Human Factors in Virtual Environments and Game Design, Vol. 50, 2022, 25–31
https://doi.org/10.54941/ahfe1002057

Mobile Cross Reality (XR) Space for Remote Collaboration

Yusuke Kikuchi[1], Ryoto Kato[1], Vibol Yem[1], Yukie Nagai[1], and Yasushi Ikei[2]

[1]Tokyo Metropolitan University, Tokyo, 1910065, Japan
[2]The University of Tokyo, Tokyo, 1138656, Japan

ABSTRACT

In this paper, we propose a mobile cross reality (XR) space where people will have a dialogue over the Internet sharing stereoscopic views of moving viewpoints at remote real places. This system is a portal to metaverses to provide especially a variety of remote experiences of the real world. The participants wear head-mounted displays and talk with remote customers/clients using a full-dome view captured by an omni-directional stereo camera carried by mobile robots and a manned two-wheeled vehicle. The face avatar of a participant is presented at the place on to the remote stereo camera for the local interlocutor wearing a pair of mixed reality glasses to see who is talking. After the system configuration is introduced, a current round-trip delay, an important factor for communication, is discussed on the essential path of the system. The result showed the delay was less than 400 ms, an acceptable amount for multimedia quality of service. As an evaluation of the XR dialogue system, the user study in which a remote spatial object was pointed and observed in three viewing conditions was conducted. The result showed that stereoscopic viewing improved accuracy of depth perception in the remote space as compared to the conventional 2D viewing.

Keywords: Cross reality, Metaverse, Stereoscopic vision, Real space, Telexperience, Viewpoint mobility

INTRODUCTION

In recent years under a pandemic fear, meeting with remote people via a video conferencing system has become a very common event. However, the sense of reality felt from the remote space rendered in a two-dimensional screen is clearly insufficient for grasping three-dimensional (3D) structure and surface detail of remote objects. This severely decreased the quality of conversation especially when 3D characteristics of the remote space was essential for the purpose of dialogue on such topics as design/maintenance of buildings and plants, mechanical structures, handling of small objects, and even the face of interlocutor as in medical diagnostics.

This paper describes the system that enabled shared viewing of remote places as well as showing the face of a talker, or an avatar, at the location of the remote cameras. The same view of a remote place may be shared by many observers (participant of the system) wearing head-mounted displays (HMDs) or waving smart-phones to watch any directions at the viewpoint

that captured a 360° full-dome image by our camera system. The camera system, TwinCam (Ikei, 2019), was invented in our previous research that transmitted a pair of full-dome live 4 K stereoscopic images captured by two omnidirectional cameras (THETA Z1, Ricoh Co., Ltd.). The TwinCam fixes the orientation of two camera lenses to the same direction in the world coordinate during a head turn of the viewer. This design removes motion blur in the received frames and an apparent delay of them when the user rotated the head. Virtual reality sickness can be reduced while the system transmits 4K-resolution 3D images with correct binocular disparity.

The camera was mounted on remote-controlled mobile robots and a manned two-wheeled vehicle (Segway i2) to send video images to the mobile cross reality (XR) metaverse portal from which streamings are distributed to the viewers. The user can choose one of the real-time views of remote places assisted by the portal operator who moves the virtual stereo camera in the portal space to guide the users and deliver binocular/monocular streaming. In such a real-time experience, it is very important to reduce delay of image transmission so that the user can communicate with remote people without noticeable unnatural response time and unpleasant visual fields.

REAL-TIME BINOCULAR AND MONOCULAR VIDEO DISTRIBUTION OF MOBILE XR SYSTEM

In the previous research, an asymmetric peer-to-peer viewing system was built where a single person could view a remote space via a binocular live streaming and the interloculator sees the avatar face. This system was extended to allow one-source to many-viewer delivery where a single participant can see the omnidirectional stereoscopic view and many others the monocular view of other directions at the same viewpoint. The stereoscopic view is possible only to one direction because the camera (TwinCam) direction must be the same as the head of a viewer who receives stereoscopic images (a master viewer), and other viewers who prefer different directions than the master viewer obtain monocular images from the either camera of the TwinCam. This design reduces the number of cameras to deploy at the real world, and enables an efficient guided tour for many participants to join with a dialogue to remote people sharing the same viewpoint.

