OXFORD CDE

Journal of Computational Design and Engineering, 2021, 8(2), 756–772

doi: 10.1093/jcde/qwab012
Journal homepage: www.jcde.org

RESEARCH ARTICLE

XR collaboration beyond virtual reality: work in the real world

Yongjae Lee[1,2] and Byounghyun Yoo[1,*]

[1]Center for Artificial Intelligence, Korea Institute of Science and Technology, 5 Hwarangro14-gil, Seongbuk-gu, Seoul 02792, South Korea and [2]School of Mechanical Engineering, Yonsei University, 50 Yonsei-ro Seodaemun-gu, Seoul 03722, South Korea

*Corresponding author. E-mail: yoo@byoo.net http://orcid.org/0000-0001-9299-349X

Abstract

Collaborating in a physically remote location saves time and money. Many remote collaboration systems have been studied and commercialized. Their capabilities have been confined to virtual objects and information. More recent studies have focused on collaborating in a physical environment and with physical objects. However, they have limitations including shaky and unstable views (scenes), view dependence, low scalability, and poor content expression. In this paper, we propose a web-based extended reality (XR) collaboration system that alleviates the aforementioned issues and enables effective, reproducible cooperation. Our proposed system comprises three parts: interaction device webization, which expands the web browser's device interfaces; unified XR representation, which describes content interoperable in both virtual reality (VR) and augmented reality (AR); and unified coordinate creation, which enables presenting physical objects' pose in world coordinates. With this system, a user in VR can intuitively instruct the manipulation of a physical object by manipulating a virtual object representative of the physical object. Conversely, a user in AR can catch up with the instruction by observing the augmented virtual object on the physical object. Moreover, as the pose of the physical object at the AR user's worksite is reflected in the virtual object, the VR user can recognize the working progress and give feedback to the AR user. To improve remote collaboration, we surveyed XR collaboration studies and proposed a new method for classifying XR collaborative applications based on the virtual–real engagement and ubiquitous computing continuum. We implemented a prototype and conducted a survey among submarine crews, most of whom were positively inclined to use our system, to convey that the system would be helpful in improving their job performance. Furthermore, we suggested possible improvements to it to enhance each participant's understanding of the other user's context within the XR collaboration.

Keywords: extended reality; virtual reality; augmented reality; XR collaboration; XR content representation; unified coordinate system; webizing

1 Introduction

The computer-supported cooperative work concept (Grudin, 1994; Lee & Paine, 2015), first advocated by Irene Greif in 1984, introduced simple collaboration tools such as e-mail and video conferencing. More recently, with the rapid development of technology, more advanced types of collaboration have been presented. In particular, virtual reality (VR) and augmented reality (AR) techniques have realized phenomena that are impossible in the physical world, enabling users to experience immersive and information-rich collaboration (Ceruti *et al.*, 2019; Fukuda *et al.*, 2019; Sun *et al.*, 2019; Soler-Domínguez *et al.*, 2020).

Extended reality (XR) collaboration stands for collaboration not only between VR applications and between AR applications

Received: 17 December 2020; **Revised:** 4 February 2021; **Accepted:** 6 February 2021

but also between VR and AR applications. XR collaboration enables users to experience an asymmetrical interaction environment. A remote assistant AR method (Gauglitz *et al*., 2014; Fakourfar *et al*., 2016; Nuernberger *et al*., 2016; Wang *et al*., 2019) that communicates the user's situation using video streaming has been widely used. However, there has been a problem in that it interferes with the viewer's context understanding due to the unstable view and dependent scene observation caused by the AR user's body movements (Kasahara *et al*., 2017). Other studies (Chen *et al*., 2015; Le Chénéchal *et al*., 2016; Lee *et al*., 2017, 2018; Gao *et al*., 2018; Lindlbauer & Wilson, 2018; Teo *et al*., 2019) have proposed XR collaboration that complements the previous problems (unstable views and view dependence) through 360 live video streaming or a VE constructed in real time. However, there are still several issues to consider with these approaches. First, it is difficult and impractical to collaborate outside of the predefined platform and system configuration. For example, when a hand tracker interface is added to a system that supports only a VR controller interface or when an AR device is changed from a hand-held type to a glass type of device, the entire system architecture needs to be modified. Second, the collaboration system and its content are not separable, and there is no proper systematic method for XR collaboration content.

To address these issues, we propose a web-based XR collaboration system. Because the web has platform-independent specifications—defined by the World Wide Web Consortium (W3C)—there is no need to consider the compatibility between the platforms on which an application runs. Furthermore, owing to the recently added WebXR specification (previously defined as WebVR; W3C, 2020b), the functions of VR and AR equipment, which were only available using native programs, have become available on the web. In addition to utilizing the advantages of the web, the expandable interaction event handling mechanism of our proposed system facilitates the easy integration of various interaction devices that are not supported by web browsers into the system. Moreover, by using a unified representation of VR and AR content, the content can be defined independently of the system. Defining a new coordinate system based on a fixed point in real space and synchronizing it to the coordinate system of a virtual world alleviate the unstable view and dependent scene observation problems that have existed in previous collaboration systems.

In the next section, we elaborate on XR and its taxonomy, previous XR collaboration studies, and previous content representation methods. In Section 3, we delineate the principal design concept of the proposed system: a device webizing mechanism, unified XR content representation, and unified world coordinate in VR and AR. In Section 4, we detail a pilot implementation of the proposed system. In Section 5, we discuss the user survey of submarine crews. In Section 6, we present applicable scenarios and discuss the limitations and potential improvements.

The main contributions of this paper are as follows:

(i) A revised expandable device webization method (Seo *et al*., 2018).
(ii) A new method for classifying XR collaborative applications, the virtual–real engagement, and ubiquitous computing continuum.
(iii) A highly engaged XR collaboration method with a unified coordinate system.
(iv) A revised unified XR content representation approach (Lee *et al*., 2020).
(v) Comparison of the proposed system with traditional systems and future direction of XR collaboration.

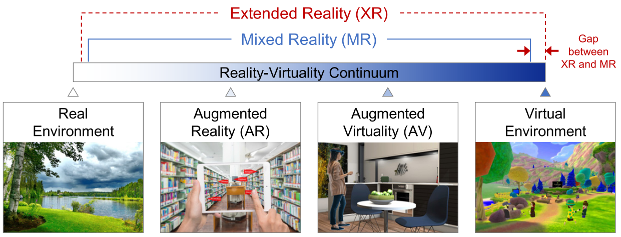


Figure 1: The XR concept expands from Milgram's MR (Milgram & Kishino, 1994).

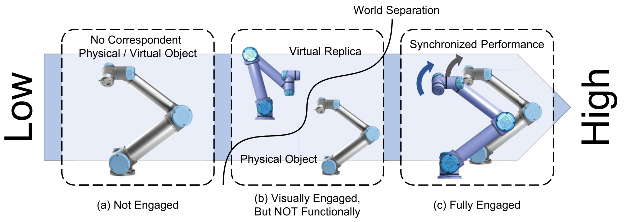


Figure 2: The virtual–real engagement continuum. (The robot arm images are under the copyright of Universal Robots.)

2 Related Work

2.1 Extended reality

In general, XR is understood to be an umbrella term encompassing VR, AR, and mixed reality (MR). In 1994, Milgram and Kishino (1994) proposed a reality–virtuality continuum where the real environment (RE) and virtual environment (VE) were located at opposite ends and AR was located somewhere between them. In his concept of the reality–virtuality continuum, RE and VE are continuously mixed, and all conceivable blended environments between these ends are understood to be MR. Mann *et al.* (2018) thought that Milgram's MR was XR, but there is another view that perceives the range encompassing MR and VR as XR (Al-Adhami *et al*., 2019), as shown in Fig. 1. Another view of XR is Mann *et al.* (2018)'s XR. In his XR, "reality" is broadened by expanding human's sensory capability with wearable computers to see what ordinary people cannot see. In 2009, Paradiso and Landay (2009) proposed cross reality, which uses ubiquitous sensor/actuator networks to influence RE and VE. Although each concept above envisages XR slightly differently, all three ideas have a common purpose in providing a sense or a perception that does not exist in reality, i.e. reality expansion.

2.2 Virtual–real engagement and XR collaboration

Newman *et al.* (2007) proposed the Milgram–Weiser continuum, classifying MR applications. His classification helps categorize applications that cover a narrow band on Milgram's continuum; however, it is insufficient to sort XR collaboration applications that often cover a broad spectrum of the continuum. In XR collaboration, not only does the user have the choice of selecting an environment that will effectively deliver useful task information to coworkers, but also the collaborative application should have the same effect, regardless of the user's environment. Thus, it is important not to provide a mixed environment at a specific ratio, but rather to an environment in which VE and RE are entangled with each other.

The more closely intertwined the VE and RE, the more thoroughly they can imitate and reflect each other. Fig. 2 shows the correlation between a virtual object and a physical object

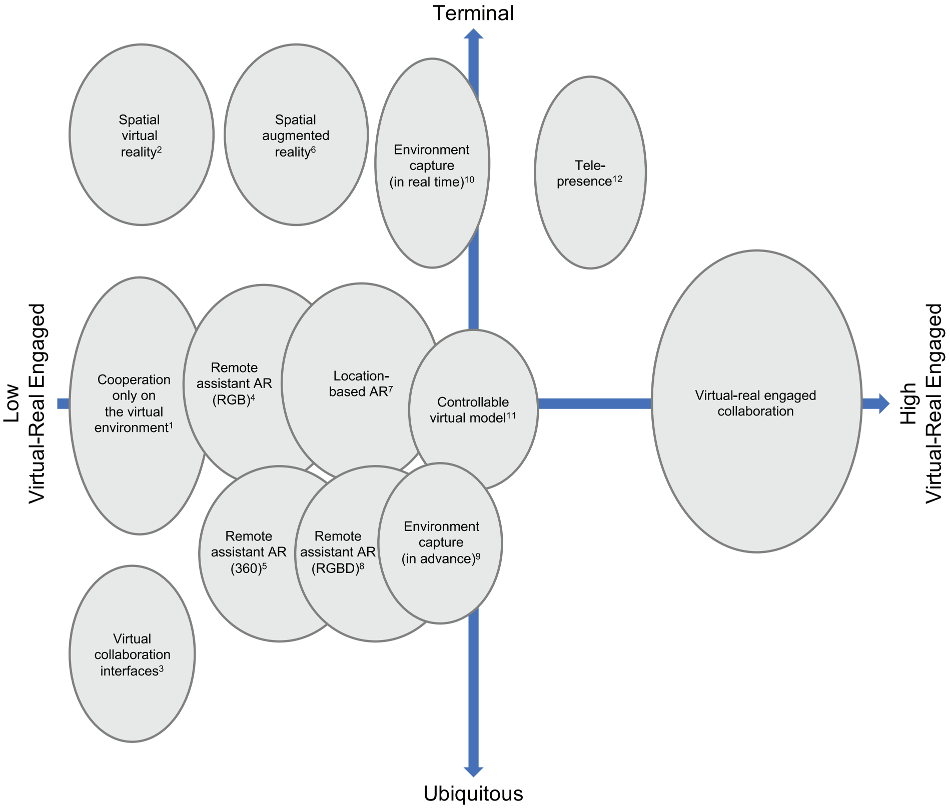

Terminal
Spatial virtual reality[2]
Spatial augmented reality[6]
Environment capture (in real time)[10]
Tele-presence[12]
Low Virtual-Real Engaged
Cooperation only on the virtual environment[1]
Remote assistant AR (RGB)[4]
Location-based AR[7]
Controllable virtual model[11]
Virtual-real engaged collaboration
High Virtual-Real Engaged
Remote assistant AR (360)[5]
Remote assistant AR (RGBD)[8]
Environment capture (in advance)[9]
Virtual collaboration interfaces[3]
Ubiquitous

Figure 3: The virtual–real engagement and ubiquitous computing continuum. (Examples of each category are listed in low to high engagement order in Table 1. Superscript number indicates the order specified in the first column of Table 1.)

based on the VE and RE degree of engagement. When the VE and RE are not engaged, only the physical object or the virtual object exists (Fig. 2a). Second Life (Linden Research Inc., 2003), which serves as the sole virtual world independent of the real world in which we live, is an unengaged world. The real world is another example that is not engaged. At a slightly higher engagement level, the virtual object resembles the physical object (Fig. 2b). While they appear identical, they are functionally not related; hence, they move separately. This type of virtual world is commonly called a mirrored world. One example is Google Earth (Google Inc., 2005), where a virtual replica reflecting the real world's shape exists. At the highest engagement level, the virtual object duplicates the physical object outwardly and functionally (Fig. 2c). They are strongly intertwined; hence, the physical object's movements are reflected in the virtual object, and vice versa. One of the application domains requiring a high degree of virtual–real engagement is the concept of a digital twin, in which a physical object and its virtual duplicate are interconnected and work as one (Tao *et al.*, 2018). The maturity of a digital twin can be considered as an index of the virtual–real engagement. Stark and Damerau (2019) proposed the eight-dimensional model for planning digital twins, but it is not to be understood as a strict maturity model. In this work, we have focused on the connectivity mode and update frequency among the eight-dimensional model to classify existing work. Integrating virtual–real engagement with Weiser's continuum, Fig. 3 and Table 1 show a new continuum for classifying cooperative XR applications—the virtual–real engagement and ubiquitous computing continuum.

