EVERYDAY PHOTO-REALISTIC SOCIAL VR: COMMUNICATE AND COLLABORATE WITH AN ENHANCED CO-PRESENCE AND IMMERSION

Simon N.B. Gunkel, Hans Stokking, Tom De Koninck, Omar Niamut

TNO, The Netherlands

ABSTRACT

While Virtual Reality (VR) retains industry interest, the overall market adoption of VR is still slow. One problem in VR is the social isolation. Many applications are still single user driven and VR services that do allow the sharing of virtual experiences with others, mostly focus on representing users as artificial avatars. This might be good in some cases (e.g. gaming) but may be too restrictive for interactions where non-verbal communication is important, such as video conferencing, watching videos together, and remote collaboration. For such cases, the importance of photo-realistic VR is acknowledged within the industry and standardization bodies have begun to address the technology gaps in order to create such services. As an example, MPEG included a generic Social VR architecture into its recent OMAF specification. Also, 3GPP recently started new work on 360-degree VR conferencing and is studying the relevance of many more VR (and AR) communication use cases and technologies to enable them in 5G. In this paper, we report on our recent experiments allowing people to interact, communicate and collaborate with each other as if they were in the same place, while sharing a virtual environment. Our VR conferencing platform is modular, web-based and allows for rapid creation of photo-realistic shared experiences. We evaluated our platform with 313 users in six experiences, for both 360-degree video and 3D volumetric VR. We describe the conditions of each experience, reflect on participant survey results, and sketch a roadmap for the future standardization of social VR services that address, both shared photo-realistic VR experiences in 360-degree, and volumetric video.

INTRODUCTION

The last few years have seen a major uptake of virtual reality technology, enabling the creation of immersive video games and training applications, and 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 large industry players mainly focus on artificial and avatar-based representations of people for use in communication applications. 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, social gatherings, presentations, watching 360-degree videos together, intense remote collaboration and many more [1]. Particularly, when it comes to talking to people like family and friends, it is important to see a natural representation of them.

To address this problem, 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, the easy creation of VR experiences that are social, and the consumption of 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.

In this paper, we present six social VR experiences which we have built with the help of our framework. We evaluated all six experiences with a total of 313 people, in unstructured testing sessions having a duration of 1-10 minutes. Each testing session was followed by a questionnaire and an informal discussion with a focus on assessing overall quality, video quality, audio quality, presence and immersion in photo-realistic social VR.

RELATED WORK

In the ’90s, much work went into creating high-end shared virtual environments. Various universities set up cave automatic virtual environments (CAVE) and CAVE-like systems which could be used to communicate remotely, of which [2] and [3] are good examples, using what they call ‘video avatars’. These environments typically used back projection and large calibrated camera rigs to produce a coherent virtual environment. Other examples such as [4] used large screens, again together with calibrated camera rigs, to also offer a sense of togetherness. Other work from this era consisted of using graphical avatars to create large shared virtual environments, of which [5-7] give some overview. In the past, the impact of avatar realism on the user's perception was studied as well [8].

Virtual reality saw renewed interest with the rise of high-quality but affordable HMDs, most notably the Oculus development kit, which carried the promise of bringing high-quality VR to the masses. This has led to new initiatives in shared and social VR experiences. Nowadays, social VR is mostly associated with artificial graphical avatars in a graphical environments. The main examples are currently: Facebook Spaces, AltSpaceVR (owned by Microsoft), BigScreen, High Fidelity, vTime, hubs by mozilla, VRChat and SteamVR. All these consist of a shared environment, using graphical avatars to represent the users, and offering various things to do, such as sharing games, screens and shared web browsing, including shared video watching, shared exploring, etc. The transfer of user motion to avatar motion is achieved by employing HMD and controller tracking. These environments offer a compelling experience of togetherness, in which immersive video is combined with spatial and 3D audio. Still, they are limited in that users do not actually see other persons, that is, “the social cues that you would normally have about someone being creepy or safe weren’t there” [1]. Recent developments show approaches that employ highly realistic animated characters that seem life-like. One of these examples can be seen in the research done by Facebook [9]. This approach however still needs a complex capture set-up to capture the user first, and currently has further limitations in its applicability.

In order to scan people in high detail and 3D, multiple systems do exist, of which an example is shown in [10]. Such volumetric capture studios offer a rich detail of capture but have severe requirements in hardware set-up and processing. Microsoft has also shown an example of how to deploy such a system for real-time communication in [11]. However, full volumetric communication approaches still require complicated and costly hardware set-ups and extensive processing, and typically do not consider data compression, transmission and scalability (like using the system with more than two people). To close this gap, we developed a framework [12-16] using video streaming and web technologies as a basis to bring people and groups together in a virtual environment. With this approach we show that shared and social VR experiences can be created by using current off-the-shelf equipment and a web-based framework.

