Real-Time 3D Reconstruction for Mixed Reality Telepresence Using Multiple Depth Sensors

Shafina Abd Karim Ishigaki[(✉)] and Ajune Wanis Ismail

Mixed and Virtual Reality Research Lab, ViCubeLab, School of Computing,
Universiti Teknologi Malaysia, 81310 Johor, Malaysia
shafina@graduate.utm.my, ajune@utm.my

Abstract. In recent years, there has been a great development in real-time three-dimensional (3D) scene reconstruction from depth sensor data, as well as the study of such data in Virtual Reality (VR) and Augmented Reality (AR) contexts. Although it has been extensively investigated and has attracted the attention of many researchers, the challenge of real-time 3D reconstruction remains a difficult research task. The majority of current techniques, target real-time 3D reconstruction for the single-view-based system rather than multi-view. In order to provide multi-view 3D reconstruction for Mixed Reality (MR) telepresence, this chapter aims to propose a multiple depth sensor capture using a marching square approach to produce a single full 3D reconstruction surface of a moving user in real-time. The chapter explains the design stage that involves setup from multiple depth sensors, surface reconstruction and merging of 3D reconstruction data for MR Telepresence. The chapter ends with results and a conclusion.

Keywords: Real-time · 3D reconstruction · MR telepresence · Depth sensor

1 Introduction

Globalization has boosted demand for immersive telepresence solutions, which are a type of technology that allows individuals who are physically separated to collaborate as if the user were in the same room [1]. By giving an alternative to travelling to presential meetings, this technology can save time and money while also decreasing environmental impact [2]. It enables users to present at a remote location through a virtual projection. The idea of telepresence arises from Marvin Minsky, one of the first to coin the term "telepresence" as the feeling of remote existence when using a teleoperator [3]. Since Marvin Minsky stated the principle, several systems have been developed to accomplish telepresence [4].

In early telepresence, due to limitations in the technology available at that time, telepresence systems were unable to acquire and transmit high-quality real-time three-dimensional (3D) reconstruction objects to distant users [5]. However, following the introduction of inexpensive commodity depth sensors such as the Microsoft Kinect

became more widely available, many 3D reconstructions at the room size were developed, such as room2room life-size telepresence [6]. The capability of 3D reconstruction that is able to capture and reconstruct real-world elements makes significant advances in telepresence [7].

However, implementing 3D reconstruction for telepresence requires real-time rendering for moving object since telepresence capture user movement in real-time. Real-time 3D reconstruction for moving objects is challenging in computer vision and computer graphics [7, 8]. The majority of current techniques target real-time 3D reconstruction for a single-view-based system, from a single depth sensor rather than multiple-view from multiple depth sensors.

With the use of a Head-mounted-display (HMD), 3D reconstruction can be viewed with a wider Field-of-View (FoV) compared to a standard phone or monitor [9]. This allows users to interact with virtual environments in a manner that is similar to being physically present in those environments [10]. Mixed Reality (MR) refers to a new category of experiences in which a layer of computer-graphics-based interaction tied to a certain activity is superimposed on top of the real environment [11]. It allows users to experience the actual world as a more integrated and better environment. The user of an MR interface sees the actual world through a portable HMD that coats graphics on the surroundings [12]. [13] were the first to demonstrate HMD, which gave rise to the concept of MR. According to [14], an MR system brings together physical and digital objects in the same space, allows for two-way communication between them, and keeps track of their location and state in real-time.

In this chapter, a depth sensor is proposed to capture real-world objects and reconstruct the data in 3D. However, the FoV of a depth sensor is insufficient to capture a full 3D object and physically impossible to gather all the input data for the object or scene at once [15]. Therefore, to produce a full 3D reconstruction, most researchers use the fusion Kinect method, which moves the depth sensor around the real-world object to capture the data by frame. This method, however, is very computationally intensive and can only be run in an offline mode [16]. Furthermore, more than half of the picture is obscured at any given time when employing a single depth sensor to reconstruct non-rigid motion, and continuous motion creates large frame-to-frame fluctuations, which may lead to inconsistency in the scene's topological structures [17].