For nonengaged virtual–real studies, collaborative tasks are conducted solely in the VE or RE. Among them, it can be seen that some studies (Kato & Billinghurst, 1999; Shen *et al.*, 2010; Wong & Gutwin, 2014; Zillner *et al.*, 2014; Higuchi *et al.*, 2015; Müller *et al.*, 2016; Dey *et al.*, 2017; Poretski *et al.*, 2018; Grandi *et al.*, 2019; Huh *et al.*, 2019; Pereira *et al.*, 2019) are located at the far left on our continuum (Fig. 3) because users can perform tasks only on virtual objects. Spatial virtual reality (SVR) is an immersive VR type, which requires a fixed space where a projector or display is installed (Bimber & Raskar, 2005a, b) and serves a terminal interface that provides services to multiple users. The studies (Febretti et al., 2013; Beck et al., 2013) employing SVR are at the upper left of Fig. 3. Meanwhile, virtual collaborative interface studies (Pauchet *et al.*, 2007; Tang *et al.*, 2010) supporting a variety of communication channels with numerous personalized devices are seen as close to the ubiquitous section of Fig. 3.

Collaborative VR/AR/MR studies are located in the middle of the terminal and the ubiquitous extremes of Fig. 3. Collaborative VR/AR/MR studies are often configured with a personal device, such as hand-held device or Head-Mounted Display (HMD). Location-based AR (Seo *et al.*, 2016) and remote assistant AR (Lipson *et al.*, 1998; Bauer *et al.*, 1999; Gurevich *et al.*, 2012; Kasahara *et al.*, 2012; Gauglitz *et al.*, 2014; Kim *et al.*, 2014; Fakourfar *et al.*, 2016; Gupta *et al.*, 2016; Nuernberger *et al.*, 2016) studies perform tasks on physical objects in the RE. These studies do not create virtual objects that directly duplicate the physical objects to be collaborated with but differ from previous studies—which did not have virtual–real engagement at all—in that they recognize the physical object in the VE and create a virtual annotation related to it. They are located to the right of the previous nonengaged studies on our continuum. Remote assistant AR studies using a 360-degree camera (Chen *et al.*, 2015; Kratz *et al.*, 2015; Nagai *et al.*, 2015; Kasahara *et al.*, 2017; Lee *et al.*, 2017, 2018;

Table 1: XR collaboration studies and those categories.

No.	Categories[a]	References[a]
1	Cooperation only on the VE	Kato & Billinghurst, 1999; Shen *et al.*, 2010; Wong & Gutwin, 2014; Zillner *et al.*, 2014; Higuch *et al.*, 2015; Müller *et al.*, 2016; Dey *et al.*, 2017; Poretski *et al.*, 2018; Grandi *et al.*, 2019; Huh *et al.*, 2019; Pereira *et al.*, 2019
2	SVR	Beck *et al.*, 2013; Febretti *et al.*, 2013
3	Virtual collaboration interfaces	Pauchet *et al.*, 2007; Tang *et al.*, 2010
4	Remote assistant AR (RGB)	Lipson *et al.*, 1998; Bauer *et al.*, 1999; Gurevich *et al.*, 2012; Kasahara *et al.*, 2012; Gauglitz *et al.*, 2014; Kim *et al.*, 2014; Fakourfar *et al.*, 2016; Gupta *et al.*, 2016; Nuernberger *et al.*, 2016
5	Remote assistant AR (360)	Chen *et al.*, 2015; Kratz *et al.*, 2015; Nagai *et al.*, 2015; Kasahara *et al.*, 2017; Lee *et al.*, 2017, 2018; Piumsomboon *et al.*, 2019
6	Spatial augmented reality	Alem & Li, 2011; Junuzovic *et al.*, 2012; Jones *et al.*, 2014; Irlitti *et al.*, 2019
7	Location-based AR	Seo *et al.*, 2016
8	Remote assistant AR (RGBD)	Sodhi *et al.*, 2013; Gao *et al.*, 2016
9	Environment capture (in advance)	Tait & Billinghurst, 2015; Gao *et al.*, 2017, 2018; Piumsomboon *et al.*, 2017, 2018; Teo *et al.*, 2019
10	Environment capture (real time)	Oda & Feiner, 2012; Tecchia *et al.*, 2012; Adcock *et al.*, 2013; Huang *et al.*, 2013; Lindlbauer & Wilson, 2018
11	Controllable virtual model	Le Chénéchal *et al.*, 2015, 2016; Aschenbrenner *et al.*, 2018; Wang *et al.*, 2019
12	Tele-presence	Petit *et al.*, 2010; Orts-Escolano *et al.*, 2016

[a] Categories are ordered in low to high engagement, and references are chronologically ordered.

Piumsomboon *et al.*, 2019) or depth camera attached to a mobile device (Sodhi *et al.*, 2013; Gao *et al.*, 2016) can convey more context of the RE with a wide field of view or greater degree of dimension information than the previously stated type of remote assistant AR studies—thus implying that they become slightly more virtual–real engaged than the previous types. Other studies (Alem & Li, 2011; Junuzovic *et al.*, 2012; Jones *et al.*, 2014; Irlitti *et al.*, 2019), which require static space to augment virtual objects in the RE, called spatial augmented reality, have terminal interfaces in the way as SVR applications do. These studies reside on the right-hand side of SVR in terms of the virtual objects that permeate the RE.

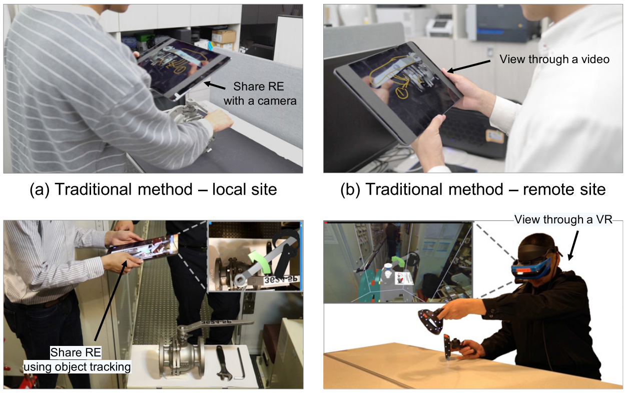
Other studies (Tait & Billinghurst, 2015; Gao *et al.*, 2017, 2018; Piumsomboon *et al.*, 2017, 2018; Teo *et al.*, 2019) also capture a part of the physical space to share real world's context (environment capture in advance). These studies had a virtual copy of the RE as a background to help understand the RE's context, rather than as an object of collaboration, as a changed RE's context is not reflected in the VE. More impressive studies (Oda & Feiner, 2012; Tecchia *et al.*, 2012; Adcock *et al.*, 2013; Huang *et al.*, 2013; Lindlbauer & Wilson, 2018) attempted to copy a part of the physical space from the RE into the VE using depth cameras (environment capture in real time). This method differed from previous methods in that the capturing was still being performed during collaboration. As this method copies the physical space during collaboration in real time, it is possible to convey the RE's changed context. This has the advantage of helping users grasp the real world's context in a more three-dimensional manner. However, it cannot discern virtual objects from the copied space semantically. At first glance, it appears to create virtual objects that duplicate physical objects. However, the data collected from the RE are as raw as voxels or points such as a pixel in a 2D image, so the copied space is not easily controllable and does not correspond to the physical objects.

Studies (Le Chénéchal *et al.*, 2015, 2016; Aschenbrenner *et al.*, 2018; Wang *et al.*, 2019) using virtually duplicated objects of physical objects can convey information more intuitively than simple types of annotations—such as direction arrows and text—by directly showing how to manage objects through virtual objects. Although virtually modeled objects that can be controlled are used, they do not reflect the physical objects' state (e.g. their position and rotation), so they appear to be less engaged than the highest engaged level.

Tele-presence studies (Petit *et al.*, 2010; Orts-Escolano *et al.*, 2016) are a case in which the collaborator's embodiment is the object of collaboration, creating a collaborator's virtual model and reflecting the collaborator's state in real time. To track a person, such studies often incorporate complex tracking system configurations and assume the form of a terminal interface. Although they serve higher engagement than the previous studies, such a terminal interface of tele-presence makes it challenging to be used in practice. People are only allowed to collaborate where the system is installed. To be able to collaborate on virtual objects as well as physical objects regardless of the user's access environment to the collaboration system, the collaboration system should serve high virtual–real engagement functionality with personal devices.

2.3 Content representation for XR

Many studies describing VR and AR content have been proposed with a declarative approach because of human readability and intelligibility that enable easy modification and management of content. Augmented Reality Markup Language (ARML; OGC, 2010a), defined by the Open Geospatial Consortium, describes AR scenes as a composite declaration of physical objects, virtual assets, and the relationships between them based on XML grammar. While ARML adopts Geography Markup Language (OGC, 2010b) to its coordinate reference system to represent physical objects, KARML adopts Keyhole Markup Language (OGC, 2008) to its coordinate reference system and supports the definition of AR content through HTML. To describe VR content on the web, Web3D Consortium (2001) standardized eXtensible 3D (X3D). X3D, the successor to the Virtual Reality Markup Language (VRML; Raggett, 1995), is designed to declaratively represent the virtual world based on the XML syntax. X3DOM (Behr



(a) Traditional method – local site

(b) Traditional method – remote site

(c) Our method – local site

(d) Our method – remote site

Figure 4: The comparison of the collaboration method within traditional (a, b) and ours (c, d).

et al., 2009, 2011), the implementation of X3D, has shown that the integration of X3D and HTML can be achieved without any browser plug-ins. By bridging and synchronizing the HTML Document Object Model (DOM) tree and X3D scene graph, X3DOM makes it possible to manipulate X3D content through a change of HTML DOM elements during runtime. Subsequently, these bridging and synchronizing capabilities are being included in the X3D v4 specification (Web3D Consortium, 2020)—X3D v4 being implemented in JavaScript by the X_ITE project. X3DOM and X_ITE have attempted to express the structure and style of three-dimensional data in a single markup language, whereas XML3D (Sons *et al.*, 2010; Jankowski *et al.*, 2013; Sutter *et al.*, 2015) has assigned roles to HTML and CSS.

3 System Design

In XR collaboration, collaboration is performed for both virtual objects and physical objects. In most traditional collaborations (e.g. remote assistant AR, environment capture in advance, or in real time) on physical objects, the local user working on physical objects shares the local environment with the remote collaborator through RGB/RGBD cameras (Fig. 4a), and the remote collaborator observes it through a VE and communicates the work instructions through verbal communication or annotations (Fig. 4b). Because these traditional collaborations have been achieved through low-level virtual–real engagement, the actual target of collaboration has not been the specific physical object, but rather the view of the local area. In addition, such traditional collaborations have problems including an unstable view, dependent scene observation, and the need for wide network bandwidth. Moreover, their scalability is low because the system design is biased to a given scenario. Our system uses object tracking technology to collaborate on physical objects rather than sharing views of the RE (Fig. 4c). As a result, VR users can grasp the field's context through a virtual scene where virtual objects are synchronized with the poses of the physical objects tracked by object tracking, not a shaking video stream with a limited view (Fig. 4d). Because remote VR users navigate in the VR world independent of the AR user's device, they are not affected by unstable view and dependent scene observation. Unlike traditional collaborations that share an entire view in a video stream, our system shares only the pose of tracked physical objects; thus, not much network bandwidth is required.

