Asymmetric MR Telecommunication System Using 3D Gaussian Splatting: Implementation and Technical Performance Evaluation

Takashi Ota
ota@cyber.t.u-tokyo.ac.jp
The University of Tokyo
Tokyo, Japan

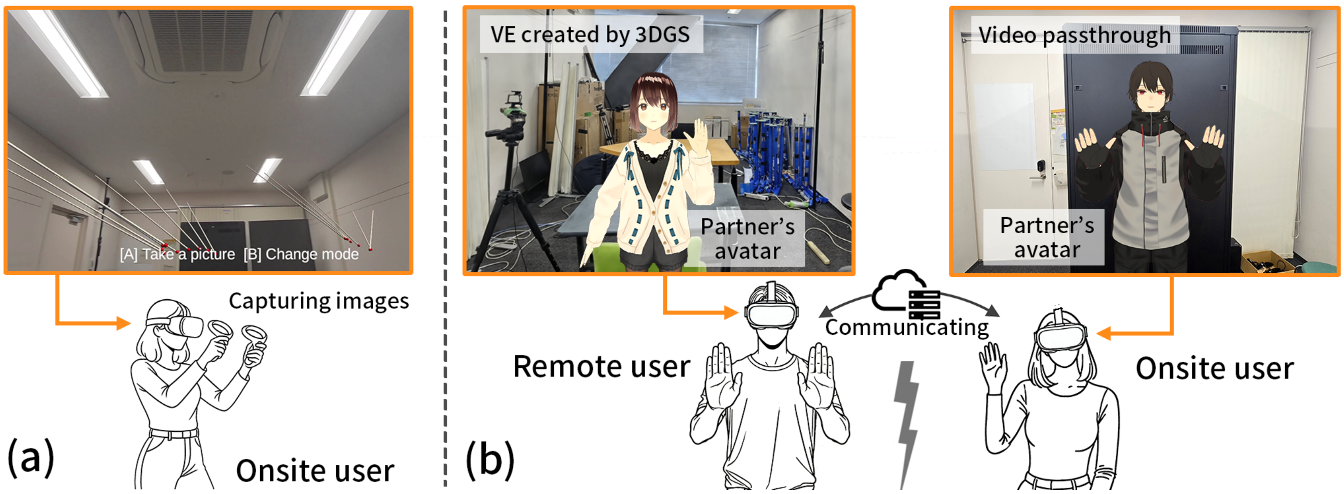

Figure 1: Concept of the proposed method. (a) Onsite user captures multiple images with an HMD. (b) The proposed system supports telecommunication between remote and onsite users.

Abstract

Telecommunication within virtual environments created through 3D reconstruction techniques is a significant research topic within Virtual Reality (VR), as it offers users an experience closely mirroring real-world interactions. We anticipated that by integrating rapid and easily prepared 3D reconstruction methods in such telecommunication systems, users would be able to engage in realistic communication with individuals in distant locations, even while traveling.

In this study, we propose an asymmetric Mixed Reality (MR) telecommunication system using 3D Gaussian Splatting and provide an implementation example. A user study was conducted to evaluate the system's effectiveness from the perspectives of processing time for 3D reconstruction, presence, usability, and incidence of simulator sickness.

UbiComp / ISWC '25 Workshops, Espoo, Finland

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

CCS Concepts

• **Computing methodologies** → **Virtual reality**; **Mixed / augmented reality**; *Computer graphics*; • **Human-centered computing** → *Collaborative and social computing*.

Keywords

3D Gaussian Splatting; Remote Communication; Mixed Reality

ACM Reference Format:
Takashi Ota. 2025. Asymmetric MR Telecommunication System Using 3D Gaussian Splatting: Implementation and Technical Performance Evaluation. In *Proceedings of UbiComp / ISWC 2025 Workshops (UbiComp / ISWC '25 Workshops)*. ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Synchronous telecommunication has become indispensable across various domains, including business, entertainment, and healthcare [7]. Compared to asynchronous communication like text chat, synchronous communication is considered crucial for fostering bonds between users due to its immediacy and greater capacity for information transfer, leading to extensive research into various interfaces. In the fields of Virtual Reality (VR) and Augmented Reality (AR), applications of physically embodied and immersed telecommunication are being investigated, utilizing avatar embodiment and visual presentation via head-mounted displays (HMDs).