As a result, multiple depth sensors are required to be positioned at different angles to obtain unobstructed data from the sides of the object in order to complete the full 3D reconstruction in real-time [18]. According to [19], multiple depth sensors, enable users to get visual information that may not be available from a certain perspective when using a single depth sensor. From the previous studies, multiple depth sensors often consist of more than two depth sensor devices. However, based on recent studies in [20–22], multiple depth sensors and multiple cameras can be achieved by utilizing two depth sensors or two cameras. Therefore, this chapter proposes a real-time 3D reconstruction method with multiple depth sensors for MR telepresence. The related work for this chapter is further discussed in the next section. The methodology of the proposed method is described in Sect. 3, and in Sect. 4 we presented the result.

2 Related Work

Real-world 3D reconstruction has long been an aim of computer vision. Numerous instruments, including as stereo cameras, laser range finders, monocular cameras, and depth sensors, have been used to correctly perceive the 3D environment [23]. With the advancement of depth sensor devices such as Microsoft Kinect and Intel RealSense, many new and exciting applications is developed. However, due to issues like noisy depth data, poor registration, camera tracking drift, and a lack of precise surface details, 3D models reconstructed by depth consumer cameras are not yet widely employed in applications. Table 1 represents the summary of related work on 3D reconstruction for MR telepresence which are utilizing depth sensor devices to enable 3D reconstruction.

Table 1. Related work of 3D reconstruction for MR telepresence using a depth sensor

Year	Researchers	Proposed method	No. of depth sensors	Drawbacks
2022	Lin et al. [24]	OcclusionFusion: calculate occlusion-aware 3D motion to guide the reconstruction	1	Unable to handle topology change
2022	Sun et al. [25]	Propose a hybrid voxel- and surface-guided sampling technique which leads to significant improvements in reconstruction quality	1	Inaccurate camera registration can affect final reconstruction quality
2021	Muhlenbrock et al. [21]	Propose a combination of a custom 3D registration with a feature detection algorithm	2	Slight distortions due to the depth camera in a larger volume
2020	Fadzli et al. [26]	Explores a robust real-time 3D reconstruction method for MR telepresence	1	MR wearable headset is heavy and has a face distortion issue

(*continued*)

Table 1. (*continued*)

Year	Researchers	Proposed method	No. of depth sensors	Drawbacks
2019	Prokopetc & Dupont [27]	Present a comparative study with an evaluation of four depth map-based Multi-View Stereo methods for dense 3D reconstruction against brand-new deep learning models	1	Dataset is hard to achieve in the settings of real medical environments due to the vast amount of specular materials and weakly-textured surfaces
2018	Papaefthymiou et al. [28]	Present a comparison of the most recent hardware and software methods for the speedy reconstruction of actual people using RGB or RGB-D images as input	1	Does not provide an evaluation of the proposed method
2017	Joachimczak et al. [10]	Proposed to render 3D reconstructions directly (real-time), inside HoloLens in MR Telepresence	2	Require stable internet connection
2016	Sing & Xie [29]	Garden: enables the player to transform their environment into a virtual garden of 3D reconstruction in MR	1	Improvements in drift correction are required

Sun et al. [25] have presented a real-time 3D reconstruction in MR by discretizing the world space into 10 cm voxels which are grouped into $8 \times 8 \times 8$ chunks. Voxels that have points that fall within its boundaries are marked occupied and rendered as cubes, like in Minecraft. While engaging with a virtual environment, the user is free to move through wide areas in the actual world without running into any barriers.