The proposed system provides an XR workspace that can collaborate on not only virtual objects but also physical objects with high virtual–real engagement. In the XR workspace, each user selects a user interaction environment suitable for himself/herself among the 3D/VR/AR options and participates in the collaboration. The content for each user interaction environment is expressed in a unified manner without the requirement to create it separately. The following subsections describe the three features of this system.

(i) An interaction device webization method that can expand the system's capability by easily connecting various new interaction devices.
(ii) An XR content representation method that describes XR content without code duplication, regardless of the user interaction environment (VR or AR).
(iii) A unified coordinate system that expresses the movement of physical objects and virtual objects in the same coordinates, regardless of the user interaction environment.

3.1 Interaction device webization

Traditional devices are not suitable for high-level tasks such as movement in 3D space (Segen & Kumar, 2000). Many researchers have tried to devise intuitive and efficient interaction methods (Segen & Kumar, 1998; Tu *et al.*, 2005; Fu & Huang, 2007) to replace the keyboard and mouse. As a result, various types of interaction devices (e.g. LeapMotion, Tobii Pro, and VIVE Tracker) have been developed and commercialized. They have provided a rich experience to users but have had compatibility problems when integrated. The Virtual Reality Peripheral Network (VRPN) project (Taylor *et al.*, 2001) was proposed for devising a device-independent and consistent access interface to these heterogeneous devices. However, its integration with the web was limited. W3C has defined the WebXR Device API (W3C, 2020b) and GamepadAPI (W3C, 2020a) interfaces, which can directly control devices from the web without plug-ins. However, they focus on directly controlling hardware rather than the handling of user interactions. Our prior work—Webizing Collaborative Interaction Space (Seo *et al.*, 2018)—extends the ubiquity, interoperability, and scalability of previous technologies, and provides synchronization of interaction events among multiple users—it was used as the basis of the interaction subsystem in our system.

The goal of webizing interaction device is to enable various interaction devices to be used on the web and to provide device-independent XR content with a consistent interface. In this system, the event negotiation process of prior work is omitted and the structure of the event object is formed in the JavaScript Object Notation (JSON) format. As every interaction event from device is treated as a DOM event in our application, negotiating the type of event needed by the application is not required. Fig. 5 shows the webization architecture of the interaction devices. The device adapters are implemented using data communication protocols or Software Development Kits (SDKs) supported by the devices and play the role of transmitting interaction event data generated by the device to the device manager. The device manager manages the metadata and configurations of connected interaction devices (see List. A1). The configuration ID is uniquely assigned from the server, and users can store and manage the metadata of their device associated with the ID. The device manager processes the event data generated from the device into a formalized JSON format using the device configuration. The event data formatted as JSON can be directly delivered

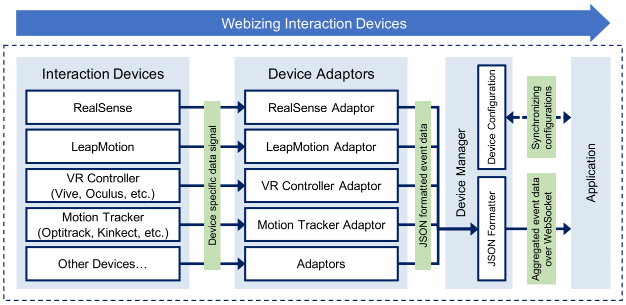
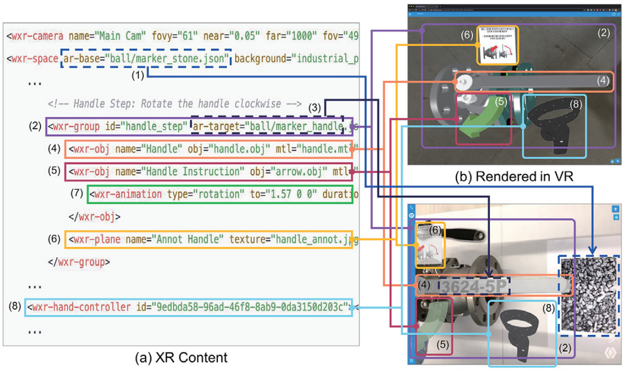

Webizing Interaction Devices
Interaction Devices
RealSense
LeapMotion
VR Controller (Vive, Oculus, etc.)
Motion Tracker (Optitrack, Kinkect, etc.)
Other Devices…
Device specific data signal
Device Adaptors
RealSense Adaptor
LeapMotion Adaptor
VR Controller Adaptor
Motion Tracker Adaptor
Adaptors
JSON formatted event data
Device Manager
Device Configuration
JSON Formatter
Synchronizing configurations
Aggregated event data over WebSocket
Application

Figure 5: The overview of interaction device webization and interaction data flow.



```
<wxr-camera name="Main Cam" fovy="61" near="0.05" far="1000" fov="49
<wxr-space ar-base="ball/marker_stone.json" background="industrial_p
              (1)
    ...
        <!-- Handle Step: Rotate the handle clockwise -->  (3)
    (2) <wxr-group id="handle_step" ar-target="ball/marker_handle.j
        (4) <wxr-obj name="Handle" obj="handle.obj" mtl="handle.mtl"
        (5) <wxr-obj name="Handle Instruction" obj="arrow.obj" mtl"
            (7) <wxr-animation type="rotation" to="1.57 0 0" duratio
            </wxr-obj>
        (6) <wxr-plane name="Annot Handle" texture="handle_annot.jpg
        </wxr-group>
    ...
(8) <wxr-hand-controller id="9edbda58-96ad-46f8-8ab9-0da3150d203c">
    ...
```

(a) XR Content

(b) Rendered in VR

(c) Rendered in AR

Figure 6: An XR content described using extended unified XR representation (a) and its rendering results in VR (b) and AR (c).

to the application or accumulated for a certain period of time, packaged, and sent at once to reduce overheads (see List. A2). Only the interaction data close to the current frame among the packaged ones are used in the application, but all of the event data are logged in the server for later analysis of user interaction patterns.

Event data consist of four fields: *ID*, *type*, *timestamp*, and *detail*. Event data are parsed as a DOM event and processed by event listeners registered in the browser. Thus, the *type* field indicates a DOM event name, and *ID* field indicates an *HTMLElement* to deliver the DOM event object. The *event_data* object in List. A2 shows an example of a *trackerDetected* event from a *Tracker*-type device. The *packaged_event_data* object shows an example in which several events are packaged and delivered. The event data *type* outside is specified as a *packagedEvent*, and the enclosed interaction events are listed in an array in the *detail* field. For a *packagedEvent*-type event, the *ID* value is ignored because there is no device that receives and processes the event shell.

3.2 XR content representation

Because traditional VR and AR content expression methods, such as VRML, X3D, XML3D, and ARML, assume which user interaction environment is to be used in advance, using them in other user interaction environments is difficult. To create XR content using this expression method, additional work apart from content authoring is required, such as creating content corresponding to each interaction environment and connecting them to operate as one. To avoid writing of the code twice for the same content and additional work of linking the codes, an integrated method of content expression independent of the user interaction environment is needed. Our previous work—Unified Representation for XR Content (Lee *et al.*, 2020)—proposed an XR content expression system and a method for solving the associated problems. This became the basis of our content describing and rendering methods.

The goal of unified representation for XR content is to facilitate content authors to readily and consistently write content without considering the user's interaction environment. In this study, additional functions necessary for XR collaboration were supplemented and extended to the previous study's tag hierarchy (see Table B1). The *wxr-peripheral* tag abstracts interaction devices. The tag corresponding to the actual interaction device is defined by inheriting the *WXRPeripheral* class and handles the interaction event specific to the device (see List. B1). Because they inherit *WXRElement*, which inherits *HTMLElement*, they can be embedded in the HTML code. The *wxr-animation* tag enables translation and rotation animations on the parent tag in the DOM. Multiple *wxr-animation* tags can be embedded under one tag to create complex animations (see List. B2).

To realize our method of collaboration on physical objects, we added the *ar-base* attribute to the *wxr-space* tag, indicating information about a static point in physical reality. Similar to the *ar-target* attribute in the *wxr-group*, this attribute has a URL pointing to feature information as a value. However, unlike *ar-target*, it is used to identify a fixed location in an AR users' working space, not to track the collaborative target, but to create a unified coordinate system of VR and AR. A detailed explanation of the principle of *ar-base* properties is provided in Section 3.3.

Fig. 6 shows the rendering result of the XR content code based on the user interaction environment. The *wxr-space* tag refers to a logical workspace unit for XR collaboration. It has a *background* attribute to be rendered in the background and an *ar-base* attribute indicating feature information at a certain point in the AR user's working space. A certain point in the working space is defined as a *space marker*, and the stone image next to the handle in the figure plays the corresponding role [the blue line, (1)]. The *wxr-group* tag groups other objects and may have an *ar-target* attribute. The *ar-target* attribute indicates the feature information of a physical object, and based on this, the AR tracking engine recognizes the physical object and augments a virtual object onto it. In the figure, the *wxr-group* tag groups virtual objects of the handle, the curved arrow, and the annotation into one group [the purple line, (2)], and its *ar-target* attribute indicates the feature information of the "3624-5P" string image attached to the handle [the dark blue line, (3)]. A *wxr-obj* tag loads a 3D model file, and a *wxr-plane* tag represents a plane. In the figure, the *wxr-obj* tag is used to load the handle model [the orange line, (4)] and the curved arrow model [the red line, (5)]—indicating the direction of turning the handle—and the *wxr-plane* tag is used to comment on the knowhow of handle manipulation [the yellow line, (6)]. The *wxr-animation* tag embedded in the curved arrow object [the green line, (7)] shows an animation in which the curved arrow object rotates 90° clockwise, delivering work directions intuitively. Finally, the *wxr-hand-controller* tag with an *ID* attribute is declared to use the interaction device in the XR content [the cyan line, (8)].

Fig. 6b and c show the rendering results of the content code (Fig. 6a) in VR and AR, respectively. Because of the special rendering process we devised, the background is not shown in the AR result. This was done with the intention that if the background appears in AR as well as in VR, the camera view showing the local area is occluded. Although the background is useful for the VR user to understand the context, it is seldom helpful for the AR user—rather, it hinders understanding of the context. Moreover, while only virtual objects associated with the *ar-target* appear in the AR result, all virtual objects appear in the VR result. The special rendering process ensures that all objects are rendered in VR regardless of whether features are detected. By using the extended unified XR content representation method, content authors can create content that responds to various user interaction environments with the same code. This feature makes the production of the XR collaborative content more efficient.

3.3 Unified coordinate system

In XR collaboration, representing virtual objects in the real world is important, but so is the other way around. In this section, we introduce a method through which users in a VE actively navigate the XR space while the space reflects the movement of physical objects in RE, so that VR users can grasp the real-world context.

Tracking information regarding physical objects can be easily used in AR, but not in VR. This is because object transformation is based on camera coordinates in AR, while world coordinates are used in VR. To utilize the tracking information of a physical object in VR, the conversion relationship between the camera coordinates and the world coordinates needs to be obtained. However, for this, the problem of indeterminacy of movement needs to be solved. Fig. 7 shows the indeterminacy problem in which the movement of a physical object observed using camera coordinates is equally produced from two different cases. As shown in Fig. 7a, when the camera does not move and the object moves in a certain direction, the camera observes the object's movement in the same direction in which the object moves, as shown in Fig. 7c. In contrast to Fig. 7a, when the object is fixed and the camera moves in the opposite direction, as shown in Fig. 7b, the camera observes the object moving in the same direction as Fig. 7a. As such, owing to the indeterminacy between the motion of the object and the motion of the camera, the conversion relationship between the camera coordinates and the world coordinates cannot be obtained using a general method.

