Computer Fraud and Security
ISSN (online): 1873-7056

Revolutionizing Telepresence: Seamless Life-Sized Holographic Imaging in Real-Time

AKM Bellal Hossain[1], Muhammad Shamsul Alam[2], Nadir Abdelrahman Ahmed Farah[3],

Khadiga Balla Omer Elfaki[4], Abakar Ibraheem Abdalla Aadam[5]

[1].Department of Information System and Cyber Security, University of Bisha, Bisha 61922, Saudi Arabia, Faculty of Computing, Universiti Teknologi Malaysia, 81310, Johor Baharu, Johor, Malaysia, bhosayn@ub.edu.sa, k.m.a@graduate.utm.my,

https://orcid.org/0000-0003-3877-7037

[2.] Department of Computer Science and Artificial Intelligence, University of Bisha, Bisha-61922 Saudi Arabia, alam@ub.edu.sa, https://orcid.org/0000-0002-9419-3928

[3.] Information System and Cyber Security Department, University of Bisha, Bisha-61922, Saudi Arabia, nfarah@ub.edu.sa

[4.] Information System and Cyber Security Department (Female Section), University of Bisha, Bisha-61922, Saudi Arabia, khbomar@ub.edu.sa

[5.] Assistance Professor, Department of Computer science and Information System, University of Bisha, Saudi Arabia, E-mail: aiaadam@ub.edu.sa

Abstract

Telepresence has emerged as a compelling research area driven by its potential to reduce travel costs, expedite communication, and enhance collaborative efforts. Among the latest developments in this field is mixed-reality telepresence, which holds significant promise. However, it has the formidable challenge of achieving strong, lifelike, and rapid 3D object reconstruction. Many current approaches emphasize real-time applications in the realm of cutting-edge technologies, particularly in the realm of immersive reality. However, these approaches often prioritize the synthesis of intermediary perspectives for specific viewing angles, rather than the comprehensive generation of complete 3D models. Introducing our ground-breaking system, aptly named "Holographic," we present a revolutionary end-to-end solution tailored for augmented and virtual reality telepresence. Our system seamlessly demonstrates the instantaneous creation of top-tier 3D models encompassing entire environments, including individuals and objects, using a sophisticated array of state-of-the-art depth cameras. What sets our innovation apart is the ability to transmit these highly detailed 3D models to remote users in real time. This enables individuals to wear virtual or augmented reality headsets to perceive, hear, and interact with remote participants in a three-dimensional space, creating an experience akin to a physical presence in the same location. This paradigm shift in audio-visual communication brings us closer to the immersive and authentic nature of face-to-face interaction. In this study, we provide a comprehensive and detailed overview of the holographic technical system, shedding light on its key interactive capabilities, the diverse range of potential application scenarios it unlocks, and initial findings from a qualitative study that delves into the unique communication medium it offers. Our primary focus is on delivering real-time 3D reconstruction that ensures the creation of accurate and realistic representations for users, paving the way for real-time holographic telepresence.

Keywords: Holographic Telepresence, 3D reconstruction, Real-time 3D reconstructions, Audio-visual perspective, face-to-face communication.

Introduction

There is a need for advanced communication technologies owing to rapid globalization. Currently, the most accessible and widely used video conferencing systems are Skype, Google Hangouts, ZOOM, and Cisco Webex, which are used by various companies and individuals. However, these systems also have several limitations. In general, webcams on laptops

or desktops are used for video conferencing. Nonetheless, these systems limit the visual representation of the participant, displaying a cropped and resized two-dimensional image on a screen. This image typically shows a person's face and only a few select body parts. This may hinder effective communication or create a communication gap, as many nonverbal gestures are not visible, although this is an essential part of human communication. Furthermore, the lack of visual continuity between the remote scene background presented on the screen and the immediate physical environment diminishes the sense of shared presence.