Meanwhile, Joachimczak et al. [10] highlighted the real-time 3D reconstruction in MR Telepresence in their study. They utilized HoloLens with Kinect devices as a depth sensor for the 3D reconstruction purpose as shown in Fig. 1. However, their study provides limited information as they did not perform the evaluation of the proposed

method for further discussion. In the research carried out by Fadzli et al. [26], the authors provided a discovery on the MR framework that might be used to transmit a full-body human from a nearby region to a more distant one. Therefore, they introduced a user telepresence capability for 3D reconstruction, allowing people to interact with an accurate 3D model of a remote user. The depth data from the depth sensor is captured and processed in real-time for the 3D reconstruction with the point cloud rendering process.
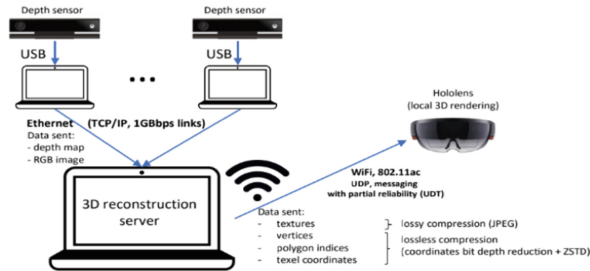
Fig. 1. Real-time 3D reconstruction using HoloLens and Kinect [10].

3   Real-Time 3D Reconstruction Method

This section elaborates on the real-time 3D reconstruction method, which consists of multiple depth sensor setup, surface reconstruction and merging of 3D reconstruction data for MR telepresence. This is explained in the next subsection.

3.1   Multiple Depth Sensor Setup

In order to produce the 3D reconstruction, depth information is required to generate the data. The conventional camera is only capable of producing two-dimensional (2D) red, green and blue (RGB) data, including photos and videos. The conventional camera is unable to collect depth data into the programmed device, the output is insufficient to enable and capture an item in 3D data. Therefore, a depth sensor is required in this chapter for capturing and converting real-world objects into 3D data. The depth sensor proposed in this chpater is the Microsoft Kinect sensor.

Microsoft Kinect sensor is a low-cost depth sensor known as a skeleton tracker that is used with the Microsoft software development kit [30]. It allows users to communicate with computers and game consoles intuitively via gestures and spoken commands without any additional peripherals [31]. It can also extract data about the subject's 3D joint locations [32].

Microsoft offers not one but two Kinect models. The first generation of Kinect sensors (Kinect v1) debuted in 2010. They were based on structured light coding technology. Following that, Microsoft made it Windows-compatible using a standard development kit (SDK) and a power conversion adaptor. Based on the time-of-flight (ToF) technology, the second generation of Kinect sensors (Kinect v2) was released in 2014 with improved specs in comparison to the first Kinect in terms of speed, accuracy, and field of vision.

However, the FoV of the Kinect sensor is constrained and physically impossible to gather all the point cloud data for the 3D object at once. Therefore, this chapter proposes two Kinect sensors, to acquire the data front and back of the user simultaneously. The setup for multiple depth sensors is shown in Fig. 2. Based on Fig. 2, the distance between the Kinect sensor and the user is fixed at 2.1 m. While the height of the Kinect sensor from the floor is fixed at 1.1 m. This setup for the multiple Kinect sensors is based on [33] research. This setup will ensure the Kinect sensor able to track the user's full body for the 3D reconstruction process.
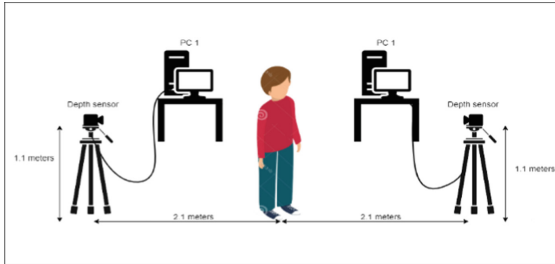

Fig. 2. Setup for multiple depths

3.2 Surface Reconstruction

Surface reconstruction is one of the steps in this chapter's real-time 3D reconstruction technique. Surface reconstruction is the process of acquiring 3D real-world object input data from a depth sensor in order to produce a mesh from the input data. For surface reconstruction, there are two methods available, explicit and implicit methods. The implicit method is proposed in this chapter to reconstruct the surface. When one of the isocontours is available near the input data, it is defined as an implicit surface. In the depiction of the reconstructed surfaces, the implicit technique employs the triangulation method.

The most intuitive form of triangulated surfaces for surface reconstruction uses k-nearest neighbours of a point to generate the connection between linked triangles made from the input points [34]. To generate the triangulated surface from the implicit method, the Marching Squares algorithm is by far the most well-known technique.

