implementing Low-Latency Holographic-Type Communication Systems. *Sensors*, 22(24):9617, Jan. 2022. Number: 24 Publisher: Multi-disciplinary Digital Publishing Institute. doi: 10.3390/s22249617
- [39] T. Piumsomboon, G. A. Lee, A. Irlitti, B. Ens, B. H. Thomas, and M. Billinghurst. 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*, CHI '19, pp. 1–17. Association for Computing Machinery, New York, NY, USA, May 2019. doi: 10.1145/3290605.3300458
- [40] A. Ranjan, J. P. Birnholtz, and R. Balakrishnan. Dynamic shared visual spaces: experimenting with automatic camera control in a remote repair task. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pp. 1177–1186. Association for Computing Machinery, New York, NY, USA, Apr. 2007. doi: 10.1145/1240624.1240802
- [41] S. Rojas, J. Zarzar, J. C. Perez, A. Sanakoyeu, A. Thabet, A. Pumarola, and B. Ghanem. Re-ReND: Real-time Rendering of NeRFs across Devices. *arXiv preprint arXiv:2303.08717*, 2023.
- [42] M. Sakashita, E. A. Ricci, J. Arora, and F. Guimbretière. RemoteCoDe: Robotic Embodiment for Enhancing Peripheral Awareness in Remote Collaboration Tasks. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1):63:1–63:22, Apr. 2022. doi: 10.1145/3512910
- [43] P. Sasikumar, M. Collins, H. Bai, and M. Billinghurst. XRTB: A Cross Reality Teleconference Bridge to Incorporate 3D Interactivity to 2D Teleconferencing. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411763.3451546
- [44] B. Thoravi Kumaravel, F. Anderson, G. Fitzmaurice, B. Hartmann, and T. Grossman. 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*, UIST '19, pp. 161–174. Association for Computing Machinery, New York, NY, USA, Oct. 2019. doi: 10.1145/3332165.3347872
- [45] R. Vertegaal. The GAZE groupware system: mediating joint attention in multiparty communication and collaboration. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, CHI '99, pp. 294–301. Association for Computing Machinery, New York, NY, USA, May 1999. doi: 10.1145/302979.303065
- [46] A. Villanueva, Z. Zhu, Z. Liu, F. Wang, S. Chidambaram, and K. Ramani. ColabAR: A Toolkit for Remote Collaboration in Tangible Augmented Reality Laboratories. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1):81:1–81:22, Apr. 2022. doi: 10.1145/3512928
- [47] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. E. Latoschik. The Impact of Avatar Personalization and Immersion on Virtual Body Ownership, Presence, and Emotional Response. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1643–1652, Apr. 2018. Conference Name: IEEE Transactions on Visualization and Computer Graphics. doi: 10.1109/TVCG.2018.2794629
- [48] M. Wang, Y.-J. Li, J. Shi, and F. Steinicke. SceneFusion: Room-Scale Environmental Fusion for Efficient Traveling Between Separate Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–16, 2023. doi: 10.1109/TVCG.2023.3271709
- [49] P. Wang, X. Bai, M. Billinghurst, S. Zhang, X. Zhang, S. Wang, W. He, Y. Yan, and H. Ji. AR/MR Remote Collaboration on Physical Tasks: A Review. *Robotics and Computer-Integrated Manufacturing*, 72:102071, 2021.
- [50] T. Wells and S. Houben. CollabAR - Investigating the Mediating Role of Mobile AR Interfaces on Co-Located Group Collaboration. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pp. 1–13. Association for Computing Machinery, New York, NY, USA, Apr. 2020. doi: 10.1145/3313831.3376541
- [51] B. Yoon, H.-i. Kim, G. A. Lee, M. Billinghurst, and W. Woo. The Effect of Avatar Appearance on Social Presence in an Augmented Reality Remote Collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 547–556. IEEEVR, Osaka, Japan, Mar. 2019. ISSN: 2642-5254. doi: 10.1109/VR.2019.8797719
- [52] J. Young, T. Langlotz, M. Cook, S. Mills, and H. Regenbrecht. Immersive Telepresence and Remote Collaboration using Mobile and Wearable Devices. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):1908–1918, 2019. doi: 10.1109/TVCG.2019.2898737
- [53] J. Young, T. Langlotz, S. Mills, and H. Regenbrecht. Mobileportation: Nomadic Telepresence for Mobile Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4(2), jun 2020. doi: 10.1145/3397331
- [54] K. Yu, U. Eck, F. Pankratz, M. Lazarovici, D. Wilhelm, and N. Navab. Duplicated reality for co-located augmented reality collaboration. *IEEE Transactions on Visualization Computer Graphics*, 28(05):2190–2200, may 2022. doi: 10.1109/TVCG.2022.3150520
- [55] K. Yu, G. Gorbachev, U. Eck, F. Pankratz, N. Navab, and D. Roth. Avatars for Teleconsultation: Effects of Avatar Embodiment Techniques on User Perception in 3D Asymmetric Telepresence. *IEEE Transactions on Visualization and Computer Graphics*, 27(11):4129–4139, Nov. 2021. Conference Name: IEEE Transactions on Visualization and Computer Graphics. doi: 10.1109/TVCG.2021.3106480
- [56] S. Zhang, B. Jones, S. Rintel, and C. Neustaedter. XRmas: Extended Reality Multi-Agency Spaces for a Magical Remote Christmas. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '21, pp. 203–207. Association for Computing Machinery, New York, NY, USA, Oct. 2021. doi: 10.1145/3462204.3481782