Holographic display technology utilizes light diffraction method to create virtual 3D images and objects in volumetric space. 3D holographic telepresence market has been growing very fast and expected to grow from USD 2.8 billion in 2021 to USD 5.7 billion by 2025, at a CAGR of 19.8%. The growing demand in medical, education, communication sector, financial sector etc are some of the factors responsible for the rise of this technology. Holography permits the quantitative measurement of different shapes and movements of objects by recording holograms through advanced digital cameras and thus reconstructing them numerically.[5]

In our current research endeavour, we employ an RGB-D sensor to capture images of individuals and convert these into real-time 3D representations. This innovative approach enhances telepresence, facilitating more engaging communication between the speaker and their audience. Our primary objective is to cultivate a profoundly captivating communication experience that fosters a stronger sense of mutual presence among participants situated in geographically separated locations. Ultimately, our research endeavours to conjure the illusion of a distant individual seamlessly inhabiting the local environment, as illustrated in Figure 1.

In our current research, we've introduced a novel method designed to efficiently create multiple textured meshes from RGB-D data streams. The existing technology faces a notable challenge when dealing with RGB-D images generated by multiple sensors. This challenge arises from the extraction of additional features from various perspectives, which often leads to delays in the 3D reconstruction process, especially in scenarios involving continuous viewpoint changes. However, by enhancing the speed of feature extraction across multiple sensors and achieving real-time frame rates, our approach aims to bolster the robustness and overall quality of 3D telepresence systems.

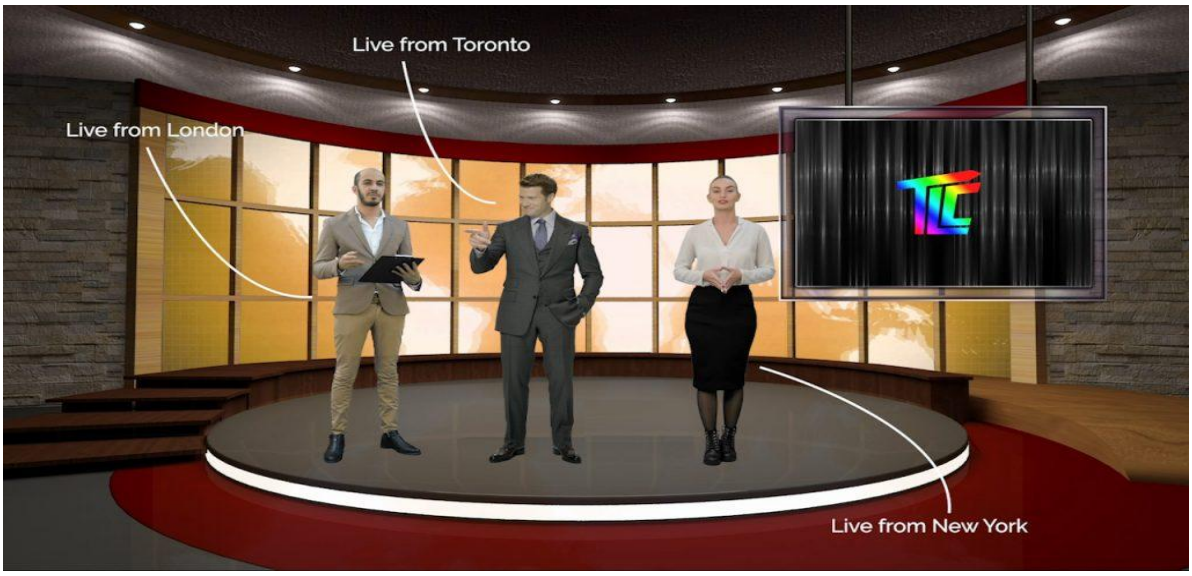

Figure 1 The 3D telecommunication system creates the illusion of a remote person being present in the local environment.

Depth processing plays a pivotal role in transforming images of objects taken from various angles into immersive 3D representations. This technique involves the meticulous registration and alignment of points from different angles to reconstruct a comprehensive 3D environment. Presently, cutting-edge devices have the capability to generate precise depth data using two primary methods: stereo cameras and structured light sensors. A stereo camera, a notable device in this context, is equipped with two or more lenses, each paired with its own image sensor. This setup emulates the human binocular vision, allowing the camera to capture 3D images effectively. To further enhance its depth-capturing abilities,

modern stereo cameras often incorporate Infrared (IR) sensors. Stereo cameras are equipped with sensors that generate crucial RGB-D data, which plays a pivotal role in performing real-time depth estimation. This process kicks off with the camera's initialization, a phase during which the hardware is linked to a computer. Subsequently, computer vision algorithms are employed to establish communication between the camera and the PC. The next step involves initiating camera streaming and initializing the post-processing stage. During post-processing, a method is applied to correlate the colour image with depth data, ultimately producing a 3D output display, as illustrated in Figure 4.