Among these, telecommunication using virtual environments created through 3D reconstruction technology is a significant research subject, as it offers users experiences unique to environments that closely reflect real-world settings. In such telecommunication, it is anticipated that utilizing 3D reconstruction methods that are both fast and easy to set up will enable users to engage in communication with remote partners, within virtual environments that accurately reflect real-world settings, even while traveling.

In recent years, advancements in machine learning algorithms and improvements in graphics hardware performance have led to the development of various 3D reconstruction techniques. For instance, notable examples include Neural Radiance Fields (NeRF) [8], a 3D reconstruction method that leverages neural networks, and 3D Gaussian Splatting (3DGS) [5], which represents 3D scenes as a superposition of Gaussian distributions. These methods are characterized by their ability to render fine details and complex lighting, distinguishing them from classical 3D reconstruction techniques.

This paper focuses on 3DGS, particularly its high-quality rendering capabilities combined with fast training times. Possessing these features, 3DGS enables scene capture in mobile settings and the sharing of 3D scenes on the go. We hypothesize that using 3DGS for instant 3D reconstruction could enable users to engage in communication with remote partners, within virtual environments that closely reflect real-world settings, even while traveling. In this study, we implemented an asymmetric Mixed Reality (MR) telecommunication system utilizing 3DGS. Furthermore, we validated the practical feasibility of our system by conducting a user study to evaluate processing time for 3D reconstruction, presence, usability, and simulator sickness within the proposed system.

2 Proposed method

This section outlines the proposed system. The system implementation and detailed explanation are available on GitHub[1].

2.1 System Overview

The proposed system supports telecommunication between an onsite and a remote user. First, the onsite user captures multiple images of the surrounding environment using the external camera of an HMD. These images are then transmitted to the remote side, where a 3D scene is reconstructed using 3DGS. After some configuration (e.g., coordinate alignment) by the remote user, both users wear HMDs to begin communication. On the onsite user's HMD, the physical environment is displayed via video passthrough, with the remote user's avatar superimposed over it. Conversely, the remote user's HMD displays a 3DGS-reconstructed virtual environment and the onsite user's avatar. This setup enables telecommunication in virtual environments that reflect real-world settings.

Image capture is implemented using the Meta Quest passthrough API, allowing the onsite user to capture the environment and engage in communication without additional equipment beyond the HMD. This system can display anchors at locations where users have captured images. This function helps users to capture images from various positions and angles, ensuring comprehensive data collection. Captured images are transmitted to a server via HTTPS and stored in the cloud. The poses of the users' HMDs and controllers are transmitted via UDP communication, enabling real-time synchronization of each user's avatar posture. To integrate the 3DGS-reconstructed scene into a Unity-based VR application, we utilized the publicly available implementation by Kleinbeck et al. [6].

Figure 2: VR view of remote users.

2.2 Comparison with Related Work

The proposed method follows the trajectory of research on asymmetric MR telecommunication systems. Here, an asymmetric MR telecommunication system is defined as one in which one user experiences a virtual environment reflecting the physical world along with their partner's avatar in VR, while the other sees their partner's avatar overlaid on the real world via AR. Prior work has explored approaches that use dedicated camera arrays [11] or 360-degree cameras [12][2] to capture the physical environment for 3D reconstruction. However, these methods often require specialized equipment, making them less accessible and more challenging to prepare. In contrast, the proposed method only requires an HMD with an external camera for the remote user. In our implementation, we employed Meta Quest 3. Thus, the present system is advantageous in that it is suitable for use in various outdoor settings. However, it should be noted that the time required for virtual scene reconstruction remains a topic for further discussion. In the next section, we present the results of an experiment evaluating the system's feasibility, including its 3D reconstruction duration.

