International Journal of Innovative Computing 10(2) 15-20

UTM UNIVERSITI TEKNOLOGI MALAYSIA
INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING
ISSN 2180-4370
Journal Homepage : https://ijic.utm.my/

A Robust Real-Time 3D Reconstruction Method for Mixed Reality Telepresence

Fazliaty Edora Fadzli & Ajune Wanis Ismail
School of Computing, Faculty of Engineering
Universiti Teknologi Malaysia
81310 UTM Johor Bahru, Johor, Malaysia
fazliaty.edora@gmail.com; ajune@utm.my

Submitted: 30/08/2020. Revised edition: 19/10/2020. Accepted: 21/10/2020. Published online: 19/11/2020
DOI: https://doi.org/10.11113/ijic.v10n2.265

***Abstract*—Mixed Reality (MR) is a technology which enable to bring a virtual element into the real-world environment. MR intends to improve reality on the virtual world immerse onto real-world space. Occasionally the MR has been improved as the display technologies advanced progressively. In MR collaborative interface context, the local and remote user work together on collaborative task while sense the immersive environment in the cooperative application. User telepresence is an immersive telepresence, where the reconstruction of a human appears in a real-life. Up till now, producing full telepresence of the life-size human body may require a high transmission bandwidth of the internet. Therefore, this paper explores on a robust real-time 3D reconstruction method for MR telepresence. This paper discusses the previous works on the reconstruction method of a full-body human and the existing research works that have proposed the reconstruction methods for telepresence. Besides the 3D reconstruction method, this paper also enlightens our recent finding on the MR framework to transport a full-body human from a local location to a remote location. The MR telepresence will be discussed, as well as the robust 3D reconstruction method which has been implemented with user telepresence feature where the user experiences an accurate 3D representation of a remote person. The paper ends with the discussion and results, MR telepresence with robust 3D reconstruction method to execute user telepresence.**

***Keywords*—Mixed Reality, 3D Reconstruction, Telepresence**

I. INTRODUCTION

The virtual world and the physical world can coexist together with Mixed Reality (MR) technology. Through MR, the virtual element able to be a part of the real world as object in physical world can occlude them as well as being fixed in the real-world position [1]. Virtual Reality (VR) shifts the user's view of being in a new environment instead of being in a real environment. The digital content can be viewed through the lenses of the Head Mounted Display (HMD), in which stereo cameras behave as the human eyes by displaying the view ahead of the HMD. Teleportation is a general concept involving a distinct movement to a destination [2]. User teleportation is a method to capture 3D objects and teleport into virtual space. Telepresence works in the early years has been conducted using an array of cameras in focus on capturing dynamics scene [3]. User teleportation or the presence of a reconstructed human in the new environment has the benefits of minimizing the travelling cost and time consumed to meet each other. In earlier researches, hardware and computational become the limitations because during those time only low-resolution of 3D models were able to be captured to be transmitted to the remote viewer by the system. Based on [4], various researches focusing on 3D telepresence systems that could bring the concept idea of user teleportation. Real-time capture, transmission, and displays are the challenges faced by the telepresence systems. Nowadays, stereo camera is equipped to provide structured light generated from the Infrared (IR) sensor. The lens sensor of the stereo camera provides RGB-Depth (RGB-D) data, which is another advanced approach for real-time depth estimation [5]. An image with estimated depth is generated by matching and patching one image to another which produced by each sensor of the stereo camera.

II. MIXED REALITY TELEPRESENCE

Telepresence has become the recent research area due to its high motivation and potential technology to cut travelling cost, speed up the time, and considering as an advanced communication approach that able to enhance the collaborative works. MR telepresence has proven to be the next technological wave. However, the reconstruction method is not an easy feat [6]. While extensively studied, the issue of realistic, robust and, importantly, fast 3d reconstruction remains a challenge task for researchers. Majority of the state-of-the-art approaches that aims for real-time applications, for instance, immersive reality, mainly discussed the problem of synthesizing intermediate views for the specified viewpoints, instead of generating a single complete 3D surface. The challenge for performing a realistic telepresence experience is the real-time positional reference devices [7]. If an avatar representation is developed to represent the real user, the avatar may be animated to imitate the movements of the user. For instance, in a telepresence session which two MR users are communicating, a viewer may perceive another user's avatar and thus create a tangible sense of the presence of the other user in the viewer's environment [8]. One of the potential solutions is the reconstruction and detection of occluded portions of the 3D model of human body using depth data from a single viewpoint [9]. However, it may not have data for a certain part of an object due to occlusion. Therefore, this paper proposes a method to use current depth sensor to capture the full human body from a single viewpoint.