To solve the indeterminacy problem, we introduce a space marker. The space marker provides a reference point for the conversion between camera coordinates and world coordinates, and through this, it is possible to distinguish the movement between the camera and the object. Because the space marker has nothing to do with the collaborative content, the transformation of the space marker is not described in the content code—only the feature information of the space marker is described. The space marker is universally used within the collaboration space; thus, it is defined in the *ar-base* attribute of the *wxr-space* tag. Similar to the *ar-target* attribute, the *ar-base* attribute has a URL for feature information as its value, and the AR tracking engine identifies a reference point based on this feature information. As the space marker's location is not specified in the content code, AR users can freely place the space marker in their working space. The transformation of the space marker on the world coordinates (model matrix) is determined when the AR tracking engine first detects the space marker and the target object simultaneously. Because the space marker provides the basis for the transformation between the camera coordinates and the world coordinates, it is assumed that it does not move once placed. Fig. 8a shows how to obtain the space marker's model matrix (M_b) that satisfies the premise above. First, the target object's model matrix (M_o), which is described in the XR content, and the target object's model-view matrix (MV_o), which is obtained from AR tracking engine in the frame where the space marker and target object were first tracked simultaneously, are known. Therefore, the view matrix (V) can be derived as follows:

$$V = MV_o \cdot M_o{}^{-1}. \tag{1}$$

In the same frame, the view matrix (V), which is derived before, and the space marker's model-view matrix (MV_b), which can be obtained from the AR tracking engine, are known. Consequently, we can obtain the space marker's model matrix (M_b) as follows:

$$M_b = V^{-1} \cdot MV_b. \tag{2}$$

By substituting the view matrix (V) in equation (2) into equation (1), the process of deriving the view matrix is omitted, as in equation (3):

$$M_b = M_o \cdot MV_o{}^{-1} \cdot MV_b. \tag{3}$$

Because it is assumed that the space marker does not move, the space marker's model matrix (M_b) obtained initially does not change in the subsequent frame sequences. Fig. 8b shows the process of obtaining the camera matrix (C') and the target object's new model matrix (M_o') in the following frame sequences. With the same principle as the view matrix (V) in the first frame, the camera matrix (C') is derived from the space marker's model matrix (M_b) and the model-view matrix (MV_b'; equation 4). With the same principle as the space marker's model matrix (M_b) in the first frame, the target object's new model matrix (M_o') is derived from the camera matrix (C') and the target object's model-view matrix (MV_o'; equation 5).

$$\begin{aligned} C' &= V'^{-1} \\ &= M_b \cdot MV_b'^{-1} \end{aligned} \tag{4}$$

$$M_o' = C' \cdot MV_o' \tag{5}$$

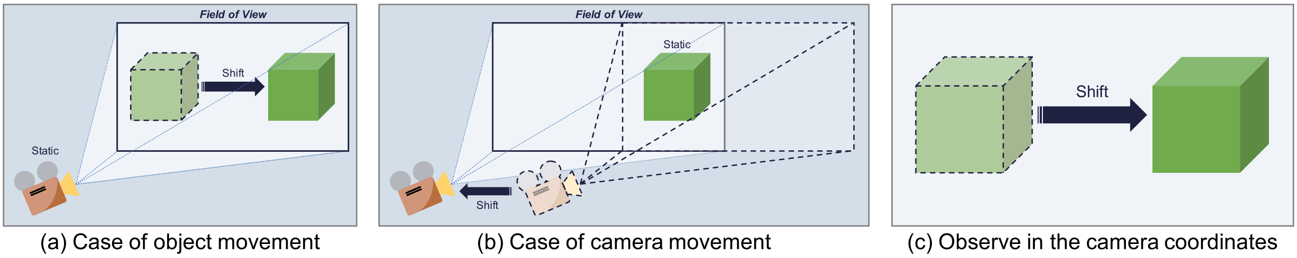
The camera matrix (C') and the target object's model matrix (M_o') present the pose of the AR device and the target object in world coordinates, respectively. By allowing both AR and VR users to use the same world coordinate system, users can grasp the context of the RE in the VE without any obstacles that existed in video stream-based collaboration.

4 Implementation

To prove our proposed concepts, we implemented a prototype, the Webized eXtended Reality (WXR) workspace system. In this system, users access the workspace, an XR collaboration space, in their desired user interaction environment, and users gathered in the workspace grasp each other's context and collaborate through XR content synchronized in real time. Fig. 9 shows an overview of the WXR workspace system.

The system is divided into two parts: server-side and client-side. The server manages the workspaces and mediates synchronization data between users. The workspace service provider of the server provides the resources and functions needed for users to collaborate in the workspace. The client can perform workspace management—such as managing workspace participants, updating the XR content, and registering the interaction devices—through the REpresentational State Transfer Application Programming Interface (REST API) provided by the workspace service provider. The event data router broadcasts interaction events and XR scene update events occurring in the workspace to synchronize all the users accessing the workspace.

When the client accesses the workspace web page, the web browser downloads the XR content and the WXR Library—which

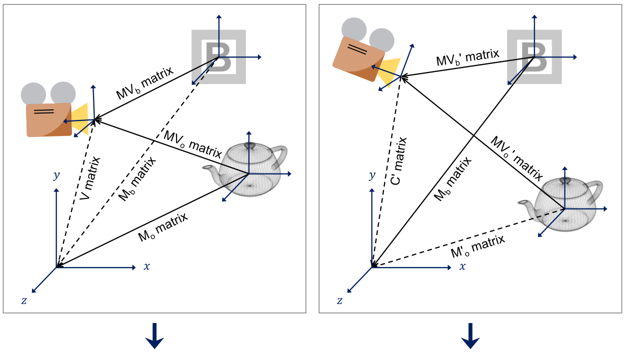


(a) Case of object movement

(b) Case of camera movement

(c) Observe in the camera coordinates

Figure 7: The indeterminacy problem when observing an object in the camera coordinates.



(a) $M_b = M_o \cdot {MV_o}^{-1} \cdot MV_{\mathrm{b}}$

(b) $C' = M_{\mathrm{b}} \cdot {MV_b'}^{-1}$
$M'_{\mathrm{o}} = C' \cdot MV_o'$

Figure 8: Derivation of the space marker's view matrix and model matrix (a) and the target object's camera matrix and model matrix (b).

is a JavaScript library rendering XR content—from the server. The unified coordinate mediator converts the object tracking information obtained by the AR tracking engine from camera coordinates to world coordinates using the method described in Section 3.3 so that object tracking information can be used in any user interaction environment. If the user does not use the AR tracking engine, the unified coordinate mediator is disabled. The device configuration manager manages the configurations of user interaction devices. Users can use their interaction devices in the workspace after being assigned device IDs by webizing their devices and registering them on the server. Interaction event data generated from interaction devices are formatted through the device manager and transferred to the event handler of the WXR Library. The event handler applies not only the interaction event received from the device manager but also the user event generated by the user's direct manipulation of the scene to the XR content, sending both event data to the event data router of the server for synchronization with other collaborators. The XR content loader loads the XR content from the server and downloads the external web resources included in the content. The XR content is composed of HTML/CSS. Feature information of physical objects described in the XR content is brought to the AR tracking engine. The WXR Library builds an XR scene graph, while the browser parses XR content code into the DOM and Cascading Style Sheets Object Model (CSSOM). The DOM tree and XR scene graph created from the XR content code are referenced and updated in each tree node unit, and they are treated as a single XR scene. The XR renderer renders the XR scene on display based on the interaction mode (3D/VR/AR) selected by the user.

The WXR Library allows users to freely select the interaction environment for XR collaboration at the web browser level. For the WXR Library to operate appropriately, a function to track physical objects needs to be supported. There are many commercial tracking engines, but they are rarely integrated with web browsers. Mobile OS vendors (e.g. Google and Apple) that initially implemented AR trackers as a native API, today, support some AR tracker functions enabled in the web browser itself (Apple Inc., 2019; W3C, 2020b); however, they do not serve the object tracking function, so XR collaboration on physical objects is difficult. To circumvent this integration problem and make the proposed system work properly, an *ad hoc* web browser, named WXR Browser, was implemented (see Fig. 10). Through the defined interface in the browser, the WXR Library and AR Tracker freely share feature information and tracking results of physical objects. If tracking functions are strengthened in the standard web browser, the WXR Library will be able to support the VE and the RE without a custom browser, similar to the WXR Browser.

Fig. 11 shows the user interface of the implemented prototype. Fig. 11a shows the main web page of the WXR workspace system. To restrict access to unauthorized users, account creation and login functions are supported (1). After signing in, users can freely create new workspaces (2) or search for existing ones (3). (4) shows a list of searched workspace, and only those that are public or to which users belong are exposed in the list. Fig. 11b shows the screen that appears after entering the workspace. Users can interactively edit XR content through the 3D view (5). The tree view of the XR scene is shown (6), and properties can be modified (7) by selecting each node. Users can change the interaction mode to 3D/VR/AR through the appropriate button (8). To change the interaction mode, the device should be equipped with an environment that can use the selected mode, such as an HMD for VR mode or an AR device for AR mode. (9) contains additional functions such as saving the workspace, registering interaction devices, and modifying XR content code.

5 Experimental Result

In this paper, we introduced a web-based XR remote collaboration system. Through this system, collaborators are physically located in remote locations, but they can communicate and perform tasks as if they were in close proximity. The following paragraphs discuss differences from traditional collaboration methods and user survey results.

5.1 Differences to traditional collaborations

These days video conferences have become popular, because system implementation is extremely simple, and many devices capable of video conferencing, such as smartphones, have proliferated. People using video conference applications use a camera to show their working location and a display to observe others.

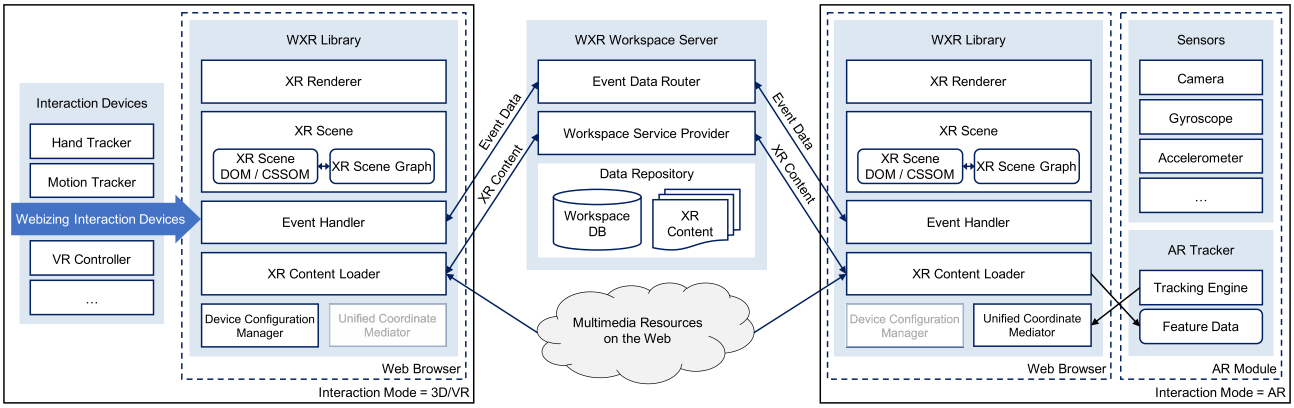


Figure 9: The overview of WXR workspace system architecture.