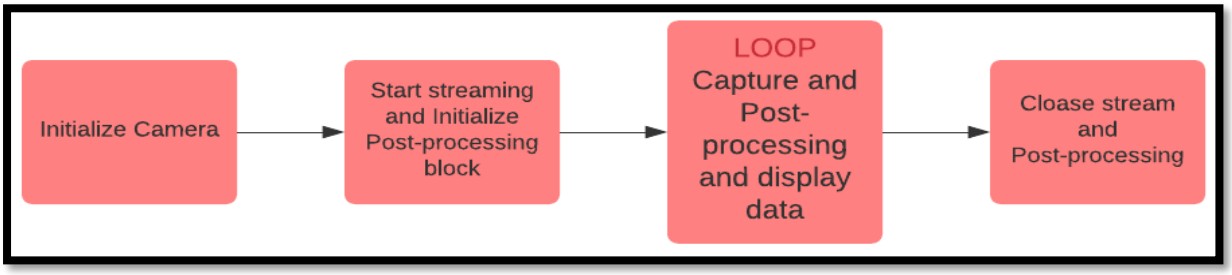

Figure 2 (a) and Figure 2 (c) depict 3D depth data captured from a unique perspective, while Figure 2 (b) and Figure 2 (d) showcase 3D data obtained by combining colour and depth information from that same distinctive viewpoint.

In Figure 3(a) and Figure 3(c), you can observe 3D depth data that has been acquired from a distinct perspective. Conversely, in Figure 3(b) and Figure 3(d), we present 3D data that has been generated by merging both colour and depth information gathered from the very same distinctive vantage point.

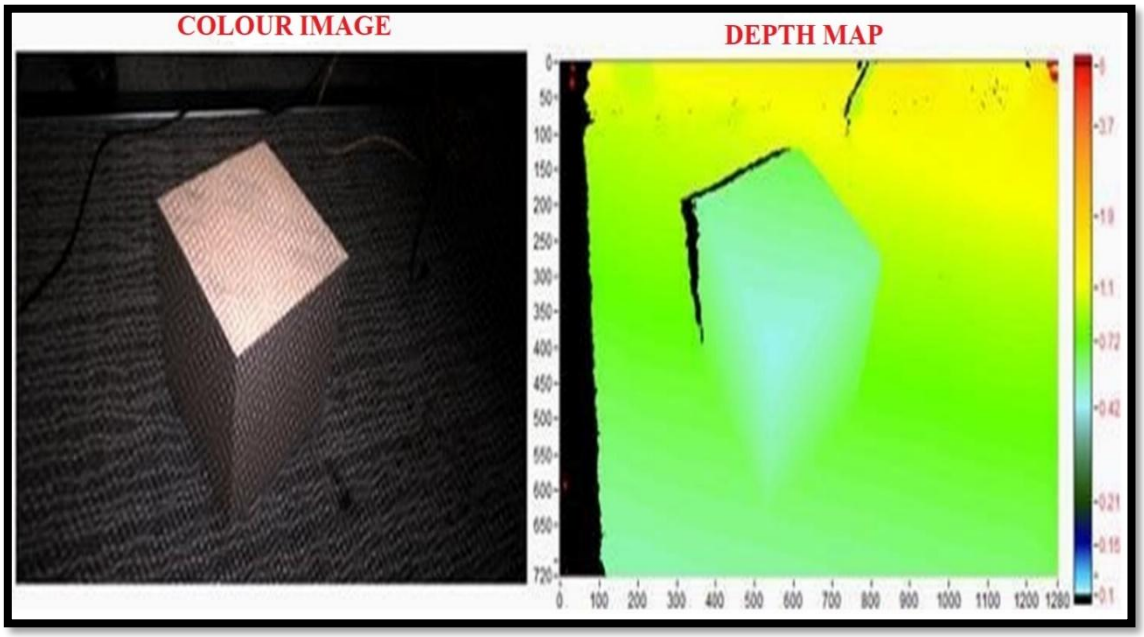

Figure 2 The input of colour image and a depth map from an RGB-D camera.

