Kachach, R., et al. A Multi-Peer, Low Cost Immersive Communication System for Pandemic Times. In: *28th IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR 3D UI Contest), March 27-April 3, 2021, Virtual.* IEEE, May 2021, Pp. 691-692

DOI: https://doi.org/10.1109/VRW52623.2021.00228

A Multi-Peer, Low Cost Immersive Communication System for Pandemic Times

Redouane Kachach Diego González Morín Francisco Pereira Pablo Perez Ignacio Benito Jaime Ruiz Ester González Alvaro Villegas

Nokia Bell Labs*

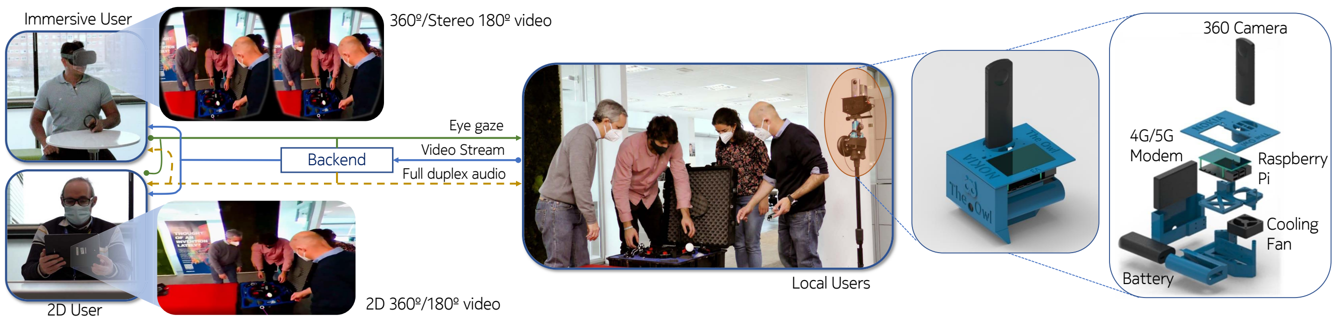

Figure 1: Overview of our multi-peer, low cost immersive communication system. Center) Local scene streamed in real-time by the immersive video capturing device. Left) Two remote users participating in the immersive conference. Right) Detailed view of our custom capturing hardware.

Abstract

We present our multi-peer, low cost immersive communication system which targets to virtually bring people closer to each other during the current self-isolation situation consequence of the pandemic outbreak. Our fully immersive conferencing system allows remote users to virtually teleport where the media capturing system is placed and interact in real-time with other local users present there. This is achieved by a fully mobile, accessible and simple to use system specifically designed to ease the use even for the less experienced users. We accomplish this by avoiding any complex 3D user interface, allowing the remote users to directly connect to the currently active session by just running the application. Furthermore, we also allow the users to share content between each other in a very seamless manner, improving the virtual collaboration tools available in the state of the art.

Index Terms: Human-centered computing—Collaborative Interaction—;——Human-centered computing—Virtual Reality—

1 Introduction

The outbreak of the coronavirus disease 2019 (COVID-19) has completely changed the human communication landscape: social, professional, and educational face to face interactions have been reduced to the minimum. In this unprecedented scenario of self-isolation, we have been forced to rely on traditional video conferences, or 2D video-based communication, which lacks essential communication information derived from the context or non-verbal behaviour, among others. Additionally, traditional video conferences restrict the user capability of perceiving the physical layout of the remote scenarios. These crucial dimensions of human communication are a necessity in many professional and teaching scenarios, such as on-site specific training (e.g.: how to use a particular tool). Besides, these limitations considerably hinder the feeling of presence which might lead the users to disengage from the remote interaction [4]. We are social animals, and there are numerous risk groups, such as the elderly or immunosupressed patients, which are still living in full isolation since the COVID-19 outbreak. Traditional video is not enough anymore, the current crisis has highlighted the necessity of a more immersive and efficient vehicle of remote communication.

The current health crisis has placed the spotlight on virtual reality (VR) technologies, as they can provide fully immersive telepresence systems which can overcome the limitations of 2D videoconferencing tools. This is shown in the increased relevance of online VR collaboration tools such as Mozilla Hubs [2]. These tools provide virtual scenarios in which several users can communicate, interact and collaborate through cartoon-like avatars. While they provide better immersiveness and presence than 2D video, they lack the interaction with the real world which provides crucial information for particular scenarios such as industrial or educational training. Besides, the absence of realism or accurate representation of the avatars and the virtual surroundings leads to user disengagement in different social situations, specially among certain population groups not used to VR interaction.

There is a recent research path which aims to overcome the mentioned limitations and can become a game changer in the field of remote human interaction: real-time immersive video conferencing. In this sort of systems, users can simply wear a VR head-set to be virtually teleported to the spot where a 360° video camera is placed and interact with the local users in real-time. Consequently, it provides a real-time fully immersive experience in which the user can grasp the feeling of being physically on site and interact with the local users. This conferencing modality has several use cases such as education, industrial training, meetings or other social events. Avatour [1] is a clear example of this novel modality of remote communication providing a consumer service for real-time 360° video conferencing, mainly focused on virtual real-time tours for different industrial use-cases.

*Authors' e-mails:{redouane.kachach, diego.gonzalez_morin, francisco.pereira_vega, pablo.perez, ignacio.benito_frontelo, jaime_jesus.ruiz_alonso, ester.gonzalez, alvaro.villegas}@nokia-bell-labs.com

In this paper we propose our custom multi-platform low-latency immersive conferencing solution as an improved and extended version of our previous work [3]. Our goal is to ensure our tool to be as accessible as possible: the simplest and most accessible user interface is having none. The main advantage of our tool is the simple and straight forward setup in comparison with similar conferencing systems. Our tool is designed so that minimum actions from users are required: once the device is configured for the first time, the user can directly run the application and start interacting with the peers at the 360º camera location. Our platform does not require any user interface to provide the remote users with the basic functionality and join the active session. In fact, it allows users with almost no VR technologies experience to use our system easily widening the range of applicable use cases and targeted users.

