Interactive Mixed-Dimensional Media for Cross-Dimensional Collaboration in Mixed Reality Environments

Balasaravanan Thoravi Kumaravel* and Björn Hartmann

Department of Computer Science, University of California, Berkeley, CA, United States

Collaboration and guidance are key aspects of many software tasks. In traditional desktop software, such aspects are well supported through built-in collaboration functions or general-purpose techniques such as screen and video sharing. In Mixed Reality environments, where users carry out actions in a three-dimensional space, collaboration and guidance may also be required. However, other users may or may not be using the same Mixed Reality interface. Users may not have access to the same information, the same visual representation, or the same interaction affordances. These asymmetries make communication and collaboration between users harder. To address asymmetries in Mixed Reality environments, we introduce Interactive Mixed-Dimensional Media. In these media, the visual representation of information streams can be changed between 2D and 3D. Different representations can be chosen automatically, based on context, or through associated interaction techniques that give users control over exploring spatial, temporal, and dimensional levels of detail. This ensures that any information or interaction makes sense across different dimensions, interfaces and spaces. We have deployed these techniques in three different contexts: mixed-reality telepresence for physical task instruction, video-based instruction for VR tasks, and live interaction between a VR user and a non-VR user. From these works, we show that Mixed Reality environments that provide users with interactive mixed-dimensional media interfaces improve performance and user experience in collaboration and guidance tasks.

Keywords: mixed reality, asymmetric interactions, collaboration, guidance, mixed-dimensional media, computer supported collaborative work

1 INTRODUCTION

Mixed Reality environments refer to computing environments which co-exist in the physical space of users. Mixed Reality environments exhibit varying proportions of *Virtuality*, as defined in the Reality-Virtuality continuum by Milgram et al. (1995), which spans all the way from an entirely real world to an entirely virtual world. In this work, we will focus on only those environments that have at least some amount of virtual content in them, i.e. Augmented Reality, Augmented Virtuality, and Virtual Reality.

Today, Mixed Reality environments have grown beyond research prototypes, to popular user platforms with a growing ecosystem of applications, games, and media. A major problem in these environments is the isolation of a user from their peers (Gugenheimer et al., 2017; Gugenheimer et al., 2019). Users find it challenging to share, co-experience and collaborate in these environments, especially if these were not programmed to be collaborative, or if peers use different types of interfaces to access a Mixed Reality environment.
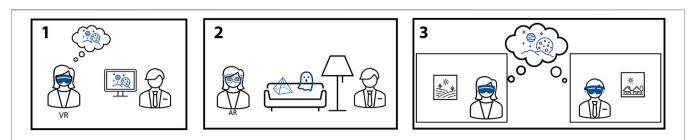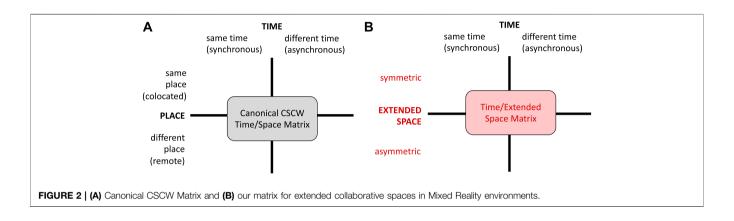
FIGURE 1 | Three example scenarios that have no meaningful place in the canonical CSCW matrix (1) A user watching a video mirror of a co-located VR user. (2) an AR user and another user who does not wear a headset, (3) two remotely located VR users who are part of the same virtual scene.

FIGURE 2 | (A) Canonical CSCW Matrix and **(B)** our matrix for extended collaborative spaces in Mixed Reality environments.

These *asymmetries* induce a communication barrier between the users. For example, most Mixed Reality systems today can "mirror" the first-person view of the Mixed Reality user to an external display (**Figure 1A**). But this only provides a partial view of the larger Mixed Reality environment, and is hard to watch because of its shaky nature. Thus, it is difficult for another user to interact with the Mixed Reality user, solely using this video feed.

The canonical CSCW matrix, introduced by Johansen (1988) has been used to categorize collaborative systems (**Figure 2A**). However, it does not address nuances of collaboration that arise in Mixed Reality spaces. The CSCW matrix considers only the physical space for categorizing different interactions. This leaves out critical details that are required for thinking about interactions in Mixed Reality environments. Some examples of interactions include (**Figure 1A–C**): 1) An artist who can see the first-person video feed of a VR user, and teaches them how to paint and sculpt in VR; 2) A spectator who talks to an AR user (Headset-based) who interacts with virtual objects scattered over a physical scene that the spectator cannot see; 3) A user in a Mixed Reality environment interacting with another user who is also in the same Mixed Reality virtual environment, but remotely located in the physical world. Such communications are asymmetric because the user in Mixed Reality environment and other interacting users do not share the same display, input capabilities or access to physical space. For example in 1) the VR user's view is blocked by the headset so they cannot see a co-located external user; in 1) and 2) the external users do not have full visibility of the virtual scene. In 3), while both users can see the virtual scene, they are not aware of each other's physical

environments and their accompanying constraints on movement. We see that none of these examples have a meaningful place in the CSCW matrix.
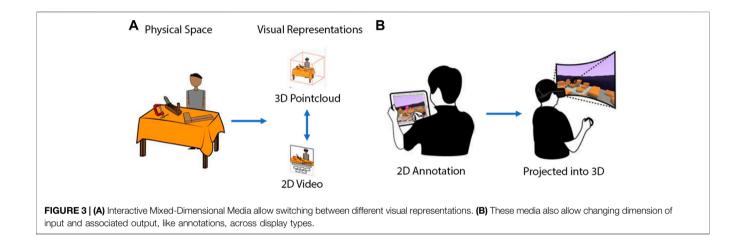
A naive solution to this would be to add more dimensions to the CSCW matrix. For instance, Virtual space can be added as an additional dimension. However, there exists a major issue in that, with such a 3D matrix representation, the common collaboration scenarios would often straddle across multiple Octants. Furthermore, the nature of these scenarios also vary based on the differences in *visual representation* with which users perceive these spaces, as well as differences in *Interaction affordances* (**Table 1**). So a full accounting could lead up to a five-dimensional matrix, which is unwieldy to think about, and reason using it as a tool.

To avoid such complex formulations, we propose a simpler modification to the matrix, to make it better suited to deal with interactions that occur in Mixed Reality spaces (**Figure 2B**). Specifically, we introduce the concept of an *extended collaborative space* which we will abbreviate as *xspace*. An *xspace* is the Mixed Reality space in the Milgram continuum that is perceived by a user, containing associated information of all objects, both physical and virtual, that are relevant to the collaboration task at hand (for this and other important definitions used throughout this paper, **Table 1**). We thus replace the notion of a co-located or remote physical space with that of an asymmetric or symmetric extended collaborative space (*xspace*).

Asymmetry in *xspaces* occur because one user can not see and/ or interact with certain parts of a Mixed Reality environment in

Term	Definition
Extended collaborative space (xspace)	The Mixed Reality space perceived by a user, containing all objects, both physical and virtual, that are relevant to the collaboration task at hand
Interactive mixed-dimensional medium	A medium that can offer different visual representations and associated interactions across and within display types
Display type	Output hardware a user employs to view a Mixed Reality environment. E.g., VR headset, AR headset, 2D screen
Visual representation	The way in which a data stream from a physical or virtual environment is captured and presented. E.g., 3D geometry, 3D point cloud, Stereo video, 2D video, 2D image
Output dimensionality	The number of dimensions inherent in a given spatial representation. E.g., 3D for point clouds, 2D for video
Input dimensionality	The degrees of freedom (DOF) available for providing input. E.g., 2D on a touch screen vs. 3D using a VR controller

FIGURE 3 | (A) Interactive Mixed-Dimensional Media allow switching between different visual representations. **(B)** These media also allow changing dimension of input and associated output, like annotations, across display types.

the same manner as another user would. Sometimes, a user has no access to these parts. For instance, in example 2), the spectator has no access to the virtual elements. However, in many other cases, the *visual representation* in which users view the spaces are different. This is seen in example 1), in which the external artist sees the virtual environment through a 2D video feed, whereas the VR user sees as well as operates in the same environment in 3D.

In such scenarios, some users can carry out 3D operations in a Mixed Reality environment, while others carry out 2D UI-based operations on a screen. Some can view information in 3D, while others only through 2D. Thus, in these asymmetric *xspaces*, there exists an uneasy co-existence of different combinations of 2D and 3D *input and output dimensionalities* (**Table 1**) for a user.

In summary, it is important to note that asymmetry in *xspaces* is a multi-faceted concept. This means that, users can belong to asymmetrical *xspaces* due to any of four possible factors: 1) Users have unequal access to the task relevant physical space 2) Users have unequal access to the task relevant physical space to task relevant virtual space 3) Difference in visual modality with which the xspace is perceived by the users. 4) Difference in Interaction affordance for the users.

We introduce *Interactive Mixed-Dimensional Media* to help bridge that gap between users operating across asymmetrical *xspaces*, using which users can switch back and forth between 2D or 3D for perception of an *xspace*. For instance, through our medium, one can view a space as a 2D video, but can also choose

to view it through various 3D representations such as a 3D pointcloud (**Figure 3A**).

Similar to user perception discussed above, Interactive Mixed-Dimensional media also allow changing between 2D and 3D while providing input to an *xspace*. For example, when a user annotates a 2D video of an *xspace*, they can choose to either project these annotations as 3D into the corresponding 3D location in the *xspace*, or be rendered as 2D over the video feed itself (**Figure 3B**).

We formally define *Interactive Mixed-Dimensional Media*, as a medium that can offer different visual representations as well as associated interactions across and within display types. Specifically, our interaction techniques give users control over exploring the *xspaces* 1) spatially, 2) temporally and most importantly to vary the 3) visual representation. In **Table 1**, we summarize the key terminologies that are used across this work.

In this paper, we synthesize lessons learned from three prior system development projects in AR and VR in our group—Loki, TutoriVR and TransceiVR by Thoravi Kumaravel et al. (2019a), Thoravi Kumaravel et al. (2019b), Thoravi Kumaravel et al. (2020)—into the novel framework of Interactive Mixed-Dimensional Media. We show that the use of *Interactive Mixed-Dimensional Media* mitigates negative effects of asymmetry and aids with collaboration and communication, in settings where full symmetry is not possible or not desirable.