3 Experiment

In this experiment, we evaluated the 3D reconstruction duration, presence, usability and simulator sickness of the proposed system to verify its effectiveness in 3D reconstruction. This experiment was conducted with the approval of our affiliated institution's ethics review board (Approval No.: UT-IST-RE-240819).

3.1 Method

Six participants were recruited for this experiment. Before the main experiment, participants completed the Simulator Sickness Questionnaire (SSQ) [4] to assess their baseline simulator sickness levels. First, participants acted as on-site users, freely capturing images within the laboratory for five minutes using the proposed system.

[1] https://github.com/t-ota0407/tele-sync

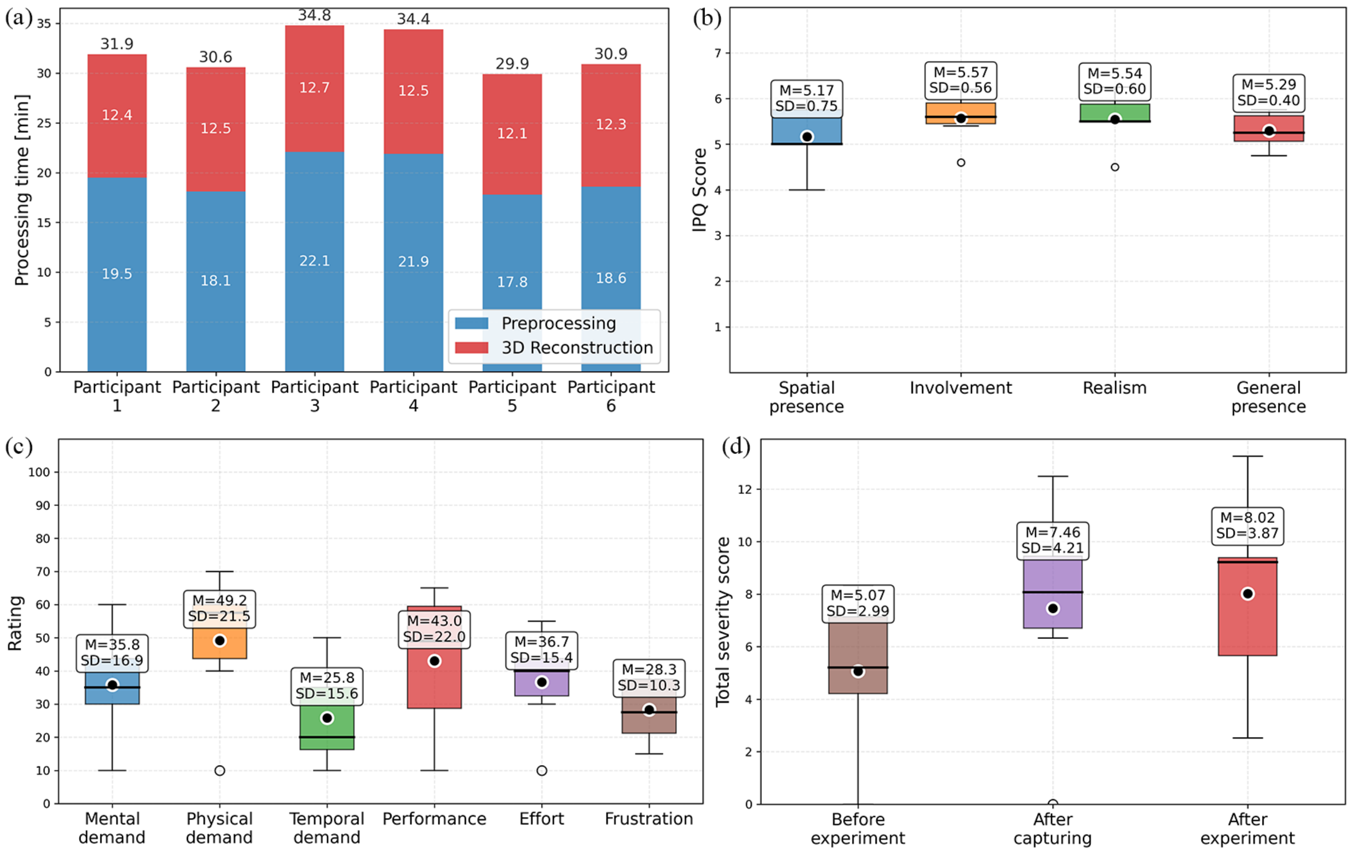

Figure 3: (a) Processing time for 3D reconstruction. (b) Presence measured by IPQ. (c) Usability measured by NASA-TLX. (d) Simulator sickness score. In the box plots, black circles represent the mean. Additionally, outliers, indicated by white circles, are data points located more than 1.5 times the interquartile range away from either the first or third quartile.

