QoE Modeling of Delay in 3D Tele-Immersion

| 1st Author Name  Affiliation  City, Country  e-mail address | 2nd Author Name  Affiliation  City, Country  e-mail address | 3rd Author Name  Affiliation  City, Country  e-mail address |
| --- | --- | --- |

# ABSTRACT

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

* ACM copyright: ACM holds the copyright on the work. This is the historical approach.
* License: The author(s) retain copyright, but ACM receives an exclusive publication license.
* Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single-spaced in Times New Roman 8-point font. Please do not change or modify the size of this text box.

Each submission will be assigned a DOI string to be included here.

## Author Keywords

Telepresence, Delay, Network Performance, QOE.

## ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous; See<http://acm.org/about/class/1998> for the full list of ACM classifiers. This section is required.

# INTRODUCTION

Communications technology plays an important role in human development. The invention of telephone made most remote communications instantaneous. From than on, more and more physical meetings were replaced by phone calls, which saves a great deal of time and money. Nowadays, video-mediated telecommunication is becoming popular. It provides convenience for teleconference [2, 3, 4], tele-collaboration [5, 6], presence remotely [7, 8, 9, 10], and so on.

Beyond that, researchers are also exploring telepresence with higher level of immersion. In the last decades, tele-immersion developed rapidly. Several 3D-reconstruction-based telepresence systems were born [11, 12, 13, 14, 15]. They aim at making up for the lack of eye contact, body language and physical presence in video-mediated telecommunications. Microsoft Research’ s Holoportation [15] was quite impressive. They presented an end-to-end system for AR/VR telepresence with high-quality, real-time reconstructions of an entire space. Because of their promising quality of service (QoS) and the fact that hardware devices are getting cheaper and more powerful, we believe that these systems will become practical in the near future.

However, previous works about tele-immersion bias to technical implementations. Only a few study works were conducted on 3D telepresence. Moreover, they either study on specific scenarios [18, 19, 20] or with pseudo-3D systems [2, 16, 17].

We argue that, fundamental studies on mapping quality of service (QoS) to quality of experience (QoE) in 3D tele-immersion is important. We have seen that the industrial standard of telephone contributes to its popularization, e.g. by avoiding network over-engineering [21]. In recent researches, the user experience (UX) studies of video-mediated telecommunications [2, 3, 4, 9, 17] are also helping its improvement. Similarly, an understanding of UX in 3D telepresence may well be helpful to both academic and industrial community.

In this paper, we focus on modeling the impact of delay, which is an important factor of QoS [5], in 3D tele-immersion. We first summarize suitable tasks from previous work. Then, we conducted a large online questionnaire (N=100) to introduce our systems, look for more candidate tasks and gather participants’ expectation. Last, we selected typical applications for our user studies.

In implementation, we do not follow the highest quality technique [22, 1] (2016) proposed by Microsoft Research, but achieve a more responsive system. Our kernel is similar to Maimone et al. ’s work [14] (2012). Supported by the recent progress of depth camera (RealSense-D435), GPU (Gtx1080 Ti) and VR device (HTC Vive), our frame rate reaches 40 FPS. Only one frame delay is necessary for transmission, so the delay is within 50ms. As several related works mentioned the importance of “shared objects” in 3D telepresence [19, ?], our system was designed to go around the common objects in both sides. Besides face-to-face telecommunications, our system provides an interactive process for non-professional users to easily set up objects-shared activities such as playing chess, piano duet and pair programming.

We have three main findings: first, some tasks with strong interaction, e.g. the finger-guessing game or piano duet, require low latency of 75ms. It breaks the “rule” in 2D telecommunication that 150ms is acceptable for most applications [5, 23]; second, participants’ expected latency of tasks based on comparison can well predict the actual needs; third, we argue that the latency requirement of a task depends on its “bottleneck”. For example, the bottleneck of most video-mediated telecommunications is *audio signals* [], which leads to an acceptable delay down to 150ms. A stronger bottleneck appears in our system as *synchronous gesture*, e.g. the gesture in the finger-guessing game. It requires a latency of 75ms.

# related work

## 3D Tele-Immersion

In order to implement a typical tele-immersion system for our studies, we first conducted a brief review of 3DTI technologies. Basically, a 3DTI system requires three processes: reconstruction, transmission and rendering [24]. We developed our reconstruction algorithm based on TSDF Volume [25] and Marching Cubes [26]. We use network line between computers for high-bandwidth transmission, but do not focus on the transmission part as [27, 19] did. In the studies, we simulated various network performance through software methods. We use head-mounted display (HTC Vive) and Unity3D engine to render 3D scenes.