We focus part of our work on Interactive-Mixed Dimensional Media to enable collaboration and guidance for single-user and

closed-source Mixed Reality applications. One strategy to avoid asymmetric interactions might be to rewrite all Mixed Reality applications as collaborative, with shared *xspaces* that eliminate or minimize asymmetries. Indeed, a rich line of prior research has done so, across varying contexts (Fraser et al., 1999; Gugenheimer et al., 2017; Nguyen et al., 2017; Gugenheimer et al., 2018; Marwecki et al., 2018; Hartmann et al., 2019; Thoravi Kumaravel et al., 2019a; Hartmann et al., 2020). However, decades of desktop and mobile application development have shown that many applications continue to be conceived as single-user experiences, and that collaboration and guidance are nevertheless necessary and common around these applications. Collaborative use of single-user applications in conventional computing environments is well established as a practice and has been studied extensively, e.g. in "social learning" Murphy-Hill et al. (2015), "occasional meetings" Izadi et al. (2003), pair programming D'Angelo and Begel (2017), and game streaming (Smith et al., 2013; Hamilton et al., 2014; Pires and Simon, 2015; Lessel et al., 2017). Similarly, we are interested in how to mitigate the drawbacks of asymmetric interactions for a wide variety of Mixed Reality software, including applications that are written for single users, and applications where source code is not available for modification. This is a core aspect of two of our prior systems—TutoriVR and TransceiVR (Thoravi Kumaravel et al., 2019b; 2020) that we will discuss later in this work.

2 RELATED WORK

There exists four bodies of prior work that are closely related to ours: Telepresence and guidance in physical tasks, collaboration in virtual reality, retrofitting and guidance for computer software, and mixed-modal tutorials and guidance.

2.1 Telepresence and Guidance for Physical Tasks

There is diverse literature investigating guidance and remote collaboration for physical tasks. Early work in the domain focused on guidance through 2D video interfaces with integrated annotations (Bauer et al., 1998; Baudisch et al., 2001; Botden and Jakimowicz, 2009). Building on these, prior work identified the value of collaboration using Mixed Reality environments Botden et al. (2008) and proposed extensions such as spatial annotations and tracked objects (Botden et al., 2008; Anderson et al., 2013; Budrionis et al. 2013). Even earlier works Ishii (1990); Ishii et al. (1994) identified that for seamless remote collaboration, it is not sufficient to have only 2D annotations, but it is also important to have access to both physical as well as digital tools, awareness of gaze and gesture and a way to manage the digital and physical workspaces. Many prior works have shown the needs and benefits of access to multiple perspectives of the remote collaborator (Ishii, 1990; Ishii and Kobayashi, 1992; Fussell et al., 2003; Henderson and Feiner, 2007; Ranjan et al., 2007; Fuchs et al., 2014).

Our idea of Interactive Mixed-Dimensional Media builds on these works and findings wherein, it supports these rich

interactions, as well as provides us with a unified framework to broadly think about systems that involve users, each of whom perceives and interacts with the same data through different visual representations and interaction affordances. In this work, we discuss our prior research systems that augment existing virtual reality systems. Specifically, these systems use data similar to those in prior work: depth maps, color video feeds and pose tracked actions. Then, we develop Interactive mixed-dimensional media using these data, and use that to facilitate collaboration in Mixed Reality environments. Our framework allows for designing and understanding user interactions that meaningfully translate across the different representations of Mixed Reality environments.
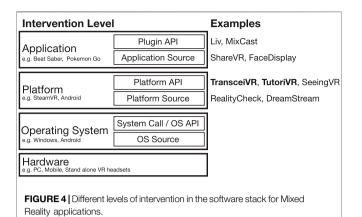
2.2 Collaboration in Virtual Reality

Multi-user interaction and collaboration in VR has been a focus for several decades (Snowdon et al., 2001; Ens et al., 2019), and continues to be actively investigated. Many of these works, enhance communication and collaboration, through reducing the asymmetry present between the users. This is done either by enabling users to co-habit the same virtual world, projection mapping the virtual environment onto a real environment (Bimber and Raskar, 2005; Benko et al., 2014); Gugenheimer et al., 2017), or even compositing the real-world into a virtual world (Hartmann et al., 2019; von Willich et al., 2019; Thoravi Kumaravel and Wilson, 2022). In contrast, our work focuses on mixed-dimensional media interactions that would not necessarily replace the asymmetry and force all users to be in a Mixed Reality environment, but rather enhance the asymmetric interactions that are a core part of the existing workflows.

Similar to prior work in telepresence and guidance for physical tasks, these systems have leveraged view sharing, object sharing and annotations (Kunert et al., 2014; Nguyen et al., 2017; Marwecki et al., 2018; Xia et al., 2018), use of multiple view ports, tracked human pose data (Ponto et al., 2012), bridging gestural references and actions across physical and virtual worlds (Kunert et al. (2014); Thoravi Kumaravel et al., 2019a), and recording and play-back of multi-user activity in Mixed Reality environments (Greenhalgh et al., 2002). These works demonstrate how effective collaboration can be supported if it is designed into the core of each application. Such approaches may lead to systems that are like "walled-gardens"; they make it easy to collaborate when all the users also have those applications installed. But otherwise, no collaboration is possible. However, single-user applications are unlikely to be rebuilt for multi-user purposes unless there are strong business needs (Cheng et al. (2004)). We take inspirations from the interactions used in prior work, but we re-think these interactions so as to retrofit them for existing VR applications. This allowed us to synthesize collaborative interactions that merge into the workflow of existing applications that were designed to be single-user.

2.3 Retrofitting and Guidance for Computer Softwares

Computer software for mixed reality applications can operate and intervene at different levels. Broadly these levels can be

FIGURE 4 | Different levels of intervention in the software stack for Mixed Reality applications.

categorized as Hardware, Operating System, Platform and Application. Systems in prior work have operated at different levels to achieve their desired functionality. These levels and some select prior work that operate at two selected levels (Platform and Application) are shown in **Figure 4**. At the Application level, a system has access to the source code and can modify the application source directly. Alternatively, some systems can provide/use an API that is included within the VR application. This can be used by external programs to interact with the application's source code. At the Platform level, a system does not have access to the source code of the application. Rather it can directly modify the source code of the VR platform that is running the VR application. Alternatively, it can leverage/create APIs for the platform that can be used by external programs to exchange data and control the platforms behavior.

Our implementation of Mixed-Dimensional Media interactions for VR applications operates at the Platform API level. Thus, it retrofits existing single-user VR applications with new communication and collaborative features. Retrofitting frequently uses a combination of available platform APIs underneath a closed source application [e.g., using UI toolkit overloading (Eagan et al., 2011) or accessibility APIs] and reverse engineering approaches to extract information where APIs are unavailable. The value of retrofitting and reverse engineering has been well established in the HCI research community. As Cheng et al. (2004) write: "Mission critical applications and legacy systems may be difficult to revise and rebuild, and yet it is sometimes desirable to retrofit their user interfaces with new collaborative features without modifying and recompiling the original code." Computer vision-based reverse engineering approaches have been used to enhance desktop software with new interaction techniques (Dixon and Fogarty 2010), automate GUI tasks Dixon and Fogarty (2010), extract reusable data from rendered information visualizations Savva et al. (2011), and improve the usability of video tutorials (Pongnumkul et al. 2011). Most commercial video conferencing tools today include the ability for a remote party to control single user software on someone else's computer using screen sharing and input event injection. Recent research works in VR systems (Zhao et al., 2019; Hartmann et al., 2019; Thoravi Kumaravel et al., 2020; Thoravi Kumaravel and Wilson, 2022) highlight the value of application-

independent compositing of information into VR. We build on these retrofitting approaches with a focus on facilitating efficient asymmetric communication in existing VR applications.

2.4 Mixed-Modal Tutorials and Guidance

Prior work explored the use of mixed-modal guidance, where different types of media are employed to guide a user's actions. Recent examples of this include MixT Chi et al. 2012), where a mix of video and text content are used to convey information; ToolClips (Grossman and Fitzmaurice 2010) where contextual video is used to guide actions in a software tool; ElectroTutor (Warner et al., 2018) uses a combination of textual instruction, interactive questions, and signals to generate tutorials on building physical computing systems; and Torta (Mysore and Guo, 2017) that uses a combination of screencast videos along with underlying OS activity traces for generating mixed-media GUI and command-line app tutorials. In our work, we employ different types of media, but in contrast to prior work, we operate with, and deliver media that spans across different dimensions of visual modality.

3 ADAPTING THE CSCW MATRIX FOR MIXED REALITY ENVIRONMENTS

The focus of this work is primarily about collaborative interactions in Mixed Reality media and it is important to understand how the interactions in a Mixed Reality medium fit into the canonical CSCW matrix by Johansen (1988). The two axes of this canonical CSCW matrix are: 1) Time/Synchronicity—Does the interaction between the users need to happen in real-time? 2) Space (Co-located/Remote)—Does the interaction between the users happen in the same physical space?

3.1 The Canonical CSCW Matrix Cannot Describe All Mixed Reality Collaborations

Through studying prior works as well as existing Mixed Reality interactions, we realized that the matrix is not sufficient to capture the nuances of the interactions that occur between the different collaborating users in Mixed Reality environments. To illustrate this, we list a few counterexamples where categorizing interactions into the canonical CSCW matrix is not particularly effective.

1. *Co-located spectator of an AR user*: Though the interaction between the two users can be considered to be co-located and synchronous, it is important to note that the spectator can not see the virtual elements seen by the AR user.
2. *Co-located spectator viewing mirror video feed of a VR user*: The two users are still co-located and synchronous. However, the VR user cannot see the external world. While the spectator as well as the VR user can see the virtual world, they do not have the same *Output and Input Dimensionality*. i. e. the spectator can only see the first-person view feed of the VR user, and can only do so through a 2D video. Furthermore, they have no interaction with the video.

3. *Co-located VR users using different VR apps*: Here, the collaborating users are co-located and synchronous, however they operate in different virtual worlds and sometimes may even be completely unaware of the context of actions of another user.