In this paper, RGB-D sensor that used to capture human and transport the human into an MR collaborative system and a methodology for the creation of accurate, realistic, full 3D reconstructions of moving foreground objects, for instance, humans, to be exploited in real-time telepresence were presented. Although the RGB-D sensor has attracted many researcher's attention however very less presented works can generate full 3D models of moving objects from multiple depth camera streams in real-time. But some researchers explored multiple depth cameras to track large-scale environment [10]. Our method is proposed to be robust to its calibration because we skip some processes and then we have proposed telepresence application to measure how it could affect the user experience in MR telepresence system using commodity depth sensors.

III. PREVIOUS WORKS

A a real sense of presence to the user by accumulating existing scene reconstruction technologies was presented by [11]. However, there has been an issue making a full-duplex system due to the hardware restriction and restriction on the display configuration. An approach where multiple cameras acquired the images of a user from various angles, then transmit it across network before being displayed on a revolving flat panel display which able to display different images depending on the viewing position orientation was suggested by [12]. However, an issue arises when a spectator walks surround the display system and could see that some positions overlapped each other as a result of the feature of the display panel. [13] also presented an immersive virtual environment framework intended for remote collaboration and training of physical activities which utilized the multi-camera system to perform real-time full-body 3D reconstruction of the user. However, the speed of the reconstruction is one of the drawbacks of the system.

The number of researches and projects applied in the area of 3D reconstruction in the telepresence system has significantly expanded with the release of affordable commodity depth sensors for instance Microsoft Kinect. [14] presented a rapidly generated detailed indoor scene 3D reconstructions by allowing a user to hold and move a Kinect camera. This system was also able to extract live depth data from a moving Kinect camera as well as generate a single 3D model that is high-quality and geometrically precise in real-time. The camera posed in six-degrees-of-freedom (DOF) were also continually tracked and new viewpoints of the scene were fused into a global surface-based representation. An advanced telepresence system was proposed by [15] which provides 3D scene capture in real-time, fully dynamic, continuous-viewpoint, and the user does not require to put on any viewing or tracking device. However, there was a multiple Kinect interference issue as same dot pattern was projected by each of the device at the same wavelength which would cause conflict as the patterns of all other units can be seen by each Kinect and it is difficult for each unit to differentiate other units from its own patterns [15]. A high-quality 3D reconstruction of people including surrounding in real-time by using a set of depth cameras were introduced by [16]. It is also an end-to-end system for AR and VR telepresence. However, they do not count on speed when they aim to achieve quality. In order to solve the speed issue, [17] has demonstrated a small-scaled telepresence which use adaptive avatar for MR remote collaboration. [18] suggested a MR workspace telepresence that enable user to interact with remote user in the provided environment using wearable MR headset and the user able to get a sense of coexistence with the remote user and objects.

Recent work in real-time 3D reconstruction for MR telepresence that wirelessly transmits the data to Microsoft's HoloLens HMD have been presented by [19]. The RGB cameras capture high definition data to be used for the reconstructed frame which represented as a polygonal mesh with polygons textured. [20] introduced a telepresence system based on simulated holographic with the aim to discard the need for HMD by using holographic display composed of commodity light projector and an acrylic square pyramid. They proceeded their work to create a visual telepresence system that is low-bandwidth and cost-effective using commodity depth sensors [21]. Remote collaboration using telepresence system which combined 360 video and 3D reconstruction together has been proposed by [22]. Over the years, there has been tremendous advancement for 3D scene reconstruction from data obtained from RGB-D sensor in real-time [23], along with the exploration of such data in VR or AR settings. Nevertheless, there are significant technological difficulties in integrating these two elements into telepresence systems.

