[DC] Multi-user (Social) Virtual Reality Communication

Simon N.B. Gunkel*

Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek (TNO)

Abstract

Virtual Reality (VR) applications are currently suffering several gaps to allow a full breakthrough in the consumer market. One aspect is the social isolation. Many VR experiences do not represent yourself or other users, making it lonely rather than a shared experience. VR applications that do offer a user representation mostly use artificial avatars. These applications may not be well suited for all communication use cases (e.g. at work or with your family). In this paper I present our current work, at TNO and in my PhD research, to create Social VR experiences where both the user and the environment are represented in photo-realistic quality. The aim is not only to allow users to experience VR together, but also to allow new ways of natural communication within VR, with an extended presence and immersion. My PhD research focuses on 3 aspects that are building a Social VR system: capture and processing, transmission and client composition. The main goal of my research is to move processing from the capture and orchestration into the cloud (and particularly into the edge network) to allow Social VR with a large set of users (100+) in one session and to support mobile end devices while maximizing the QoS and QoE.

Keywords: Virtual Reality; VR; Social VR; Requirements; WebRTC; WebVR; Immersive Virtual Environments

Index Terms: Information systems—World Wide Web—Web conferencing; Information systems—Multimedia information systems— Multimedia streaming

1 Introduction

The last few years have seen a major uptake of virtual reality technology, enabling the creation of immersive videogames and training applications, but also paving the way for new forms of video entertainment. One key challenge that many of those VR experiences face is the social barrier. That is, the apparent discrepancy between the physical separation of wearing a head-mounted display (HMD) and the human need for sharing their experiences. This can also be seen by large investments into Social VR from key industry companies such as Facebook, Microsoft and HTC. However, currently the efforts of these big companies mainly focus on artificial and avatar-based user representations. Even though this is good for some use cases, avatar-based approaches may be too restrictive for interactions where non-verbal communication is important, such as video conferencing, presentations, watching 360-degree videos together, intense remote collaboration and many more [1]. Particularly, when it comes to talking to your family and friends, you like to see a natural representation of them.

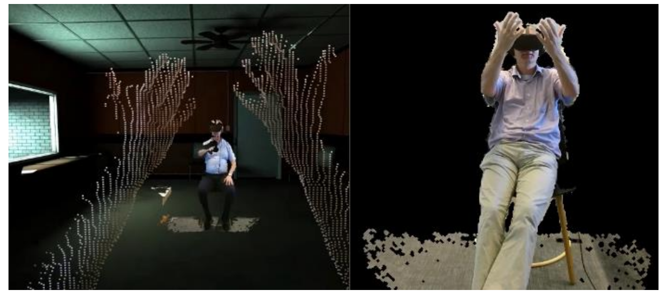

Figure 1: Social VR Application: Environment with other user and self-representation (left) and user capture (right).

When it comes to solutions and use cases addressing photo-realistic representations, we can distinguish between two types: A. capturing and transmitting the user and environment together (e.g. recorded with an omnidirectional camera) and B. capturing and transmitting users independently from the environment (e.g. with a depth camera). With a focus on the later, we developed a VR framework that extends current video conferencing capabilities with new VR functionalities. Our framework is modular, based on web technologies and allows both to easily create VR experiences that are social and to consume them with off-the-shelf hardware. With our framework, we aim to allow users to interact or collaborate while being immersed in interactive VR content. Further, our framework allows us to quickly create different VR experiences both in 3D and 360-degree environments and use different types of user representations. Users are captured with a depth-based camera in real-time and extracted from their background (see Figure 1) to be transmitted and blended into another user's VR environment. Examples of some of the experiences we created are presented in [2][3][4][5] and include 4 types of experiences addressing 3DOF and 3DOF+ VR experiences (360-degree VR for 2 users playing a game or watching a movie, 3 users communicating and a 3D experience where 2 users are represented in 3D in a full 3D environment).

2 Research

Remote collaboration was extensively researched in the past and is also currently addressed by some research in VR [7][8]. However, current research and technology has several gaps in understanding how to build robust end-to-end systems that can support multiple user scenarios as well as cater for different limitations in end devices and the network. Looking at Social VR systems in general (see Figure 2) we can identify 3 main blocks of components that form such systems: i) capture, ii) transmission and iii) rendering. In my PhD research I will investigate these components in more detail, including individual sub-components, with the following main aim:

Moving processing from the capture and orchestration into the cloud (and particularly into the edge network) will allow photo-realistic Social VR with a large set of users in one session (100+ users) as well as the support of Social VR on mobile end devices while maximizing the QoS and QoE.

