2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) | 978-1-6654-4057-8/20/$31.00 ©2021 IEEE | DOI: 10.1109/VRW52623.2021.00209

# Remote Assistance with Mixed Reality for Procedural Tasks

Manuel Rebol* American University
Colton Hood† George Washington University
Claudia Ranniger‡ George Washington University
Adam Rutenberg§ George Washington University
Neal Sikka¶ George Washington University
Erin Maria Horan‖ American University
Christian Gütl** Graz University of Technology
Krzysztof Pietroszek†† American University

## Abstract

We present a volumetric communication system that is designed for remote assistance of procedural tasks. The system allows a remote expert to visually guide a local operator. The two parties share a view that is spatially identical, but for the local operator it is of the object on which they operate, while for the remote expert, the object is presented as a mixed reality "hologram". Guidance is provided by voice, gestures, and annotations performed directly on the object of interest or its hologram. At each end of the communication, spatial is visualized using mixed-reality glasses.

**Keywords:** Remote Procedure, Augmented Reality, Virtual Presence.

**Index Terms:** Human-centered computing—Interaction design; Computing methodologies—Mixed / augmented reality

## 1 Introduction

Remote collaboration is playing an increasingly important role in people's everyday lives. With more people working remotely, video communication has become a *de facto* standard for remote collaboration. Yet, in some applications, due to a lack of spacial information, a limited field of view, a lack of context, and a limited transmission of non-verbal cues and body language, the affordances provided by video communication are insufficient. One example where traditional video communication is not sufficient is remote assistance in procedural tasks.

To address the deficiencies of video communication, previous work has explored domain specific systems for remote assistance using virtual, augmented, or mixed reality. Examples include remote assistance during space flight emergency [1], remote surgery [6], robotics operation [2, 3], telehealth [5], and many others. The proposed remote assistance systems show promise, but are typically expensive and highly-specialized with their domains of use.

In this work, we introduce a general-purpose volumetric communication system that can be used for remote assistance across many domains. To increase the accessibility, we develop our communication system using off-the-shelf hardware. While in our example application, we focus on remotely guided assembly procedures, our system can be deployed for other remote communications situations in which a remote expert guides a local operator. Currently, our system supports two users in a one-to-one communication scenario, but we plan to support one-to-many and many-to-many communication scenarios. Compared to previous domain specific systems, our volumetric communication system consists of widely available off-the-shelf hardware. Moreover, our system is developed such that it can be used for different applications without hardware changes. We focus on high portability to allow for fast deployment whenever a remote procedure needs to be conducted. To provide an intuitive collaborative experience, we virtually place the remote expert and the local operator side-by-side similar to how they would interact if the remote expert was physically present.

*e-mail: mrebol@american.edu
†e-mail: chood@mfa.gwu.edu
‡e-mail: cranniger@mfa.gwu.edu
§e-mail: arutenberg@mfa.gwu.edu
¶e-mail: nsikka@mfa.gwu.edu
‖e-mail: ehoran@american.edu
**e-mail: c.guetl@tugraz.at
††e-mail: pietrosz@american.edu

Figure 1: Top: Remote expert view, including annotations. Bottom: Local operator's view. The expert is adding annotations which are visible to local operator.

## 2 Mixed-Reality Remote Assistance

Our system supports audiovisual, spatially-enhanced communication and interaction between two parties, the local operator and the remote expert. The local operator works on a procedural task while being assisted by a remote expert providing audio and visual guidance.

**Remote Expert View** When using traditional video communication, the essential problem the remote expert is facing, when training or assisting local operators, is insufficient spatial information. Thus, the main functionality of our system is to provide spatial information in real time. To do so, our system continuously scans the object of interest, tracks the position of the hands and tools used

978-1-6654-4057-8/21/$31.00 ©2021 IEEE
DOI 10.1109/VRW52623.2021.00209

by the local operator, and presents this information as a holographic visualization to the remote expert. The "holographic" rendering of the object of interest is as realistic as a two-dimensional video recording, but it additionally includes the shape of the object (Fig. 1). Importantly, spatial sound of the local operator's environment is also communicated to the remote expert, increasing immersion of the expert in the local operator's environment.

Using the spatially-rendered view of the local operator's object of interest and its context, the remote expert visually guides the local operator in performing a procedure. The expert uses speech, pointing and annotating topology of the object of interest.

An example of the view of the expert is shown in Fig. 1. As illustrated, the remote expert can annotate the view and that annotation is visible to the local operator. In our example scenario, the procedure is to replace a part in desktop computer, so the remote expert annotates the slot in the motherboard where the replacement part should be placed.

**Local Operator View** Since the local operator is present at the physical place of execution of the procedure, he/she sees a physical object of interest in front of themselves, and therefore does not need to be presented with the "hologram" of the object of interest. However, the spatial information about the shape of the object of interest is used on the local operator's side to enable displaying remote expert's annotation directly on the physical object of interest. Additionally, the local operator can annotate the physical object of interest and these annotations are visible to the expert, providing two way annotation-based communication between the local and the remote party.

