Towards Scalable Sharing of Immersive Live Telepresence Experiences Beyond Room-scale based on Efficient Real-time 3D Reconstruction and Streaming

Patrick Stotko, Stefan Krumpen, Reinhard Klein, Michael Weinmann
Institute of Computer Science II, University of Bonn, Germany
{stotko,krumpen,rk,mw}@cs.uni-bonn.de

Abstract

We present a framework for sharing immersive live telepresence experiences to groups of remote users for arbitrary-sized environments. Our framework builds upon RGB-D data capture of the local environment (by a person or robot) and involves real-time 3D reconstruction, scalable data streaming and visualization to a multitude of remote users at modest bandwidth requirements and low latency while preserving the visual quality of current real-time reconstruction approaches.

1. Introduction

Sharing immersive live telepresence experiences has received increasing attention in recent years with applications in entertainment, teleconferencing, remote collaboration, site exploration, robotics, medical rehabilitation and education. The impression of telepresence – defined as the subjective experience of being in an environment that may differ from the user's actual local physical surrounding – heavily relies on the ability of users to interactively explore the respective scene while avoiding motion sickness which requires scene visualization at high framerates and low latencies. Purely video-based solutions strongly restrict the scene exploration to views in the vicinity of the camera poses during scene capture and are not suitable for live scenarios [3, 6]. Therefore, sharing live-captured scenes as needed for teleconferencing [5] or remote collaboration [9, 1, 4, 7] typically involves data capture, 3D reconstruction, data streaming and visualization. All these steps have to be achieved within strong real-time constraints while preserving the visual quality of the scene, considering typically available network bandwidth and client-side hardware scenarios. Whereas capturing a small fixed-sized region of interest as typical for teleconferencing [5] and small-scale remote collaboration [9, 1] allows the exploitation of expensive well-calibrated setups with statically placed cameras, the efficient capturing, data representation and streaming become significantly more challenging when capturing scenes of arbitrary size with a moving camera [7].

In the scope of this extended abstract, we address the task of sharing immersive live telepresence experiences for arbitrary-sized environments with groups of remote users based on efficient large-scale real-time 3D reconstruction, data streaming and visualization (see Figure 1). By design, our system shifts the hardware requirements from the involved users towards the cloud. In addition, the system runs at low/modest bandwidth requirements with low latency and can handle network interruptions. As a result, our approach allows a re-thinking of well-established exploration processes towards (1) VR-based remote consulting/collaboration, where experts may save the travel time and costs, (2) VR-based remote exploration of contaminated scenes and thereby avoiding the exposure of people to danger, or (3) VR-based education scenarios, where expensive excursions may be replaced with immersive group-scale telepresence in the respective environments. Most of the ideas of this abstract have been published [7] or are currently under review [8].

2. Methodology

As illustrated in Figure 1, our framework involves (1) a local user (or robot) capturing the local environment based on RGB-D sensors as present in mobile phones or the Kinect, (2) a cloud-based real-time reconstruction framework for on-the-fly/online scene capture and camera localization based on volumetric fusion, (3) a cloud server to manage the global scene model and to control the data transmission according to the requests by remotely connected users, (4) the visualization component that updates the locally generated meshes for the individual remote users according to the already transmitted data, and (5) remote users that may independently and immersively explore and interact with the live-captured scene while communicating with the person (or robot) capturing the scene. In the following, we provide details regarding the major components and discuss the key enabling steps towards immersive group-scale