4. *Social VR apps*: Perhaps the most common of scenarios is the realm of social VR apps. Here, the collaborating users might be in remotely located physical spaces, but are in the came co-located virtual space. Moreover, they may all be seeing the virtual world in 3D or be participating through a 2D web interface.

5. *Mixed-Reality Telepresence*: Collaborative interactions in Mixed-Reality environments is a well-studied topic in the literature. Here, a multitude of interactions may not fit the conventional CSCW matrix. Consider Loki, in case of a remote guidance task carried out by two users, A help-seeker user (learner) wears an AR headset, whereas a remote expert wears a VR headset. Though they are located remotely, both of them view the same physical environment of the help-seeker in 3D. While the help-seeker views it directly through lenses of the AR headset, the remote expert views a live 3D reconstruction of the same.
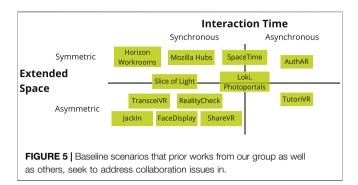
Prior work (Kunert et al., 2014; Gugenheimer et al., 2017; Cheng et al., 2017; Gugenheimer et al., 2018, Hartmann et al., 2019, Thoravi Kumaravel et al., 2019a,Thoravi Kumaravel et al., 2019b; Hartmann et al., 2020; Thoravi Kumaravel et al., 2020; Wang et al., 2020) have built multi-user systems and interactions, that fit one or more the above listed examples, allows for collaboration and communication between users in Mixed Reality environments. However, these systems do not have a meaningful place in the canonical CSCW matrix.

3.2 Extended Collaborative Space

To address the shortcomings of the canonical CSCW matrix towards modelling interactions in Mixed Reality environments, in this work, we propose a modified *CSCW matrix for Mixed Reality environments*. We redefine the notion of a space from a "Physical Space" to an *Extended Collaborative Space*, abbreviated as *xspace*. We define *xspace* as the Mixed Reality space in the Milgram continuum that is perceived by a user, containing associated information of all objects, both physical and virtual, that are relevant to the collaboration task.

3.2.1 Task Relevancy of *Xspaces*

To design successful interactions, it is important to note that the *xspace* varies with the collaborative task at hand. In an Augmented Reality Space where the virtual elements actively interact with the physical world (e. g. a virtual pet jumping on a couch), all the virtual elements as well as the physical elements constitutes its *xspace*. In a VR activity where the user works only with virtual objects, the *xspace* consists of the virtual environment only. In the same VR activity, however, if the VR user has to interact with some elements of their physical surroundings too [e.g. works by Cheng et al. (2017); Gugenheimer et al. (2018)], then those elements are also part of the VR user's *xspace*.

FIGURE 5 | Baseline scenarios that prior works from our group as well as others, seek to address collaboration issues in.

3.2.2 Visual Representations of *Xspaces*

The collaborating users may each experience and interact with a given *xspace* through a different *modality (s)*, that act as a lens to the *xspace*. For instance, in a mixed reality telepresence system, one may see another user's *xspace*, directly through their own eyes. However, if the other user is located in a physically remote place, then they may see the *xspace* through a live 2D video (**Figure 8E**), a live 3D reconstruction (**Figure 8B**). They may also see it through a combination of both, such as in Hologlyphs in Loki. In contrast, the local user may see the same *xspace* by viewing the physical world directly through their eyes (or video see-through), and possibly any other virtual objects on top of them rendered through an AR device. These *modalities* for perceiving *xspaces*, can be broad in terms of other sensory characterizations such as haptic, auditory, visual. However, in this work we will focus only on the *Visual modality* which is the dominant one in current Mixed Reality environments. In this work, we explore the different visual representations of the visual modality, e.g., Reconstructed 3D point clouds, 3D interactions, stereo video, regular 2D video. These are illustrated in **Figure 8**.

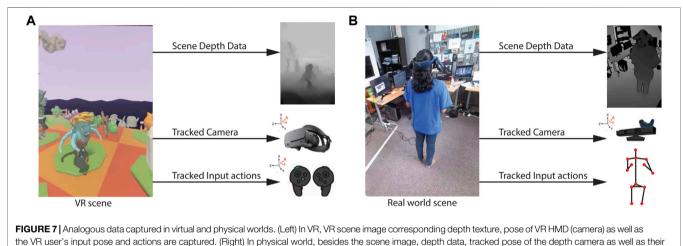### 3.2.3 What Makes *Xspaces* Asymmetrical Across Users?

The interaction between users who occupy *xspaces* can be classified as symmetrical or asymmetrical. We identify four elements, which if different (or absent for a user) between the two users, can lead to asymmetry. Later, in this paper, we have tabulated the nature of asymmetry for our prior works that we discuss (**Figures 14**, **15**, **17**).

- Task-relevant aspects of Virtual parts of the *xspace*
- Task-relevant aspects of the Physical parts of the *xspace*
- Visual-representation used to perceive the physical/virtual parts of the *xspace*
- Interaction Affordance of physical/virtual parts of the *xspace*

Empowered by this notion of *xspace*, we can now categorize prior works based on the baseline mixed-reality scenarios in which they try to address collaboration issues in. This is shown in **Figure 5**. Typically these prior works try and bridge the asymmetry across the *xspaces* that users operate in. A few selected ones are tabulated in **Figure 6**.

System & Purpose	Access to Physical Space		Access to Virtual Space		Visual Fidelity		Interaction Affordance	
	Without System	With System	Without System	With System	Without System	With System	Without System	With System
TutoriVR Asynchronous Learning by one VR user from another VR user	NA	NA	Asymmetric. VR tutor has full access to their virtual space. The VR learner sees only what gets recorded in the video.	Mostly same as without system. But the VR learner can also see all the controller interactions that may not be present in the video frame.	VR tutor sees their environment in VR and perceives 3D. The VR learner sees only a 2D video.	Better than without system. The VR learner can see the video as well as controller interactions in 3D.	Asymmetric. The VR learner has no affordance beyond navigating the video.	Better than without system. The VR learner can take different perspective and access additional information about controller interactions
RealityCheck Collaboration between a co-located VR user and non-VR user for a task in the physical world	Asymmetric. VR user cannot see the physical space	VR user can see the physical space in 3D	NA	NA	Asymmetric - Because the VR user does not see the physical space.	Better than Asymmetric - The VR user can now see the real world in 3D through 3D point clouds.	Asymmetric. The VR user cannot interact with the physical world because they cannot see it.	Symmetric. The VR user and the non-VR user can interact with the physical world in the same capacity
TransceiVR Collaboration between a VR user and non-VR user for a VR task	NA	NA	Asymmetric. External user only sees what the VR user sees	Better than without system. External user can see along directions beyond what the VR user sees.	Asymmetric	Asymmetric	Asymmetric. The external user has no affordance.	Better than without system. External users can now spawn direction arrows, annotations, shared screens and trigger controller highlights in the scene.
SpaceTime Collaboration between two VR users for 3D modeling	NA	NA	Symmetric	Symmetric	Symmetric	Symmetric	Asymmetric. Only one user can control and manipulate a virtual object at a time.	Symmetric. Both users can work with same set of the virtual objects, at the same time, in a similar manner.

FIGURE 6 | Sources of *xspace* asymmetry addressed by select by systems in select prior works.

FIGURE 7 | Analogous data captured in virtual and physical worlds. (Left) In VR, VR scene image corresponding depth texture, pose of VR HMD (camera) as well as the VR user's input pose and actions are captured. (Right) In physical world, besides the scene image, depth data, tracked pose of the depth camera as well as their actions (through skeletal tracking) are captured.

While the physical space and time aspects of collaborative interactions have been studied in the CSCW literature, interactions between users spanning different *xspaces*, using different modalities have been studied to a lesser extent. In our work, we demonstrate that, systems that deploy Interactive Mixed-dimensional Media techniques, allow for effective collaboration between users of Mixed Reality environments. These users might span multiple sectors of this new CSCW matrix, i.e. across different *xspaces* and synchronicity. Our proposed Interactive Mixed-Dimensional Media techniques allow for efficient communication of information and collaboration across these sectors.

To enable this, in our work, we are going to make assumptions about what kind of data is available about an *xspace* in order to characterize it. Interactive mixed-dimensional media that we have used in our works involve capturing four specific types of data: 1) RGB video frames, 2) Corresponding depth data, 3) Tracked Camera, and 4) Position tracked input actions by the user in the *xspace*. These are illustrated in **Figure 7**.

4 INTERACTIVE MIXED-DIMENSIONAL MEDIA

We define *Interactive Mixed-Dimensional Media* as a medium that has components of varying visual dimensional representation. Different representations can be chosen automatically, based on context, or through associated interaction techniques that give users control, over exploring spatial, temporal, and dimensional levels of detail. It ensures that any information or interaction makes sense across different dimensions, interfaces and spaces.