The configuration of the system is shown in Figure 1. The video streaming (4K $\times$ 2 channels) from the TwinCam mounted on the avatar robot/manned vehicle in the field is forwarded to the WebRTC server via a 5G/WiFi channel. The server acts as a selective forwarding unit (SFU) for one-to-many delivery. A single observer, who determines the direction of the camera and obtains a stereoscopic view, sends out the angle of head rotation to the remote TwinCam camera to rotate to the direction of line of sight. Other observers can see the same binocular images when looking in the same direction as the master viewer. Otherwise, the presented image to the observers is switched to be monocular. The terminal software is implemented using WebXR (Jones et al. 2022) which renders the image to an HMD using a browser. A full-dome view from the TwinCam camera on the robot is presented to terminals such as Meta Quest2, iPhones, Android smartphones, and PCs.

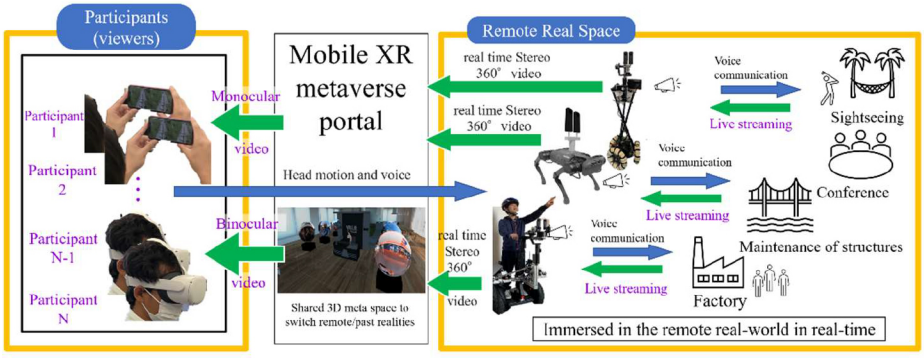


Figure 1: Mobile XR system.

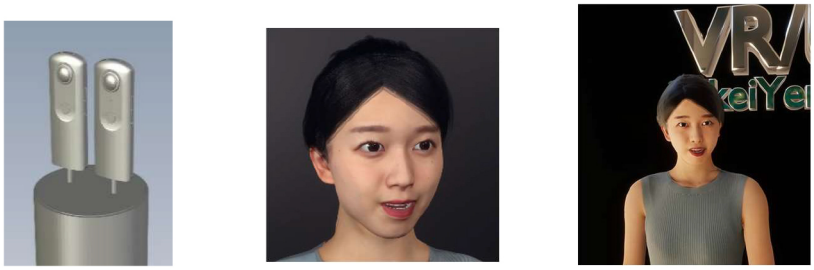
An operator of the mobile XR metaverse portal controls the position and motion of a virtual camera system that handles the video connection between remote real cameras and participants with a media channel for the viewing direction. The motion of the avatar robot is controlled either by the portal operator or a participant.

XR METAVERSE PORTAL AND VIRTUAL CAMERA

A 3D space of the XR metaverse portal is shown in Figure 2. The portal is a space-gallery where many spheres are displayed for a participant to enter the remote or past real spaces. Currently, the viewpoint of a participant is attached to a virtual camera (see Figure 3a) that is moved by the portal operator during a space tour. The virtual camera simulates the function of the real TwinCam camera to provide binocular disparity in the specific direction. When the virtual camera runs into one of the real world spheres, the camera is placed at the center of the sphere on which remote image streaming for left/right eyes are projected. The streaming is recaptured by the cameras before distribution to the participants. The past-space spheres hold the video of places which were recorded with the omnidirectional cameras. A stereoscopic view of the past scenes is available when the participants look at objects in the same direction as the TwinCam recorded them in the front view. Otherwise, participants can see the scenes in the monocular presentation.

First, a participant is connected to the virtual camera of the XR portal, then a realistic guide avatar (see Figure 3b) will introduce the portal at the reception desk (see Figure 3c). Then, the participant viewpoint is moved to the sphere of interest by the portal operator. In the remote real space sphere, the participants take the viewpoint of the avatar robot walking under the guidance of the portal operator or a single participant. A privileged participant wearing the HMD presenting a stereoscopic view can control the position of the avatar robot. Having an omnidirectional view at a remote workplace or an event spot, the participant talks with a coworker in charge of the remote space. The coworker wearing a pair of mixed reality (MR) glasses conversely see an avatar face of the participant presented in an augmented reality way at the camera position on the robot. In case of discussion on a 3D model of

Figure 2: View of the mobile XR metaverse portal.



(a) (b) (c)

Figure 3: (a) Virtual camera, (b) Guide avatar, (c) Avatar at the reception desk of mobile XR metaverse portal.

a product, a shared virtual object can be presented in the visual field of both sides. They can talk about the object using a 3D pointer in the field. The participant may give instructions to the coworker about the spatial product in the 3D space.