The initial stage in this module involves obtaining depth streams, a task that necessitates meticulous calibration of both intrinsic and extrinsic factors to calculate the camera parameters accurately. Following this, a standard calibration procedure is executed to ensure uniform and consistent colour information across all RGB cameras. This involves individually adjusting the white balance by utilizing a colour calibration chart, as illustrated in Figure 3. After meticulously fine-tuning the colours for each RGB camera individually, a particular RGB camera is designated as the reference point. The other cameras are then calibrated to align with this reference using a linear mapping technique. This meticulous

procedure guarantees uniform signal consistency across all RGB cameras. It is crucial to note that this calibration process takes place offline, and the resulting linear mapping adjustments are later implemented during runtime.

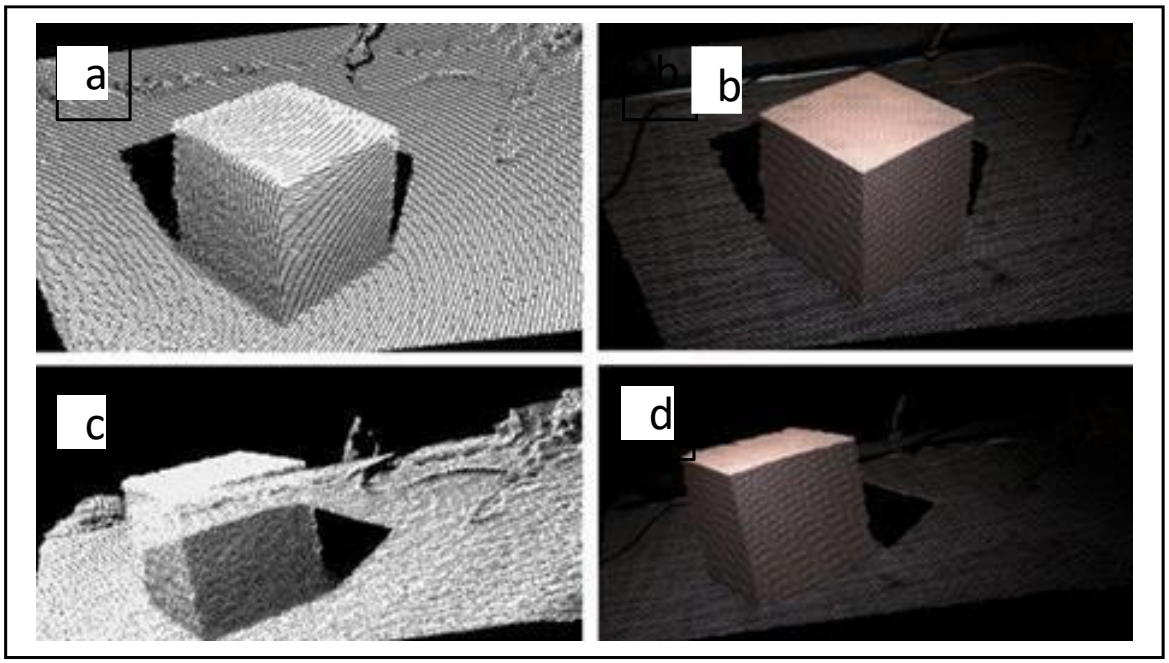

Figure 3 Results of the colour image and depth data

The process begins with a continuous loop, running until the user decides to stop the method. Following this, the employed system employs cloud rendering to capture the user's physical body and process it into the appliance, thereby enabling user telepresence. This concept is illustrated in Figure 5 (a) and Figure 5 (b), where human presence is maintained within a real environment, even if individuals are located at different places. In our research, we have developed a 3D reconstruction method aimed at efficiently compressing the data transmitted over the network. Importantly, we refrain from using green screens in our setups to ensure that the system can function seamlessly in natural environments, reflecting real-world scenarios. Our system relies on depth cameras to capture the user's perspective and enhances it by integrating virtual objects. Hossain and colleagues proposed a hybrid FCM-CLSTM model for efficient brain MRI segmentation. The method integrates threshold and region-based techniques to enhance accuracy while reducing noise sensitivity, computational complexity, and overfitting problems found in traditional approaches.