SOCIAL VR WEB FRAMEWORK

In this section we explain our platform called TogetherVR, i.e. a web-based framework to build and consume shared and social VR experiences. Our main motivation to utilise web technology is to allow for easy and widespread deployment and low entry burden for end users and developers, and to allow rapid prototyping of new experiences. In this way we only use off-the-shelf hardware and currently-available web technologies. Our work is based on open source software frameworks, primarily SimpleWebRTC, Dash.js, AngularJS, Node.js and A-Frame. The entry point to our social VR application is offered by the TogetherVR web server backend, which is based on Node/Express. WebVR enabled web-browsers connect to this backend to obtain the client application, which is based on AngularJS.

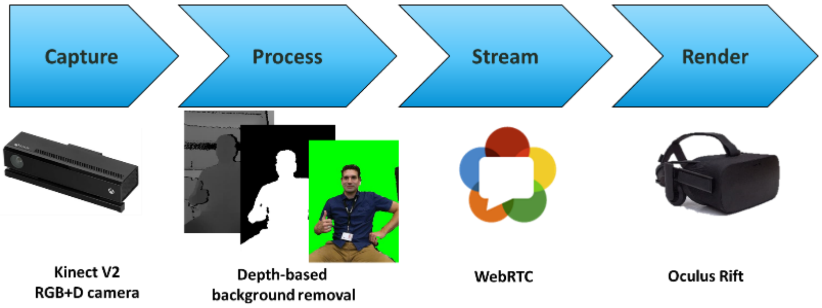

Figure 1 – Simplified technology pipeline

Our system includes a mechanism to alpha-blend people into the environment based on WebGL shaders. Our complete end-to-end pipeline is shown in Figure 2. We first record users with a Kinect 2 RGB-plus-depth camera, replace the background with an chroma-colour before transmission (over WebRTC), and after reception apply alpha-blending to remove the background in the receiving browser, leaving us with a transparent image showing just the user without his/her physical background. Currently for capture and transmission we use a resolution of 960x540 pixels, as this matches the depth resolution of the Kinect sensor. However, our platform also supports the integration any other depth sensor (e.g. Intel RealSense) and using higher resolutions.

USER EXPERIENCES

We have employed the TogetherVR system in a series of user experiments with the aim of evaluating our social VR framework. However, we also conducted this set of user experiments to better understand social VR experiences in general, i.e. in terms of requirements, immersion and presence. In this section we introduce six experiences and how we tested them with 313 users. Each of these experiences follows a similar user set-up. Each set-up has multiple specific and similar user places, each user place includes a VR-capable laptop (e.g. MSI GT62VR) or desktop PC, Oculus Rift HMD (CV1), a Kinect v2 camera, and an audio headset. It is important to note that the physical environment of the user is aligned with the virtual environment, so that if a user looks into the camera, she or he will look at the other person in VR.

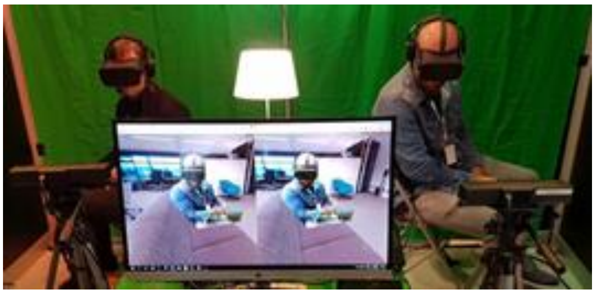

Figure 3 – Two users in VR experience

Two users playing a game

Our first experience is based on a classical social TV use case, i.e. watching apart together. Within the experiment users were either able to engage in a game of pong (GAME) or were watching a movie together (MOVIE) in VR (see Figure 3). More details of the set-up can be found in [13]. We held a 3-day experiment session in an informal and uncontrolled setting at a conference space, collecting feedback through a short questionnaire from 51 participants (avg. age of 31,35 and 22% female) playing the game (GAME) and from another 24 participants (avg. age of 36,56 and 21% Female) watching the video (MOVIE). Users rated the system with a good overall experience of 4.01 (SD = 0.76), a video quality of 3.59 (SD = 0.81) and an audio quality of 3.45 (SD = 0.91) on a 5-point Likert scale. Users did not communicate as much as anticipated, because of the limited field of view of the HMD. A user cannot see the other user on the side and see the movie or game in front at the same time. Overall however, people expressed a high level of interaction and immersion.

