Point Cloud Stream on Spatial Mixed Reality

Toward Telepresence in Architectural Field

Tomohiro Fukuda[1], Yuehan Zhu[2], Nobuyoshi Yabuki[3]
[1,2,3] Division of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University
[1,3] {fukuda|yabuki}@see.eng.osaka-u.ac.jp [2] zhu@it.see.eng.osaka-u.ac.jp

In remote meetings that involve the study of buildings and cities, sharing three-dimensional (3D) virtual spatial of buildings and cities is just as necessary as sharing the appearances and voices of meeting participants. Because of this, system development and pilot projects have attempted to share 3D virtual models via the internet in real-time but is still insufficient compared with face-to-face meeting. Therefore, this research explores the applicability of a spatial mixed reality (MR) system that displays point cloud streams to realize 3D remote meeting in architecture and urban fields. MR is a new technology that enables 3D presentations of various information, combining the physical and virtual worlds. One MR method is telepresence, which is expected to give people a way to communicate remotely as if face to face in a realistic way. We first developed a MR system named PcsMR (Point cloud stream on mixed reality) to display point cloud streams. The PcsMR system's operation consists of generating and transferring a point cloud stream and then rendering a point cloud stream using MR. The PcsMR acquired the point cloud stream in real-time using Kinect for Windows v2 and transferred it to Microsoft HoloLens, which uses optical see-through MR. Then we constructed two prototypes based on PcsMR and carried out pilot projects. Through observing the experiments, application possibilities for architecture and urban fields are found in meetings and communications that share real-time 3D objects and include the movement of remote participants and objects. The proposed method was evaluated feasible and effective.

Keywords: *Telepresence, Mixed reality, Point cloud stream, Remote meeting, Real time*

INTRODUCTION

The development of broadband internet and cloud computing technologies have led to the realization of remote synchronous communication (different places at the same time such as in video conferencing) and remote asynchronous communication (different places at different times such as in e-mail). However, even in cases where participants can take part remotely, some still sacrifice time, incur cost, and add to CO_2 emissions by carrying out face-to-face meetings (same place and same time) to communicate directly. One reason for this is that current remote synchronous communication technologies cannot replace face-to-face communication. In a conference using a remote synchronous communication system, the feeling of having a conference through the screen remains, which is uncomfortable compared to a face-to-face meeting. Telepresence is one concept and technique intended to solve this problem, as it gives people a way to remotely communicate in a realistic way similar to a face-to-face meeting (Buxton 1992).

In remote meetings that involve the study of buildings and cities, sharing three-dimensional (3D) virtual spatial of buildings, cities, furniture, or other objects is just as necessary as sharing the appearances and voices of meeting participants. Because of this, system development and pilot projects have attempted to share 3D virtual models via the internet in real-time. One example of this is a study conducted to share a 3D virtual architecture and urban design space in real-time during remote video conferences (Shen and Kawakami 2010). In addition, the progress of the discussion during that remote conference led to a later study of rea-time sharing of sketches and figures drawn in a 3D virtual space (Sun et al. 2014).

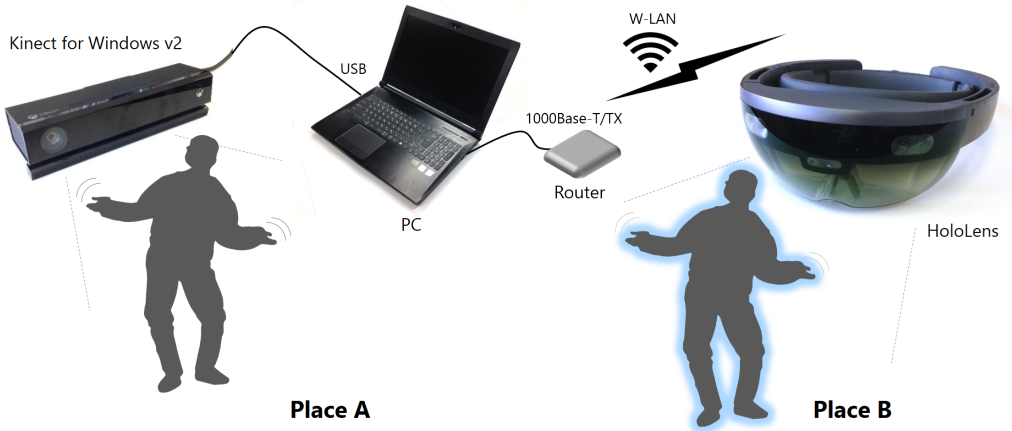
The shared 3D virtual model's modeling method not only uses 3D computer-aided design (CAD) and building information modeling (BIM) software, it also uses a method of creating point clouds of existing space and objects using a laser scanner (Marchand et al. 2017). The previous studies mentioned above were realized through virtual reality (VR), which defines a 3D virtual world using a 3D virtual model. The challenge of VR in the construction and urban development fields is the increase in man-hours needed to create objects such as buildings and terrain around a design target. In addition, although VR can be made to resemble the real world by defining a 3D virtual object, it consists of a virtual world separate from the real world. Therefore, VR is not suitable for connection with the real world (Fukuda et al. 2012).