2 System Description

The system has been designed as a multi-platform ecosystem which includes 3 different groups of users: local, immersive and 2D users. The local users are those physically located where the 360° video capture system is placed. The device captures and streams the full local environment to the immersive users. These users will be wearing the VR head mounted displays and will have the feeling of being present at the exact point where the camera is located, enabling the real-time interaction between the local and remote users. Finally, we define the 2D experience participants who join the session from regular mobile devices, tablets or laptops. This option provides a much lower level of immersiveness in comparison with the second set of users. However, it allows the users to rotate the view around the 360 scene in real-time and get a better grasp of the physical layout of the local site and body language of the local users in a very inexpensive and accessible manner.

In both the immersive and 2D experience applications the real-time feed can be enriched with the representation of the other individual remote users, as real-time 2D video or virtual avatars. To achieve the latter, the real-time gaze of the remote users is continuously transmitted along with the audio feed. On top of the basic functionality, the users can also share content (e.g.: slide presentation) which appears in floating screens with configurable size and position. The content to be shared can be selected and placed with the simple actions of grabbing and dragging using the VR remote controllers or the touch screen.

2.1 Setup Description

The proposed system is designed as a multi-peer communication application with centralized architecture. The central server receives all data flows from all peers and replicates and sends each one as required. The scalability of the approach is guaranteed through the careful design of the backend, where the most intensive process is implemented through a native video processing application that can scale horizontally. The backend is designed in a layered architecture of three levels: session management, control and data. The first one handles the database for individual users and end devices, including their roles, permissions and unique ids. The control level handles the signalling between all applications and users, managing all sources and destination of data streams, the stream life cycle and media control (pause/play). Both the session management and control layers are implemented in Python. Finally, the data level manages the distribution and replication of all video, audio and data streams. Following the commands from the control plane this process receives, replicates and sends the streams to all the target devices. This layer is implemented as a native process written in C to comply with the required data handling efficiency.

Our immersive teleconference system is prepared to support several models of 360° video cameras. We currently support both the Ricoh Theta-V camera and the Vuze XR which can stream in both 360° and stereo 180°. The video quality can be manually adapted to the network quality, being 4K the maximum supported resolution for 360° video. While video encoding is performed within the cameras hardware (using our custom firmware to encode 4K video within the Ricoh Theta-V) the communication with the backend is handled by a Raspberry Pi 4. The video is MPEG encoded with support for both H.264 and H.265 and transmitted over UDP with RTP headers. For both audio capture and playback a Jabra speaker is used, being the audio communication fully managed by Mumble library. The entire hardware setup is powered by a 10000 mAh battery. The sessions are configured and launched using a simple user interface displayed on a small touchscreen. The connectivity is also multi-modal, allowing Ethernet, WiFi and mobile network connections, using a 4G or 5G portable modem. The proposed system is fully portable as it is mounted on a custom 3D-printed structure (see Fig. 1) which can be attached to a mobile tripod or simply placed on a flat surface.

The immersive application has been developed in Unity and our custom libraries have been cross-compiled for different operating systems. Currently, we have a multi-platform VR application that has been tested in different VR devices such as the Oculus Quest 2, Oculus Rift, Oculus Go and Android smartphones with cardboard-style VR goggles. A slightly modified version of the same Unity application is used for the 2D experience participants. All users participating in an immersive conference need to connect to the backend using a Virtual Private Network(VPN). This approach provides secure communication, firewall transparency and access control. The system has been optimized to achieve a latency suitable for two-way communications. To test the described functionality, the system was used for a remote immersive conference between sites located 2000 km apart. Setting the video quality to 2K at a stream rate of 6 Mbps we achieved an end to end latency of 400 milliseconds.

3 Conclusions

In this paper we introduced our multi-peer low cost immersive communication system which provides the users with a fully immersive conferencing system using a simple and accessible setup, widening the range of potential users during the current health crisis and beyond. To fulfil the latter goal, the use of a 3D user interface was directly substituted by the real-time immersive video feed, considerably simplifying our tool usage. We also presented the hardware and software implementation details of our fully functional prototype along with some initial performance measurements. We plan to extend the current boundaries of our tool incorporating full stereo 180° conferencing support. Furthermore, we are currently implementing an Augmented Reality based solution to represent the remote users' avatars in the local scenario, improving the two-way communication experience.

Acknowledgments

This work has been partially supported by the Spanish Administration Agency CDTI under Project IDI-20200225 (TARDIS). This work has also received funding from the European Union (EU) Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie ETN TeamUp5G, grant agreement No. 813391.

References

[1] Avatour. `https://avatour.co/`. Accessed: 2020-12-26.

[2] Mozilla hubs. `http://github.com/mozilla/hubs`. Accessed: 2020-12-26.

[3] R. Kachach, P. Perez, A. Villegas, and E. Gonzalez-Sosa. Virtual tour: An immersive low cost telepresence system. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 504–506, 2020.

[4] M. Massimi and C. Neustaedter. Moving from talking heads to newlyweds: Exploring video chat use during major life events. In *Proceedings of the 2014 Conference on Designing Interactive Systems*, DIS '14, p. 43–52. Association for Computing Machinery, New York, NY, USA, 2014.