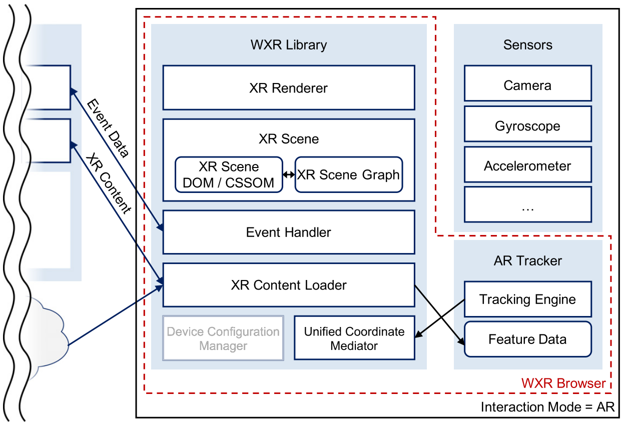
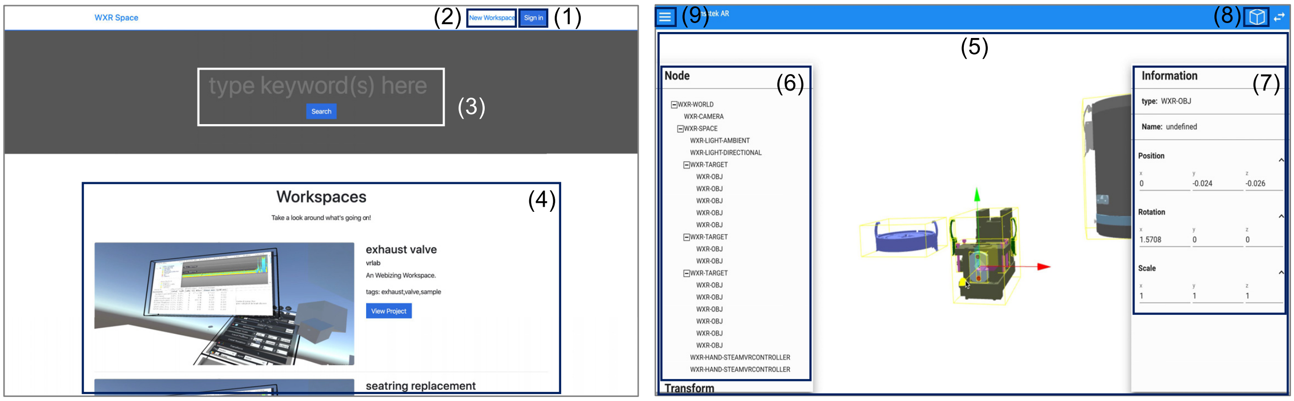


Figure 10: The implementation of the WXR web browser. This supplements the lack of AR functions of a native web browser.

In this case, it is difficult for participants to use gestures because the camera does not show them. In addition, as the instruction is conveyed only in words, more and more detailed descriptions would be needed to articulate their meaning. The more detailed the explanations are, the higher the degree of understanding would be required of the task performer; however, the communication time increases proportionally.

The method considered next to video conferencing is remote assistant AR. As it can be used in devices of the same type as those used for video conferencing, the use of remote assistant AR is rapidly spreading. In collaboration using the remote assistant AR, when a local worker points at the operational target with the camera, the remote person delivers an instruction by drawing a picture on the worker's video image with words—the instructor augments the 3D model of the operational target and shows the simulation to work directly through this 3D model. This method allows workers to understand more intuitively than listening to explanations in words and to convey information using fewer explanations than in video conferencing applications. However, as the only way an instructor can view that the local worker's site is through the worker's camera footage, this type of remote collaboration is used mainly in simple tasks because it is time consuming and is difficult for the instructor to comprehensively understand the context in complex environments and with large tasks.

Unlike traditional collaboration methods, XR collaboration allows the instructor to move the view at will. Like the remote assistant AR, the local worker points at the camera to the operational target and follows the simulation of the augmented 3D model, but the instructor experiences it in VR, not through a video image. At this stage, as the 3D model in VR is in synchronization with the physical object owing to the local worker's unified coordinate mediator and AR tracking engine, the



(a) Main page

(b) Workspace page

Figure 11: The user interface of the WXR workspace system.

instructor can freely move in VR and grasp the context of the local site. Therefore, communication with the worker to grasp the context of the local site significantly decreases. Thus, we believe that XR collaboration shortens the time to collaborate compared with the two existing methods.

5.2 User study

We conducted user tests to determine the usefulness of this system in domains requiring XR collaboration. Owing to COVID-19, we experimented on a limited number of workers with similar backgrounds in their work environments.

5.2.1 *Background*

The experiment and survey were conducted with submarine crew members under the memorandum of understanding between this study's governing institution and the navy. After experiencing and observing our system along with the XR collaboration scenario described in Section 5.2.2, they responded to 25 questions. The experiment was carried out in the submarine in which they were working. By experimenting with the space in which they were initially performing their duties, we offered a situation in which they could compare solving problems through XR collaboration with solving problems through existing technology without significant bias. 15 people participated in the survey, 8 in their 20s, 6 in their 30s, and 1 in their 40s, all male. 11 participants had former VR experience, and 6 participants had AR experience. Three participants had experienced remote maintenance using video conference. Questions pertaining to usability were evaluated on a five-point Likert scale.

5.2.2 *Scenario*

XR collaboration enables efficient remote collaborative repairs. Prior to XR collaboration, there remained ambiguity in communications because language was used as a means of communication. By contrast, XR collaboration helps intuitive understanding by communicating through visual expression. Ball valve replacement is an example of an XR remote collaborative maintenance scenario. To replace the damaged ball in the valve, a series of tasks, disassembly, replacement, and assembly must be performed in the correct sequence and direction. For disassembly, first, the valve is closed by turning the handle, and then the six bolt/nut pairs that fasten the top and bottom covers surrounding the ball are released. For the replacement, the top cover—which does not have a handle attached to it—is removed to expose the ball, after which the damaged ball is replaced with a new one. Notably, the handle and the ball should be aligned, else the ball does not rotate correctly when the handle is turned. As opposed to the disassembly process, the assembly process requires closing the top cover and tightening the bolt/nut pairs to secure the cover, before turning the handle to open the valve. Replacing the ball by oneself is extremely complicated for a layman with no knowledge of the process.

Fig. 12 shows the configuration of an XR collaboration for ball valve replacement. The remote instructor and local worker equip the VR and AR hardware, respectively, and access the WXR Workspace through a web browser. They participate in the collaboration by selecting the interaction mode (VR and AR) corresponding to their hardware. Fig. 13 shows the first step that instructs the user to turn the handle and follow instructions. The local worker recognizes the handle, which is the operating target, and the space marker together. When the AR device recognizes the space marker and handle, the handle model is augmented onto the physical handle (Fig. 13a). The remote instructor then demonstrates turning the handle model in VR, teaching the local worker the correct process. This remote instructor's demonstration is shown as an augmented handle model's movement on the local worker's device (Fig. 13b). The local worker then follows the instruction (Fig. 13c). At this point, owing to the space marker, it is possible to differentiate whether the handle is moving or the AR device is moving—the augmented handle model moves along with the physical handle's movement. The augmented handle model's movement is shown similarly in VR, allowing the remote instructor to recognize the local worker's working progress.

5.2.3 *Result and analysis*

Overall, the survey participants were positive about using our system. Responding to the questions about technology affinity, 11 and 6 people answered that they had experienced technology in VR and AR, respectively. They responded positively to the question "What do you think of adopting VR and AR technology in the submarine?" (Q5). Crews often perform dangerous missions, so they tend to be conservative in introducing new technologies. Considering the lack of AR experience, many crew members' positive responses to the question were attributed to the system being judged to be useful and safe for their missions. They showed the highest response rate to "Maintenance" in the question "Select the field that you think is the most necessary to adopt VR and AR" (Q6). This could be because a submarine consists of highly advanced technologies and components; therefore, it is not possible to manage all aspects of a submarine with a limited number of crew members, and the inability of maintaining submarines affects the working of the entire crew. For similar reasons, it is difficult to incorporate professional medical personnel into a regular mission, so the proportion of "Medicine" that provides diagnosis and emergency medical treatment appears to follow that of "Maintenance." With regard to "Training," as the training is already imparted to the crew using a simulator, the crew members appear to have a relatively low preference for other methods. However, owing to issues such as the cost of maintaining a training center or the cost of establishing a new system, a training management department is expected to be more responsive than the crew members. The answer to the question "How many times a year do you need remote expert help on the submarine?" (Q21) supports this idea. Because of having to perform a mission with only limited resources, the crew prepares thoroughly before sailing. However, 13 respondents reported that situations requiring expert assistance still occurred more than twice a year.

No respondent answered "No" to the question "Is it helpful to perform the work while looking at the augmented object compared to the existing verbal explanation or the manual?" (Q17). It appears that it is more effective to understand how to perform a task visually than to understand it through language. To the question "Do you have any experience using video calls/conferences for remote maintenance assistance?" (Q19), three respondents answered "Yes" among them, two personnel reported that they could solve the problem 80% of the time, and one person reported that he could solve the problem 60% of the time. In response to the question "Is it convenient to use the program?" (Q24), there were eight "Neutral" responses, four "Agree," and three "Strongly Agree" responses, indicating that all respondents were favorably disposed to the use of the prototype system (see Fig. 14).

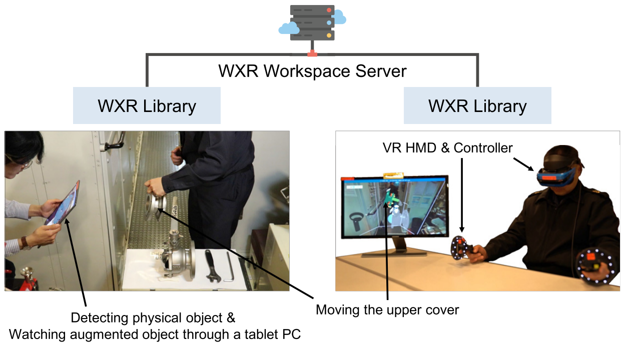


Figure 12: The configuration of the remote maintenance scenario.

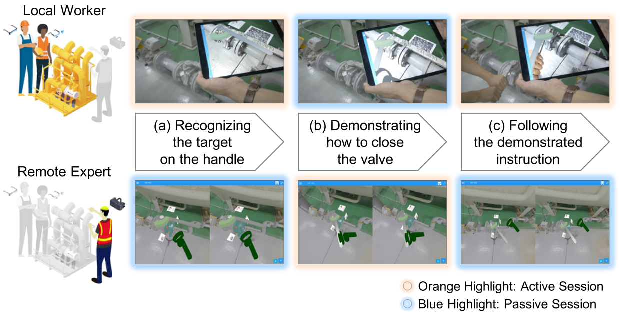

Local Worker
Remote Expert
(a) Recognizing the target on the handle
(b) Demonstrating how to close the valve
(c) Following the demonstrated instruction
Orange Highlight: Active Session
Blue Highlight: Passive Session

Figure 13: The process of ball valve decomposition.

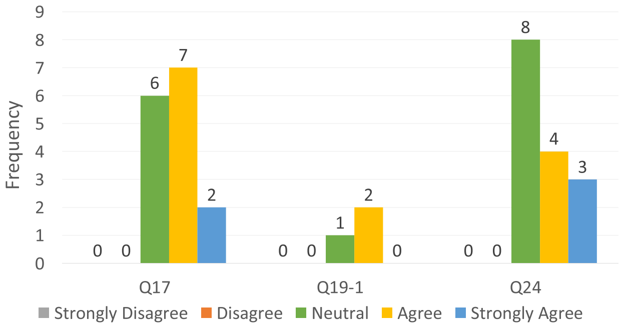


Figure 14: The usability results in the survey of the submarine crews.

Besides, we conducted an additional test to measure how long time a user's action appears in another user's display. This test's network topology was configured by connecting a server and two machines (one for user 1 in VR and one for user 2 in 3D) to a router. We measured the time difference from the moment user 1 moving a virtual object with a VR controller to the moment the virtual object's movement appearing in user 2's display. We conducted this test in varying network bandwidth conditions (no limit, 1 Mbps, 256 Kbps, and 128 Kbps). We took the video (30fps) of user 1's movement and user 2's display on the same frame. The calculation of delay (d) is the following: (the frame number that user 1 starts to move a virtual object) − (the frame number that the movement starts to appear in the display of user 2) + (frame sampling error). The frame sampling error is two (one for the first term and one for the second term in the calculation). The test result was $d \leq 0.2$ s in no limit, $d \leq 0.26$ s in 1 Mbps, $d \leq 0.66$ s in 256 Kbps, and $d > 1$ s in 128 Kbps. This measurement includes the time a sensor captures the movement of the VR controller, the time the sensor's signals are processed and passed to the user 1's event handler, the time the event data reach user 2's device via the WXR Workspace Server and the router, and the time the event data are reflected in the XR scene and rendered on user 2's display. Thus, the result shows that the proposed system has acceptable latency even in a low network quality (256 Kbps).