They received a brief explanation of 3DGS beforehand and were instructed that capturing images from various positions and angles was desirable for successful 3D reconstruction. As shown in the top-left photo of Figure 1, the system displayed captured points with Quest Passthrough API. Therefore, participants could take multiple images from various perspedctives by using this function as needed. Subsequently, a 3D scene was reconstructed from the captured dataset using 3DGS, and the time taken for this process was measured. Next, participants assumed the role of remote users and communicated with an experimenter from a separate laboratory using the system. Figure 2 shows the VR view for the remote user. Finally, participants completed the Igroup Presence Questionnaire (IPQ) [3] to evaluate presence, the NASA Task Load Index (NASA-TLX) [1] to assess usability, and the SSQ again. They also participated in a short interview about their experience with the system.

The computational resources used in the 3D reconstruction consisted of a machine with the following specifications: CPU: Intel Core i9-10900X, GPU: NVIDIA RTX A6000.

3.2 Results and Discussion

3.2.1 Processing time for 3D reconstruction. Figure 3 (a) shows the time required for 3D scene reconstruction using the collected datasets. The 3D reconstruction process using 3DGS generally consists of two steps: preprocessing (camera pose estimation with COLMAP [9]) and 3DGS training. Therefore, we measured the processing time for each step. The average time required for preprocessing was 19.7 minutes, and the average time for 3DGS training was 12.4 minutes. The total latency for 3D reconstruction across the entire system was 32.1 minutes. This result indicates that the duration is practical enough for a user to capture a scene at a travel destination and wait for the 3D reconstruction to finish while relaxing at a cafe.

3.2.2 Presence. Figure 3 (b) illustrates the presence scores measured by the IPQ. The mean General Presence score was 5.29, Spatial Presence was 5.17, Involvement was 5.57, and Realism was 5.54. Therefore, it was confirmed that the proposed system provides a sufficiently high level of presence.

We compared the mean scores for each metric with values reported in prior research on asymmetric MR communication systems using 360-degree cameras [12]. The Realism score in this study was slightly lower. This may be because, in some user trials, parts of the environment not well-captured in the user's input image (such as the ceiling) appeared dark, leading to a less realistic impression for users. Conversely, all metrics other than Realism were largely consistent with those reported in the prior study, suggesting that the proposed method can offer a very high level of presence with a simple setup.

3.2.3 Usability. Figure 3 (c) illustrates the ratings measured by NASA-TLX. Among the task load indices, Physical Demand, Performance, and Effort showed relatively high values. The high ratings for Physical Demand and Effort are likely due to the need to walk around the room when capturing real-world environments. The high Performance rating is probably because users were concerned about the quality of the captured data when photographing real-world environments. Indeed, many interview comments indicated that users made an effort to capture images from various positions as instructed when photographing the real-world environment.

3.2.4 Simulator sickness. Figure 3 (d) displays the total severity score of the SSQ. Compared to pre-experiment levels, scores were slightly elevated after the real-world shooting task and at the end of the entire experiment. However, no users exceeded a score of 20, which is considered a significant criterion for VR sickness [10], neither after the real-world shooting task nor at the conclusion of the experiment. This suggests that the present system does not induce excessive VR sickness.

4 Limitation and Future work

All experiments in this study were conducted indoors. However, since outdoor use is anticipated, as explained at the outset, we plan to conduct experiments outdoors to demonstrate the system's effectiveness. Additionally, the current experiment had too few participants for a robust statistical comparison. Therefore, we plan to conduct the outdoor experiments with a sufficient number of participants.

