360-Degree Photo-realistic VR Conferencing

Simon N.B Gunkel, Marleen D.W. Dohmen, Hans Stokking, Omar Niamut*

Netherlands Organisation for applied scientific research (TNO)

Abstract

VR experiences are becoming more social, but many social VR systems represent users as artificial avatars. For use cases such as VR conferencing, photo-realistic representations may be preferred. In this paper, we present ongoing research into social VR experiences with photo-realistic representations of participants and present a web-based social VR framework that extends current video conferencing capabilities with new VR functionalities. We explain the underlying design concepts of our framework and discuss user studies to evaluate the framework in three different scenarios. We show that people are able to use VR communication in real meeting situations and outline our future research to better understand the actual benefits and limitations of our approach, to fully understand the technological gaps that need to be bridged and to better understand the user experience.

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

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

1 Introduction

While Virtual Reality (VR) increases its market adoption, VR experiences are becoming more social. This can be seen particularly with many social and collaborative VR systems representing participants as artificial avatar, like Facebook spaces, AltspaceVR, High Fidelity, vTime, Rec Room, VR chat, Big Screen, and many more.

Even though representing users as non-realistic avatars may be effective for some use cases (e.g. games), it can break the immersion and presence in many different use cases, like business meetings, family calls and others. Therefore, with the promise of more natural interactions, some services emerged that represent users in a photo-realistic manner, like Mimesys. Initial research incorporating recent developments in VR hardware and examining photo-realistic VR communication in different use cases also emerged [2][4][5]. However, as the current advantages and limitations of photo-realistic VR are still not fully explored, both in terms of technological gaps and in terms of impact towards the user experience, more research is necessary.

When it comes to photo-realistic social VR we can distinguish between three types: i) capturing the user and environment at the same time, e.g. with omni-directional cameras that is stitched into one image; ii) capturing the user alone, e.g. with depth or stereo cameras and applying foreground/background segmentation; and iii) capturing a full volumetric 3D representation of the user. Particularly a full volumetric capture of a user is difficult and demands complex setups with at least four cameras [1][6]. We are interested in solutions that are easy to use and deploy on a mass scale. Therefore, the work in this paper focuses on capturing the user alone with a single RGB+depth camera and a web-based system to transmit and display users in a virtual environment.

In a recent Social VR requirement gathering and analysis, VR users and experts were asked about their interest in different application contexts for Social VR [3]. In this study the highest interest was shown in video conferencing, and collaboration with others in a virtual environment. In this paper we outline our work on building a VR conferencing platform that allows users to communicate and collaborate in VR, and present initial results of evaluating VR conferencing in three user tests.

* email address – {firstname}.{lastname}@tno.nl

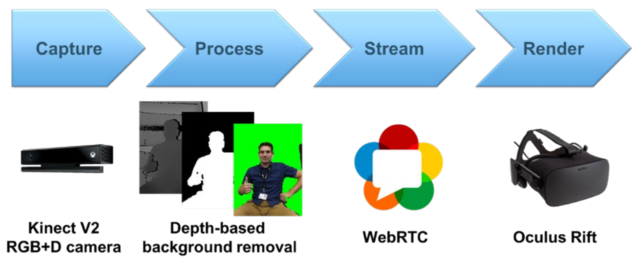

Figure 1: Simplified technology pipeline

2 VR Conferencing System

Our social VR framework extends current video conferencing capabilities with new VR functionality. Our framework is modular, based on web technology and allows easy creation of VR experiences that are social and consumption with off-the-shelf hardware. With our framework, we aim to allow users to interact or collaborate while being immersed in interactive VR content. Figure 1 displays the pipeline for our framework. In order to capture a photo-realistic representation of a user, we first need to capture a color and depth image using an RGB+depth camera (e.g. Kinect v2 or Realsense). In a second step the image information from the depth camera is analyzed and the user is separated from his/her background (the background is replaced with a chroma-key). The resulting image is injected into the web client and is transmitted as video using WebRTC to another web client. We use the SimpleWebRTC library[1] to support direct WebRTC-based peer-to-peer communication between users, including audio and video. At this moment, voice communication is monaural and spatially positioned in the receiver client, and the video image size can be adapted as needed. Currently for capture and transmission we use a resolution of 960x540 pixels. The complete environment (as a 3D scene or a 360-degree image background) is rendered in the browser and the other user is placed into the environment (making the chroma-key background transparent) so that he/she naturally blends into the surroundings. Thus, each end-point consists of a VR capable PC or laptop, a single depth camera, audio headset and a VR HMD.

Figure 2: VR Conferencing meeting between 3 users

3 User Study

We tested our system in three small scale user experiment sessions presented in this section. As an example, Figure 2 shows the view of a user communicating with two co-workers in VR.