Based on the review of previous works, there are research questions to be answered. There are:

- How to speed up the 3D reconstruction in the MR framework during real-time 3D scene capturing from a single continuous-viewpoint?
- How to implement the proposed 3D reconstruction method by considering two users from different

locations are required to interact in a shared environment?

- How to achieve 3D telepresence that commoditized depth sensors?

IV. Proposed Method

Depth processing allows objects captured from various perspective to be backed up into 3D. Then, all the points will be registered and aligned to be able to completely restructure the environment. Depth data can be obtained by the current RGB-D devices and there are two methods to so, first by stereo cameras, and second by structured light sensors. In order to capture local user in the workspace, user telepresence requires stereo camera, which is a camera that consists of two or more lens with an image sensor for each of those lenses, to provide depth data. By replicating the human binocular vision, the camera able to delivers the ability to acquire 3D images. To enhance the depth data capturing capacity, the stereo camera nowadays has been equipped with an Infrared (IR) sensor.

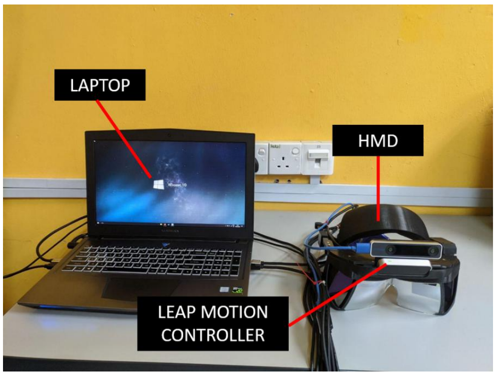

Fig. 1. Experiments setup

RGB-D data are provided from the lens sensor of the stereo camera. RGB-D data is an alternative advanced tool for real-time depth estimation. An image with estimated depth by match and patch one image to another was generated by each sensor in the stereo camera. Fig. 1 illustrates the stereo camera along with other devices that has been used in this research to capture RGB-D data.

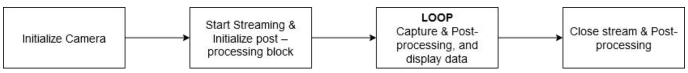

Fig. 2. Point cloud rendering process

As shown in Fig. 2, the process of point cloud rendering was performed to visualized the data into 3D. This is important in order to map the RGB-D data from the depth camera into the output display. This process starts with the initialization of a camera where the hardware is attached to the PC, which has run the computer vision algorithm in order to identify the camera input with the PC. Then, the camera is streamed and the post-processing block is initialized, which is the colour and depth mapping as in Fig. 3 that later will be used as the next process input.

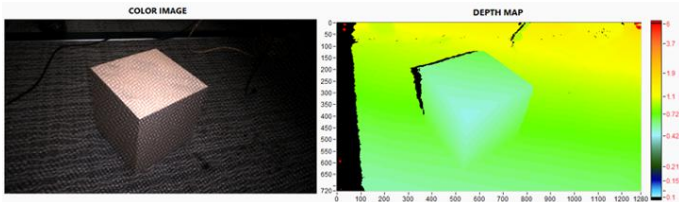

Fig. 3. Colour image and depth maps

Next, the post-processing method for mapping the colour image and depth data was called in order to obtain the output display in 3D data. The outcome of this process shown in Fig. 4 as the result obtained from the previous colour image and depth data input. The 3D data of depth data from a different angle are illustrated in Fig. 4 (a) and Fig. 4 (c), while Fig. 4 (b) and Fig. 4 (d) demonstrate the 3D data from colour and depth data from a different angle.