Figure 4 Results of the point cloud rendering process

4.3 Expected Result

To achieve a comprehensive 360° view, we employ a system comprising eight camera pods strategically positioned around the room's periphery, all directed inwards to capture distinct perspectives of the subject. Each of these pods comprises a robust setup consisting of two Near-Infrared (NIR) cameras, complemented by a colour camera securely mounted atop an optical bench. The stability of this setup is ensured to maintain precision.

For generating a pseudo-random pattern, we incorporate a diffractive optical element along with a laser. To filter out the visible light spectrum, we integrate NIR filters into the system.
Our camera setup incorporates a cluster of 24 Grasshopper Point Grey cameras, each capable of producing a synchronized pair of streams: one in RGB color and the other in depth. These streams are meticulously aligned for accuracy, employing sophisticated stereo matching techniques. The cameras operate at a consistent frame rate of 30 frames per second (fps), capturing high-resolution 4MP images. This synchronization is achieved through an external trigger system, ensuring precise coordination among all camera pods.

The initial step of this process involves generating depth streams, which necessitates thorough intrinsic and extrinsic calibration of the cameras. We employ the calibration method proposed by Zhang et al. (2000) to compute these vital camera parameters. Furthermore, we conduct a standard calibration procedure to ensure consistent and uniform colour information across all the RGB cameras.
This process involves performing individual white balance adjustments using a colour calibration chart. Afterward, one of the RGB cameras is chosen as the reference, and the remaining cameras are calibrated to match this reference through linear mapping. This calibration procedure is carried out offline, and during runtime, linear mapping is applied to ensure consistent signals across all the RGB cameras. This seamless operation is essential for maintaining uniformity in the system's performance.

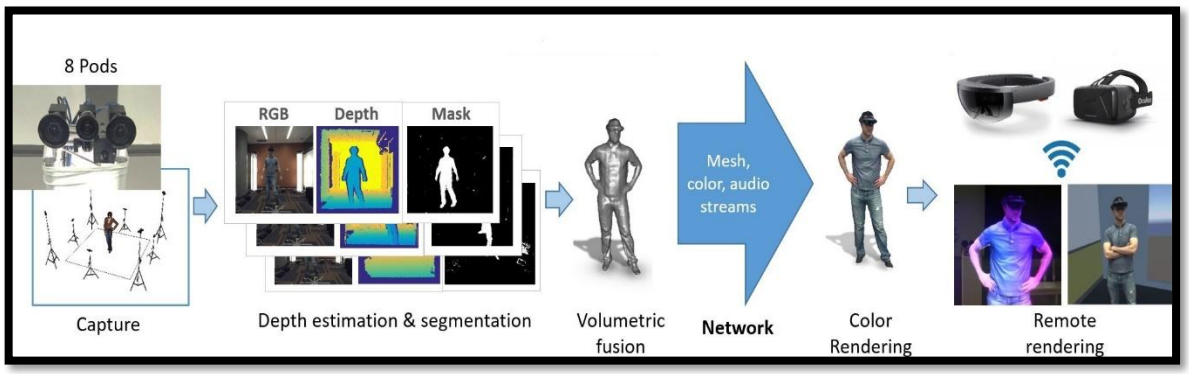

Figure 5 Depth estimation for a proposed 3D reconstruction method

The next phase involves depth estimation using RGB stereo technology, which excels in providing real-time depth information. This process relies on a pair of rectified images, where passive stereo techniques are employed to calculate depth by meticulously matching patches of pixels in one image with their corresponding counterparts in the other image. However, it's worth noting that this method may struggle to estimate depth accurately on texture-less surfaces. To overcome this limitation, active stereo systems are utilized for depth estimation. In this configuration, each camera setup comprises two Near-Infrared (NIR) cameras and one or more random Infrared (IR) dot pattern projectors. The depth estimation employs the Patch Match stereo approach, which is known for producing high-quality dense depth maps. Patch Match stereo employs a randomized correspondence algorithm that alternates between generating random depth values and propagating depth information. To achieve real-time performance, it assumes front-parallel windows and reduces the number of depth propagation iterations. Figure 7 provides examples of the resulting depth maps.