6 Future Work

This system alleviates the limitations of other XR remote collaboration systems, such as dependent view and unstable view that causes VR sickness (Chang *et al.*, 2020). However, as being a prototype, it is still inadequate for general use. In this section, we propose several remote collaboration scenarios that may be made possible through the proposed XR collaboration system. In addition, the shortcomings and directions for improvements in the prototype are addressed.

6.1 Example scenarios

XR collaboration can be used not only in the engineering field but also in other fields such as medicine and education. Herein, we present some scenarios in which high virtual–real engagement XR collaboration can be used. Remote collaborative maintenance has already proven its usefulness through the prototype and survey. Remote collaborative surgery and remote education are hypothetical scenarios that have not yet been tested.

6.1.1 Remote collaborative surgery

The XR collaboration scenario in surgery is a fusion of AR surgery (Vávra *et al.*, 2017) and the remote surgery (Gupta *et al.*, 2019) methods. In AR surgery, a surgeon can view the patient's CT or MRI image and immediately check the patient's vital signs, such as heart rate and body temperature, augmented on the patient. In remote surgery, the surgeon and patient are physically separated. The surgeon in the remote location checks the patient's video images and vital signs through a monitor and controls the remote robot on the patient's side to conduct the surgery. When the patient's lesions are complex, multiple specialists are required. In this case, XR collaboration can be useful. In XR collaboration in surgery, the leading specialist with AR obtains the patient's vital signs instantly through the visualized information augmented over the patient, while the fellow specialist in the remote location receives the same information from VR. The remote specialist supports the process that is not the area of expertise of the leading specialist by conveying his or her findings regarding the operation's progress.

6.1.2 Remote education

VR and AR have been highlighted in areas that require much experience for learners to get knowledge, such as medical education. The introduction of VR and AR is recently actively studied (Moro *et al.*, 2017, 2021; Birt *et al.*, 2018). Moreover, traditional collaboration methods are not suitable for them because it is difficult for learners to understand class content and for professors to grasp the level of understanding of students. XR collaboration complements these shortcomings and allows students to accomplish high educational outcomes with active and interactive learning experiences, such as Marker-based AR education (Fleck *et al.*, 2015). In the XR collaboration scenario in education, students recognize the learning material and the space marker through AR. The educator then intuitively teaches them to practice by incorporating a virtual model of the learning material in

VR. The students can understand what they need to do to manipulate the learning material through augmented virtual model movements. When a student practices using the learning material, the virtual model of the learning material moves along with the physical one, allowing the professor in VR to grasp the student's progress and to give appropriate feedback.

6.2 Limitations on system implementation

Our proposed XR collaboration system enhances the VR user's understanding of the AR user's context and enhances the AR user's intuitive understanding of the work process. However, there are still improvements that can be made to our prototype, such as blindness to ambient changes, missing communication cues, and 3D model preparation.

6.2.1 Blindness to ambient changes

The changes in the AR user's environment are not reflected in the VE in real time. In this system, because of the limitation that we should attach markers for tracking target movement, AR systems and remote collaborators cannot identify the object that constitutes the surrounding environment without the marker. Similarly, physical deformations on the tracking target are not noticeable at remote locations, as AR systems only identify the movement of the tracking target through the marker. This lack of ambient information can be supplemented by a video from 360 cameras or real-time streamed point clouds using depth cameras. Neural point graphics (Aliev *et al.*, 2020), which provides a realistic view of point clouds using natural rendering, can provide remote collaborators with sufficient insight into the AR user's environment changes.

6.2.2 Missing communication cues

Our system does not convey communication cues such as the user's gaze, facial expressions, or body gestures. People consciously exchange voice information when communicating face to face in their daily lives, but unconsciously use nonvoice information such as facial expressions and gestures. Communication through virtual spaces does not often convey this nonvoice information, thus failing to achieve efficient communication. Photorealistic Facial Animation (Schwartz *et al.*, 2020), a study on how to reconstruct human facial expressions and eyes in VR in real time through cameras attached to an HMD, can solve the issue of missing facial communication cues. Motion tracking systems (e.g. VIVE Tracker) can be used to communicate the user's body pose information. Alternatively, a deep learning approach, such as xR-EgoPose (Tome *et al.*, 2019), can be used to obtain body poses from cameras mounted on HMDs.

6.2.3 3D model preparation

A 3D model of the augmented object to identify the AR user's working progress in VR should be prepared before collaboration begins. In this system, we prepared 3D models for collaboration in advance and synchronized the collaboration target pose through the model. However, this method is unavailable if 3D models are not prepared beforehand. If arbitrary models, such as primitives, are used to represent the collaboration targets, users in VR would confuse the very object to be worked on. Moreover, when dealing with multiple collaboration objects, VR users find it difficult to grasp the relationship between 3D models. Therefore, a 3D reconstruction method can be used to create 3D models for collaboration targets either immediately before or after the commencement in quasi-real time.

6.2.4 Comparative study

We have implemented a system based on the proposed XR collaboration method and conducted a user study. While the survey participants were positive about using our system, the study lacks comparative tools or prototypes. The validation of the highly engaged XR collaboration method requires a few prototypes that look at different aspects of the method to make the results relevant. Another prototype implementation of the proposed XR collaboration method, WXR Library version 2.0, is based on an open-source web framework—A-Frame (Marcos *et al.*, 2020). It is valuable to conduct a comprehensive user study with comparative tools and prototypes.

7 Conclusion

In this paper, we proposed a web-based XR remote collaboration system. By recognizing the working target and the space marker with AR, the local worker communicates the working progress to the remote collaborator in real time. The remote collaborator identifies the working progress through the movement of the 3D model synchronized with the working target in VR or 3D mode and manipulates the 3D model to deliver the work instructions to the local worker. The local worker understands the instructions through the augmented 3D model's movement on the working target and continues to perform the necessary work. The interaction device webization enables event data from various interaction devices, not supported in the web browser by default, to be used by the DOM via a unified interface. XR content representation allows VR and AR content to have a single representation without separating them. Therefore, the content author does not have to duplicate content code for VR and AR and perform additional work to associate them. The unified coordinate system makes it possible to define the physical object's pose in the world coordinate system by differentiating its movement from the device's movement through the space marker.

We implemented the prototype, provided experiences for submarine crew members, and gathered their opinions on the system. Many survey participants responded that the system was convenient to use and reported that the most necessary field for XR collaboration was "Maintenance" among others ("Command & Control," "Training," and "Medicine"), for the submarine operational domain. Finally, we discussed the limitations of the prototype. Our next task is to solve these problems. We believe that the XR collaboration method presented in this paper provides an effective way to collaborate with objects that actually exist in the physical world.

Acknowledgements

This work was supported by the Industrial Technology Innovation Program (20012462) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea) and the KIST under the Institutional Program (Grant Numbers 2V08640 and 2E31051).

Conflict of interest statement

None declared.

References

Adcock, M., Anderson, S., & Thomas, B. (2013). RemoteFusion: real time depth camera fusion for remote collaboration on physical tasks. In *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applica-*

tions in Industry - VRCAI '13(pp. 235–242). Association for Computing Machinery. https://doi.org/10.1145/2534329.2534331.

Al-Adhami, M., Wu, S., & Ma, L. (2019). Extended reality approach for construction quality control. In *International Council for Research and Innovation in Building and Construction (CIB)*.

Alem, L., & Li, J. (2011). A study of gestures in a video-mediated collaborative assembly task. *Advances in Human–Computer Interaction*, *2011*, 1–7. https://doi.org/10.1155/2011/987830.

Aliev, K.-A., Sevastopolsky, A., Kolos, M., Ulyanov, D., & Lempitsky, V. (2020). Neural point-based graphics. https://arxiv.org/abs/1906.08240.

Apple Inc.(2019). AR quick look. Retrieved July 20, 2020, from https://developer.apple.com/augmented-reality/quick-look.

Aschenbrenner, D., Li, M., Dukalski, R., Casper Verlinden, J., Dukalski, R., Verlinden, J., & Lukosch, S. (2018). Exploration of different augmented reality visualizations for enhancing situation awareness for remote factory planning assistance. In *Fourth IEEE VR International Workshop on 3D Collaborative Virtual Environments (3DCVE 2018)*(pp. 3–7). https://doi.org/10.13140/RG.2.2.14819.66083.

Bauer, M., Kortuem, G., & Segall, Z. (1999). "Where are you pointing at?" A study of remote collaboration in a wearable video conference system. In *Digest of Papers. Third International Symposium on Wearable Computers*(pp. 151–158). IEEE. https://doi.org/10.1109/ISWC.1999.806696.

Beck, S., Kunert, A., Kulik, A., & Froehlich, B. (2013). Immersive group-to-group telepresence. *IEEE Transactions on Visualization and Computer Graphics*, *19*(4), 616–625. https://doi.org/10.1109/TVCG.2013.33.

Behr, J., Eschler, P., Jung, Y., & Zöllner, M. (2009). X3DOM: A DOM-based HTML5/X3D integration model. In *Proceedings of the 14th International Conference on 3D Web Technology - Web3D '09*(pp. 127–135). Association for Computing Machinery. https://doi.org/10.1145/1559764.1559784.

Behr, J., Jung, Y., Drevensek, T., & Aderhold, A. (2011). Dynamic and interactive aspects of X3DOM. In *Proceedings of the 16th International Conference on 3D Web Technology - Web3D '11*(pp. 81–87). Association for Computing Machinery. https://doi.org/10.1145/2010425.2010440.

Bimber, O., & Raskar, R. (2005a). Modern approaches to augmented reality. In *ACM SIGGRAPH 2005 Courses - SIGGRAPH '05*(pp. 1–265). Association for Computing Machinery. https://doi.org/10.1145/1198555.1198711.

Bimber, O., & Raskar, R. (2005b). *Spatial augmented reality: Merging real and virtual worlds*. (1st ed.). A K Peters/CRC Press. https://doi.org/10.1201/b10624.

Birt, J., Stromberga, Z., Cowling, M., & Moro, C. (2018). Mobile mixed reality for experiential learning and simulation in medical and health sciences education. *Information*, *9*(2), 14. https://doi.org/10.3390/info9020031.

Ceruti, A., Marzocca, P., Liverani, A., & Bil, C. (2019). Maintenance in aeronautics in an Industry 4.0 context: The role of augmented reality and additive manufacturing. *Journal of Computational Design and Engineering*, *6*(4), 516–526. https://doi.org/10.1016/j.jcde.2019.02.001.

Chang, E., Kim, H. T., & Yoo, B. (2020). Virtual reality sickness: A review of causes and measurements. *International Journal of Human–Computer Interaction*, *36*(17), 1658–1682. https://doi.org/10.1080/10447318.2020.1778351.

Chen, H., Lee, A. S., Swift, M., & Tang, J. C. (2015). 3D collaboration method over HoloLens^{TM} and Skype^{TM} end points. In *Proceedings of the Third International Workshop on Immersive Media Experiences - ImmersiveME '15*(pp. 27–30). Association for Computing Machinery. https://doi.org/10.1145/2814347.2814350.

Dey, A., Piumsomboon, T., Lee, Y., & Billinghurst, M. (2017). Effects of sharing physiological states of players in a collaborative virtual reality gameplay. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*(pp. 4045–4056). Association for Computing Machinery. https://doi.org/10.1145/3025453.3026028.

Fakourfar, O., Ta, K., Tang, R., Bateman, S., & Tang, A. (2016). Stabilized annotations for mobile remote assistance. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*(pp. 1548–1560). Association for Computing Machinery. https://doi.org/10.1145/2858036.2858171.

Febretti, A., Nishimoto, A., Thigpen, T., Talandis, J., Long, L., Pirtle, J. D., Peterka, T., Verlo, A., Brown, M., Plepys, D., Sandin, D., Renambot, L., Johnson, A., & Leigh, J. (2013). CAVE2: A hybrid reality environment for immersive simulation and information analysis. In *The Engineering Reality of Virtual Reality 2013*(Vol. *8649*, pp. 9–20). SPIE. https://doi.org/10.1117/12.2005484.

Fleck, S., Hachet, M., & Bastien, J. M. C. (2015). Marker-based augmented reality: Instructional-design to improve children interactions with astronomical concepts. In *Proceedings of the 14th International Conference on Interaction Design and Children - IDC '15*(pp. 21–28). Association for Computing Machinery. https://doi.org/10.1145/2771839.2771842.