3.1 Round Table

We built a multi-user 360-degree experience, allowing 2-4 people to sit around a virtual table in VR. The view of each user is the same: each user sees the other users on the opposite side of the table and optionally a video or presentation on the top of the table. In case a video is displayed on the table it is rotated so that each user sees it "normally" (not head down). To evaluate this experience, we held a 1-day experiment session in an informal and uncontrolled setting at our lab facilities. We collected feedback through a short questionnaire from 54 participants (avg. age of 33.09 and 43% Female) were 3 people are communicating and watching a video. Overall, people expressed a high level of interaction and immersion. Users rated the system with a good overall quality of 4.35 (SD = 0.51) and video quality of 3.65 (SD = 0.77) on a 5-point Likert scale.

3.2 Stand-up

To evaluate the system in a real-world use case we developed a stand-up experience. Users at an external company used this experience within their regular briefing session in their normal work setting and communication. In total, ten users (30% female) participated in this experience and we collected feedback through a short questionnaire in an informal and uncontrolled setting. Overall, users were able to conclude a full meeting of about 25 minutes with our system, similar as to their usual face to face meetings, with 80% of the participants agreeing that "The concept of doing stand-ups in VR is a promising concept that should be further pursued". Participates rated the overall quality with 3.5 (SD = 0.48), and video quality with 3 (SD = 0.85) on a 5-point Likert scale. Asked for improvements, people expressed discomfort with wearing the HMD for this duration (due to heat), a lack of interacting with the environment (pointing), and a wish for a better video quality.

3.3 Remote Conferencing

Extending on the "stand-up" scenario (3.2), we demonstrated the same experience remotely between two office locations during a social event. We collected feedback from thirteen participants (no demographics were collected) through a short questionnaire in an informal and uncontrolled setting. Most participants (69%) indicated they would like to use VR for conferencing purposes, versus only one participant expressing no interest in VR conferencing. In addition, 54% of participants believed VR conferencing to be a good alternative for video conferencing. Strong points, as indicated by participants, were amongst others the feeling of presence and the experienced naturalness of the interaction. The possibility to interact with the environment (e.g. whiteboard use) was stated by 46% of participants as the first thing that should be developed, followed by HMD removal (31%), a more natural positioning of the self in the room (15%), and the possibility to take notes (7%). None of the participants stated that self-view should be developed first.

4 Conclusion & Future work

In this paper we present our photo-realistic VR framework that allows users to have remote meetings and collaborate in virtual spaces. Further, we present three scenarios in which we evaluated the system under different conditions. Overall, we show that people are able to use VR communication in real work situations (stand-ups) and see a benefit of VR from traditional video conferencing setups (even with the current technological gaps). In order to understand the actual benefits and limitations of our approach, as well as the technological gaps, more research is needed; both in terms of enhancing and evaluating the technology and in terms of better understanding the user experience.

Our future technology development focuses on i) scalability, allowing larger groups of people to join one session. This year we will scale our system to allow up to ten users in one communication session; ii) facial reconstruction by replacing the HMD with the real face of the user.; and iii) investigating how to extend and map our current framework to support AR and mobile display devices. To evaluate our technology and to get a better understanding of the user experience, our research focuses on conducting long term trials, allowing people to use our system in everyday work situations for remote meetings, and conducting a comparative user study on collaboration in different communication settings (face-to-face, Skype, and VR) to identify the benefits of photo-realistic VR compared to video conferencing.

Acknowledgements

This paper was partly funded by the European Commission as part of the H2020 program, under the grant agreement 762111 (VRTogether, http://vrtogether.eu/).

References

[1] S. Beck and B. Froehlich. 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, March 23-24, 2015), pages 89-96. IEEE, 2015.

[2] S. N. B. Gunkel, M. J. Prins, H. M Stokking, and O. A. Niamut. Social VR platform: Building 360-degree shared VR spaces. In Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video, ACM, 2017.

[3] S. N. B. Gunkel, H. M. Stokking, M. J. Prins, O. A. Niamut, E. Siahaan, and P. S. Cesar Garcia. 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.

[4] M. J. Prins, S. N. B. Gunkel, H. M. Stokking, and O. A. Niamut. TogetherVR: A Framework for photorealistic shared media experiences in 360-degree VR. SMPTE Motion Imaging Journal 127.7:39-44, August 2018.

[5] J. W. Woodworth, S. Ekong, and C. W. Borst, C.W. Virtual field trips with networked depth-camera-based teacher, heterogeneous displays, and example energy center application. In 2017 IEEE Virtual Reality (VR), pages 471-472. IEEE, 2017.

[6] N. Zioulis, D. Alexiadis, A. Doumanoglou, G. Louizis, K. Apostolakis, D. Zarpalas, and P. Daras, 3D Tele-immersion platform for interactive immersive experiences between remote users. In 2016 IEEE International Conference on Image Processing (ICIP) (Phoenix, Arizona, September 25-28, 2016). IEEE 2016.