Two users watching a movie

In order to understand the requirements of social VR further, we performed a requirement gathering and analysis extending on the MOVIE experience. We conducted our requirements gathering at a large VR exhibition, the VR Days Europe 2017 in Amsterdam (Figure 4). In this way, we ensured that our participants at this stage are people who at least had an interest in VR, and/or had experience using VR applications. In total we gathered feedback from 91 participants (22% female and age ranging from 18 to 60). Users rated the system with a good overall quality of 5.15 (SD = 1.04), a video quality of 4.46 (SD = 1.30) and an audio quality of 3.87 (SD = 1.20) on a 7-point Likert-type scale. More details of this study can be found in [14]. One of the outcomes of this analysis is that people identified video conferencing and education as the most interesting use cases for social VR experiences. Further people see a clear benefit in using social VR as a tool for communication

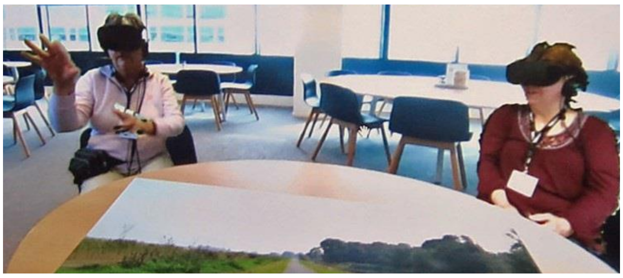

Figure 5 – FoV of 3-way VR conferencing

Three-user VR conferencing

In this experience, three people sit around a round table in VR (THREE). Each user sees the other users on the opposite side of the table and a video playing on the top of the table (see Figure 5). We held a 1-day experiment in an informal and uncontrolled setting at our lab facilities, using a short questionnaire with 54 participants (avg. age of 33.09 and 43% female). 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. We observed more interaction between participants in this experience compared with the previous two in their feedback. People indicated they found it a natural conversation setting, due to seeing people from the front and seeing the video playback and the other people at the same time.

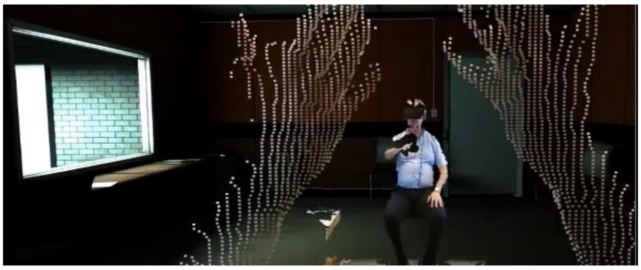

Figure 6 – User's FoV in a 3D experience

Multi-user full 3D Experience

In this experience (see Figure 6), we include a 3D environment and use both the colour and depth information of the user capture to display users in 3D (as point cloud or mesh) [16]. We map the depth image from the depth sensor into a grayscale and transmit the image as 2D video, including RGB and depth frames side-by-side. To display this image as a point cloud or mesh we developed an optimized WebGL shader that maps each pixel into coordinates in the 3D space. This approach is also used to display a self-representation to each user. To evaluate this experience, we held a 2-day experiment session in an informal and uncontrolled setting at a conference venue. We collected feedback through a short questionnaire from 25 participants (avg. age of 35.26 and 12% female) where two people communicated both in a 3D and a 360-degree space. Overall, people expressed a clear preference for the 3D experience (60% 3D, 28% 2D, 12% no clear preference) with a good overall quality of 6.92 (SD = 1.55) on a 9-point Likert-type scale.

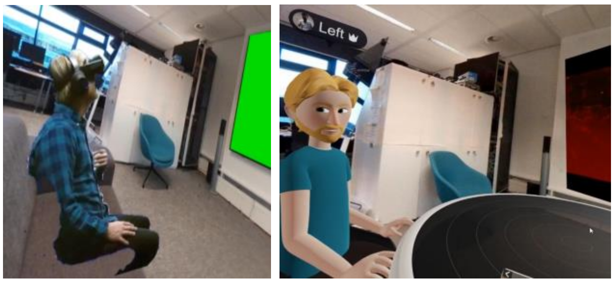

Figure 7 – User in Social VR (left) and Facebook spaces (right)

Comparing face-to-face vs. avatar vs. photo realistic social VR watching a movie

Based on the MOVIE experience, [17] compares our social VR system with an artificial avatar-based system (Facebook spaces) and a face-to-face condition. The three conditions are compared in structured experiments with 16 pairs of users (avg. age of 31.06 and 53% women) watching movie trailers. The experiments showed that our photo-realistic social VR is comparable to face-to-face (no statistical difference) in terms of interaction and social connectedness, for the given use case. Furthermore, avatars limit the perceived quality of interaction, which is also in-line with existing studies in literature that correlate an avatar's realism to the quality of the (mediated) communication.