Figure 6 Top: RGB stream. Bottom: depth stream

Following the depth estimation process, the next step involves segmentation, which generates 2D profiles of the regions of interest. This segmentation step is of utmost importance as it serves a dual purpose: ensuring the temporal reliability of 3D reconstructions and compressing the data for efficient transmission over the network. The process begins by combining the gathered depth data, resulting in a cohesive volumetric representation. Subsequently, this volumetric data serves as the foundation for generating a polygonal 3D model, a task accomplished through the utilization of the marching cubes algorithm. Subsequently, this 3D model is textured using eight input RGB images. A straightforward texturing method is employed to compute the colour for each pixel by blending information from all eight RGB images (as illustrated in Figure 8).

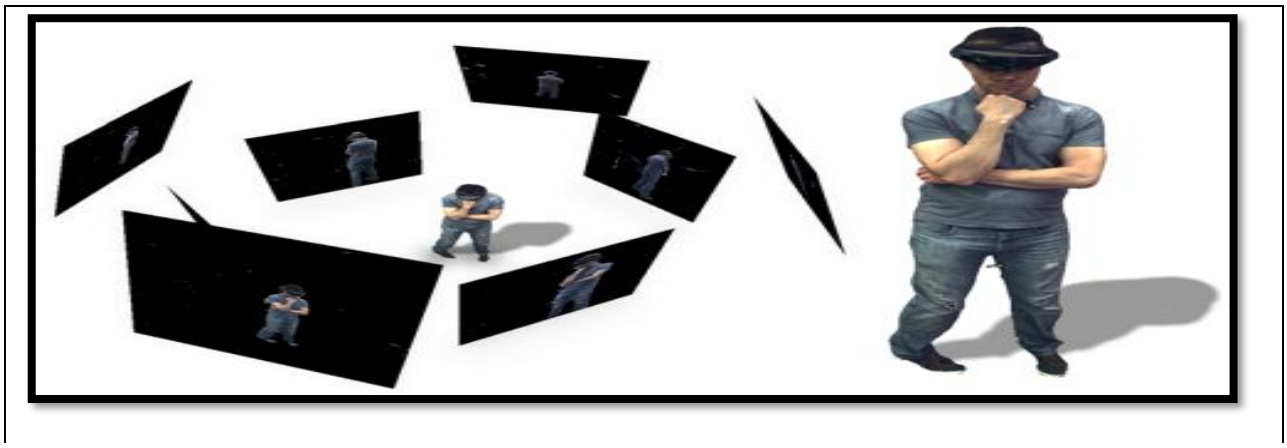

Figure 7 Projective Texturing: segmented-out RGB images and reconstructed colour.

Effective communication in a visual context relies heavily on synchronized audio output, ensuring a seamless and engaging experience. In the realm of holographic telepresence, precise alignment between visual and auditory cues is paramount. To achieve this, each remote audio source undergoes synchronization. This involves capturing audio from a remote user using an integrated monaural microphone. The captured audio samples are then divided into 20ms frames and seamlessly integrated with the user's head position data within their local room coordinate system. Subsequently, these audio frames, paired with relevant head position information, are transmitted to remote users through multiple individual connections. Notably, the orientation of a sound source in relation to the listener can often impact its amplitude. To address this, audio spatialization is employed through HRTF (Head-Related Transfer Function) audio processing, utilizing the XAUDIO2 framework within the Windows 10 operating system. This approach ensures an immersive and synchronized audio-visual experience in holographic telepresence.

Figure 8 Volumetric fusion with skeleton reconstruction in real-time

In order to effectively manage the substantial volume of data generated through the activities mentioned above, it becomes imperative to compress this data before transmitting it across a wide area network to a remote location for analysis. To ensure real-time interpretation with the highest level of quality, we employ a lightweight real-time compression method and a straightforward TCP-based wire format. This system is specifically designed to support 4-5 viewing clients connecting from multiple capture locations via a single 10 GBPS link, particularly in point-to-point teleconferencing scenarios. To achieve this goal, we transmit a standard triangle mesh enriched with additional color information captured from various camera viewpoints. The reduction in the size of raw data is achieved through various transformations applied to the per-frame results obtained during the capture and fusion stages.

The system captures and transmits audio and user position data to all remote participants, enabling the creation of spatial audio sources that match their precise locations, particularly for users utilizing AR or VR headsets. Each audio stream is single-channel, sampled at 11 KHz, and recorded in 16-bit PCM format. This results in a broadcast stream with a total bandwidth of 176 Kbps, complemented by an additional 9.6 Kbps for pose data. The audio data is buffered to ensure smooth playback at the receiving end. Communication between distant participants occurs bidirectionally and independently, with each pair of users exchanging audio and pose data. The communication system operates on a peer-to-peer basis, with headsets directly connecting with one another. It's crucial to note that this audio communication system is entirely distinct from the visual communication system.