Below are brief reviews of reconstruction and rendering technologies for 3DTI systems:

### 3D Reconstruction

In early works, researchers used an array of cameras to capture the dynamic scenes [28, 29]. For a given camera view, these systems create a polygonal model that will look correct. That is, they do not construct stand-alone 3D model from physical world.

TELEPORT [30] can composites video-textured surfaces within 3D geometric models. The only one camera limits its construction quality. In 2002 and 2003, researchers started to design immersive 3D video acquisition and rendering environment with multiple cameras [31, 32]. However, their 3D reconstruction output was only point cloud but not polygon mesh. In 2008, Kurillo et al. presented a framework for remote collaboration and training of physical activities [11]. This work tried a reconstruction method with triangulation, but only reached the frame rate of about 5-7 FPS. [12] and [33] for the first time presented compelling real-time reconstruction techniques with multiple cameras. However, the lack of depth dimension indicated their modeling with only silhouette boundaries.

Researchers have made great progress of 3DTI system in the last decade. Both the development of hardware and algorithm made contributions to the real-time performance of high-quality reconstruction. In October 2011, Maimone et al. presented a 3DTI system with Kinects [13]. They developed a pixel-based mesh generation algorithm and reached a frame rate of 30 FPS. This work was followed by Beck el al.’s group-to-group telepresence system [19]. In the same month, however, Microsoft introduced voxel-based [25, 26] system KinectFusion [34] and achieved a better reconstruction quality. Though the volumetric methods were invented about 30 years ago, developed depth cameras and GPUs made them practical. In the next year (2012), Maimone et al. also turned to the volumetric methods [14] for improved quality.

In 2016, Microsoft proposed reconstruction pipeline Fusion4D [22], which is highly robust to occlusions, large frame-to-frame motions and topology changes. “The fourth dimension” in this paper was the time dimension, indicating that it leverages the temporally coherence of physical scenes. In the same year, Microsoft integrated fusion4D into their 3DTI system Holoportation [1]. However, Fusion4D is extremely complex and not open-source. Even with expensive devices, Holoportation has an end-to-end latency of 60ms, which can not be ignored in our study. In this paper, we apply a 3D-reconstruction method similar to the one proposed by Maimone et al. [14]. It is a satisfactory system with high quality, responsive interaction and can be easily set up by commercial devices.