Comparison face-to-face vs. Skype vs. photo realistic social VR in a negotiation task

Based on the THREE experience, we created a user experience in a neutral office setting with four users sitting around a table. We tested this experience in three conditions (face-to-face, Skype and our system) in a negotiation task in order to better understand how photo-realistic social VR compares with state-of-the-art video conferencing. This is with a focus on a conferencing and collaboration communication task. We tested the experience with 9 groups of 4 (36 participants with mean age of 38.3 and 55% female) in structured experiments followed by questionnaires and interviews. Our initial results show that face-to-face is significantly different and preferred by the users, with no clear differences between Skype and social VR. Further, multiple users remarked on the importance of eye-gaze in such a negotiation task, which is currently not present in the social VR system. However, users also remarked that they do see a benefit in perceiving more conversational cues (like body language) in the social VR setting. In this way users identified (in a questionnaire) the HMD removal as top priority (24/36) for future development, followed by note-taking (19/36), interaction with the virtual environment (4/36), a more natural VR environment (4/36) and self-representation (1/36).

DISCUSSION & STANDARTISATION

Even though users expressed different levels of interaction, immersion and presence in all six experiences, the overall level of interaction was high. In this way the experiences indeed allowed people to communicate while being immersed in VR content. Other than through the questionnaire data, this is also what we observed during the different testing sessions, where we saw a high degree of activity and interaction between users.

From our experiments and proof-of-concept applications, we see that VR can be made social by leveraging real-time communication in VR settings, and in a scalable and low-cost manner by using the multimedia functionalities already available in modern web browsers. Our approach of capturing users with one camera and blending them in the VR space proves promising. The current limitations (e.g. a resolution of 960x540 pixels for user representations and users not being able to see eye gaze when others wear a headset) are not a factor to hinder participants from both communicating and consuming the immersive experience. However, multiple technical gaps need to be closed by the industry to allow a wide spread deployment of such a VR communication platform successful. Currently we see the following challenges for social VR applications:

Eye-gaze: to look each other in the eye. Currently we are working on replacing the HMD in a user representation with a photo-realistic model of the face in real-time.
Scalability: allowing large groups of people to join one session. Currently we work on solutions to scale our system to allow up to 10 users in one communication session, by using a centralised mixing facility, i.e. a VR conference bridge.
AR and mobility: as both augmented reality and mobile devices are becoming more important, we are currently investigating how to extend and map our current framework to support AR devices and mobile display devices.
Long term trials: we are currently setting up long term trials allowing people to use our system in their normal work environment for remote collaboration and meetings.
NBMP (network-based processing): moving processing (like foreground background removal, image processing, object detection and network-based media synchronisation) into the cloud will allow higher resolution and higher quality video streams, while addressing the resource limitations of end devices.
QoE (Quality of Experience): to better understand the benefits of social VR we need to further analyse the influencing factors of conferencing and collaboration in photo-realistic VR compared to traditional video conferencing and face-to-face. After adding HMD removal and the ability to take notes, we aim to repeat the controlled experiment in comparing VR to Skype and face-to-face.

Besides these technical gaps, for implement interactive VR and AR systems at a large scale, interoperability is key. We consider standardisation as the preferred approach towards wide-scale industry adoption. That is, collaborative AR/VR applications media and communication standardisation activities are essential, e.g. within 3GPP and MPEG. While MPEG introduced a generic social VR architecture in MPEG-I [18], currently the interest in social VR has slowed down. However, MPEG is currently actively looking into new aspects for immersive media streaming particularly for volumetric video data which is needed for future photo-realistic multi-user AR and VR experiences and applications. In 3GPP however, new immersive communication scenarios are highly relevant and discussed. For example, 3GPP SA4 recently started work on a 360-degree conferencing allowing users in VR to connect into remote meeting rooms [19]. Further 3GPP SA4 currently studies extended reality (XR) in 5G (XR5G)[20]. In the XR5G study item, 16 out of the 22 proposed use cases deal with conversational aspects. Particularly the promise of 5G to offer high bandwidth and low delay will allow many new and complex XR use cases and application. We see communication in VR (and AR) as one of the most promising applications for 5G, while the technologies of 5G are essential to fulfil the technical requirements and in order to allow a wide spread deployment of such services.