An Interactive Mixed-Dimensional medium allows for presenting an *xspace* in various visual representations—3D point cloud, Reconstructed using abstract 3D graphics, Stereo video and 2D video. Some of these are illustrated in **Figure 8**. We have used different combinations of these representations in our works. For instance, an *xspace* can be presented as a 2D video, 360 videos, as a stereoscopic video, as a 3D point cloud, as an abstract 3D representation etc,. It may also be presented
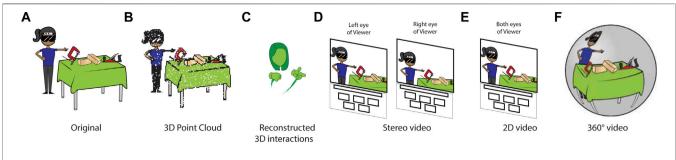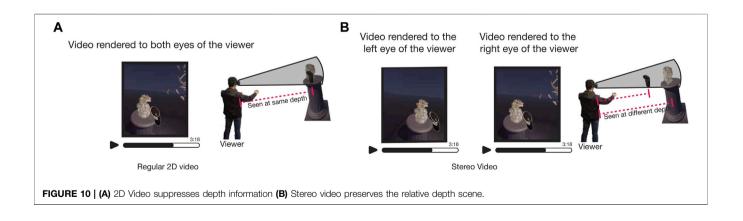
FIGURE 8 | Examples of a few possible visual representations of *xspaces* that are used in prior work. **(A)** Original, **(B)** 3D point cloud, **(C)** Reconstructed 3D interactions, **(D)** Stereo video, **(E)** 2D video, and **(F)** 360° video.
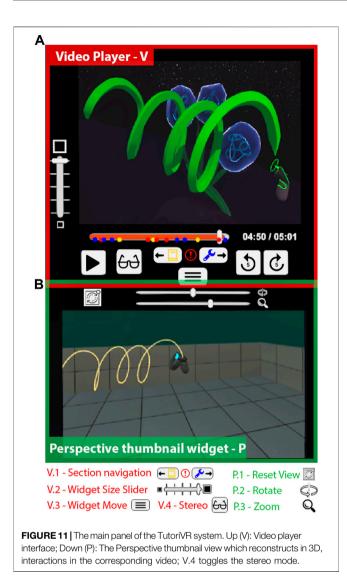
Factor	Level	Relevant Physical Space N/A	Relevant Virtual Space Instructor's VR space that includes all components of the VR design
Visual Representation	Same	N/A	
	Limited Representation - Same Output Dimensionality		With TutoriVR - Stereo Video, 3D interactions
	Limited Representation - Reduced Output Dimensionality		
	Video Only		Without TutoriVR (Recorded VR mirror)
	Not Visible		
Interaction Affordance	Same	N/A	
	Limited Representation - Same Input Dimensionality		With TutoriVR - Perspective exploration
	Limited Representation - Reduced Input Dimensionality		
	No Input		Without TutoriVR (Recorded VR mirror)

FIGURE 9 | Role of Interactive Mixed Dimensional Media in enhancing learning of VR tasks from 2D videos.

FIGURE 10 | (A) 2D Video suppresses depth information **(B)** Stereo video preserves the relative depth scene.

as a collage of the above forms, providing users with alternative sources of information to better perceive the underlying data, that is at the root of all these representations. For instance, In Loki (**Figure 9**), a collage

of live 2D videos and point clouds is used to present instructional information.

A key property of Interactive Mixed-Dimensional Media is that Interactions made over one representation of the *xspace*, can

FIGURE 11 | The main panel of the TutoriVR system. Up (V): Video player interface; Down (P): The Perspective thumbnail view which reconstructs in 3D, interactions in the corresponding video; V.4 toggles the stereo mode.

be made to translate to other representations and instances of the same *xspace*, in the different forms. For instance, in TransceiVR, annotations made over a 2D video feed (**Figure 10A**), can either be projected to the 3D space (**Figure 10B**) in a meaningful manner, or can be shared as such in a 2D format if required. Depending on the context, either one of the interactions might be useful. While 3D projected annotations, can be of use for referring to objects in the VR environment, sometimes annotations projected directly in the 3D environment may not make sense. For instance, in **Figure 11** where the annotations made over a video frame are used for shared discussions. This analogous to the interactions that happen with desktop screen-sharing tools today.

In our work, we implemented and studied interactions that enable three specific kinds of exploration of the underlying *xspace* in an Interactive Mixed-Dimensional Medium.

Perspective exploration: Allows a user to get different perspectives of the *xspace*.

Temporal exploration: Allows a user to navigate across and interact with the historical trace of the *xspace*.

Dimensional-detail exploration: Allows a user to explore and enhance different dimensions of the underlying *xspace*. This is similar to the level-of-detail (LOD) effects in computer graphics (Neider et al., 1993), in which one provides different representations of a single artifact each with a different level of complexity. In this work, we use a similar analogy to the context of the dimensional representation.

In the following sections we review how prior systems built in our research group use the time/xspace matrix to analyze issues in collaborations when using different Mixed-Reality environments. We will then see how mixed-dimensional media can be used to address those. Broadly, we will use the following higher-level methodology for designing collaborative interactions for mixed-reality environments. First, we seek to identify the fundamental factors that cause issues in collaboration, and we do that through a series of steps that leverage our extended CSCW matrix.

Step 1: Identify the extended (xspace) for the collaboration task. The xspace is defined as the mixed-reality space perceived by a user, containing those physical as well as virtual objects, that are relevant to the collaboration task at hand. Sometimes, it may seem that a task may seem to have more than one xspace. But in our experience, such a task can usually be broken down to smaller sub-task tasks that each have a single xspace.

Step 2: Identify how each user perceives and interacts with the xspace with the interface that they use in the baseline scenario. Use that to determine which quadrant of the CSCW matrix they fall into.

Step 3: If they fall into either of the asymmetric quadrants, identify the potential sources of asymmetry as one or more of the four factors defined above [i.e. 1) Access to task relevant physical space, 2) Access to task relevant virtual space, 3) Difference in visual modality with which the xspace is perceived by the users, 4) Difference in Interaction affordance for the users.]

Step 4: For the specific task, gauge the impact of each asymmetry on collaboration efficiency between the users.

Once these asymmetries are identified and prioritized, the next step is to design and develop interfaces to mitigate the key asymmetries of concern. This can be any interface, but in our work, we propose interactive mixed-dimensional media as a candidate for solving collaboration issues that arise due to such asymmetry in xspaces. These media have a set of common interaction techniques and patterns that facilitate perspective, temporal and dimensional-detail exploration. These techniques happen to be well suited to broadly mitigate the various asymmetries that we identified through our extended CSCW matrix.

5 TUTORIVR—ENHANCING 2D VIDEO CAPTURES OF VR ENVIRONMENTS WITH 3D INFORMATION

TutoriVR focuses on asynchronous instruction of design activities carried out in VR, specifically Virtual Reality painting. VR painting is a form of 3D-painting done in a VR space. Most users learn the activity through 2D-videos posted by an instructor on the internet. Even though the
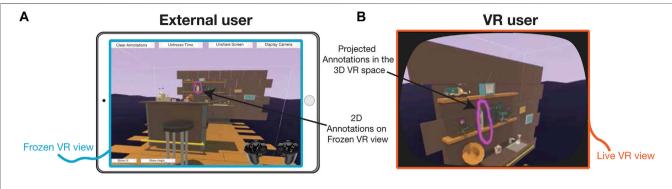
FIGURE 12 | (A): Annotations made over a 2D video for referring to an object in the VR space; **(B)**: Corresponding annotations projected to the right 3D position in the VR space.

Factor	Level	Relevant Physical Space N/A	Relevant Virtual Space VR user's space that includes all virtual elements in it
Visual Representation	Same	N/A	
	Limited Representation - Same Output Dimensionality		
	Limited Representation - Reduced Output Dimensionality		With TransceiVR - Spatial exploration through choice of Viewports
	Video Only		Without TransceiVR (VR Mirror)
	Not Visible		
Interaction Affordance	Same	N/A	
	Limited Representation - Same Input Dimensionality		With TransceiVR - Annotate and project to VR space, share 2D annotations
	Limited Representation - Reduced Input Dimensionality		
	No Input		Without TransceiVR

FIGURE 13 | Role of Interactive Mixed Dimensional Media in facilitating live communication between a user operating a 2D UI and VR user.

instructor as well as the learner operate in a VR environment, this is an example of an asymmetric *xspace* interaction. This is because the relevant parts of the instruction task here are present in the instructor's virtual environment. While, the instructor sees these parts in 3D through a VR headset, the learner sees these only through the visual representation of a first-person view, 2D video.

Formative studies were carried out, in which participants learnt VR painting from 2D videos, that were produced by expert artists. We realized that these videos by themselves fail in delivering crucial details that are required by a learner to understand actions in a VR space. We found that users had major issues in three aspects—1) Depth Judgement, 2) VR Interaction Understanding and 3) Missing out on instructions when following alongside. To address these, TutoriVR deploys interactive Mixed Dimensional Media to enhance this asynchronous learning from 2D video-based Tutorials. Each issue is addressed by what we termed a *VR-embedded widget*. These widgets are interfaces that TutoriVR overlays and embeds inside any existing VR application.

First issue is that of depth judgement. It is hard for users to instantly perceive depth from a 2D video. This happens because the video suppresses, and flattens the z-dimensional depth information in the scene. The user has to rely on other depth cues like occlusion, lighting etc, and these may not be readily present in the screen-captures of a VR-design application. For instance, consider here, video showing an instructor painting a VR sculpture. It was hard for users to quickly figure out the relative depth of the different elements in the scene, e.g. How far is the VR controller from the different parts of the sculpture? The regular 2D video here degrades the z-dimensional depth information in the scene (**Figure 12A**).

To enhance the learner's perception of the z-dimensional perception of the instructor's *xspace*, TutoriVR allows a learner to view the video in stereo (Figure 12B), through a toggle in the UI (**Figure 13**–V.4). To do this, TutoriVR's capture application, while recording the instruction, records the feed rendered to each of the eyes of the instructor. Note that, this stereo rendering is a strong depth cue for the instructor themselves in their own VR environment. During playback, the

learner can choose to view the captured video in stereo, in which the captured video feed of the instructor's left eye is rendered to the learner's left eye. The same is done for the right eye. The rendering is corrected for the IPD of the learner's HMD and the principle of operation is similar to that of a 3D TV. In this way, TutoriVR implicitly provides access to the depth of the scene without a source code access to the original VR application of the instructor.

This stereo-capable video widget is an example of a Mixed-dimensional medium, that allows for dimensional-detail exploration of a user's *xspace*.

Second set of issues that were identified broadly relate to difficulties in understanding the instructor's interactions in the VR environment. These issues can be further sub-categorized by the source of difficulties—1) Perspective of the video capture, 2) Complex Hand Motions by the instructor, 3) Rich Input space of VR controllers and 4) Visibility of Controllers. These are described in detail in TutoriVR (Thoravi Kumaravel et al., 2019b).

To solve these issues, TutoriVR uses another mixed dimensional medium, the perspective thumbnail widget (**Figure 13**-P). This is placed directly below the video player and it provides perspective, temporal as well as dimensional-detail exploration of actions carried out by the VR instructor in their *xspace*. It renders in 3D, a separate set of VR controllers that mimics all the actions carried out by the VR instructor in the video capture. So, whenever a button is pressed in the video, the corresponding buttons on these virtual controllers light up. It is like a fish tank, with walls, and 3D strokes being performed in it as they begin to appear in the video. It offers alternate visual representations that provides additional dimensional-details of these stroke motions. These details are absent in its video counterpart.