Marching Squares is a computer graphics approach for recovering 2D isocontours from a grid of sample points. The goal is to analyze each grid cell individually and to calculate the contours that pass through it depending on the values at its corners. The initial stage in this procedure is to identify the edges that are linked by contours depending on whether the corners are above or below the contour value. To keep things simple, the evaluation is only considering contours along with the value 0, focusing on the positive or negative corners. The Marching Squares lookup table is shown in Fig. 3 Based on Fig. 3, there is a total of $2^4 = 16$ distinct examples which consists of Case 0 until Case 15.
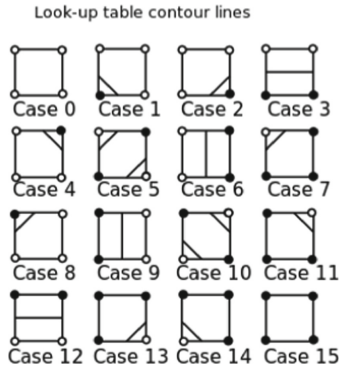
Fig. 3. Marching squares lookup table

The pseudocode for the marching square algorithm is shown in Table 2 as referred to [35]. Based on the Marching Squares algorithm the triangulation process is implemented by including the process of searching for vertex candidates from the value of the contour based on the Marching Squares lookup table. Hence, the triangulation is formed according to the vertex candidates lists.

Table 2. Marching squares algorithm

Marching Squares algorithm
1. function TRIANGULATION
2.   for all vertex candidates do
3.   current disparity coordinate ß visited
4.   block size = 2 x 2 FIND_BLOCK (current
5.   disparity coordinate, block size, current four
6.   vertex candidates per block)
7.   four vertex candidates ß current four vertex
8.   candidates per block
9.   FIND_TRIANGLE_CANDIDATE (four vertex
10.   candidates, current triangle candidate lists)
11.   triangle candidate lists ß current triangle
12.   candidates lists
13.   CREATE_TRIANGLE (triangle candidate lists,
14.   triangle)
15.   triangle lists ß triangle
16.   end for
17. end function

This process of triangulation using the Marching Squares algorithm is performed after the point cloud data is acquired from the depth sensor as shown in Fig. 4. Based on Fig. 4. After the triangulation has been completed and successfully produced the mesh. The mesh will be mapped with texture mapping. Texture mapping is the final stage to complete the 3D reconstruction look. The texture for the mesh is generated by composing the segmented images from the depth sensor as shown in Fig. 4. The images contain vertices that consist of a single colour per vertex. The colours are acquired to produce the texture and mapped to the mesh. After the texture has been mapped to the mesh, the 3D reconstruction final output is ready to display as a 3D representation for MR telepresence.
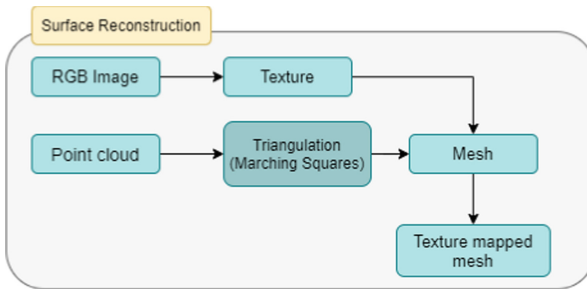
Fig. 4. Surface reconstruction process

3.3   Merging of 3D Reconstruction Data for MR Telepresence

After the mapped mesh texture has been completed, the next phase is to merge the 3D reconstruction data from multiple depth sensors for MR telepresence. The drawback of the Kinect sensor is one personal computer (PC) can only connect to one Kinect sensor. Therefore, a network is required to merge the data between the multiple Kinect sensors. The network used for this chapter is a local network. The local network will transmit the 3D reconstruction data from each Kinect at the remote user site to the receiver site, which is the local user. Figure 5 shows the process of merging 3D reconstruction data for the MR telepresence system.