Mixed reality (MR) is a technology that enables 3D presentations of various information, combining the physical and virtual worlds (Milgram et al. 1994). MR makes it possible to use a real world environment rather than modeling environmental objects as is the case in VR. Thus, we can maintain a connection with the real world. One MR method is telepresence (Buxton 1992).

This research focuses on creating 3D virtual models using point clouds, which are capable of realistic 3D modeling of existing space and objects. Point clouds generally define 3D coordinates and the red, green, and blue (RGB) color values of each 3D coordinate. Using generated point clouds makes it possible to preserve and observe architecture and city structure, and apply them to building stock renovation design projects. Generating meshes from point clouds thus enhances reusability of CAD and BIM design models. Therefore, studies of polygon optimization based on objects are being advanced (Hidaka et al. 2018). Point clouds are created by a method using a laser scanner (Datta and Beynon 2005) and a method known as structure from motion (SfM; Agarwal et al. 2009). SfM has become popular in the construction and urban development fields because it makes it possible to generate 3D point clouds using familiar digital cameras and inexpensive software. However, the amount of time it takes to acquire real spaces through either laser scanning or photography and generate 3D point clouds presents a time-gap problem.

In contrast, a method using a RGB camera and depth sensor (hereinafter, RGB-D camera) can acquire the point clouds in real-time (hereinafter, point

Figure 1
PcsMR System Configuration



cloud stream). Previous research constructed an MR system that included a point cloud stream of life-size 3D participant avatars acquired by an RGB-D camera, and tried to use these avatars in remote conferences between two groups (Beck et al. 2013). Another previous study proposed a remote sharing system that included point cloud streams of objects such as people and furniture (Orts-Escolano et al. 2016). Both of these previous studies involved state-of-the-art MR systems meant to achieve telepresence remote meetings and act as references for this research. However, there is not sufficient research on MR system applicability in architecture and urban fields.

We therefore aim to explore the applicability of a MR system that displays point cloud streams in architecture and urban fields. To accomplish this, we first developed a simple MR system to display point cloud streams, and then considered its potential application by using and observing the prototype systems.

DEVELOPMENT OF POINT CLOUD STREAM ON MIXED REALITY

The point cloud stream on mixed reality (PcsMR) system we developed is a MR system that can display point cloud streams using a simple and inexpensive system configuration (see Figure 1). The PcsMR system's operation consists of generating and transferring a point cloud stream and then rendering a point cloud stream using MR. We acquired the point cloud stream in real-time using Kinect for Windows v2 (hereinafter, Kinect) and transferred it to Microsoft HoloLens, which uses optical see-through MR. Kinect generates point cloud streams within 0.5-8.0 m (human skeleton detection 0.5-4.5 m) from the Kinect main body, a 70° horizontal and 60° vertical angle range using the Kinect's equipped RGB-D camera. Section 2.1 outlines the equipment used and section 2.2 shows the proposed system flow.

Equipment

The devices constituting the system are:

- PC running Windows 10
- HoloLens
- Kinect
- Router (required for Wi-Fi transmission between PC and HoloLens: IEEE 802.11ac up to 867 Mbps; IEEE 802.11n up to 300 Mbps)

System Flow

To generate point cloud streams, the Kinect's RGB-D camera scans real-world objects at 30 fps (frames per second). This generated point cloud stream is then transferred to the connected PC in real-time.