This medium visualizes the motion trails of the controller during stroke creation. These fade out gradually, giving users enough time to absorb the information. Furthermore, when the video is paused, the medium renders the entire stroke that is currently being drawn in the video. The users can then pan and zoom (**Figure 13**-P.1–3) to view the stroke independently from different perspectives. This medium enables seeing an isolated and static visual of an ephemeral, dynamic action. They can then resume playing to see the creation from that chosen perspective. This allows for focus-context exploration of such spatio-temporal actions. Upon pausing the video, the medium provides a focused information of the stroke being drawn. It removes other spatial elements in it, as well as draws out the stroke across all time. In contrast the video in the video player provides the context for this stroke.

The medium also solves the issue of understanding spatial motions from the shaky first person video feeds. This is because the walls and grid lines rendered in the medium provide important pictorial and perspective depth cues. These cues may not be there in the original VR environment that the VR instructor operates in. The rendering also offers an effectively increased field of view compared to the video, and stabilizes it by suppressing the noisy and drastic head movements present in the First-person video recording. All these collectively aid in enhancing the learner's understanding of the motion of the VR controllers with reference to the instructor's environment.

We had evaluated the TutoriVR system with an exploratory user study, where we compared it with the video player (with 2D video) only baseline condition (**Figure 13**-P). We designed two tutorial tasks in VR. Participants were asked to watch a tutorial video in VR and were required to replicate the final results in the video as quickly and as accurately as possible. With the presence of TutoriVR, participants achieved significantly more critical steps of the process when using our system compared to the baseline system.

On average across the tasks, users were able to complete 49.2% of instructional steps that involved intricate 3D strokes, 55.4% involved relative 3D depth, and 63.8% of 6DOF controller interactions, in comparison to the baseline condition where the corresponding numbers were 28.3, 31.7, and 41.3% respectively. Qualitative evidence backed this up. Out of 10 participants, eight users felt the Stereo Mode and the Perspective Thumbnail widget helped them in the tasks. Stereo Mode helped users in getting better task awareness and assessing relative 3D structure and depth of the painting. Perspective Thumbnail helped users understand intricate 3D shapes and/or controller interactions. More details of this study can be found in the full paper of TutoriVR.
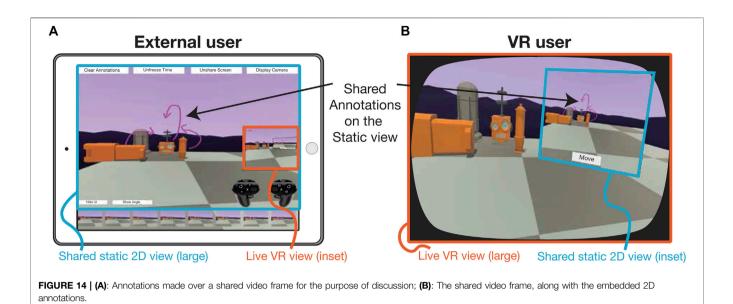
In summary, the mixed-dimensional medium and interactions here, highlights the relevant virtual parts of the instructor's *xspace* that is important for the instruction task, and provides interactions to facilitate its spatial, temporal as well as dimensional-detail exploration. These are summarized in **Figure 14**.

6 TRANSCEIVR—TRANSLATING 2D UI INTERACTIONS OVER TO A 3D MIXED REALITY ENVIRONMENT

In contrast to TutoriVR, the TransceiVR system focuses on synchronous collaboration, where one user is present in VR, while another user is co-located but outside of VR. Such situations are commonplace in collaboration and review on technical projects, showcasing and teaching VR applications in public venues, as well as to friends, and instruction of VR activities.

The collaboration involves carrying out actions with virtual entities in the VR space. The user outside VR, also referred to as an external user, often has access to a live 2D video-feed of what the VR user sees. Despite whether the external user is physically co-located or remote, such an interaction is an example of one where the users are in asymmetric *xspaces*. This is because both users see the virtual space with different visual representations and interaction affordance. This is described in **Figure 15**.

Our prior work, TransceiVR proposes a novel mixed dimensional medium that aims to bridge this asymmetry of *xspaces* in live communication between a VR user and an external user who may or may not be physically co-located. In TransceiVR, formative interviews were carried out with experts who carry out such interactions regularly. It was seen that the current status quo for communication with a VR user, is the VR

FIGURE 14 | (A): Annotations made over a shared video frame for the purpose of discussion; **(B)**: The shared video frame, along with the embedded 2D annotations.

mirror that shows the live 2D video feed of what the VR user is doing. Broadly, there were two main problems with interactions using this feed.

First was due to the shaky first person feed that is being observed by the external user. A number of problems arise due to this—since the mirror shows the VR user's view, the object being referred to is only visible when the VR user is fixating on it. It may not be visible long enough to be seen by the external user if the VR user accidentally looks in another direction. Worse, if external users were co-located, the VR user may instinctively respond to verbal comments, occasionally turning towards them, thereby changing views much more frequently.

In general, frequent head motion changes by the VR user cause two main problems: 1) They affect the ability to easily reference things in the environment, and 2) The external user looses their view of an object of interest they wish to inspect. To solve this problem, an external user may request the VR user position their head to look stably at the desired object, so they can discuss it. Though this is inconvenient, due to lack of alternatives, this is currently being used in scenarios involving joint review in domains such as medicine, architecture, art, and film, where both users need to pay closer attention to details in the relevant parts of the VR scene.

Second set of issues were broadly about the difficulties relating to talking about VR scene elements of differing types—1) transient elements, 2) Elements that cannot be verbally described, 3) Gestural Elements, 4) Directional and Attentional elements.
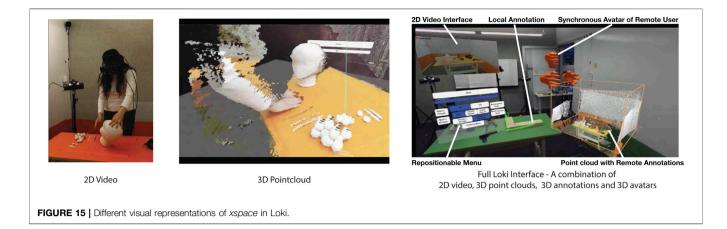
To solve these issues, TransceiVR augments existing VR applications with an interactive mixed dimensional medium. This medium builds over existing live 2D video mirror, but also provides the ability for the external user to do three actions with the 2D video frames—1) *freeze* the live frame, navigate these frames 2) temporally, as well as 3) spatially. The *freeze* action and temporal navigation are analogous to

pausing and seeking in videos. The spatial navigation allows the external user to view the last seen frames along different directions in the VR. It enables the external users to observe the VR environment surrounding the VR user.

With TransceiVR, the external users can not only view the video frames that are spread out across time and space, but they can also interact with them through making annotations over them or by sharing them with the VR user. The medium here allows users to specify the *output dimensionality* of the annotations, either to be 1) Projected directly into the 3D VR space (**Figures 10A, B**-L,R), or 2) as 2D sketches over the video frame, which can then be shared with the VR user as an annotated image (**Figures 11A, B**). Each of these choices have their own advantages. The direct 3D projection can be useful for quick and easy referencing of objects in the VR scene. When the VR user is not looking at the right part of the scene, they are guided to it through 3D arrows and spatial audio cues. The annotated image, on other hand, can be used for more detailed discussion, especially about virtual objects that maybe non-existent or dynamic in the VR scene.

Besides annotations and video frame sharing, the medium also features a 2D UI-based controller panel, through which the external users can refer to and indicate button actions on the VR controllers which are always in motion. Whenever, the external user presses a button on the controller panel, the medium highlights to the VR user, the corresponding button on the controller.

We evaluated Mixed Dimensional media interactions in TransceiVR through a co-located user study. Our goal was to investigate if these interactions can aid users in a scenario where an external user is guiding a VR user, and with an application that was only designed for single-user usage. We compared TransceiVR to the baseline condition—standard VR mirror, through metrics—task success, the task completion time, error rate, and other subjective metrics. With TransceiVR, user pairs

FIGURE 15 | Different visual representations of *xspace* in Loki.

were able to complete tasks faster, with better performance and had a better rate of completing the tasks. All these were statistically significant. When asked to rate the overall experience, eighteen out of twenty of our participants agreed that TransceiVR made the communication easy and efficient. Participants rated (on a 5-point Likert scale the usefulness, higher the better), of individual components of TransceiVR. The highest scoring component were 1) the ability of the medium to perform 2D to 3D projection of annotations, followed by 2) sharing Screen (Median = 4) with embedded 2D annotations. More details of this study can be found in the full paper of TransceiVR.

In summary, during a live interaction, the mixed dimensional medium allows an external user to better perceive the VR user's virtual environment. It also meaningfully translates the external user's interactions across dimensions, from a 2D video to the 3D space of the VR user. These are summarized in **Figure 15**.

Post-TransceiVR, we realized that some types of conversation that were still hard. These conversations typically involved information that need to be shown through objects part of the physical environment of external user, e.g. gestural information, reference objects etc,. Now the external user's *xspace* of the interaction not only involves the Virtual environment, but also the physical environment of the external user, because it has details relevant to the collaboration task at hand. To facilitate this, we later added an ability for the external user, to stream live web-camera feed of themselves into the scene of VR user.

7 LOKI—BI-DIRECTIONAL SYNCHRONISATION OF INTERACTIONS ACROSS DIFFERENT *XSPACES*

The third work, Loki is a Mixed Reality system for facilitating remote instruction of physical tasks using symmetric and two-way mixed reality telepresence. Just like in collaboration over tasks in 3D virtual environments, video-only approach may not be sufficient for physical tasks. Loki leverages video, audio and spatial capture along with mixed-reality presentation methods to allow users to explore and annotate the local and remote environments, and record and review their own performance as well as their peer's. In contrast to TutoriVR and TransceiVR,

the mode of collaboration here can be synchronous as well as asynchronous.