Based on Fig. 5, before the 3D reconstruction data is transmitted through the network, the 3D reconstruction data that consists of vertices, triangles, colours and UVs will undergo compression to reduce the data size. Reducing the data size can reduce the bandwidth required for data transmission. After the receiver PC has received the 3D reconstruction data from both sender PC, the 3D reconstruction data will decompress and reconstruct the 3D data. The merging of 3D reconstruction data is then executed at the receiver site to be displayed at the MR telepresence. We manually adjust the position and rotation of the 3D reconstruction data from sender 1 and sender 2 using unity tools software, to be aligned in a full 3D reconstruction.
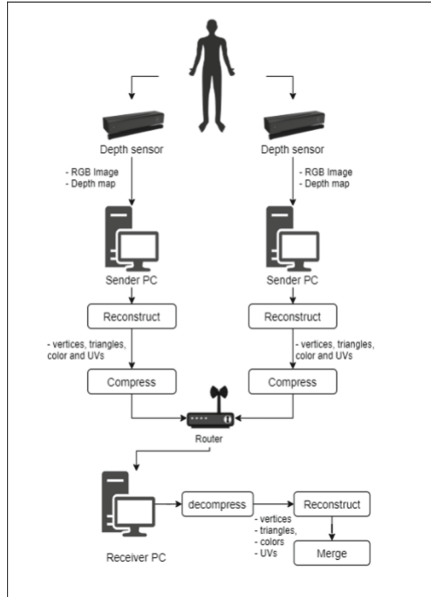
Fig. 5. Merging process at the receiver
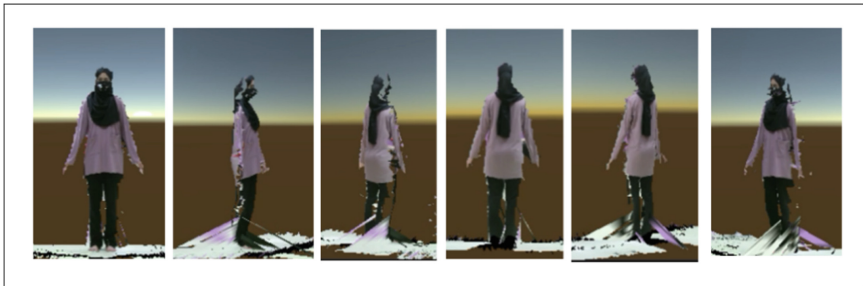
4 Results

This section continues to discuss the results of the proposed method using multiple depth sensors for 3D reconstruction in real-time for MR Telepresence. Figure 6 shows the final setup for the remote user workspace. Based on Fig. 6, the remote user is placed in between the Kinect sensor and the setup for the Kinect sensor is as discussed in the previous section.
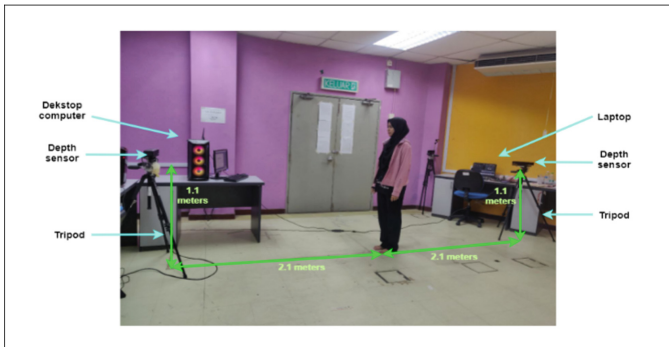
Fig. 6. Final setup for the remote user using multiple depth sensors

Next, Fig. 7 shows the results after the 3D reconstruction data has been merged from multiple depth sensors. In Fig. 7, we present a full real-time 3D reconstruction view of

the remote user at six different angles. As shown in Fig. 7, the front and behind views of the 3D reconstruction remote user have been aligning together into one complete full 3D reconstruction. However, there is a small gap between the 3D reconstruction data from the front and behind. This issue rises in this chapter and requires further improvement. The recommendation for improvement by [36] is to align and stitch two 3D meshes by the boundaries to complete 3D object models without prior registration using the Joint Alignment and Stitching of Non-Overlapping Meshes (JASNOM) algorithm. Recent studies by [37], also suggest a mesh stitching method with their algorithm that includes pre-processing process, to ensure the alignment is reasonable.