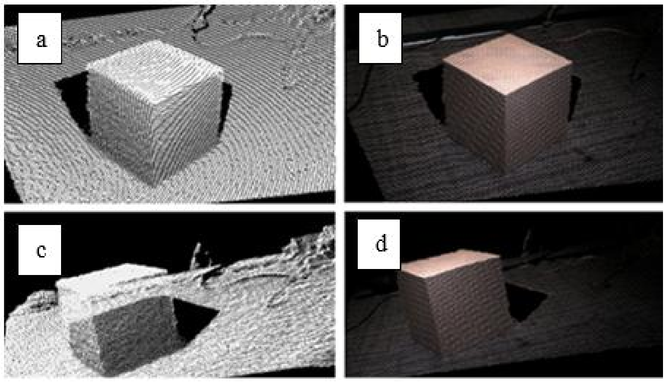

Fig. 4, Results of the colour image and depth data

The process will continue to loop and it is depending on the user to decide on when to stop the process. The proposed system acquired the local user's body and process the data into the application obtained from the point cloud rendering process, in order to achieve the user telepresence. As shown in Fig. 5 (a) and Fig. 5 (b), the presence of human exists in a real environment, even they are in a different location. This work proposed the fast 3D reconstruction method to allow for compressing the data sent over the network. However, green screen setups are not used in this research to make sure that the system is allowed to be used in natural environments that appear in realistic scenarios.

Fig. 5. Results of the point cloud rendering process

For this research, both local and remote users will be in different environment which are MR and VR respectively. Hence, there will be different method for both of the users to interact with the application. However, both users will share the same collaborative task, which are to spawn, manipulate, and delete the 3D object. Thus, there will be different executable program to respond to each user for the collaborative interface. Object manipulation interaction share the same process in MR and VR sides. The interaction consists of tap on the 3D button and for translate and rotate manipulation were by grabbing the virtual 3D object which made possible by the collision detection between the rigid finger body from the input device with the 3D button and object. Once the rigid body collides with the 3D button, the object was spawned in a shared space.

Fig. 6 shows the process of interaction to manipulate the object. In this research, basic user interaction was used and collision detection was performed to spawn object and grab the object in the VR environment. Rigid body was applied to enable the collider, so that once the user can hit the desired object and move the object.

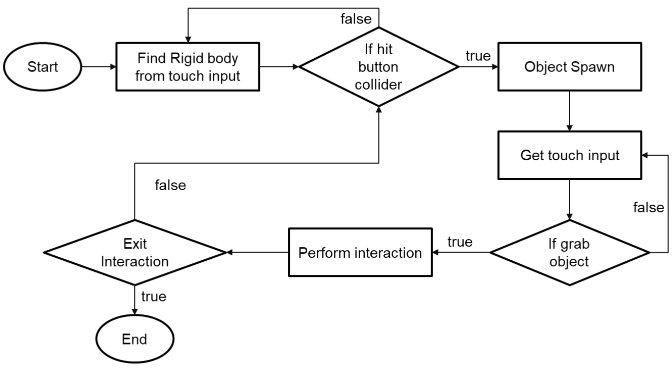

Fig. 6. Interaction with 3D objects in MR

For local user who was in MR environment to perform interaction of grabbing 3D virtual object, the input device tracked their index and thumb coordinate to identify them as a pinch gesture. Once the pinch gesture was detected by the system, the local user enables to perform translation or rotation of the 3D virtual object. After performing the manipulation, the user can choose to repeat the interaction or end the process. While for remote user who was in VR environment, the interaction was performed by using joystick as the interaction input.