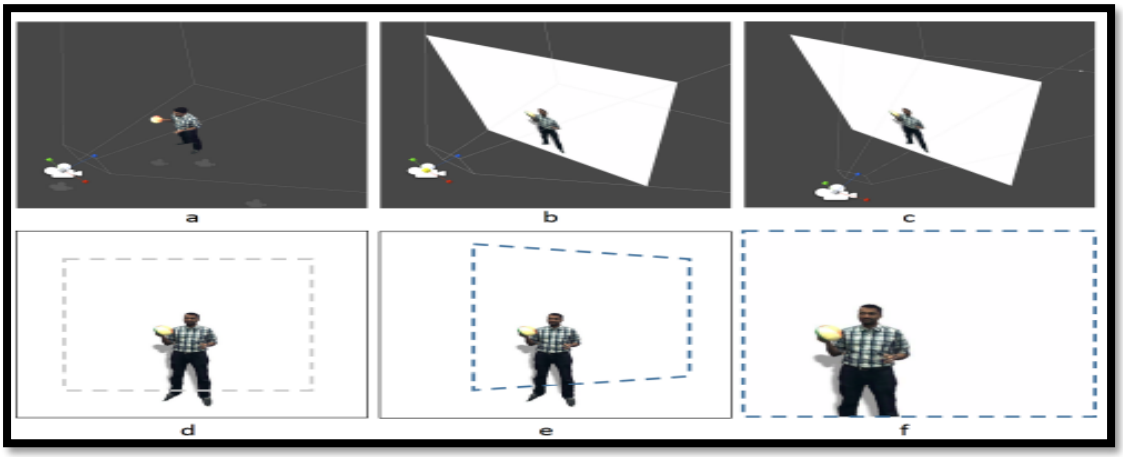

Figure 9 Rendering: (a) Rendering predicted pose on PC, (b) Image encoding through rendering, (c) Rendering with actual pose on HMD, (d) Predicted view on PC, along with its over-rendering with enlarged FOV, (e) Pixels in decoded video stream, (f) Re-projected actual view.

From the user's perspective, creating a detailed and immersive 3D model on a standalone VR or AR headset like HoloLens can be both costly and detrimental to the overall experience. This is primarily due to the increased delay in rendering. To address this issue, we've implemented a solution that shifts the rendering workload to a dedicated desktop PC on the receiving end, a process known as remote rendering. This approach offers several advantages, including maintaining a consistent framerate, reducing perceived latency, and preserving the device's battery life. Additionally, it allows us to leverage high-end rendering capabilities that may not be readily available on mobile GPUs.

In our setup, the rendering PC is connected to the untethered HMD via WiFi. The HMD continually transmits the user's six degrees of freedom (6DoF) pose data to the rendering PC. At the time of rendering, the system predicts the headset's pose and generates the scene for each eye accordingly (as illustrated in Fig. 10a). Subsequently, these rendered scenes are encoded into a file (as depicted in Fig. 10b) and transmitted, along with the corresponding poses, back to the HMD. On the HMD side, the received video streams are interpreted and displayed for each eye as textured quads, positioned based on the predicted rendered pose (as shown in Fig. 10c). These images are then re-projected to align with the most up-to-date user pose (as seen in Fig. 10f).

To mitigate potential inaccuracies in pose prediction and account for any latency between the PC and HMD, we have implemented a speculative rendering system on the desktop side. This system allows us to dynamically adjust the rendered content based on the user's predicted pose. Furthermore, to address issues related to orientation mispredictions, we have adopted a strategy of rendering a wider field of view (FoV) that centers around the anticipated user direction, as illustrated in Figure 10d. It's worth noting that we have observed a minor rotation misprediction when the HMD renders the textured quad within the actual display field of view, as depicted in Figure 10e. However, we effectively manage positional mispredictions through the use of view interpolation techniques.