Fig. 7. Results of merging 3D reconstruction data at the receiver site

Furthermore, Fig. 8 shows the final output for real-time 3D reconstruction for the MR telepresence system using multiple depth sensors. Figure 8 is a captured image from the local user's point-of-view (POV). In Fig. 8, the remote user appears in 3D reconstruction in the MR environment. The interaction between the local user and the remote user is in real-time. The local user is able to interact and move around the MR telepresence simultaneously with the remote user.

Fig. 8. Final output real-time 3D reconstruction for MR telepresence

During the real-time interaction, the data that has been compressed which consists of vertices, triangles, colors and UVs from the remote user are transmitted to the local user using a stable local network. Since the data transmitted is required a local network, the bandwidth of the local network may influence the speed of the data transmission. The higher, the bandwidth of the local network, the faster the compressed data can be transmitted to the local user. However, if the local network bandwidth is low, network latency will occur and reduce the speed of the data transmission which can affect the real-time performance. Figure 9(a) shows the local user view of the remote user from the side. While Fig. 9(b) shows the side of the remote user in closer range.
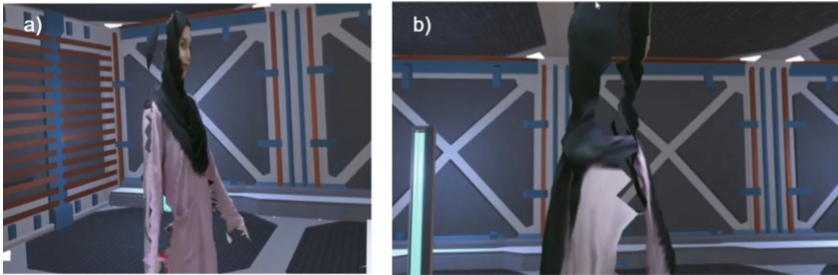
Fig. 9. Local user view of the remote user (a) side view (b) closer range of the side view

5 Conclusion

Based on the literature study, real-time 3D reconstruction technology and MR HMDs may now be used to enable a new potential for telepresence and teleoperation systems. Users can interact with a distant individual or operate remote devices by perceiving their 3D virtual representations as a part of their nearby surroundings.

In this chapter, we suggest employing multiple depth sensors to provide a real-time 3D reconstruction for MR telepresence. We utilized two depth sensors to produce depth data for the 3D reconstruction process. First, we captured depth data to form triangulation using a marching square algorithm and produce a surface reconstruction. The process continues with generating texture from the RGB image captured by the depth sensor to be mapped on the mesh. As the process is completed, the 3D reconstruction of the remote user is presented in real-time within the MR environment that was enabled on the HMD worn by the local user.

As for future work, we suggest addressing the gap issues that we found during the merging of the 3D reconstruction from multiple depth sensors to produce a more realistic and volumetric presentation. Other than that, enhancement for the interaction between the local user and remote user can be studied and improved by adding whiteboard interaction, pointing ray and object manipulation in MR telepresence.

References

1. Itani, O.S., Hollebeek, L.D.: Light at the end of the tunnel: visitors' virtual reality (versus in-person) attraction site tour-related behavioral intentions during and post-COVID-19. Tour. Manag. **84**, 104290 (2021). https://doi.org/10.1016/J.TOURMAN.2021.104290

2. Luevano, L., Lopez de Lara, E., Quintero, H.: Professor avatar holographic telepresence model. Hologr. Mater. Appl. (2019). https://doi.org/10.5772/INTECHOPEN.85528

3. Zahorik, P., Jenison, R.L.: Presence as being-in-the-world. Presence Teleop. Virtual Environ. **7**, 78–89 (1998). https://doi.org/10.1162/105474698565541

4. Sakashita, M., Minagawa, T., Koike, A., Suzuki, I., Kawahara, K., Ochiai, Y.: You as a puppet: evaluation of telepresence user interface for puppetry. In: UIST 2017 - Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, pp. 217–228 (2017). https://doi.org/10.1145/3126594.3126608