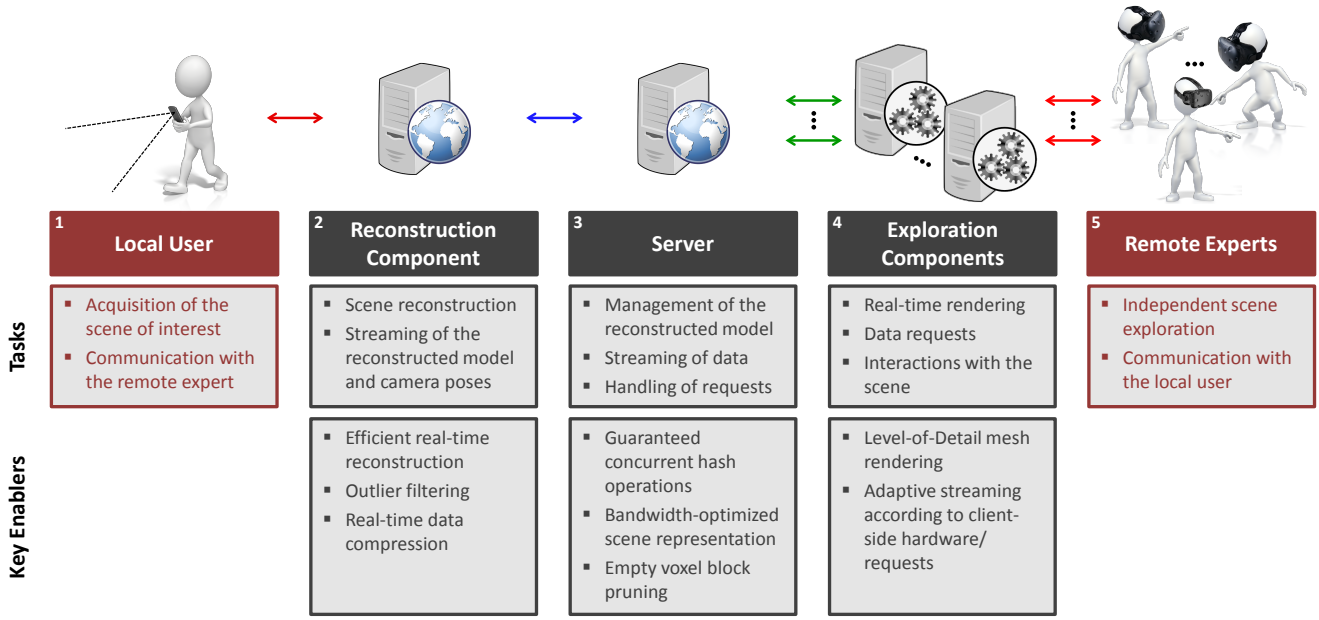

Figure 1. Overview of the framework and its components. Images are partially provided by PresenterMedia.com.

telepresence in live-captured scenes.

Reconstruction Component The RGB-D data captured by a local user (or robot) are streamed to a reconstruction component that reconstructs a dense 3D model in real-time [8] and streams reconstructed parts to the server component [7, 8]. During reconstruction, voxel blocks which fall outside the current camera frustum are queued for streaming and asynchronously compressed and transmitted. Furthermore, the current camera pose is also streamed to the server and broadcasted to the exploration components.

Server The server maintains the global 3D scene model and controls the streaming to connected remote users' exploration components [7]. Each exploration component is assigned a stream set based on GPU hash data structures supporting guaranteed thread-leveled concurrent insertion, retrieval and removal of entries which is crucial for the efficient and reliable management of the updated voxel blocks. The scene model is converted to a highly bandwidth-efficient representation based on Marching Cubes (MC) indices [2] to support pruning empty blocks [8] and enable efficient streaming to a large number of connected exploration components (see Figure 1).

Exploration Component To update the locally generated scene representation, the exploration component sends requests to the server [7]. Such requests can be chosen adaptively based on either the remote expert's current viewing perspective representing the current region of interest, without any preference for prefetching data or using other advanced strategies. The received scene data are grouped into larger mesh blocks for which triangle mesh data as well as three levels of detail are generated asynchronously for efficient rendering. Besides the immersive live exploration of the scene, the remote expert can interact with the scene by measuring distances and marking objects and can also collaborate with other connected experts (e.g. via VoIP).

3. Results

The evaluation comprises the analysis of our approach regarding its performance, its visual quality and a study of respective user experiences.

Performance Analysis For performance assessment, we measured the bandwidth for streaming the data to the exploration components as well as the streaming latency and component scalability using several datasets captured with off-the-shelf RGB-D cameras [8]. Even with low streaming rates of 512 blocks/request at a request rate of 12Hz, a low latency was observed. Furthermore, the required mean (and maximum) bandwidth was around 13MBit/s (and 25 MBit/s respectively) while the server was able to handle 24 components simultaneously using standard consumer hardware without introducing further latency (see Figure 2).

Visual Quality To ensure a high degree of immersion, we also evaluated the visual quality of the reconstructed 3D models. As a result of filtering outliers in the input and model data, our reconstruction component generates higher-

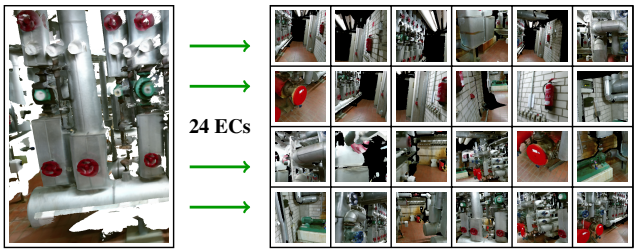

Figure 2. Illustration of our scalable telepresence system which enables sharing live telepresence experiences of high-quality scene reconstructions to more than 24 exploration components.