Our proposed MR system uses a PC as a transfer station. To receive the point cloud stream on a PC during the transfer process, we developed a system using the Kinect for Windows software development kit 2.0 (hereinafter, Kinect SDK) (see Table 1). We used the Unity game engine to adjust the position, angle, and scale of the point cloud stream to accurately place it in the MR system. We then transferred the point cloud stream to HoloLens by Wi-Fi (TCP/IP).

If the point cloud stream transferred to the HoloLens is directly displayed during the MR rendering process, unnecessary point clouds appear as noise around the displayed point cloud stream. Also, a highly granular point cloud can lead to data volume becoming too large, making real-time rendering difficult. Therefore, to reduce unnecessary noise, we implemented a software filter based on previous research (Kowalski et al. 2015). This algorithm finds n neighboring points for each point (target point) in the point cloud. If the distance between the target point and a neighboring point is greater than the threshold t, the neighboring point is considered an outer point (noise). The n and t values can be set manually depending on the project. With this filtering function, the PcsMR system can display a point cloud stream after deleting unnecessary points and can thus render more quickly.

EXPERIMENTS AND RESULTS

To consider the potential application of the proposed PcsMR, we constructed two prototype systems with the following common conditions:

- We used one Kinect and one HoloLens in each prototype. Audio functions (microphone and speaker) were technically possible but not implemented.
- The router used the IEEE 802.11n Wi-Fi standard. The volume of point cloud stream traffic between the router and HoloLens was about 4 MB / sec.

MR with Point Cloud Stream of People and BIM

We developed our first prototype (PcsMR-1) for university events in June 2017 for high school juniors and seniors who planned to go on to university. The event was held on a different campus from one of our laboratory, so we created a system to allow a virtual laboratory experience for students who could not visit our laboratory.

We used a notebook PC with an Intel Core i7-7700HQ CPU, 8 GB of RAM, and a NVIDIA GeForce GTX 1060 GPU with 6 GB memory. We connected the PC and router using a gigabit Ethernet cable. We adjusted the position, orientation, and scale of the BIM model and point cloud stream using the Unity game engine installed on the notebook PC.

PcsMR-1 generated a point cloud stream of a seated user (A) and synthesized a 3D model of our laboratory (about 150 m^2) manually created using

Table 1
Kinect SDK
Functions used for
PcsMR

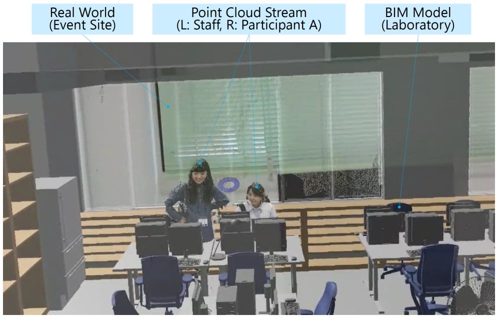
Function	Feature
GetDefaultKinectSensor(IKinectSensor)	Get current Kinect device
IKinectSensor::open() Kinec	Operate the Kinect device
IKinectSensor::get_CoordinateMapper	Get coordinate converter
IKinectSensor::get_*FrameSource	Get 3D point source frame by frame
I*FrameSource::OpenReader(I*FrameReader *)	Create a frame reader for each source
I*FrameReader::AcquireLatestFrame(I*Frame *)	Request new data frame in the main loop

*: wildcard character (Depth, Color, BodyIndex)

Figure 2
PcsMR-1: Experimental photo (2 Scenes)

BIM in other processes manually. Another user (B) could experience MR using HoloLens. At the event, participant A might sit in front of the Kinect and move, and their appearance and actions were shown as a point cloud stream as if they were sitting in a chair in the laboratory BIM model. Participant B could move freely around the event venue by wearing the HoloLens and could explore the 3D virtual laboratory where 3D participant A was seated using MR (see Figure 2). Figure 3 shows a screen shot of HoloLens. When observing the user's experience, participant A performed actions, such waving hands, according to participant B's reaction. Participant B seemed to understand the laboratory's scales of and the MR felt more realistic because of the 3D rendering of participant A staying in the unknown laboratory and moving.