Loki seems different from works such as TutoriVR or TransceiVR, in that it does not focus on virtual tasks, and that the both users use a Mixed Reality interface. However, it still suffers from similar issues in scenarios where users work on a virtual task in an entirely virtual Mixed Reality environment. 1) Asymmetry exists here, wherein the users are present and perceive different physical spaces that they work on. 2) The task is a physical task, and the closed-source nature is embodied in the physical environment itself. Users manipulate, and work over existing tools and objects in their physical environments. With Loki, these are not modified (additionally instrumented with sensors/trackers) for the specific purpose of collaboration, just like how VR softwares are not modified for the purpose of collaboration. Depth cameras, and VR trackers capture the users actions and the physical environment in a non-intrusive manner. In contrast, an alternate approach would require embedding within the different objects of the environment, sensors that can be used to track or even actuate any changes to them. This is analogous to "source-code" access solutions in VR, and there is a vast literature of works that takes an IoT sensing and actuation approach to facilitate remote collaboration.

The primary mixed dimensional medium deployed in Loki is the *Hologlyph*. It contains a point cloud representation and a 2D video of both users. This allows for dimensional detail exploration. Depending upon the nature of what the user seeks, they may choose to view one form over the other, or both (**Figure 9**). This may also facilitate focus-context exploration, For instance, for a fine-grained task, the point cloud can give the context, while the video provides a focused information. For inherently spatial tasks, video can provide the context, and point cloud provides the focused information.

Naturally, hologlyphs allow for spatial exploration, through providing users, with the ability to position, scale and navigate their partners space and gain different perspectives. They can do this in AR too, where they can also see their own space as well as in VR, where they can focus only on their partner's space. This provides the means to a user to navigate to the
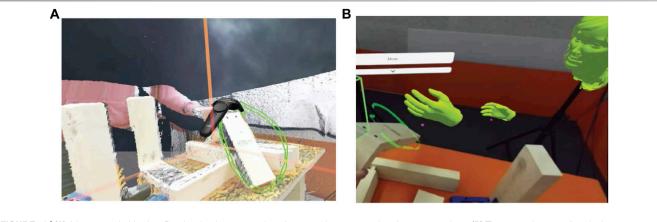
FIGURE 16 | (A) A learner co-habits the 3D point cloud representation of a remote instructor and performs annotations; **(B)** The annotation as well as the learner rendered as a 3D avatar (in AR) in the corresponding position in the instructor's space.

Factor	Level	**Relevant Physical Space** Work bench of the remote Instructor	**Relevant Virtual Space** Virtual Annotations and avatars anchored to the Instructor's workbench
Visual Representation	Same		With Loki - Virtual annotations and avatars are synced across the point clouds and the corresponding physical space
	Limited Representation - Same Output Dimensionality	With Loki - Live 3D point clouds and 3D avatars of users	
	Limited Representation - Reduced Output Dimensionality		
	Video Only	Without Loki	
	Not Visible		Without Loki
Interaction Affordance	Same		With Loki - Interface for accessing and interacting with virtual elements is symmetrical for both users
	Limited Representation - Same Input Dimensionality	With Loki - Free 3D exploration in the space, but cannot manipulate the physical objects	
	Limited Representation - Reduced Input Dimensionality		
	No Input	Without Loki	Without Loki

FIGURE 17 | Role of Interactive Mixed Dimensional Media in facilitating Mixed Reality telepresence for instruction of physical tasks.

specific *xspace* of interest, which varies depending upon the goals of the collaboration—e.g. giving/receiving instruction, discussion etc,. Besides the point clouds, either of the users can also toggle between video views provided by underlying data. This offers an alternative means of spatial exploration, in the video data.

Hologlyphs also allow for temporal exploration, in which user can record data of any environment. This can be played-back in absence of the other user, leading to an asynchronous mode of collaboration. Users can browse this recorded data in a shared VR space, leading to a synchronous collaboration over the recorded data.

One can either view the Hologlyph of a specific physical environment, from a bird's eye perspective, or also "jump in" to these to see it VR on a 1:1 scale. They acquire an embodied virtual presence in that physical environment. They can move around in it, and interact with them through annotations. In case

of a live Hologlyph, the spatial presence of a user immersed in a Hologlyph, as well as their annotations are translated and rendered in a meaningful manner into the corresponding physical space. This can be seen in **Figure 16**, where the annotations made over a point cloud of a space are also rendered and viewable through AR in the corresponding physical space. Similar transformation is applied to indicate the spatial presence of a user, where they are represented as abstract avatars. **Figure 16**. Thus, Hologlyphs enable users to flexibly, transition and customize the *xspaces* that a user wants to view, and also the visual representation they want to view it through (e.g. video, reconstructed point cloud).

We evaluated Loki, through a qualitative evaluation in which participants used the system to learn hot-wire foam carving from a remote instructor.

Participants varied in the way they used the different visual representations provided by Loki. Some liked to keep the 3D

point cloud small and kept it on the side as a reference material, while others preferred it in 1:1 scale directly opposite or beside them. They reasoned about the trade-offs between the point clouds and the videos; "the point cloud was good because if I miss something in real time, I could just turn around and see a slightly different perspective... and if you're in a video, you don't want to switch between perspectives, toggle between several videos just to find the right one."; as well as how those trade-offs affect the usage of other features like annotations and collaborative playback review; P5: "point cloud has benefit, you get more 3D perception... you can annotate it in context of the 3D scene."

We found that participants appreciated Loki's unified workflow, to allow for exploration of the different visual representations such as videos, 3D models and point clouds as well as the different interaction affordances such as annotations, collaborative review and playback. Participants found it easy to access these affordances and explore the different visual representations. Most of them also felt that the system helped them engage better with their partners in the one-on-one learning setting of the study and made the learning process enjoyable.

Loki provides an information-rich mixed dimensional medium that leverages 3D point clouds, 3D avatars, videos, annotations amongst others, that helps connect people as they teach and learn real-world tasks. Asymmetric xspaces in Loki, and the value of bridging these are summarized in **Figure 17**.

8 REFLECTIONS ON MIXED-DIMENSIONAL MEDIA INTERACTIONS

From our experience working on these projects, we now reflect on key strengths and challenges of mixed-dimensional media.

8.1 Strengths of Mixed-Dimensional Media

We note three key strengths specific to mixed-dimensional media interactions, that are valuable for guidance and collaboration in Mixed Reality environments: 1) Dimensional Exploration, 2) Inter-dimensional translation, 3) *xspace* Mutual awareness.

8.1.1 Dimensional Exploration

Mixed-dimensional media interactions, allows a user to vary and explore the different dimensions through the different visual representations of the underlying data. Each representation has its merits. Today, 2D videos are popularly understood, and a well supported visual representation to scalably store, transmit and exchange information. Almost any media device today supports recording and playback of video-based information. Hence, it is inevitable to avoid their usage. However, in the context of conveying information about a Mixed Reality space, they suffer from numerous issues such as shaky first-person view feeds, information occlusion, lack of depth information, lack of ambient spatial awareness and the inability to transition to alternate perspectives. In our works, we have explored other visual representations beyond 2D video that solve some of these issues: alternate scene view ports, stereo 3D videos, reconstructed 3D interactions of a user and 3D point cloud

reconstructions of a scene. These representations allow for communication of information that maybe ambiguous in a 2D video only format.

In TutoriVR, we have a recorded video of an instructor performing activities in VR. There is a lot of 3D interaction that the instructor performs in the VR space, as well as 3D information that is present in the scene. Some of these information is lost when converted to a 2D video. For a learner in VR, TutoriVR uses Stereo 3D to enhance the depth information in regular 2D video captures. It also reconstructs in 3D, the key user gestures and their interactions with their VR controller. During the reconstruction, it adds stabilization to the VR user's movements, and leaves out any other occluding and unnecessary objects that maybe present in the original video.

In TransceiVR, we have one user in VR who is capable of 3D perception and interaction. The other user is external to VR, operating a 2D UI, and is not easily capable of 3D perception or interaction. During a live interaction, TransceiVR allows an external user to independently take alternate view ports into the scene, and interact with them. External users can choose to perform annotations in a 2D format, or a 3D format.

And Finally, Loki allows a user to view multiple 2D video feeds, as well as a 3D point cloud reconstruction of another user's space. Users can freely explore the 3D point cloud at different scales. They can miniaturize it and get a bird's eye perspective, as well as "jump" into it, and experience it on a full first-person scale (1:1). Coupled with dimensional exploration, spatial and temporal exploration are also valuable in all these scenarios. But the latter two have been well studied in prior works, and are not specific to Mixed Reality environments.

8.1.2 Inter-Dimensional Translation

We have different visual representations of an *xspace*, and a user may be operating on any of the representations. It then becomes important to ensure that any interaction carried out over a specific visual representation is translated in a meaningful manner to other visual representations.

In TutoriVR, key information relevant to the VR activity is recorded through stereo captures and logging the VR HMD and controller poses. This is used to render meaningful, and valuable visual representations for the learner—stereo 3D, reconstructed 3D interactions and 2D video. Since, this is an asynchronous and uni-directional communication, translation is required only from the VR instructor to the VR learner, and not the other way.

In case of TransceiVR and Loki, this is not the case. They involve synchronous and bi-directional communication between two users. Annotations made in TransceiVR, when in 3D format needs to projected into the correct 3D location in the VR scene. Alternatively, 2D annotations over specific frames, need to be rendered as a shared screen in the VR scene.

Similarly in Loki, annotations made over a virtual 3D point cloud representation of a space is translated and rendered at the right position, when the corresponding space is viewed through AR. This is also the case for user presence in a space. If a user is viewing the point cloud of a space from a specific viewpoint, then they are rendered as a virtual avatar in the AR view of the space.

8.1.3 Mutual Awareness

Mixed-dimensional media is valuable in facilitating communication and guidance between users operating across asymmetric *xspaces*. In all our works, users operated in asymmetric *xspaces*, and we found that due to this, it is important to always provide an ambient awareness of activities carried out by another user. In TutoriVR, this is provided by an awareness widget that anchors the instructor's tutorial video to the FOV of a user as they wander in their own *xspace*. In Loki, this was provided by miniaturized Hologlyphs that always showed the live feed of the other user's *xspace*. In TransceiVR, when the external user is not primarily viewing the live feed of the VR user, such as when that share and annotate screens, then live view of the VR user is shown through an inset overlaid over the shared screen (**Figure 11B**). These mutual awareness mechanisms, while not core to mixed-dimensional media interactions, are however essential when users operate across asymmetric *xspaces*.