Fu, Y., & Huang, T. S. (2007). hMouse: Head tracking driven virtual computer mouse. In *2007 IEEE Workshop on Applications of Computer Vision (WACV '07)*(pp. 30–30). IEEE. https://doi.org/10.1109/WACV.2007.29.

Fukuda, T., Yokoi, K., Yabuki, N., & Motamedi, A. (2019). An indoor thermal environment design system for renovation using augmented reality. *Journal of Computational Design and Engineering*, *6*(2), 179–188. https://doi.org/10.1016/j.jcde.2018.05.007.

Gao, L., Bai, H., Lee, G., & Billinghurst, M. (2016). An oriented point-cloud view for MR remote collaboration. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications - SA '16*(pp. 1–4). Association for Computing Machinery. https://doi.org/10.1145/2999508.2999531.

Gao, L., Bai, H., Lindeman, R., & Billinghurst, M. (2017). Static local environment capturing and sharing for MR remote collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications - SA '17*(pp. 1–6). Association for Computing Machinery. https://doi.org/10.1145/3132787.3139204.

Gao, L., Bai, H., He, W., Billinghurst, M., & Lindeman, R. W. (2018). Real-time visual representations for mobile mixed reality remote collaboration. In *SIGGRAPH Asia 2018 Virtual & Augmented Reality - SA '18*(pp. 1–2). Association for Computing Machinery. https://doi.org/10.1145/3275495.3275515.

Gauglitz, S., Nuernberger, B., Turk, M., & Höllerer, T. (2014). World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology - UIST '14*(pp. 449–459). Association for Computing Machinery. https://doi.org/10.1145/2642918.2647372.

Google Inc.(2005). Google earth. Retrieved December 3, 2020, from https://earth.google.com/web.

Grandi, J. G., Debarba, H. G., & Maciel, A. (2019). Characterizing asymmetric collaborative interactions in virtual and augmented realities. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*(pp. 127–135). IEEE. https://doi.org/10.1109/VR.2019.8798080.

Grudin, J. (1994). Computer-supported cooperative work: History and focus. *Computer*, *27*(5), 19–26. https://doi.org/10.1109/2.291294.

Gupta, K., Lee, G. A., & Billinghurst, M. (2016). Do you see what i see? The effect of gaze tracking on task space remote collaboration. *IEEE Transactions on Visualization and Computer Graphics*, 22(11), 2413–2422. https://doi.org/10.1109/TVCG.2016.2593778.

Gupta, R., Tanwar, S., Tyagi, S., & Kumar, N. (2019). Tactile-internet-based telesurgery system for healthcare 4.0: An architecture, research challenges, and future directions. *IEEE Network*, 33(6), 22–29. https://doi.org/10.1109/MNET.001.1900063.

Gurevich, P., Lanir, J., Cohen, B., & Stone, R. (2012). TeleAdvisor: A versatile augmented reality tool for remote assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '12*(pp. 619–622). Association for Computing Machinery. https://doi.org/10.1145/2207676.2207763.

Higuchi, K., Chen, Y., Chou, P. A., Zhang, Z., & Liu, Z. (2015). ImmerseBoard: Immersive telepresence experience using a digital whiteboard. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*(pp. 2383–2392). Association for Computing Machinery. https://doi.org/10.1145/2702123.2702160.

Huang, W., Alem, L., & Tecchia, F. (2013). HandsIn3D: Supporting remote guidance with immersive virtual environments. In *Human–Computer Interaction – INTERACT 2013*(Vol. *8117*, pp. 70–77). Springer. https://doi.org/10.1007/978-3-642-40483-2_5.

Huh, S., Muralidharan, S., Ko, H., & Yoo, B. (2019). XR collaboration architecture based on decentralized web. In *The 24th International Conference on 3D Web Technology - Web3D '19*(pp. 1–9). Association for Computing Machinery. https://doi.org/10.1145/3329714.3338137.

Irlitti, A., Piumsomboon, T., Jackson, D., & Thomas, B. H. (2019). Conveying spatial awareness cues in XR collaborations. *IEEE Transactions on Visualization and Computer Graphics*, 25(11), 3178–3189. https://doi.org/10.1109/TVCG.2019.2932173.

Jankowski, J., Ressler, S., Sons, K., Jung, Y., Behr, J., & Slusallek, P. (2013). Declarative Integration of Interactive 3D Graphics into the World-Wide Web: Principles, current approaches, and research agenda. In *Proceedings of the 18th International Conference on 3D Web Technology - Web3D '13*(pp. 39–45). Association for Computing Machinery. https://doi.org/10.1145/2466533.2466547.

Jones, B., Shapira, L., Sodhi, R., Murdock, M., Mehra, R., Benko, H., Wilson, A., Ofek, E., MacIntyre, B., & Raghuvanshi, N. (2014). RoomAlive: Magical experiences enabled by scalable, adaptive projector-camera units. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology - UIST '14*(pp. 637–644). Association for Computing Machinery. https://doi.org/10.1145/2642918.2647383.

Junuzovic, S., Inkpen, K., Blank, T., & Gupta, A. (2012). IllumiShare: Sharing any surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '12*(pp. 1919–1928). Association for Computing Machinery. https://doi.org/10.1145/2207676.2208333.

Kasahara, S., Heun, V., Lee, A. S., & Ishii, H. (2012). Second surface: Multi-user spatial collaboration system based on augmented reality. In *SIGGRAPH Asia 2012 Emerging Technologies - SA '12*(pp. 1–4). Association for Computing Machinery. https://doi.org/10.1145/2407707.2407727.

Kasahara, S., Nagai, S., & Rekimoto, J. (2017). JackIn Head: Immersive visual telepresence system with omnidirectional wearable camera. *IEEE Transactions on Visualization and Computer Graphics*, 23(3), 1222–1234. https://doi.org/10.1109/TVCG.2016.2642947.

Kato, H., & Billinghurst, M. (1999). Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *Proceedings Second IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*(pp. 85–94). IEEE. https://doi.org/10.1109/IWAR.1999.803809.

Kim, S., Lee, G., Sakata, N., & Billinghurst, M. (2014). Improving co-presence with augmented visual communication cues for sharing experience through video conference. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*(pp. 83–92). IEEE. https://doi.org/10.1109/ISMAR.2014.6948412.

Kratz, S., Avrahami, D., Kimber, D., Vaughan, J., Proppe, P., & Severns, D. (2015). Polly wanna show you: Examining viewpoint-conveyance techniques for a shoulder-worn telepresence system. In *Proceedings of the 17th International Conference on Human–Computer Interaction with Mobile Devices and Services Adjunct - MobileHCI '15*(pp. 567–575). Association for Computing Machinery. https://doi.org/10.1145/2786567.⟨?PMU?⟩2787134.

Le Chénéchal, M., Duval, T., Gouranton, V., Royan, J., & Arnaldi, B. (2015). The stretchable arms for collaborative remote guiding. In *International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments (ICAT-EGVE 2015)*. Kyoto. https://doi.org/10.2312/egve.20151322.

Le Chénéchal, M., Duval, T., Gouranton, V., Royan, J., & Arnaldi, B. (2016). Vishnu: Virtual immersive support for HelpiNg users - An interaction paradigm for collaborative remote guiding in mixed reality. In *The Third International Workshop on Collaborative Virtual Environments (3DCVE 2016)*. IEEE. https://doi.org/10.1109/3DCVE.2016.7563559.

Lee, C. P., & Paine, D. (2015). From the matrix to a model of coordinated action (MoCA): A conceptual framework of and for CSCW. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*(pp. 179–194). Association for Computing Machinery. https://doi.org/10.1145/2675133.2675161.

Lee, G. A., Teo, T., Kim, S., & Billinghurst, M. (2017). Mixed reality collaboration through sharing a live panorama. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications - SA '17*(pp. 1–4). Association for Computing Machinery. https://doi.org/10.1145/3132787.3139203.

Lee, G. A., Teo, T., Kim, S., & Billinghurst, M. (2018). A user study on MR remote collaboration using live 360 video. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*(pp. 153–164). IEEE. https://doi.org/10.1109/ISMAR.2018.00051.

Lee, Y., Moon, C., Ko, H., Lee, S.-H., & Yoo, B. (2020). Unified representation for XR content and its rendering method. In *The 25th International Conference on 3D Web Technology - Web3D '20*(pp. 1–10). Association for Computing Machinery. https://doi.org/10.1145/3424616.3424695.

Linden Research Inc.(2003). Second life. Retrieved December 3, 2020, from https://secondlife.com.

Lindlbauer, D., & Wilson, A. D. (2018). Remixed reality: Manipulating space and time in augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*(pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3173574.3173703.

Lipson, H., Shpitalni, M., Kimura, F., & Goncharenko, I. (1998). Online product maintenance by web-based augmented reality. *Proceedings of CIRP Design Seminar on New Tools and Workflow for Product Development*, 131–143.

Mann, S., Furness, T., Yuan, Y., Iorio, J., & Wang, Z. (2018). All reality: Virtual, augmented, mixed (X), mediated (X,Y), and multimediated reality. https://arxiv.org/abs/1804.08386.

Marcos, D., McCurdy, D., & Ngo, K. (2020). A-Frame. Retrieved December 3, 2020, from https://aframe.io.

Milgram, P., & Kishino, F. (1994). Taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, *E77-D*(12), 1321–1329.

Moro, C., Štromberga, Z., Raikos, A., & Stirling, A. (2017). The effectiveness of virtual and augmented reality in health sciences and medical anatomy. *Anatomical Sciences Education*, *10*(6), 549–559. https://doi.org/https://doi.org/10.1002/ase.1696.

Moro, C., Phelps, C., Redmond, P., & Stromberga, Z. (2021). HoloLens and mobile augmented reality in medical and health science education: A randomised controlled trial. *British Journal of Educational Technology*, *52*(2), 680–694. https://doi.org/https://doi.org/10.1111/bjet.13049.

Müller, J., Rädle, R., & Reiterer, H. (2016). Virtual objects as spatial cues in collaborative mixed reality environments: How they shape communication behavior and user task load. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*(pp. 1245–1249). Association for Computing Machinery. https://doi.org/10.1145/2858036.2858043.

Nagai, S., Kasahara, S., & Rekimoto, J. (2015). LiveSphere: Sharing the surrounding visual environment for immersive experience in remote collaboration. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction - TEI '15*(pp. 113–116). Association for Computing Machinery. https://doi.org/10.1145/2677199.2680549.

Newman, J., Bornik, A., Pustka, D., Echtler, F., Huber, M., & Schmalstieg, D. (2007). Tracking for distributed mixed reality environments. In *Proceedings of IEEE Virtual Reality Workshop on Trends and Issues in Tracking for Virtual Environments*.

Nuernberger, B., Lien, K.-C., Hollerer, T., & Turk, M. (2016). Anchoring 2D gesture annotations in augmented reality. In *2016 IEEE Virtual Reality (VR)*(pp. 247–248). IEEE. https://doi.org/10.1109/VR.2016.7504746.

Oda, O., & Feiner, S. (2012). 3D referencing techniques for physical objects in shared augmented reality. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*(pp. 207–215). IEEE. https://doi.org/10.1109/ISMAR.2012.6402558.

OGC(2008). Keyhole markup language. Retrieved November 12, 2020, from https://www.ogc.org/standards/kml.

OGC(2010a). Augmented reality markup language. Retrieved July 17, 2020, from https://www.ogc.org/standards/arml.

OGC(2010b). Geography markup language. Retrieved November 12, 2020, from https://www.ogc.org/standards/gml.

Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., Kim, D., Davidson, P. L., Khamis, S., Dou, M., Tankovich, V., Loop, C., Cai, Q., Chou, P. A., Mennicken, S., Valentin, J., Pradeep, V., Wang, S., Kang, S. B., & Izadi, S. (2016). Holoportation: Virtual 3D teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology - UIST '16*(pp. 741–754). Association for Computing Machinery. https://doi.org/10.1145/2984511.2984517.

Paradiso, J. A., & Landay, J. A. (2009). Guest editors' introduction: Cross-reality environments. *IEEE Pervasive Computing*, *8*(3), 14–15. https://doi.org/10.1109/MPRV.2009.47.