In the current experiment, COLMAP was used for camera pose estimation. However, with the system configuration proposed in this study, it may be possible to skip camera pose estimation for captured images by substituting Quest 3's self-localization information as camera pose data. If realized, this would significantly reduce the preprocessing time required for 3D reconstruction. In the current experiment, the average preprocessing time was 19.7 minutes, and the average training time for 3DGS was 12.4 minutes. Thus, skipping preprocessing would considerably shorten the overall waiting time for 3D reconstruction. Therefore, we plan to investigate methods for substituting Quest 3's self-localization information for camera pose information in future work.

5 Conclusion

In this study, we proposed an asymmetric MR telecommunication system that leverages 3DGS for virtual environment construction, based on the idea that incorporating 3DGS could enable the practical use of such systems in mobile settings. We also presented an implementation example of the proposed system. Through the user study, we confirmed that 3D reconstruction can be completed within approximately 30 minutes, demonstrating the potential effectiveness of the system. In the future, substituting Quest 3's self-localization information for camera pose data is projected to lead to a substantial reduction in 3D reconstruction processing time. This system is expected to considerably facilitate communication in virtual environments that closely reflect real-world conditions, even when users are traveling.

Acknowledgments

This study was supported by JSPS KAKENHI (24KJ0926) and a research grant from the Tateisi Science and Technology Foundation.

References

[1] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload* 1, 3 (1988), 139–183.

[2] Xincheng Huang, Michael Yin, Ziyi Xia, and Robert Xiao. 2024. VirtualNexus: Enhancing 360-Degree Video AR/VR Collaboration with Environment Cutouts and Virtual Replicas. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 55, 12 pages. doi:10.1145/3654777.3676377

[3] igroup.org. 2001. igroup presence questionnaire (ipq). https://www.igroup.org/pq/ipq/index.php

[4] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology* 3, 3 (1993), 203–220. doi:10.1207/s15327108ijap0303_3

[5] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics* 42, 4 (July 2023). https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/

[6] Constantin Kleinbeck, Hannah Schieber, Klaus Engel, Ralf Gutjahr, and Daniel Roth. 2025. Multi-Layer Gaussian Splatting for Immersive Anatomy Visualization. *IEEE Transactions on Visualization and Computer Graphics* (2025), 1–11. doi:10.1109/TVCG.2025.3549882

[7] Nick Majedi, Muhammad Naeem, and Alagan Anpalagan. 2016. Telecommunication integration in e-healthcare: technologies, applications and challenges. *Transactions on Emerging Telecommunications Technologies* 27, 6 (2016), 775–789.

[8] Ben Mildenhall, Pratul P. Srinivasan, and Matthew et al. Tancik. 2021. NeRF: representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 1 (dec 2021), 99–106. doi:10.1145/3503250

[9] Johannes L. Schönberger and Jan-Michael Frahm. 2016. Structure-from-Motion Revisited. In *2016 IEEE CVPR*. 4104–4113. doi:10.1109/CVPR.2016.445

[10] Kay M Stanney, Robert S Kennedy, and Julie M Drexler. 1997. Cybersickness is not simulator sickness. In *Proceedings of the Human Factors and Ergonomics Society annual meeting*, Vol. 41. SAGE Publications Sage CA: Los Angeles, CA, 1138–1142.

[11] Theophilus Teo, Ashkan F. Hayati, Gun A. Lee, Mark Billinghurst, and Matt Adcock. 2019. A Technique for Mixed Reality Remote Collaboration using 360 Panoramas in 3D Reconstructed Scenes. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology* (Parramatta, NSW, Australia) *(VRST '19)*. Association for Computing Machinery, New York, NY, USA, Article 23, 11 pages. doi:10.1145/3359996.3364238

[12] Faisal Zaman, Craig Anslow, Andrew Chalmers, and Taehyun Rhee. 2023. MRMAC: Mixed Reality Multi-user Asymmetric Collaboration. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 591–600. doi:10.1109/ISMAR59233.2023.00074

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009