5. Stotko, P., Krumpen, S., Hullin, M.B., Weinmann, M., Klein, R.: SLAMCast: large-scale, real-time 3D reconstruction and streaming for immersive multi-client live telepresence. IEEE Trans. Vis. Comput. Graph. **25**, 2102–2112 (2019). https://doi.org/10.1109/TVCG.2019.2899231

6. Pejsa, T., Kantor, J., Benko, H., Ofek, E., Wilson, A.: Room2Room: enabling life-size telepresence in a projected augmented reality environment. In: Proceedings of ACM *Conference* on *Computer Supported Cooperative Work*, CSCW, vol. 27, pp. 1716–1725 (2016). https://doi.org/10.1145/2818048.2819965

7. Ingale, A.K.: Real-time 3D reconstruction techniques applied in dynamic scenes: a systematic literature review. Comput. Sci. Rev. **39**, 100338 (2021). https://doi.org/10.1016/J.COSREV.2020.100338

8. Cai, H., Feng, W., Feng, X., Wang, Y., Zhang, J.: Neural Surface Reconstruction of Dynamic Scenes with Monocular RGB-D Camera (2022)

9. Teo, T., Hayati, A.F., Lee, G.A., Billinghurst, M., Adcock, M.: A technique for mixed reality remote collaboration using 360 panoramas in 3D reconstructed scenes. In: Proceedings of ACM Symposium on Virtual Reality Software and Technology, VRST, vol. 11 (2019). https://doi.org/10.1145/3359996.3364238

10. Joachimczak, M., Liu, J.: Real-time mixed-reality telepresence via 3D reconstruction with HoloLens and commodity depth sensors (2017). https://doi.org/10.1145/3136755.3143031

11. John, B., Wickramasinghe, N.: A review of mixed reality in health care. In: Wickramasinghe, N., Bodendorf, F. (eds.) Delivering Superior Health and Wellness Management with IoT and Analytics. HDIA, pp. 375–382. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-17347-0_18

12. Harborth, D.: A systematic literature review on augmented reality augmented reality in information systems research: a systematic literature review (2017)

13. Sutherland, I.E.: A head-mounted three dimensional display (1968)

14. Azuma, R.T.: A survey of augmented reality. Presence Teleop. Virtual Environ. **6**, 355–385 (1997)

15. Li, J., Huang, S., Cui, H., Ma, Y., Chen, X.: Automatic point cloud registration for large outdoor scenes using a priori semantic information. Remote Sens. **13**, 3474–3474 (2021). https://doi.org/10.3390/RS13173474

16. Jia, Q., et al.: Real-time 3D reconstruction method based on monocular vision. Sensors **21**, 5909–5909 (2021). https://doi.org/10.3390/S21175909

17. Nor'a, M.N.A., Fadzli, F.E., Ismail, A.W.: A review on real-time 3D reconstruction methods in dynamic scene. Int. J. Innov. Comput. **12**, 91–97 (2022). https://doi.org/10.11113/IJIC.V12N1.317

18. Clark, R.A., Mentiplay, B.F., Hough, E., Pua, Y.H.: Three-dimensional cameras and skeleton pose tracking for physical function assessment: a review of uses, validity, current developments and Kinect alternatives. Gait Posture. **68**, 193–200 (2019). https://doi.org/10.1016/J.GAITPOST.2018.11.029

19. Meruvia-Pastor, O.: Enhancing 3D capture with multiple depth camera systems: a state-of-the-art report. In: Rosin, P.L., Lai, Y.-K., Shao, L., Liu, Y. (eds.) RGB-D Image Analysis and Processing. ACVPR, pp. 145–166. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-28603-3_7

20. Mühlenbrock, A., Fischer, R., Schröder-Dering, C., Weller, R., Zachmann, G.: Fast, accurate and robust registration of multiple depth sensors without need for RGB and IR images. Vis. Comput. **38**, 3995–4008 (2022). https://doi.org/10.1007/s00371-022-02505-2

21. Muhlenbrock, A., Fischer, R., Weller, R., Zachmann, G.: Fast and robust registration of multiple depth-sensors and virtual worlds. In: Proceedings of 2021 International Conference on Cyberworlds, CW 2021, pp. 41–48 (2021). https://doi.org/10.1109/CW52790.2021.00014