EXPERIMENTAL EVALUATION OF COMMUNICATION DELAY

One of crucial problems in remote collaboration is a communication delay both in an audio/video channel and a motion control signal channel. The delay is also problematic when the user presents him/herself by an avatar, especially by a high quality avatar that requires a large data traffic in the future. For a real-time video communication, ITU-T recommends in G.1010 (ITU, 2001) that the maximum delay between terminals should be less than 400 ms in one-way direction as a criterion. We verified if our system worked within this criterion of delay by measurement. We assumed that the streaming of video data taken at a remote scene is distributed from a cloud computer to an office PC at an institution.

A model communication path was built for the experiment purpose as shown in Figure 4. The delay time was calculated when the data was transmitted bidirectionally at a local PC as follows.

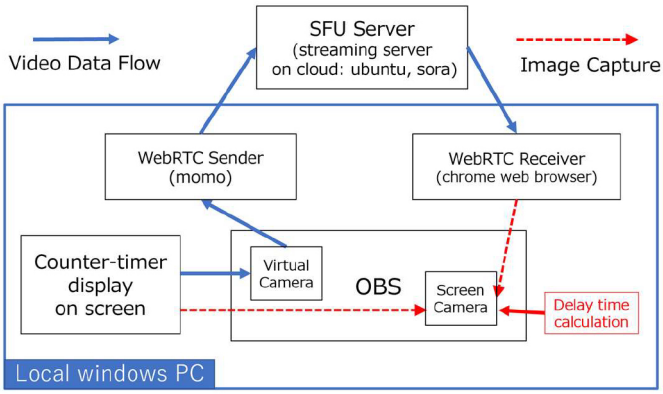


Figure 4: Data flow of video transmission during measurement.

Table 1. SFU server and local PC specifications.

	SFU server		Local PC
OS	Ubuntu 20.04 (Microsoft Azure)		Window 10
CPU	2.35~3.35 GHz (AMD EPYC 7452)		Intel Core i9 9900K
Memory	8 GB		64 GB
GPU	NA		Geforce RTX 2080 (Nvidia)
Network	1000 Mbps	← The Internet →	1000base-T

1. A figure of a digital counter timer in ms is displayed on a screen of the PC
2. The timer window is captured using Open Broadcaster Software (OBS)
3. The video of the timer is transferred to the SFU server using the WebRTC protocol
4. The video steam at the SFU server is returned to the local PC's web browser
5. Two figures of the timer source and the returned stream are compared after a screen recording

The video stream was in 4K resolution, 30 frames per second (FPS), H.264 encoding. The screen recording was performed at 30 FPS for 30 seconds. The devices used were shown in Table 1 (Mamccrea, 2022).

The results are shown in Figure 5. The mean delay time was 341 ms according to the figure comparison of the screen camera images. The delay time was within the ITU-T G.1010 recommendation for quality of service to ensure smooth conversation between remote users. The actual transmission delay depends on the Internet state, encoding processing, inside camera delay, etc.

EVALUATION OF REMOTE SHARED OBJECT PERCEPTION

For collaboration on a shared real space between the remote users who communicate via the metaverse portal, it is often required to point to the feature

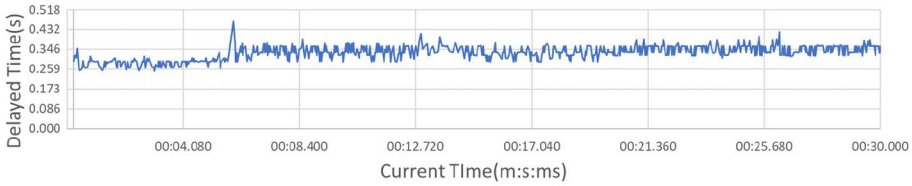


Figure 5: Delay time for a roundtrip from timer capture to appear in browser window.