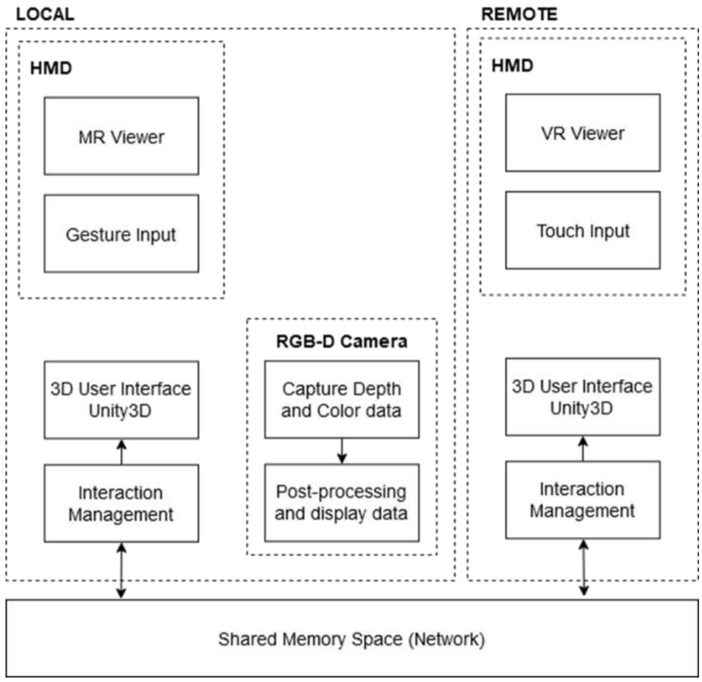

Fig. 7. Telepresence Remote user and local user connection

The integration of the overall process both on user telepresence and the collaborative interface was done in order for both parts to be working together. Fig. 7 demonstrates the overview of the collaborative interface components, which involved all the hardware required to made this application works. Networking was crucial in this research. Intel Realsense D435 was utilized to enable the user telepresence. This research requires two separate spaces to test the user telepresence and remote collaborative interface. Thus, this application will have local and remote users.

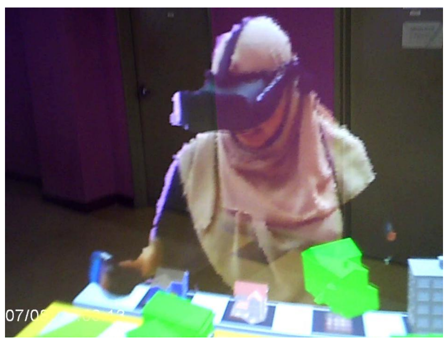

Fig. 8. Local user's perspective in MR environment

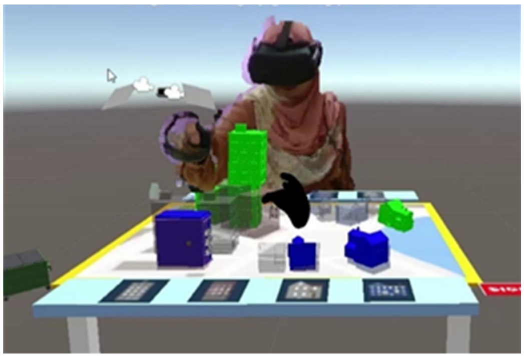

Fig. 9. Remote User Interaction in VR space

TABLE 1. Wilcoxon Testing to Compare MR and VR user experiences

MR	VR	Sign	Abs	Rank	Signed Rank
5.5	6.6	-1	1.1	12	-12
6.1	6.7	-1	0.6	6.5	-6.5
6.2	6.6	-1	0.4	4	-4
6	6.8	-1	0.8	9	-9
5.8	6.7	-1	0.9	10.5	-10.5
6.2	6.7	-1	0.5	5	-5
5.9	6.8	-1	0.9	10.5	-10.5
5.9	6.6	-1	0.7	8	-8
2.4	1.8	1	0.6	6.5	6.5
6.4	6.6	-1	0.2	2	-2
6.1	6.2	-1	0.1	1	-1
6.3	6.6	-1	0.3	3	-3

Fig. 8 shows the local user's perspective who can see the remote user in MR telepresence performing the interaction using the joystick and wearing HMD. While Fig. 9 shows the user in a VR view. To integrate user telepresence with the collaborative interface across a shared network, both local and remote user application needs to execute concurrently. Remote user experience in VR viewer using HMD and RGB-D camera to capture and transmitted 3D data to the local user who experience in MR viewer using wearable MR headset. The RGB-D camera acquired the remote user's depth and colour data and these data were processed by executing the Post-Processing. In this research, the one who experienced the user telepresence of the remote user in 3D mesh on the display view is the local user. The remote user only perceived the social presence of local users through the 3D avatar visual cues on their display view. The 3D UI were controlled and synchronized at the network by the interaction management. The synchronization between users were controlled by the network which also performed as shared memory space.

V. RESULTS AND ANALYSIS

Testing results are evaluated using the Wilcoxon test to compare two results from user experience in MR and VR view. This test measures the difference between each pair set and analyses these differences, as seen in Table 1.

According to the results in the Table 1, the MR and VR side will determine the mean value for each linear scale responses. The second and third column identified as set 1 and set 2 values. The difference between set 1 and set 2 value specified in the "Sign" column, where the answers were either positive or negative. The absolute value of the difference between the two sets were signified by the "Abs" column. The rank of the absolute value for each value were indicated in the "Rank" column and the column "Signed Rank" refers to the multiplication of input from Rank and Sign column. Since there is no significant difference is found based on the result, it was suggested for both MR and VR applications to be enhanced for better enjoyment and comfortability. Most of the respondent has rate they felt comfortable and enjoy on both applications.