Conclusion

We have introduced Holoportation, a comprehensive system designed for the seamless real-time capture, transmission, and rendering of people, environments, and objects within fully immersive 3D settings. This ground-breaking technology empowers users to not only see and hear but also engage with distant colleagues in three-dimensional space through innovative capture techniques paired with mixed reality displays. The result is an audio-visual experience that closely simulates physical presence, revolutionizing how we connect and collaborate. Our demonstrations have showcased various interactive scenarios, including one-on-one communication and one-to-many broadcasts, enabling live, real-time interactions. Furthermore, our technology has the capability to capture and recreate vivid "living memories." We expect that both practitioners and researchers will explore the vast potential of this live 3D capture technology, opening up new horizons for its applications.

References:

[1] Anjos, R. K. D., Sousa, M., Mendes, D., Medeiros, D., Billinghurst, M., Anslow, C., & Jorge, J. (2019, November). Adventures in Hologram Space: Exploring the Design Space of Eye-to-eye Volumetric Telepresence. In 25th ACM Symposium on Virtual Reality Software and Technology (pp. 1-5).

[2] Bove, V. M. (2011). Engineering for live holographic tv. SMPTE Motion Imaging Journal, 120(8), 56-60.

[3] Collet, A., Chuang, M., Sweeney, P., Gillett, D., Evseev, D., Calabrese, D., Hoppe, H., Kirk, A., and Sullivan, S. High-quality streamable free-viewpoint video. ACM TOG 34, 4 (2015), 69.

[4] Dalvi AA, Siddavatam I, Dandekar NG, Patil AV (2015) 3D holographic projections using prism andhand gesture recognition. In: Proceedings of the 2015 international conference on advanced research incomputer science engineering & technology (ICARCSET 2015). ACM, p 1

[5] RamMohan Reddy Kundavaram. (2022). Cloud-Based Data Analysis Identifying Key Financial Influencers at Scale. European Economic Letters (EEL), 12(2), 234–241. Retrieved from https://www.eelet.org.uk/index.php/journal/article/view/3389

[6] Elmorshidy, A. (2010). Holographic projection technology: the world is changing. arXiv preprint arXiv:1006.0846.

[7] FRAeS, B., & HFAvn, L. (2020). Holographic Telepresence and Collaborative Learning Platforms: The Future Reality of Aviation Education.

[8] Fairchild, A. J., Campion, S. P., García, A. S., Wolff, R., Fernando, T., & Roberts, D.

[9] J. (2016). A mixed reality telepresence system for collaborative space operation. IEEE Transactions on Circuits and Systems for Video Technology, 27(4), 814-827.

[10] Hossain, M., et al. (2025). An efficient threshold and region-based approach. Nanotechnology Perceptions, 21(1), 118–143.

[11] Kwon, O. H., Koo, S. Y., Kim, Y. G., & Kwon, D. S. (2010, October). Telepresence robot system for English tutoring. In 2010 ieee workshop on advanced robotics and its social impacts (pp. 152-155). IEEE.

[12] Li, N., & Lefevre, D. (2020). Holographic teaching presence: participant experiences of interactive synchronous seminars delivered via holographic videoconferencing. Research in Learning Technology, 28.

[13] Lu, X., Shen, J., Perugini, S., & Yang, J. (2015, December). An immersive telepresence system using rgb-d sensors and head mounted display. In 2015 IEEE International Symposium on Multimedia (ISM) (pp. 453-458). IEEE.

[14] Luevano, L., de Lara, E. L., & Quintero, H. (2019). Professor Avatar Holographic Telepresence Model. In Holographic Materials and Applications. IntechOpen.

[15] Marques, L. F., Tenedório, J. A., Burns, M., Romão, T., Birra, F., Marques, J., & Pires, A. (2017). Cultural Heritage 3D Modelling and visualisation within an Augmented Reality Environment, based on Geographic Information Technologies and mobile platforms.

[16] 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. In ISMAR (Vol. 11, No. 2011, pp. 127-136).

[17] Pates, D. (2020). The Holographic Academic: Rethinking Telepresence in Higher Education. In Emerging Technologies and Pedagogies in the Curriculum (pp. 215-230). Springer, Singapore.

[18] Zhao, M., Tan, F., Fu, C. W., Tang, C. K., Cai, J., & Cham, T. J. (2013, July). High-quality Kinect depth filtering for real-time 3D telepresence. In 2013 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1-6). IEEE.