22. Espinosa, R., Ponce, H., Gutiérrez, S., Martínez-Villaseñor, L., Brieva, J., Moya-Albor, E.: Application of convolutional neural networks for fall detection using multiple cameras. Stud. Syst. Decis. Control **273**, 97–120 (2020). https://doi.org/10.1007/978-3-030-38748-8_5/COVER

23. Li, J., Gao, W., Wu, Y., Liu, Y., Shen, Y.: High-quality indoor scene 3D reconstruction with RGB-D cameras: a brief review. Comput. Vis. Media **83**(8), 369–393 (2022)

24. Lin, W., Zheng, C., Yong, J.-H., Xu, F.: OcclusionFusion: occlusion-aware motion estimation for real-time dynamic 3D reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1736–1745 (2022)

25. Sun, J., Chen, X., Wang, Q., Li, Z., Averbuch-Elor, H., Zhou, X., Snavely, N.: Neural 3D reconstruction in the wild. In: Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings, pp. 1–9 (2022)

26. Fadzli, F.E., Ismail, A.W.: A robust real-time 3D reconstruction method for mixed reality telepresence. Int. J. Innov. Comput. **10**, 15–20 (2020). https://doi.org/10.11113/IJIC.V10N2.265

27. Mandal, S., et al.: Lyft 3D object detection for autonomous vehicles. In: Artificial Intelligence for Future Generation Robotics, pp. 119–136 (2021). https://doi.org/10.1016/B978-0-323-85498-6.00003-4

28. Papaefthymiou, M., Kanakis, M.E., Geronikolakis, E., Nochos, A., Zikas, P., Papagiannakis, G.: Rapid reconstruction and simulation of real characters in mixed reality environments. In: Ioannides, M. (ed.) Digital Cultural Heritage. Lecture Notes in Computer Science, vol. 10605, pp. 267–276. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75826-8_22

29. Das, S., et al.: Advance machine learning and artificial intelligence applications in service robot. In: Artificial Intelligence for Future Generation Robotics, pp. 83–91 (2021). https://doi.org/10.1016/B978-0-323-85498-6.00002-2

30. Moon, S., Park, Y., Ko, D.W., Suh, I.H.: Multiple kinect sensor fusion for human skeleton tracking using Kalman filtering. **13** (2017). https://doi.org/10.5772/62415

31. Mukhopadhyay, M., et al.: Facial emotion recognition based on textural pattern and convolutional neural network. In: 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), pp. 1–6 (2021). https://doi.org/10.1109/GUCON50781.2021.9573860

32. Mocanu, C., Mocanu, I.: Human body posture recognition using a kinect sensor (2015)
33. Fadzli, F.E., Ismail, A.W., Ishigaki, S.A.K., Nor'a, M.N.A., Aladin, M.Y.F.: Real-time 3D reconstruction method for holographic telepresence. Appl. Sci. **12**, 4009–4009 (2022). https://doi.org/10.3390/APP12084009
34. Khatamian, A., Arabnia, H.R.: Survey on 3D surface reconstruction (2016). https://doi.org/10.3745/JIPS.01.0010
35. Jang, G.R., et al.: Real-time polygon generation and texture mapping for tele-operation using 3D point cloud data. J. Inst. Control Robot. Syst. **19**, 928–935 (2013). https://doi.org/10.5302/J.ICROS.2013.13.8012
36. Brandão, S., Costeira, J., Costeira, J.P., Veloso, M.: Effortless Scanning of 3D Object Models by Boundary Aligning and Stitching mobile phone dataset analysis View project Point matching View project Effortless Scanning of 3D Object Models by Boundary Aligning and Stitching (2014)
37. Naik, R., Singh, P., Kalra, P.: Putting jewellery and accessories on a 3D face model generated from 2D image. In: Babu, R.V., Prasanna, M., Namboodiri, V.P. (eds.) NCVPRIPG 2019. CCIS, vol. 1249, pp. 229–238. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-8697-2_21