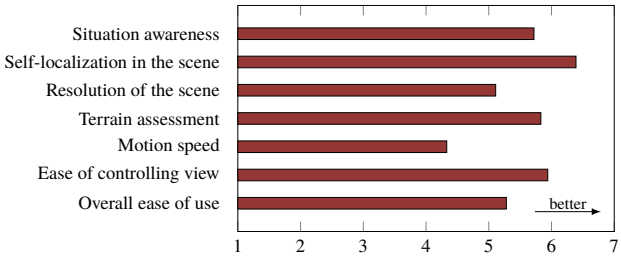

Figure 3. Assessment of user experience based on mean ratings on a 7-point Likert scale.

quality and smaller models in comparison to standard 3D reconstruction which also benefits streaming bandwidth and scalability (see Figure 2). This may be explored for guiding the user to perform a more thorough scene acquisition resulting in more complete 3D models with higher accuracy.

Evaluation of User Experience To evaluate the practicality of our framework for telepresence in live-captured scenes, we immersed 18 subjects (mean age of 28.0 years) into an on-the-fly captured scene based on standard VR devices, where the current local scene model was visualized according to the participant's current pose. This way, the users were able to interactively inspect the scene independent from the camera's current pose. The user ratings (see Figure 3) indicate that the users experienced a high degree of situation awareness and self-localization in the simultaneously captured scene and could easily assess the terrain for navigation purposes. Furthermore, they reported the controls for scene interaction (i.e. teleporting, performing distance measurements, etc.) to be intuitive. The ratings regarding the resolution of the reconstructed model and the speed of movements were slightly lower. Future improvements regarding texture resolution and regarding the overall model quality as well as the increasing availability of affordable VR devices, and with it the higher familiarity to the control mechanisms, may address these aspects.

4. Conclusions

With our approach towards scalable sharing of immersive live telepresence experiences beyond room-scale based on efficient real-time 3D reconstruction and streaming, we hope to foster further research addressing the current challenges regarding model quality, streaming efficiency and collaborative capturing. In addition, we envision the spreading of such telepresence applications into other domains such as robotics, education and collaboration.

Acknowledgements

This work was supported by the DFG projects KL 1142/11-1 (DFG Research Unit FOR 2535 Anticipating Human Behavior) and KL 1142/9-2 (DFG Research Unit FOR 1505 Mapping on Demand).

References

[1] A. J. Fairchild, S. P. Campion, A. S. García, R. Wolff, T. Fernando, and D. J. Roberts. A Mixed Reality Telepresence System for Collaborative Space Operation. *IEEE Trans. on Circuits and Systems for Video Technology*, 27(4):814–827, 2016. 1

[2] W. E. Lorensen and H. E. Cline. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. In *Proc. of the 14th Annual Conf. on Computer Graphics and Interactive Techniques*, pages 163–169, 1987. 2

[3] B. Luo, F. Xu, C. Richardt, and J. Yong. Parallax360: Stereoscopic 360 Scene Representation for Head-Motion Parallax. *IEEE Trans. on Visualization and Computer Graphics*, 24(4):1545–1553, April 2018. 1

[4] A. Mossel and M. Kröter. Streaming and Exploration of Dynamically Changing Dense 3D Reconstructions in Immersive Virtual Reality. In *Proc. of IEEE Int. Symp. on Mixed and Augmented Reality*, pages 43–48, 2016. 1

[5] S. Orts-Escolano et al. Holoportation: Virtual 3D Teleportation in Real-time. In *Proc. of the Annual Symp. on User Interface Software and Technology*, pages 741–754, 2016. 1

[6] A. Serrano, I. Kim, Z. Chen, S. DiVerdi, D. Gutierrez, A. Hertzmann, and B. Masia. Motion parallax for 360 RGBD video. *IEEE Trans. on Visualization and Computer Graphics*, 25(5):1817–1827, May 2019. 1

[7] P. Stotko, S. Krumpen, M. B. Hullin, M. Weinmann, and R. Klein. SLAMCast: Large-Scale, Real-Time 3D Reconstruction and Streaming for Immersive Multi-Client Live Telepresence. *IEEE Trans. on Visualization and Computer Graphics*, 25(5):2102–2112, 2019. 1, 2

[8] P. Stotko, S. Krumpen, M. Weinmann, and R. Klein. Efficient 3D Reconstruction and Streaming for Group-Scale Multi-Client Live Telepresence. *Proc. of IEEE Int. Symp. on Mixed and Augmented Reality (conditionally accepted)*, 2019. 1, 2

[9] R. Vasudevan, G. Kurillo, E. Lobaton, T. Bernardin, O. Kreylos, R. Bajcsy, and K. Nahrstedt. High-Quality Visualization for Geographically Distributed 3-D Teleimmersive Applications. *IEEE Trans. on Multimedia*, 13(3):573–584, 2011. 1