8.2 Challenges to Mixed-Dimensional Media

8.2.1 Access to Calibrated 3D Data and Multiple Camera Viewports

A major challenge to facilitating Mixed Dimensional media interactions is a streamlined way to access the 3D information in a virtual or physical environments. Physical environments are required to be instrumented with depth cameras such as Kinect, iPhone depth cameras. To acquire multiple view ports into the scene, one may need to install more than one cameras, and these need to be calibrated to get spatially synchronised 3D meshes. In virtual environments, there are currently no straightforward and universally accepted mechanisms to get depth information from a VR scene, or to spawn spectator cameras at arbitrary locations in the scene.

Beside access to the depth data, another concern is in obtaining a noise-free 3D reconstruction. Today, most depth-based reconstructions suffer from the issue of depth shadows and noise. Hence, there is a need for high precision depth data and algorithmic techniques that could improve the quality of reconstructions.

8.2.2 Sensing Human Activity

When users operate across asymmetric *xspaces*, we see that it is important to keep track of user's actions, so that they can be communicated to other users. In our works, for activities carried out in a virtual space, we leveraged the HMD and hand pose tracking that is available through today's VR systems. However, this misses out on the nuances face and body expressions and poses. For activities carried out in a physical space, we use the Kinect's 3D point cloud as a proxy. This however suffers from noise, depth shadows and occlusions. Solutions such as full-body tracking suit can go a long way in capturing data on human activity, but these are also inconvenient and expensive. Recent works that explore tracking and reconstruction of human activities through RGB cameras Cao et al. (2019); Ma et al. (2021) would prove to be valuable to the future of Mixed Dimensional Media interfaces.

8.2.3 Information Density vs. Cognitive Load

When interacting with mixed-dimensional media, a user tends access to much more information and UI when compared to just a video, i. e different dimensional representations of a data and its associated UI to browse through and interact with them. When not designed correctly, the UI for these could bloat easily, hamper the user experience. Depending upon the nature of the collaboration task, one needs to carefully balance the trade-off between the density of information presented to the users, and the cognitive load required to perceive and interact with them. This balancing is usually done through identifying the minimum information required for accomplishing the goals of a collaboration task, and then designing the media interactions around that.

8.2.4 Will There Be Asymmetric *Xspaces* in the *Metaverse* Future?

Mixed-dimensional media interactions help bridge the gap in guidance and collaboration across asymmetric *xspaces*. With emergence of online virtual space such as *Horizon Workrooms*, *Mozilla Hubs*, one may speculate a Metaverse future, where there is no asymmetry; in which every one uses Mixed Reality displays, operates in virtual worlds, does virtual tasks and views all information in 3D. We believe that such a future is not possible. There will always be tasks that will be carried out in the physical world, where one may need to collaborate with users over the internet. Similarly, 2D media and displays will continue to play an active role in many tasks. As Buxton (2007) states: "Everything is best for something and worst for something else". The new challenge is to figure out how these 3D virtual worlds will interface with our 2D digital interfaces and 3D physical worlds. We hypothesize that Interactive Mixed-dimensional Media will pave way for this.

9 CONCLUSION

In this paper we introduced Interactive Mixed-Dimensional Media. We motivated the need for these Media through several examples that show that a canonical CSCW time-and-space analysis is insufficient for collaboration in Mixed Reality settings. We then proposed a modified CSCW matrix that introduces the notion of symmetric or asymmetric extended spaces (xspaces). Interactive Mixed-Dimensional Media address problems of asymmetric interaction in Mixed Reality Space by allowing users access to different visual representations and associated interaction techniques. We discussed three different systems that instantiate the concept of Mixed-Dimensional Media. We hope this conceptual framing will allow other researchers to analyze, design and create useful and usable media for other types of collaborative tasks in Mixed Reality environments.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

BK—Primary Author, Doctoral Candidate BH—Doctoral Advisor.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frvir.2022.766336/full#supplementary-material

REFERENCES

Anderson, F., Grossman, T., Matejka, J., and Fitzmaurice, G. (2013). "YouMove: Enhancing Movement Training with an Augmented Reality Mirror," in Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology, St. Andrews, Scotland, United Kingdom (New York, NY, USA: Association for Computing Machinery), UIST'13. 311–320. doi:10.1145/2501988.2502045

Baudisch, P., Good, N., and Stewart, P. (2001). "Focus Plus Context Screens Combining Display Technology with Visualization Techniques," in Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology, Orlando, FL (New York, NY, USA: Association for Computing Machinery), UIST '01. 31–40. doi:10.1145/502348.502354

Bauer, M., Heiber, T., Kortuem, G., and Segall, Z. (1998). "A Collaborative Wearable System with Remote Sensing," in Digest of Papers. Second International Symposium on Wearable Computers (Cat. No. 98EX215) (IEEE), 10–17.

Benko, H., Wilson, A. D., and Zannier, F. (2014). "Dyadic Projected Spatial Augmented Reality," in Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, Honolulu, HI (New York, NY, USA: Association for Computing Machinery), UIST '14. 645–655. doi:10.1145/2642918.2647402

Bimber, O., and Raskar, R. (2005). Spatial Augmented Reality: Merging Real and Virtual Worlds. Natick, MA, USA: A. K. Peters, Ltd.

Botden, S. M. B. I., Buzink, S. N., Schijven, M. P., and Jakimowicz, J. J. (2008). Promis Augmented Reality Training of Laparoscopic Procedures Face Validity. Simulation Healthc. 3, 97–102. doi:10.1097/sih.0b013e3181659e91

Botden, S. M. B. I., and Jakimowicz, J. J. (2009). What Is Going on in Augmented Reality Simulation in Laparoscopic Surgery. Surg. Endosc. 23, 1693–1700. doi:10.1007/s00464-008-0144-1

Budrionis, A., Augestad, K. M., Patel, H. R., and Bellika, J. G. (2013). An Evaluation Framework for Defining the Contributions of Telestration in Surgical Telementoring. Interact J. Med. Res. 2, e14. doi:10.2196/ijmr.2611

Buxton, B. (2007). Multi-Touch Systems that I Have Known and Loved. Microsoft Res. 56, 1–11.

Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., and Sheikh, Y. (2019). Openpose: Realtime Multi-Person 2d Pose Estimation Using Part Affinity Fields. IEEE Trans. Pattern Anal. Mach Intell. 43, 172–186. doi:10.1109/TPAMI.2019.2929257

Cheng, L.-P., Marwecki, S., and Baudisch, P. (2017). "Mutual Human Actuation," in Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, Quebec, QC, Canada (New York, NY, USA: Association for Computing Machinery), UIST '17. 797–805. doi:10.1145/3126594.3126667

Cheng, L.-T., Rohall, S. L., Patterson, J., Ross, S., and Hupfer, S. (2004). "Retrofitting Collaboration into Uis with Aspects," in Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, Chicago, IL (New York, NY, USA: Association for Computing Machinery), CSCW '04. 25–28. doi:10.1145/1031607.1031612

Chi, P.-Y., Ahn, S., Ren, A., Dontcheva, M., Li, W., and Hartmann, B. (2012). "MixT: Automatic Generation of Step-by-Step Mixed Media Tutorials," in Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology - UIST '12 (New York, New York, USA: ACM Press), 93. doi:10.1145/2380116.2380130

D'Angelo, S., and Begel, A. (2017). "Improving Communication Between Pair Programmers Using Shared Gaze Awareness," in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO (New York, NY, USA: ACM), CHI '17. 6245–6290. doi:10.1145/3025453.3025573

Dixon, M., and Fogarty, J. (2010). "Prefab: Implementing Advanced Behaviors Using Pixel-Based Reverse Engineering of Interface Structure," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1525–1534.

Eagan, J. R., Beaudouin-Lafon, M., and Mackay, W. E. (2011). "Cracking the Cocoa Nut: User Interface Programming at Runtime," in Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, Santa Barbara, CA (New York, NY, USA: Association for Computing Machinery), UIST '11. 225–234. doi:10.1145/2047196.2047226

Ens, B., Lanir, J., Tang, A., Bateman, S., Lee, G., Piumsomboon, T., et al. (2019). Revisiting Collaboration Through Mixed Reality: The Evolution of Groupware. Int. J. Human-Computer Stud. 131, 81–98. doi:10.1016/j.ijhcs.2019.05.011

Fraser, M., Benford, S., Hindmarsh, J., and Heath, C. (1999). "Supporting Awareness and Interaction Through Collaborative Virtual Interfaces," in Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology, Asheville, NC (New York, NY, USA: ACM), UIST '99. 27–36. doi:10.1145/320719.322580

Fuchs, H., State, A., and Bazin, J.-C. (2014). Immersive 3d Telepresence. Computer 47, 46–52. doi:10.1109/MC.2014.185

Fussell, S. R., Setlock, L. D., and Kraut, R. E. (2003). "Effects of Head-Mounted and Scene-Oriented Video Systems on Remote Collaboration on Physical Tasks," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Ft. Lauderdale, FL (New York, NY, USA: Association for Computing Machinery), CHI '03. 513–520. doi:10.1145/642611.642701

Greenhalgh, C., Flintham, M., Purbrick, J., and Benford, S. (2002). "Applications of Temporal Links: Recording and Replaying Virtual Environments," in Virtual Reality, 2002. Proceedings. IEEE (IEEE), 101–108.

Grossman, T., and Fitzmaurice, G. (2010). "ToolClips: An Investigation of Contextual Video Assistance for Functionality Understanding," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, GA (New York, NY, USA: ACM), CHI '10. 1515–1524. doi:10.1145/1753326.1753552

Gugenheimer, J., Mai, C., McGill, M., Williamson, J., Steinicke, F., and Perlin, K. (2019). "Challenges Using Head-Mounted Displays in Shared and Social Spaces," in Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, United Kingdom (New York, NY, USA: Association for Computing Machinery), CHI EA '19. 1–8. doi:10.1145/3290607.3299028

Gugenheimer, J., Stemasov, E., Frommel, J., and Rukzio, E. (2017). *ShareVR: Enabling Co-Located Experiences for Virtual Reality Between HMD and Non-HMD Users*. New York, NY, USA: Association for Computing Machinery, 4021–4033.