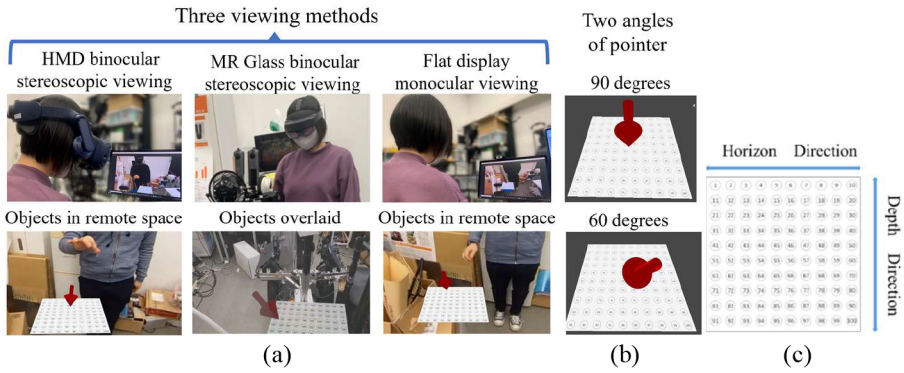
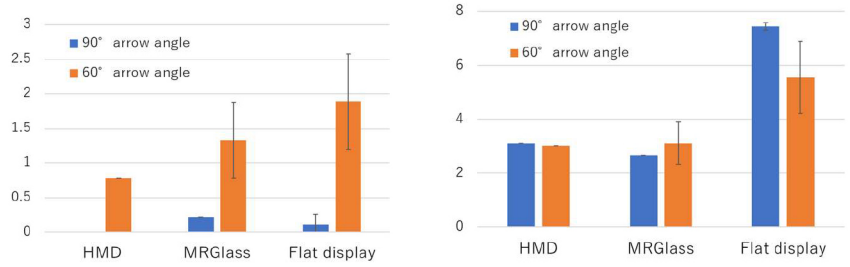


Figure 6: Comparison of viewing methods during spatial pointing in a remote space.

(vertices and surfaces) of a 3D object including a virtual design model, a small real object, and a part of a large structure of industrial plants, machines, or buildings. In such a case, we have to use a virtual 3D pointing device which can indicate a specific 3D location on a target object in the conversation to share both side intentions. A scene of collaboration was simulated with the XR system for the evaluation. The accuracy of 3D visual perception of a shared object presented by the XR system was measured. The participants were nine university (graduate) students with an average of 23.2 years. The participants observed two 3D virtual objects floating in the real space, in three conditions: through an HMD (Vive Pro) with the field of view fed from the two full-dome camera, and via a pair of MR glasses (HoloLens 2, Microsoft) in the vicinity of the cameras. To compare the effect of viewing conditions, the virtual objects were presented also in a monocular flat display (27 inch) to the participant as shown in Figure 6a. The virtual objects were a numbered plate (see Figure 6c, 200×200 mm^2) and a 3D arrow (83 mm in length) to point to the sheet from 100 mm above (see Figure 6b). The 3D arrow took two configurations of an angle of 60 and 90 degrees (two levels of pointing angles). The arrow was set randomly to point to any number on the sheet by the experimenter before the participant orally answers the number. The 2D distance between two numbers was measured as pointing error.

Results and Discussion

The results of pointing errors are shown in Figure 7. The error in (a) horizontal direction to the participant was smaller than (b) depth direction because of its easier view. The error in horizontal direction at the 90 degree angle



(a) Horizontal error (b) Depth error

Figure 7: Pointing error on the virtual plate (mm).

was markedly smaller than the 60 degree angle, however in depth error, the angle difference was not significant. This suggests that stereoscopic viewing enabled correct perception of spatial configuration of the 3D arrow at the both angles. The 2D viewing with the flat display did not provide a good cue for depth perception. The HMD and HoloLens viewings allows not only binocular disparity but also motion parallax when the participant changed the head position. These results indicate that an HMD and HoloLens are effective for spatial perception which is required to perform remote control or collaboration in virtual and remote shared spaces.

CONCLUSION

We have developed a mobile XR space as metaverse portal for use in remote collaboration. This system can connect with various remote worlds to be immersed in especially real spaces where mobile robots are arranged. A real-time binocular view to any direction is delivered to a single master user while other users who prefer different direction than the master can see the monocular full-dome scene. The results of the user study showed that 3D viewing improved accuracy of depth perception ensuring the user the base for sense of immersion as compared to popular 2D monocular viewing condition.

ACKNOWLEDGMENT

This work was supported by the MIC/SCOPE #191603003, JSPS KAKENHI Grant Number 18H04118, 21K19785, 18H03283 and TMU local-5G project.

REFERENCES

Ikei, Y., Yem, V., Tashiro, K., Fujie, T., Amemiya, T., Kitazaki, M. (2019). Live Stereoscopic 3D Image With Constant Capture Direction of 360 Cameras for High-Quality Visual Telepresence, IEEE Virtual Reality and 3D User Interfaces.

International Telecommunication Union (ITU). (2001) ITU-T. G.1010: End-user multimedia QoS categories.

Jones, B., Goregaokar, M., Cabanier, M. (Feb 08, 2022) WebXR Device API. W3C Website: https://www.w3.org/TR/webxr/

Mamccrea. (Jan 04, 2022) Dav4 and Dasv4-series Virtual machines spec. Microsoft Website: https://docs.microsoft.com/en-us/azure/virtual-machines/dav4-dasv4-series.