Based on the user preferences, as shown in Fig. 10, the user prefers to stay in VR view compared to the MR view. They felt the MR environment is not realistic because only the frontal body of the remote user is displayed; they much prefer to be in VR because they can see a different perspective. The survey involves eight undergraduate students with their background in Computer Graphics and familiar with MR/AR technologies. Instruction manual for the MR application was given and the respondents were guided by the instructor before they can work with the application independently. Data were collected from questionnaires and the results presented in the bar graph. Besides, the analysis using Wilcoxon testing between VR and MR views has been presented in the previous table.

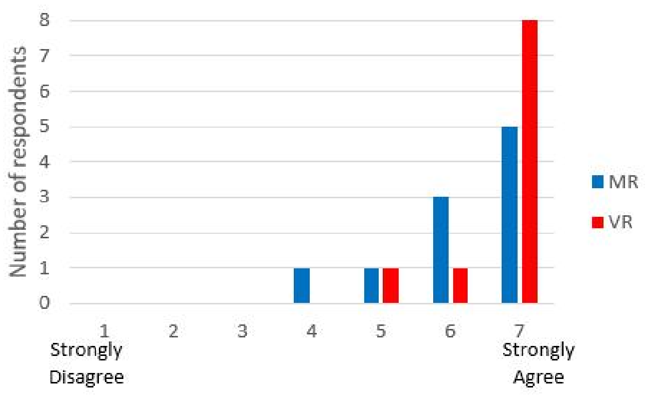

Fig. 10. User preferences based on the survey

VI. CONCLUSION AND LIMITATION

This paper has discussed the integration of robust 3D reconstruction method into collaborative MR interface that enable local and remote user to work on collaborative task. The early works in 3D reconstruction for telepresence system has also been presented in this paper. After 3D reconstruction of remote user was acquired by RGB-D sensor, the data were transmitted over network to be display on the remote site. The collaborative interface provided a shared environment for both local and remote users who were in MR and VR setting respectively.

There are, however, some limitations faced by MR user as the MR wearable headset was felt heavy on the user. The hardware also faced distortion issue that requires the users to somehow cross their eye in order to see MR view properly. The user's height position required to be initialized first for better experience of the MR telepresence. It is suggestable to provide a flexible way for height calibration and setting viewing point of the MR wearable headset .

ACKNOWLEDGMENT

Deepest gratitude and appreciation to Mixed and Virtual Reality Laboratory (mivielab) at Vicubelab, Faculty of Engineering, Universiti Teknologi Malaysia (UTM).

REFERENCES

[1] Nor' a, M. N. A., and Ismail, A. W. (2019). Integrating Virtual Reality and Augmented Reality in a Collaborative User Interface. *International Journal of Innovative Computing,* 9(2).

[2] Rahimi, K., Banigan, C., and Ragan, E. D. (2020). Scene Transitions and Teleportation in Virtual Reality and the Implications for Spatial Awareness and Sickness. *IEEE Transactions on Visualization and Computer Graphics*, 26(6), 2273-2287.

[3] Fuchs, H., State, A., and Bazin, J.-C. (2014). Immersive 3D Telepresence, 47.

[4] Orts, S., Rhemann, C., Fanello, S., Kim, D., Kowdle, A., Chang, W., *et al.* (2016). Holoportation: Virtual 3D Teleportation in Real-time.

[5] Fekri, A. M. Y., & Wanis, I. A. (2019, August). A Review on Multimodal Interaction in Mixed Reality Environment. *IOP Conference Series: Materials Science and Engineering,* 551(1), 012049, IOP Publishing.

[6] Fadzli, F. E., Ismail, A. W., Aladin, M. Y. F., & Othman, N. Z. S. (2020, May). A Review of Mixed Reality Telepresence. *IOP Conference Series: Materials Science and Engineering,* 864(1), 012081, IOP Publishing.