Figure 3
PcsMR-1: HoloLens Screen Shot



MR using Point Cloud Stream of People and Physical Model

Our next prototype (PcsMR-2) anticipated a scenario where participant B, who was in a remote place, wore a HoloLens to observe participant A's presentation while handling 3D models. PcsMR-2 generated both participant A and a city model (1:500 scale) as a point cloud stream, and participant B could see both participant A and the city model while using a MR system with HoloLens (see Figure 4). Because the scale model was on a desk, another desk was also placed beside participant B and positions were aligned as if the scale model was on this desk. Figure 5 and 6 show screen shots of RGB-D image captured by Kinect and HoloLens respectively.

We used a desktop PC with an Intel i5-7500 CPU and 16 GB of RAM and connected this PC with the router using gigabit Ethernet We adjusted the position, orientation, and scale of the point cloud stream using the Unity game engine installed on the desktop PC.

In this experiment, participant B could freely move about in the meeting room and confirm the participant A's movement and the scale model from the 3D point cloud stream. When participant B moved, the appearance of Participant A and the scale model changed from participant B's viewpoint in real-time, which is impossible with a 2D display such as a video conference system (see Figure 7).

Figure 4
PcsMR-2:
Experimental photo

Figure 5
RGB-D image
captured by Kinect
(inversion
condition)

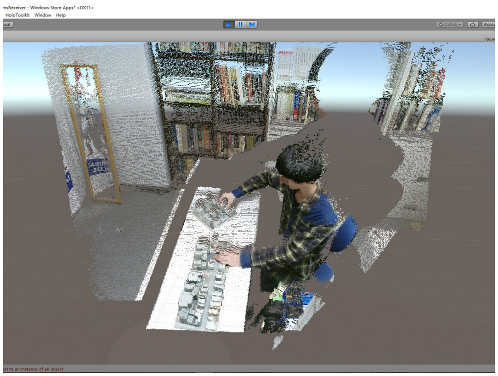


Figure 6
Point Cloud Stream
on Game Engine
(Unity)

Discussion

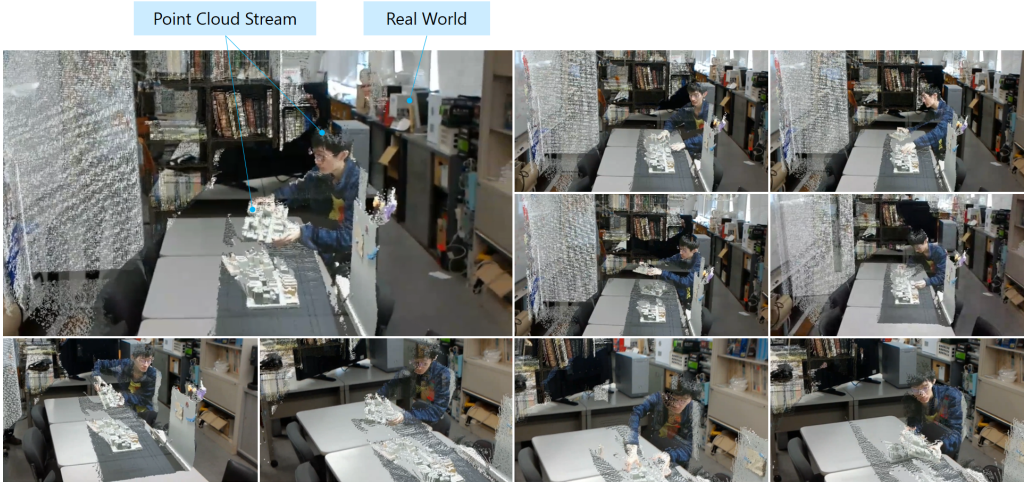
As described in sections 3.1 and 3.2, the PcsMR system could display the point cloud stream in the MR space in real time. Using the PcsMR system, a participant could observe other participants' behavior and objects, such as a physical model of a remote place, in 3D and at actual size from arbitrary viewpoints. If necessary, the PcsMR system can also synthesize 3D virtual models created using BIM software.

However, we found some problems.

The first is related to wide area network (WAN) environments. The two experiments we conducted could be carried out only in local area network (LAN) environment. We planned for the the PcsMR-1 project to be constructed in a WAN environment and for the laboratory to create models by BIM and by point cloud stream. This is because a point cloud stream can realistically represent features of the laboratory's state, such as student behavior and papers placed on the desk, which are hard to express using a BIM model. However, the transfer of point cloud streams experienced a large latency over a WAN environment, thus, we had to construct the laboratory model using only BIM.