### 2.1 Hardware

In our system, both users are equipped with a lightweight, standalone, high-resolution mixed-reality head-mounted display. The head mounted display is used to present, in real-time, a rendering of three-dimensional data *registered* to real objects in the users' environment. For our prototyping, we choose the Hololens 2, but in principle any mixed-reality headset capable of freehand interaction could be used.

The local operator's system is equipped with volumetric capture camera that continuously feeds high-fidelity spatial data to the remote expert. We use the volumetric capture of the low-cost time-of-flight depth camera Kinect v4 Azure with a separate processing unit. In our setup, the volumetric capture is performed from the position where a supervising operator would have position themselves, if they were co-located with the local operator.

Our choice of hardware is informed by a desire to increase accessibility of spatial communication. We aim to keep the overall system cost as low as possible, as compared to the existing industry-standard telepresence systems. The total cost of our system with the above described components is around $5,000 per user.

### 2.2 Implementation

At the local side, the scene is captured with the Kinect Azure (Kinect v4) volumetric camera. The color information is retrieved by the RGB camera and the depth by the time-of-flight sensor. We process both data streams using the Azure Kinect SDK. Once the streams are processed, they are forwarded to the Mixed Reality WebRTC [4] libraries that handle the peer-to-peer connection between the remote and local operator, avoiding most of the issues with firewalls and NAT traversals. The camera sends both depth and color data in real time over network to the remote expert. The camera sends only the depth data to local operator. The color frames are compressed using VP9 video compression codec. To save on data bandwidth, for depth data we compare frames over time and only send those parts of the depth that have changed.

The WebRTC client running on the head mounted display of the remote and local operator receives the depth streams. Additionally, the remote expert receives the color stream. For the remote view, we assemble the depth and color information and render a 3D mesh. As a result, the local scene is visualized with high spatial fidelity on the Hololens 2 head mounted display of the remote expert.

Besides viewing the local scene, the expert also gives verbal and visual instructions to the local operator. Using their hands, he/she can provide pointing directives and annotations. We track the hand with the kinect camera built in the Hololens 2. We use the Mixed-Reality Toolkit to recognize the hand gestures, delineating between pointing and annotation gestures. The gestures and the annotations are sent to the Hololens 2 of the local operator. The annotations and gestures provided by local operators are relayed to the remote expert in the same manner.

We are able to keep the end-to-end latency of the system below 500 milliseconds. In our pilot studies, the delay is not noticeable, because no instant interaction between the operators takes place on the visual level.

### 2.3 Evaluation

The feasibility of the volumetric system prototype has been tested in a pilot study within our research group. Initial implementation suffered from over 3 seconds latency. However, by introducing above-mentioned compression and switching to WebRTC for network communication, this insufficiency has been resolved. We plan to perform a full user study for the system in medical remote assistance in a near future, when the Covid-19 pandemic subsides allowing us back on campus.

## 3 Conclusion

We present a real-time volumetric communication system designed for remote assistance. Our system allows a distant remote expert to guide a local operator on a procedural task through virtual gestures and annotations. The guidance is given using mixed-reality headsets which support intuitive and efficient communication. Our system is low cost and can be used in various industries including healthcare, logistics and education.


## Acknowledgments

The work is supported by National Science Foundation grant no. 2026505 and 2026568. We would also like to thank Ilia Kowalzik for their help in developing the system.


## References


[1] D. Datcu, M. Cidota, S. Lukosch, D. M. Oliveira, and M. Wolff. Virtual co-location to support remote assistance for inflight maintenance in ground training for space missions. In *Proceedings of the 15th International Conference on Computer Systems and Technologies*, pp. 134–141, 2014.

[2] P. Gurevich, J. Lanir, B. Cohen, and R. Stone. Teleadvisor: a versatile augmented reality tool for remote assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 619–622, 2012.

[3] D. Mourtzis, V. Zogopoulos, and E. Vlachou. Augmented reality application to support remote maintenance as a service in the robotics industry. *Procedia CIRP*, pp. 46 – 51, 2017.

[4] G. Suciu, S. Stefanescu, C. Beceanu, and M. Ceaparu. Webrtc role in real-time communication and video conferencing. In *2020 Global Internet of Things Summit (GIoTS)*, pp. 1–6, 2020.

[5] S. Wang, M. Parsons, J. Stone-McLean, P. Rogers, S. Boyd, K. Hoover, O. Pastor, M. Gong, and A. Smith. Augmented reality as a telemedicine platform for remote procedural training. *Sensors (Basel, Switzerland)*, 17, 2017.

[6] E. L. Wisotzky, J.-C. Rosenthal, P. Eisert, A. Hilsmann, F. Schmid, M. Bauer, A. Schneider, and F. C. Uecker. Interactive and multimodal-based augmented reality for remote assistance using a digital surgical microscope. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1477–1484. IEEE, 2019.