[7] Hilty, D. M., Randhawa, K., Maheu, M. M., McKean, A. J., Pantera, R., & Mishkind, M. C. (2020). A Review of Telepresence, Virtual Reality, and Augmented Reality Applied to Clinical Care. *Journal of Technology in Behavioral Science*, 1-28.

[8] Miller, S. A. (2015). U.S. Patent No. 9,215,293. Washington, DC: U.S. Patent and Trademark Office.

[9] Jang, I. Y., Cho, J. H., & Lee, K. H. (2012). 3D Human Modeling from a Single Depth Image Dealing with Self Occlusion. *Multimedia Tools and Applications*, 58(1), 267-288.

[10] P. Stotko, S. Krumpen, M. B. Hullin, M. Weinmann and R. Klein. (2019). SLAMCast: Large-Scale Real-Time 3D Reconstruction and Streaming for Immersive Multi-Client Live Telepresence. *IEEE Trans. on Visualization and Computer Graphics*, 25(5), 2102-2112.

[11] Towles, H., Chen, W. C., Yang, R., Kum, S. U., Kelshikar, H. F. N., Mulligan, J., ... & Holden, L. (2002). 3d Tele-collaboration over Internet2. *International Workshop on Immersive Telepresence, Juan Les Pins*.

[12] Tanikawa, T., Suzuki, Y., Hirota, K., & Hirose, M. (2005, December). Real World Video Avatar: Real-time and Real-size Transmission and Presentation of Human Figure. *Proceedings of the 2005 International Conference on Augmented Tele-existence*, ACM, 112-118.

[13] Kurillo, G., Bajcsy, R., Nahrsted, K., & Kreylos, O. (2008, March). Immersive 3d Environment for Remote Collaboration and Training of Physical Activities. *2008 IEEE Virtual Reality Conference*, IEEE, 269-270.

[14] Newcombe, R. A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A. J., ... & Fitzgibbon, A. W. (2011, October). Kinectfusion: Real-time Dense Surface Mapping and Tracking. ISMAR, 11(2011), 127-136.

[15] Maimone, A., & Fuchs, H. (2011, October). Encumbrance-free Telepresence System with Real-time 3D Capture and Display Using Commodity Depth Cameras. *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, IEEE, 137-146.

[16] Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., ... & Tankovich, V. (2016, October). Holoportation: Virtual 3d Teleportation in Real-time. *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, ACM, 741-754.

[17] Piumsomboon, T., Lee, G. A., Hart, J. D., Ens, B., Lindeman, R. W., Thomas, B. H., & Billinghurst, M. (2018, April). Mini-me: an Adaptive Avatar for Mixed Reality Remote Collaboration. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ACM, 46.

[18] Parikh, V., & Khara, M. (2018, May). A Mixed Reality Workspace Using Telepresence System. *International Conference on ISMAC in Computational Vision and Bio-Engineering,* Springer, Cham, 803-813.

[19] Joachimczak, M., Liu, J., & Ando, H. (2017, November). Real-time Mixed-reality Telepresence via 3D Reconstruction with HoloLens and Commodity Depth Sensors. *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, ACM, 514-515.

[20] Córdova-Esparza, D. M., Terven, J., Jiménez-Hernández, H., Herrera-Navarro, A., & García-Huerta, A. V. C. J. (2018). Telepresence System based on Simulated Holographic Display. arXiv preprint arXiv:1804.02343.

[21] Córdova-Esparza, D. M., Terven, J. R., Jiménez-Hernández, H., Herrera-Navarro, A., Vázquez-Cervantes, A., & García-Huerta, J. M. (2019). Low-bandwidth 3D Visual Telepresence System. *Multimedia Tools and Applications,* 78(15), 21273-21290.

[22] Teo, T., Lawrence, L., Lee, G. A., Billinghurst, M., & Adcock, M. (2019, April). Mixed Reality Remote Collaboration Combining 360 Video and 3D Reconstruction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ACM, 201.

[23] Stotko, P., Krumpen, S., Hullin, M. B., Weinmann, M., & Klein, R. (2019). SLAMCast: Large-scale, Real-time 3D Reconstruction and Streaming for Immersive Multi-client Live Telepresence. *IEEE Transactions on Visualization and Computer Graphics*, 25(5), 2102-2112.