Pauchet, A., Coldefy, F., Lefebvre, L., Picard, S. L. D., Bouguet, A., Perron, L., Guerin, J., Corvaisier, D., & Collobert, M. (2007). Mutual awareness in collocated and distant collaborative tasks using shared interfaces. In *Human–Computer Interaction – INTERACT 2007*(pp. 59–73). Springer. https://doi.org/10.1007/978-3-540-74796-3_8.

Pereira, V., Matos, T., Rodrigues, R., Nobrega, R., & Jacob, J. (2019). Extended reality framework for remote collaborative interactions in virtual environments. In *2019 International Conference on Graphics and Interaction (ICGI)*(pp. 17–24). IEEE. https://doi.org/10.1109/ICGI47575.2019.8955025.

Petit, B., Lesage, J.-D., Menier, C., Allard, J., Franco, J.-S., Raffin, B., Boyer, E., & Faure, F. (2010). Multicamera real-time 3D modeling for telepresence and remote collaboration. *International Journal of Digital Multimedia Broadcasting*, *2010*, 1–12. https://doi.org/10.1155/2010/247108.

Piumsomboon, T., Day, A., Ens, B., Lee, Y., Lee, G., & Billinghurst, M. (2017). Exploring enhancements for remote mixed reality collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications - SA '17*(pp. 1–5). Association for Computing Machinery. https://doi.org/10.1145/3132787.3139200.

Piumsomboon, T., Lee, G. A., Hart, J. D., Ens, B., Lindeman, R. W., Thomas, B. H., & Billinghurst, M. (2018). Mini-Me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*(pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3173574.3173620.

Piumsomboon, T., Lee, G. A., Irlitti, A., Ens, B., Thomas, B. H., & Billinghurst, M. (2019). On the shoulder of the giant: A multi-scale mixed reality collaboration with 360 video sharing and tangible interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*(pp. 1–17). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300458.

Poretski, L., Lanir, J., & Arazy, O. (2018). Normative tensions in shared augmented reality. *Proceedings of the ACM on Human–Computer Interaction*, 2(CSCW), 1–22. https://doi.org/10.1145/3274411.

Raggett, D. (1995). Extending WWW to support platform independent virtual reality. In *Proceedings of the Internet Society/European Networking*(pp. 242).

Schwartz, G., Wei, S.-E., Wang, T.-L., Lombardi, S., Simon, T., Saragih, J., & Sheikh, Y. (2020). The eyes have it: An integrated eye and face model for photorealistic facial animation. *ACM Transactions on Graphics*, *39*(4), 91. https://doi.org/10.1145/3386569.3392493.

Segen, J., & Kumar, S. (1998). Gesture VR: Vision-based 3D hand interface for spatial interaction. In *Proceedings of the Sixth ACM International Conference on Multimedia - MULTIMEDIA '98*(pp. 455–464). Association for Computing Machinery. https://doi.org/10.1145/290747.290822.

Segen, J., & Kumar, S. (2000). Look Ma, no mouse! *Communications of the ACM*, *43*(7), 102–109. https://doi.org/10.1145/341852.341869.

Seo, D., Yoo, B. E., & Ko, H. (2016). Webizing mixed reality for cooperative augmentation of life experience. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion - CSCW '16 Companion*(pp. 401–404). Association for Computing Machinery. https://doi.org/10.1145/2818052.2869078.

Seo, D., Yoo, B., & Ko, H. (2018). Webizing collaborative interaction space for cross reality with various human interface devices. In *Proceedings of the 23rd International ACM Conference on 3D Web Technology - Web3D '18*(pp. 1–8). Association for Computing Machinery. https://doi.org/10.1145/3208806.3208808.

Shen, Y., Ong, S., & Nee, A. (2010). Augmented reality for collaborative product design and development. *Design Studies*, *31*(2), 118–145. https://doi.org/10.1016/j.destud.2009.11.001.

Sodhi, R. S., Jones, B. R., Forsyth, D., Bailey, B. P., & Maciocci, G. (2013). BeThere: 3D mobile collaboration with spatial input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*(pp. 179–188). Association for Computing Machinery. https://doi.org/10.1145/2470654.2470679.

Soler-Domínguez, J. L., de Juan, C., Contero, M., & Alcañiz, M. (2020). I walk, therefore I am: A multidimensional study on the influence of the locomotion method upon presence in virtual reality. *Journal of Computational Design and Engineering*, *7*(5), 577–590. https://doi.org/10.1093/jcde/qwaa040.

Sons, K., Klein, F., Rubinstein, D., Byelozyorov, S., & Slusallek, P. (2010). XML3D: Interactive 3D graphics for the web. In *Proceedings of the 15th International Conference on Web 3D Technology - Web3D '10*(pp. 175–184). Association for Computing Machinery. https://doi.org/10.1145/1836049.1836076.

Stark, R., & Damerau, T. (2019). *Digital twin*. Springer. https://doi.org/10.1007/978-3-642-35950-7_16870-1.

Sun, C., Hu, W., & Xu, D. (2019). Navigation modes, operation methods, observation scales and background options in UI design for high learning performance in VR-based architectural applications. *Journal of Computational Design and Engineering*, *6*(2), 189–196. https://doi.org/10.1016/j.jcde.2018.05.006.

Sutter, J., Sons, K., & Slusallek, P. (2015). A CSS integration model for declarative 3D. In *Proceedings of the 20th International Conference on 3D Web Technology - Web3D '15*(pp. 209–217). Association for Computing Machinery. https://doi.org/10.1145/2775292.2775295.

Tait, M., & Billinghurst, M. (2015). The effect of view independence in a collaborative AR system. *Computer Supported Cooperative Work (CSCW)*, *24*(6), 563–589. https://doi.org/10.1007/s10606-015-9231-8.

Tang, A., Pahud, M., Inkpen, K., Benko, H., Tang, J. C., & Buxton, B. (2010). Three's company: Understanding communication channels in three-way distributed collaboration. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work - CSCW '10*(pp. 271–280). Association for Computing Machinery. https://doi.org/10.1145/1718918.1718969.

Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., & Sui, F. (2018). Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology*, *94*(9), 3563–3576. https://doi.org/10.1007/s00170-017-0233-1.

Taylor, R. M., Hudson, T. C., Seeger, A., Weber, H., Juliano, J., & Helser, A. T. (2001). VRPN: A device-independent, network-transparent VR peripheral system. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology - VRST '01*(pp. 55–61). Association for Computing Machinery. https://doi.org/10.1145/505008.505019.

Tecchia, F., Alem, L., & Huang, W. (2012). 3D helping hands: A gesture based MR system for remote collaboration. In *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry - VRCAI '12*(pp. 323–328). Association for Computing Machinery. https://doi.org/10.1145/2407516.2407590.

Teo, T., Lawrence, L., Lee, G. A., Billinghurst, M., & Adcock, M. (2019). Mixed reality remote collaboration combining 360 video and 3D reconstruction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*(pp. 1–14). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300431.

Tome, D., Peluse, P., Agapito, L., & Badino, H. (2019). xR-EgoPose: Egocentric 3D human pose from an HMD camera. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*(pp. 7727–7737). IEEE. https://doi.org/10.1109/ICCV.2019.00782.

Tu, J., Huang, T., & Tao, H. (2005). Face as mouse through visual face tracking. In *The Second Canadian Conference on Computer and Robot Vision (CRV'05)*(pp. 339–346). IEEE. https://doi.org/10.1109/CRV.2005.39.

Vávra, P., Roman, J., Zonča, P., Ihnát, P., Němec, M., Kumar, J., Habib, N., & El-Gendi, A. (2017). Recent development of augmented reality in surgery: A review. *Journal of Healthcare Engineering*, *2017*, 1–9. https://doi.org/10.1155/2017/4574172.

W3C(2020a). Gamepad. Retrieved September 14, 2020, from https://w3c.github.io/gamepad.

W3C(2020b). WebXR device API. Retrieved September 25, 2020, from https://immersive-web.github.io/webxr.

Wang, P., Bai, X., Billinghurst, M., Zhang, S., Han, D., Lv, H., He, W., Yan, Y., Zhang, X., & Min, H. (2019). An MR remote collaborative platform based on 3D CAD models for training in industry. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*(pp. 91–92). IEEE. https://doi.org/10.1109/ISMAR-Adjunct.2019.00038.

Web3D Consortium(2001). X3D. Retrieved July 17, 2020, from https://www.web3d.org/x3d/what-x3d.

Web3D Consortium(2020). X3D version 4. Retrieved December 3, 2020, from https://www.web3d.org/x3d4.

Wong, N., & Gutwin, C. (2014). Support for deictic pointing in CVEs: Still fragmented after all these years. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '14*(pp. 1377–1387). Association for Computing Machinery. https://doi.org/10.1145/2531602.2531691.

Zillner, J., Rhemann, C., Izadi, S., & Haller, M. (2014). 3D-Board: A whole-body remote collaborative whiteboard. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology - UIST '14*(pp. 471–480). Association for Computing Machinery. https://doi.org/10.1145/2642918.2647393.

Appendix A: Interaction Device Webization

Listing A1: An example of device configuration.

```
let device_configuration = {
  "id": "d7905a86-40b9-4e56-b37e-af5186763e90",
  "device": "Optitrack",
  "name": "My Optitrack",
  "connectionType": "VRPN",
  "deviceType": "Tracker",
  "status": "connected"
}
```

Listing A2: An example of event data and packaged event data.

```
let event_data = {
  "id": "d7905a86-40b9-4e56-b37e-af5186763e90",
  "type": "trackerDetected",
  "timestamp": 1529853124,
  "detail": {}
}

let packaged_event_data = {
  "type": "packagedEvent",
  "timestamp": 1529875350,
  "detail": [
    {
      "id": "d7905a86-40b9-4e56-b37e-af5186763e90",
      "type": "trackerMoved",
      "timestamp": 1529875337,
      "detail": {
        "position": [...],
        "quaternion": [...]
      }
    },
    {
      "id": "9edbda58-96ad-46f8-8ab9-0da3150d203c",
      "type": "HandControllerMoved",
      "timestamp": 1529875345,
      "detail": {
        "position": [...],
        "quaternion": [...]
      }
    }
  ]
}
```

Appendix B: XR Content Representation

Listing B1: An example of implementing *WXRHandController* class for a hand controller-type interaction device.

```
class WXRHandController extends WXRPeriphaeral {
  constructor() {
    super();

    /* Initializing properties here... */
  }

  static get is() {
    return "wxr-hand-controller";
  }

  /* Be executed when wxr-handcontroller tag is added into DOM. */
  connectedCallback() {
    super.connectedCallback();

    /* Attaching event handlers. */
    this.addEventListener("HandControllerDetected",
        this.onHandControllerDetected);
    this.addEventListener("HandControllerMoved",
        this.onHandControllerMoved);
    this.addEventListener("HandControllerMissed",
        this.onHandControllerMissed);
  }

  /* Defining the evnet handlers here... */
}
customElements.define(WXRHandController.is, WXRHandController);
```

Listing B2: An example of declaring a set of animations.

```
<wxr-box name="anim_box">
  <wxr-animation type="translation" to="1 0 0"
      duration="1000"></wxr-animation>
  <wxr-animation type="rotation" to="1.57 0 0"
      duration="1000"></wxr-animation>
</wxr-box>
```

Table B1: The basic and extended WXR features to represent unified XR interoperable in VR and AR.

Feature	Type	Description
wxr-element	Tag	The base tag of WXR Tag hierarchy. This performs elemental behaviors that add or remove a 3D object into a 3D scene graph when corresponding WXR tag is attached to or detached from DOM Tree. Every WXR tag is defined by inheriting this.
wxr-peripheral	Tag	The abstract tag for interaction device. Different interaction devices extend this and implement algorithms for handling their specific interaction event.
wxr-animation	Tag	An element tag for animating an object. An object embedding of a set of this tag shows a movement according to the rules defined by the set.
ar-base	Attribute	The attribute of wxr-space. The value of this attribute is a URL containing feature data for a fixed location in the real world.
ar-target	Attribute	The attribute indicating feature data of physical object. This attribute can be set only in wxr-group and wxr-geometry tag and its inheritances.