6 CONCLUSIONS

In this paper, we present a web-based social VR framework, that allows us to rapidly develop, test and evaluate social VR experiences. Based on this framework we present an evaluation of six user experiences in both 360-degree and 3D volumetric VR. Overall, with our current web framework we offer a general testbed, in order to quickly execute more studies, to evaluate different types of technology, and their impact on user perception (i.e. immersion, interaction and quality). Furthermore, it has the potential to simplify the efforts needed to create social VR experiences for widespread consumption of immersive media together with other people in the near future. However, more research is necessary to better understand the impact of individual factors towards the communication and QoE in photo-realistic social VR systems, as well as getting all aspects of the technology ready.

ACKNOWLEDGEMENT

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. Leibrick, Suzanne, “Why women don't like social vr”, https://extendedmind.io/social-vr
2. Hirose, Michitaka, Tetsuro Ogi, and Toshio Yamada. "Integrating live video for immersive environments." IEEE MultiMedia 6.3 (1999): 14-22.
3. Ogi, T., Yamada, T., Tamagawa, K., Kano, M., & Hirose, M. (2001, March). Immersive telecommunication using stereo video avatar. In Virtual Reality, 2001. Proceedings. IEEE (pp. 45-51). IEEE.
4. Kauff, P., & Schreer, O. (2002, September). An immersive 3D video-conferencing system using shared virtual team user environments. In Proceedings of the 4th international conference on Collaborative virtual environments (pp. 105-112). ACM.
5. Benford, S., Greenhalgh, C., Rodden, T., & Pycock, J. (2001). Collaborative virtual environments. Communications of the ACM, 44(7), 79-85.

6. Schroeder, R. (Ed.). (2012). The social life of avatars: Presence and interaction in shared virtual environments. Springer Science & Business Media.
7. Leigh, J., Johnson, A. E., DeFanti, T. A., Brown, M., Ali, M. D., Bailey, S., ... & Curtis, J. (1999, March). A review of tele-immersive applications in the CAVE research network. In Virtual Reality, 1999. Proceedings., IEEE (pp. 180-187). IEEE.
8. Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., & Sasse, M. A. (2003, April). The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 529-536). ACM.
9. Facebook is building the future of connection with lifelike avatars, Facebook tech blog, https://tech.fb.com/codec-avatars-facebook-reality-labs/
10. Collet, Alvaro, Ming Chuang, Pat Sweeney, Don Gillett, Dennis Evseev, David Calabrese, Hugues Hoppe, Adam Kirk, and Steve Sullivan. "High-quality streamable free-viewpoint video." ACM Transactions on Graphics (ToG) 34, no. 4 (2015): 69.
11. Orts-Escolano, Sergio, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim et al. "Holoportation: Virtual 3d teleportation in real-time." In Proceedings of the 29th Annual Symposium on User Interface Software and Technology, pp. 741-754. ACM, 2016.
12. 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.
13. Gunkel, Simon, Martin Prins, Hans Stokking, and Omar Niamut. "WebVR meets WebRTC: Towards 360-degree social VR experiences." In 2017 IEEE Virtual Reality (VR), pp. 457-458. IEEE.
14. 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.
15. Gunkel, S. N., Stokking, H. M., Prins, M. J., van der Stap, N., Haar, F. B. T., & Niamut, O. A. (2018, June). Virtual Reality Conferencing: Multi-user immersive VR experiences on the web. In Proceedings of the 9th ACM Multimedia Systems Conference. ACM.
16. 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.
17. De Simone, F, Li, J, Galvan Debarba, H, El El Ali, A, Gunkel, S.N.B, & César Garcia, P.S. (2019). Watching videos together in social Virtual Reality: An experimental study on user's QoE. In Proceedings of 2019 IEEE Virtual Reality (VR) Proceedings.
18. MPEG-I, https://mpeg.chiariglione.org/standards/mpeg-i
19. ITT4RT (Support of Immersive Teleconferencing and Telepresence for Remote Terminals), 3GPP S4-190503, ITT4RT Permanent Document - Requirements, Working Assumptions and Potential Solutions (v0.1.1)
20. FS_XR5G, Study Item on eXtended Reality (XR) in 5G (FS_XR5G), S4-190526, FS_XR5G permanent document 0.4.0