* simongunkel@googlemail.com

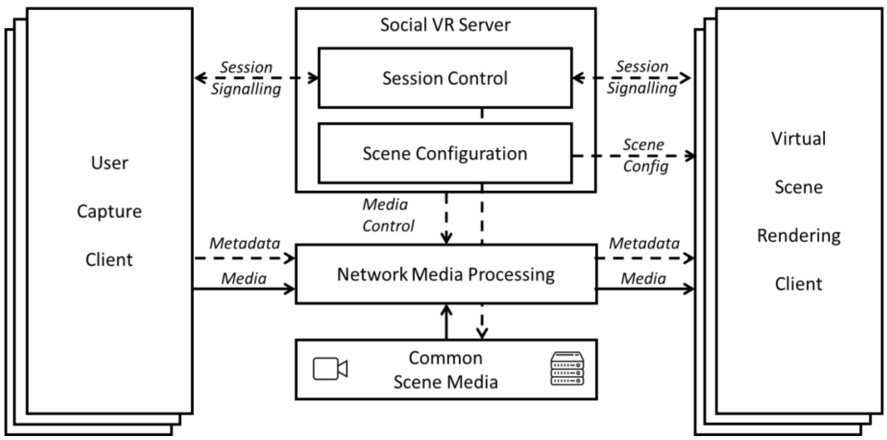

Figure 2: Generic Social VR Architecture (MPEG-I Architecture [10])

We can map this goal into 3 main research questions following the 3 Social VR building blocks:

[RQ1] Capture: Can we optimize the user capture (move processing into the cloud, increase quality, reduce delay) to allow a variety of end devices with different constraints (e.g. mobile devices) to participate in and fully benefit from Social VR experiences?

[RQ2] Transmission: Can we control (synchronize, transmit, process) immersive media formats (point cloud, mesh, stereo billboards, 360-degree video) in a Social VR system (allocate bandwidth, reduce delays) in order to allow 100+ users in one communication session?

[RQ3] Client/Rendering: Can we optimize the composition in the client of different media objects, user representation and VR environments in order to reduce the complexity and support different constrains of end devices depending on differ use cases?

To address these research questions, we will execute a series of simulations based on criterion sampling. The simulations aim to get a better understanding of the building blocks (including sub components) that compose Social VR systems under different constraints of the end devices, the system and the network. In this way we can balance the processing across different computing nodes (in the end devices and network cloud) as well as optimizing transmission flows. For example, one such processing step can be the real-time removal of the HMD from a VR participant with a facial reconstruction [9]. The in-depth knowledge and strategies will be further evaluated in true experiments where we deploy strategies in real-world use cases. To evaluate our system and strategies, we will utilize QoE and QoS performance indicators that are established to evaluate video conferencing systems (e.g. delay, CPU/memory usage, bandwidth and video quality). This will both allow us to compare differences in our strategies and to compare them with existing VR and video communication systems.

3 DISCUSSION

With our web-based approach to Social VR we have developed a framework in which we can rapidly test and evaluate VR experiences. From our previous experiments and proof-of-concept applications, we see that VR can be made more social by leveraging real-time communication in VR, using multimedia functionalities already available in modern web browsers. The approach of capturing users with a camera and blending them in the VR space proves promising. Users have a high degree of immersion, high degree of presence and overall natural communication interactions. Even though this framework already proves useful more research is necessary in the different parts of the system to get a better understanding of the individual (sub-) components that compose such a Social VR System (i.e. capture, transmission and rendering). This is, our web framework offers a general testbed, that allows us to quickly execute more studies and to address the research questions outlined in this paper.

4 CONCLUSION

With the technological basic framework in place, individual components can be enhanced to address the proposed research. This is, my PhD research will compare individual components that form Social VR systems under different constrains. This will help to build more reliable Social VR systems in the future supporting a large set of users, while optimizing the QoS and QoE under different limitations of the end devices and network.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. Pablo Cesar, Dr. Omar Niamut, Hans Stokking and Prof. Dick Bulterman for their guidance and support in his Ph.D. studies. This paper was partly funded by the European Commission as part of the H2020 program, under the grant agreement 762111 (http://vrtogether.eu/).

REFERENCES

[1] Leibrick, Suzanne, “Why women don't like social vr”, https://extendedmind.io/social-vr

[2] Gunkel, SNB. et al. “Experiencing Virtual Reality Together: Social VR Use Case Study”. In Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video. ACM, 2018.

[3] Gunkel, SNB, et al. "Virtual Reality Conferencing: Multi-user immersive VR experiences on the web." Proceedings of the 9th ACM Multimedia Systems Conference. ACM, 2018.

[4] Prins, Martin J., et al. "TogetherVR: A framework for photorealistic shared media experiences in 360-degree VR." SMPTE Motion Imaging Journal 127.7 (2018): 39-44.

[5] Gunkel SN, Prins M, Stokking H, Niamut O. Social “VR Platform: Building 360-degree Shared VR Spaces”. Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video 2017 Jun 14. ACM, 2017.

[6] Beck, S., Froehlich, B. “Volumetric Calibration and Registration of Multiple RGBD-Sensors into a Joint Coordinate System.”In Proceedings of IEEE Symposium on 3D User Interfaces (3DUI), Arles, France, pp. 89-96, March 2015.

[7] Maimone A, Yang X, Dierk N, State A, Dou M, Fuchs H. General-purpose telepresence with head-worn optical see-through displays and projector-based lighting. In2013 IEEE Virtual Reality (VR) 2013 Mar 18 (pp. 23-26). IEEE.

[8] Ekong, S, et al. "Teacher-student vr telepresence with networked depth camera mesh and heterogeneous displays." International Symposium on Visual Computing. Springer, Cham, 2016.

[9] Burgos-Artizzu, Xavier P., et al. "Real-time expression-sensitive hmd face reconstruction." SIGGRAPH Asia 2015 Technical Briefs. ACM, 2015.

[10] MPEG-I Architecture Document