### 3D Rendering

## Studies in Telepresence

## QoE Measurement

# REFERENCES

1. Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., ... & Tankovich, V. (2016, October). Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (pp. 741-754). ACM.
2. Higuchi, K., Chen, Y., Chou, P. A., Zhang, Z., & Liu, Z. (2015, April). ImmerseBoard: Immersive telepresence experience using a digital whiteboard. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*(pp. 2383-2392). ACM.
3. Nemiroff, G. (1989). Beyond" talking heads": Towards an empowering pedagogy of women's studies. *Atlantis: Critical Studies in Gender, Culture & Social Justice*, *15*(1).
4. Marlow, J., Van Everdingen, E., & Avrahami, D. (2016, February). Taking Notes or Playing Games?: Understanding Multitasking in Video Communication. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 1726-1737). ACM.
5. Donovan, A., Alem, L., Huang, W., Liu, R., & Hedley, M. (2014, September). Understanding How Network Performance Affects User Experience of Remote Guidance. In *CYTED-RITOS International Workshop on Groupware* (pp. 1-12). Springer, Cham.
6. Avellino, I., Fleury, C., & Beaudouin-Lafon, M. (2015, April). Accuracy of deictic gestures to support telepresence on wall-sized displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2393-2396). ACM.
7. Nakanishi, H., Tanaka, K., & Wada, Y. (2014, April). Remote handshaking: touch enhances video-mediated social telepresence. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2143-2152). ACM.
8. Misawa, K., & Rekimoto, J. (2015, April). ChameleonMask: Embodied physical and social telepresence using human surrogates. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 401-411). ACM.
9. Rae, I., & Neustaedter, C. (2017, May). Robotic Telepresence at Scale. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 313-324). ACM.
10. Neustaedter, C., Venolia, G., Procyk, J., & Hawkins, D. (2016, February). To Beam or not to Beam: A study of remote telepresence attendance at an academic conference. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 418-431). ACM.
11. Kurillo, G., Bajcsy, R., Nahrsted, K., & Kreylos, O. (2008, March). Immersive 3d environment for remote collaboration and training of physical activities. In *Virtual Reality Conference, 2008. VR'08. IEEE* (pp. 269-270). IEEE.
12. Petit, B., Lesage, J. D., Menier, C., Allard, J., Franco, J. S., Raffin, B., ... & Faure, F. (2010). Multicamera real-time 3d modeling for telepresence and remote collaboration. *International journal of digital multimedia broadcasting*, *2010*.
13. Maimone, A., & Fuchs, H. (2011, October). Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras. In *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*(pp. 137-146). IEEE.
14. Maimone, A., & Fuchs, H. (2012, October). Real-time volumetric 3D capture of room-sized scenes for telepresence. In *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2012* (pp. 1-4). IEEE.
15. Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., ... & Tankovich, V. (2016, October). Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (pp. 741-754). ACM.
16. Kuster, C., Ranieri, N., Zimmer, H., Bazin, J. C., Sun, C., Popa, T., & Gross, M. (2012, October). Towards next generation 3D teleconferencing systems. In *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2012* (pp. 1-4). IEEE.
17. Boustila, S., Capobianco, A., & Bechmann, D. (2015, November). Evaluation of factors affecting distance perception in architectural project review in immersive virtual environments. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology* (pp. 207-216). ACM.
18. Benko, H., Jota, R., & Wilson, A. (2012, May). MirageTable: freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 199-208). ACM.
19. Beck, S., Kunert, A., Kulik, A., & Froehlich, B. (2013). Immersive group-to-group telepresence. *IEEE Transactions on Visualization and Computer Graphics*, *19*(4), 616-625.
20. Pejsa, T., Kantor, J., Benko, H., Ofek, E., & Wilson, A. (2016, February). Room2room: Enabling life-size telepresence in a projected augmented reality environment. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing* (pp. 1716-1725). ACM.
21. Bergstra, J. A., & Middelburg, C. A. (2003). ITU-T Recommendation G. 107: The E-Model, a computational model for use in transmission planning.
22. Dou, M., Khamis, S., Degtyarev, Y., Davidson, P., Fanello, S. R., Kowdle, A., ... & Kohli, P. (2016). Fusion4d: Real-time performance capture of challenging scenes. *ACM Transactions on Graphics (TOG)*, *35*(4), 114.
23. ITU-T, I. T. U. T. (2003). Recommendation G. 114. *One-Way Transmission Time, Standard G*, *114*.
24. Fuchs, H., State, A., & Bazin, J. C. (2014). Immersive 3d telepresence. *Computer*, *47*(7), 46-52.
25. Curless, B., & Levoy, M. (1996, August). A volumetric method for building complex models from range images. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (pp. 303-312). ACM.
26. Lorensen, W. E., & Cline, H. E. (1987, August). Marching cubes: A high resolution 3D surface construction algorithm. In *ACM siggraph computer graphics* (Vol. 21, No. 4, pp. 163-169). ACM.
27. Pece, F., Kautz, J., & Weyrich, T. (2011, September). Adapting standard video codecs for depth streaming. In *Proceedings of the 17th Eurographics conference on Virtual Environments & Third Joint Virtual Reality* (pp. 59-66). Eurographics Association.
28. Fuchs, H., Bishop, G., Arthur, K., McMillan, L., Bajcsy, R., Lee, S., ... & Kanade, T. (1994, September). Virtual space teleconferencing using a sea of cameras. In *Proc. First International Conference on Medical Robotics and Computer Assisted Surgery* (Vol. 26).
29. Kanade, T., Rander, P., & Narayanan, P. J. (1997). Virtualized reality: Constructing virtual worlds from real scenes. *IEEE multimedia*, *4*(1), 34-47.
30. Gibbs, S. J., Arapis, C., & Breiteneder, C. J. (1999). TELEPORT–Towards immersive copresence. *Multimedia Systems*, *7*(3), 214-221.
31. Towles, H., Chen, W. C., Yang, R., Kum, S. U., Kelshikar, H. F. N., Mulligan, J., ... & Holden, L. (2002). 3d tele-collaboration over internet2. In *In: International Workshop on Immersive Telepresence, Juan Les Pins*.
32. Gross, M., Würmlin, S., Naef, M., Lamboray, E., Spagno, C., Kunz, A., ... & Strehlke, K. (2003, July). blue-c: a spatially immersive display and 3D video portal for telepresence. In *ACM Transactions on Graphics (TOG)* (Vol. 22, No. 3, pp. 819-827). ACM.
33. Loop, C., Zhang, C., & Zhang, Z. (2013, July). Real-time high-resolution sparse voxelization with application to image-based modeling. In *Proceedings of the 5th High-Performance Graphics Conference* (pp. 73-79). ACM.
34. Newcombe, R. A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A. J., ... & Fitzgibbon, A. (2011, October). KinectFusion: Real-time dense surface mapping and tracking. In *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on* (pp. 127-136). IEEE.