Gugenheimer, J., Stemasov, E., Sareen, H., and Rukzio, E. (2018). "FaceDisplay: Towards Asymmetric Multi-User Interaction for Nomadic Virtual Reality," in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada (New York, NY, USA: ACM) CHI '18. 54, 1–13. doi:10.1145/3173574.3173628

Hamilton, W. A., Garretson, O., and Kerne, A. (2014). "Streaming on Twitch: Fostering Participatory Communities of Play Within Live Mixed Media," in Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, Toronto, ON, Canada (New York, NY, USA: ACM), CHI '14. 1315–1324. doi:10.1145/2556288.2557048

Hartmann, J., Holz, C., Ofek, E., and Wilson, A. D. (2019). "Realitycheck: Blending Virtual Environments with Situated Physical Reality," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, United Kingdom (New York, NY, USA: ACM), CHI '19. 12. 347, 1–347. doi:10.1145/3290605.3300577

Hartmann, J., Yeh, Y.-T., and Vogel, D. (2020). "Aar: Augmenting a Wearable Augmented Reality Display with an Actuated Head-Mounted Projector," in Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (New York, NY, USA: Association for Computing Machinery), UIST '20. 445–458. doi:10.1145/3379337.3415849

Henderson, S. J., and Feiner, S. K. (2007). *Augmented Reality for Maintenance and Repair (Armar)*. Columbia Univ New York Dept of Computer Science. Tech. Rep. doi:10.21236/ada475554

Ishii, H., Kobayashi, M., and Arita, K. (1994). Iterative Design of Seamless Collaboration Media. *Commun. ACM* 37, 83–97. doi:10.1145/179606.179687

Ishii, H., and Kobayashi, M. (1992). "ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Monterey, CA (New York, NY, USA: Association for Computing Machinery), 525–532. CHI '92. doi:10.1145/142750.142977

Ishii, H. (1990). "Teamworkstation: Towards a Seamless Shared Workspace," in Proceedings of the 1990 ACM Conference on Computer-Supported Cooperative Work, Los Angeles, CA (New York, NY, USA: Association for Computing Machinery), CSCW '90. 13–26. doi:10.1145/99332.99337

Izadi, S., Brignull, H., Rodden, T., Rogers, Y., and Underwood, M. (2003). "Dynamo: A Public Interactive Surface Supporting the Cooperative Sharing and Exchange of Media," in Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology, Vancouver, Canada (New York, NY, USA: Association for Computing Machinery), UIST '03. 159–168. doi:10.1145/964696.964714

Johansen, R. (1988). *Groupware: Computer Support for Business Teams*. Free Press.

Kunert, A., Kulik, A., Beck, S., and Froehlich, B. (2014). "Photoportals: Shared References in Space and Time," in Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work amp; Social Computing, Baltimore, MA (New York, NY, USA: Association for Computing Machinery), CSCW '14. 1388–1399. doi:10.1145/2531602.2531727

Lessel, P., Vielhauer, A., and Krüger, A. (2017). "Expanding Video Game Live-Streams with Enhanced Communication Channels," in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO (New York, NY, USA: ACM), CHI '17. 1571–1576. doi:10.1145/3025453.3025708

Ma, S., Simon, T., Saragih, J., Wang, D., Li, Y., De La Torre, F., et al. (2021). "Pixel Codec Avatars," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 64–73. doi:10.1109/cvpr46437.2021.00013

Marwecki, S., Brehm, M., Wagner, L., Cheng, L.-P., Mueller, F. F., and Baudisch, P. (2018). "Virtualspace - Overloading Physical Space with Multiple Virtual Reality Users," in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada (New York, NY, USA: ACM), CHI '18. 241, 1–241. doi:10.10.1145/3173574.3173815

Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1995). "*Augmented Reality: A Class of Displays on the Reality-Virtuality Continuum*" in Telemanipulator and Telepresence Technologies, International Society for Optics and Photonics. 2351, 282–292.

Murphy-Hill, E., Lee, D. Y., Murphy, G. C., and Mcgrenere, J. (2015). How Do Users Discover New Tools in Software Development and Beyond. *Comput. Supported Coop. Work* 24, 389–422. doi:10.1007/s10606-015-9230-9

Mysore, A., and Guo, P. J. (2017). "Torta: Generating Mixed-Media GUI and Command-Line App Tutorials Using Operating-System-Wide Activity Tracing," in Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, Quebec, QC, Canada (New York, NY, USA: Association for Computing Machinery), UIST '17. 703–714. doi:10.1145/3126594.3126628

Neider, J., Davis, T., and Woo, M. (1993). *OpenGL Programming Guide*, Vol. 478. Reading, MA: Addison-Wesley.

Nguyen, C., DiVerdi, S., Hertzmann, A., and Liu, F. (2017). "CollaVR: Collaborative In-Headset Review for VR Video," in Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, Quebec, QC, Canada (New York, NY, USA: ACM), UIST '17. 267–277. doi:10.1145/3126594.3126659

Pires, K., and Simon, G. (2015). "YouTube Live and Twitch: A Tour of User-Generated Live Streaming Systems," in Proceedings of the 6th ACM Multimedia Systems Conference, Portland, OR (New York, NY, USA: ACM), MMSys '15. 225–230. doi:10.1145/2713168.2713195

Pongnumkul, S., Dontcheva, M., Li, W., Wang, J., Bourdev, L., Avidan, S., et al. (2011). "Pause-and-Play: Automatically Linking Screencast Video Tutorials with Applications," in Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, Santa Barbara, CA (New York, NY, USA: Association for Computing Machinery), UIST '11. 135–144. doi:10.1145/2047196.2047213

Ponto, K., Kohlmann, J., and Gleicher, M. (2012). Effective Replays and Summarization of Virtual Experiences. *IEEE Trans. Vis. Comput. Graphics* 18, 607–616. doi:10.1109/TVCG.2012.41

Ranjan, A., Birnholtz, J. P., and Balakrishnan, R. (2007). *Dynamic Shared Visual Spaces: Experimenting with Automatic Camera Control in a Remote Repair Task*. New York, NY, USA: Association for Computing Machinery, 1177–1186.

Savva, M., Kong, N., Chhajta, A., Fei-Fei, L., Agrawala, M., and Heer, J. (2011). "ReVision Automated Classification, Analysis and Redesign of Chart Images," in Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, Santa Barbara, CA (New York, NY, USA: ACM), UIST '11. 393–402. doi:10.1145/2047196.2047247

Smith, T., Obrist, M., and Wright, P. (2013). "Live-Streaming Changes the (Video) Game," in Proceedings of the 11th European Conference on Interactive TV and Video, Coma, Italy (New York, NY, USA: ACM), EuroITV '13. 131–138. doi:10.1145/2465958.2465971

Snowdon, D., Churchill, E. F., and Munro, A. J. (2001). "Collaborative Virtual Environments: Digital Spaces and Places for CSCW: An Introduction," in Collaborative Virtual Environments. Editors E. F. Churchill, D. N. Snowdon, and A. J. Monro (London: Springer), 3–17. Chap. 1. doi:10.1007/978-1-4471-0685-2_1

Thoravi Kumaravel, B., Anderson, F., Fitzmaurice, G., Hartmann, B., and Grossman, T. (2019a). "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 (New York, NY, USA: Association for Computing Machinery), UIST '19. 161–174. doi:10.1145/3332165.3347872

Thoravi Kumaravel, B., Nguyen, C., DiVerdi, S., and Hartmann, B. (2020). "TransceiVR : Bridging Asymmetrical Communication Between VR Users and External Collaborators," in Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (New York, NY, USA: Association for Computing Machinery), UIST '20. 182–195. doi:10.1145/3379337.3415827

Thoravi Kumaravel, B., Nguyen, C., DiVerdi, S., and Hartmann, B. (2019b). *TutoriVR: A Video-Based Tutorial System for Design Applications in Virtual Reality*. New York, NY, USA: Association for Computing Machinery, 1–12.

Thoravi Kumaravel, B., and Wilson, A. D. (2022). "Dreamstream: Immersive and Interactive Spectating for VR," in Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems CHI '22, New Orleans, LA (New York,

NY, USA: Association for Computing Machinery). doi:10.1145/3491102.
3517508

von Willich, J., Funk, M., Müller, F., Marky, K., Riemann, J., and Mühlhäuser, M.
(2019). "You Invaded My Tracking Space! Using Augmented Virtuality for
Spotting Passersby in Room-Scale Virtual Reality," in Proceedings of the 2019
on Designing Interactive Systems Conference, San Diego, CA (New York, NY,
USA: ACM), DIS '19. 487–496. doi:10.1145/3322276.3322334

Wang, C.-H., Tsai, C.-E., Yong, S., and Chan, L. (2020). "Slice of Light: Transparent
and Integrative Transition Among Realities in a Multi-Hmd-User
Environment," in Proceedings of the 33rd Annual ACM Symposium on
User Interface Software and Technology, Quebec, QC, Canada (New York,
NY, USA: Association for Computing Machinery), UIST '20. 805–817. doi:10.
1145/3379337.3415868

Warner, J., Lafreniere, B., Fitzmaurice, G., and Grossman, T. (2018). "ElectroTutor:
Test-Driven Physical Computing Tutorials," in Proceedings of the 31st Annual
ACM Symposium on User Interface Software and Technology, Berlin,
Germany (New York, NY, USA: Association for Computing Machinery),
UIST '18. 435–446. doi:10.1145/3242587.3242591

Xia, H., Herscher, S., Perlin, K., and Wigdor, D. (2018). "Spacetime: Enabling Fluid
Individual and Collaborative Editing in Virtual Reality," in Proceedings of the
31st Annual ACM Symposium on User Interface Software and Technology,
Berlin, Germany (New York, NY, USA: ACM), UIST '18. 853–866. doi:10.1145/
3242587.3242597

Zhao, Y., Cutrell, E., Holz, C., Morris, M. R., Ofek, E., and Wilson, A. D. (2019). "SeeingVR:
A Set of Tools to Make Virtual Reality More Accessible to People with Low Vision," in
Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems,
Glasgow, Scotland, United Kingdom (New York, NY, USA: Association for
Computing Machinery), CHI '19. 1–14. doi:10.1145/3290605.3300341