The second problem is improving the point

Figure 7
PcsMR-2: HoloLens Screen Shots



cloud stream's granularity (present: 512 × 424) and expanding the generation area (present: 0.5-8.0 m from the Kinect, and within 70° horizontal and 60° vertical).

The last problem is that only one Kinect can be used per PC; therefore, it would be necessary to synchronize multiple Kinects and PCs when acquiring point cloud streams from many directions.

CONCLUSION AND FUTURE WORK

This research has the following contributions.

- To realize telepresence in architecture and urban fields, we developed a simple MR system known as PcsMR that displays point cloud streams. It consists of a Kinect, a PC, a router, and a HoloLens, and uses a noise elimination function on its point cloud stream.
- We constructed two prototypes based on PcsMR and carried out experiments. Application possibilities for architecture and urban fields include meetings and communications that share real-time 3D objects and include the movement of remote participants and objects.

Future tasks include adapting PcsMR to wide area networks and the internet environment and improving point cloud stream graininess and acquisition ranges.

ACKNOWLEDGEMENT

This research has been partly supported by the research grant of Nohmura Foundation for Membrane Structure's Technology, and by JSPS KAKENHI Grant Number JP16K00707.

REFERENCES

Agarwal, S, Snavely, N, Simon, I, Seitz, SM and Szeliski, R 2009 'Building Rome in a Day', *IEEE 12th International Conference on Computer Vision (ICCV)*, pp. 72-79

Beck, S, Kunert, A, Kulik, A and Froehlich, B 2013, 'Immersive group-to-group telepresence', *Visualization and Computer Graphics, IEEE Transactions*, 19(4), pp. 616-625

Buxton, W 1992 'Telepresence: integrating shared task and person spaces', *Proceedings of Graphics Interface*, pp. 123-129

Datta, S and Beynon, D 2005, 'A Computational Approach to the Reconstruction of Surface Geometry from Early Temple Superstructures', *International Journal of Architectural Computing*, 3(4), pp. 471-486

Fukuda, T, Zhang, T and Yabuki, N 2012 'Availability of Mobile Augmented Reality System for Urban Landscape Simulation', *9th International Conference CDVE 2012*, Osaka, pp. 231-238

Hidaka, N, Michikawa, T, Motamedi, A, Yabuki, N and Fukuda, T 2018, 'Polygonization of Point Cloud of Tunnels Using Lofting Operation', *International Journal of Automation Technology (IJAT)*, 12(3), pp. 356-368

Kowalski, M, Naruniec, J and Daniluk, M 2015 'LiveScan3D: A Fast and Inexpensive 3D Data Acquisition System for Multiple Kinect v2 Sensors', *In 3D Vision, 2015 International Conference on*, Lyon, France

Marchand, EB, Han, X and Dorta, T 2017 'Immersive retrospection by video-photogrammetry', *Proceedings of the 35th eCAADe Conference, Vol. 2*, Rome, pp. 729-738

Milgram, P, Takemura, H, Utsumi, A and Kishino, F 1994 'Augmented Reality: A Class of Displays on the Reality-Virtuality Continuum', *Proc. SPIE, Vol.2351*, pp. 282-292

Nagakura, T, Tsai, D and Choi, J 2015 'Capturing history bit by bit - Architectural database of Photogrammetric Model and Panoramic Video', *Proceedings of the 33rd eCAADe Conference - Volume 1*, Vienna, pp. 685-694

Orts-Escolano, S, Rhemann, C, Fanello, S, Chang, W, Kowdle, A, Degtyarev, Y, Kim, D, Davidson, PL, Khamis, S, Dou, M, Tankovich, V, Loop, C, Cai, Q, Chou, PA, Mennicken, S, Valentin, J, Pradeep, V, Wang, S, Kang, SB, Kohli, P, Lutchyn, Y, Keskin, C and Izadi, S 2016 'Holoportation: Virtual 3D Teleportation in Real-time', *Proc. of the 29th Annual Symposium on User Interface Software and Technology*, pp. 741-754.

Shen, Z and Kawakami, M 2010, 'An online visualization tool for Internet-based local townscape design', *Computers, Environment and Urban Systems*, 34(2), pp. 104-116

Sun, L, Fukuda, T and Resch, B 2014, 'A synchronous distributed cloud-based virtual reality meeting system for architectural and urban design', *Frontiers of Architectural Research*, 3(4), pp. 348-357