Augmented telecommunication in factory setting

Narek Minaskan, Bastian Krayer, Alain Pagani, Didier Stricker

German Research Center for Artificial Intelligence

Narek Minaskan: nareg.minaskan_karabid@dfki.de

Keywords: XR system architecture, Collaborative and distributed XR, Industrial applications

Introduction

As a result of the COVID-19 pandemic, the possibility of remote work or "Home Office" has been normalized by many companies. Although the existing telecommunication tools are sufficient for many mainstream tasks, they lack the capabilities of 3D interaction which is necessary for complex tasks which require physical presence for engaging in problem solving.

As an early implementation of collaborative 3D communication tools, Microsoft [1] developed the HoloLens system, which introduces a novel interaction model for supporting collaboration between a head-mounted display (HMD) user and remote participants. The HoloLens allows remote companions to join the AR space by hitching onto the view of the primary HMD user through Skype-enabled devices, such as tablets or PCs. This system facilitates asynchronous interaction in a shared 3D space with digital objects, allowing remote parties to contribute to tasks and have their inputs reflected back to the primary user in real-time, thus enabling new scenarios for remote collaboration.

Further advancements in remote collaboration systems have focused on complex tasks like environmental pollution analysis, which require expertise from multiple fields. One such system was designed by Mahmood et al. [2]. It uses mixed reality to support co-presence and collaborative analysis, demonstrating improved remote analysis through shared user and data spaces. Drey et al. [3] explored how the benefits of pair-learning and virtual reality (VR) can be combined by comparing symmetric systems, where both peers use VR, and asymmetric systems, where only one peer uses VR and the other uses a tablet. They found that the symmetric system significantly enhanced presence, immersion, and reduced cognitive load, which are important for learning. However, both systems resulted in similar learning outcomes, demonstrating that both symmetric and asymmetric setups are effective for co-located VR pair-learning.

In industrial and technical settings, operating machinery often requires assistance or training that can be difficult to acquire with traditional documentation or voice/video calls alone. These methods often fail to convey spatial relationships, leading to miscommunication and repeated explanations. To address these challenges, we start by 3D scanning the machines and environments ahead of time, to have them available when running the application. The technician on site is assisted by a remote expert, with the option for additional observers, using multi-device support. Depending on available hardware, participants join the session through their respective devices (PC, XR headset), with the software adapting to features like webcams and tracking.

In this setup (Figure 1), the expert views a virtual representation of the object or environment and can track the technician's pose to better understand what they are looking at. The expert can place and manipulate 3D annotations in the scene and provide additional guidance via voice and video. The technician sees the actual scene through a webcam or XR headset with the expert's annotations superimposed, matching the 3D position. This setup enables efficient collaboration between the expert and technician to solve complex problems more effectively.

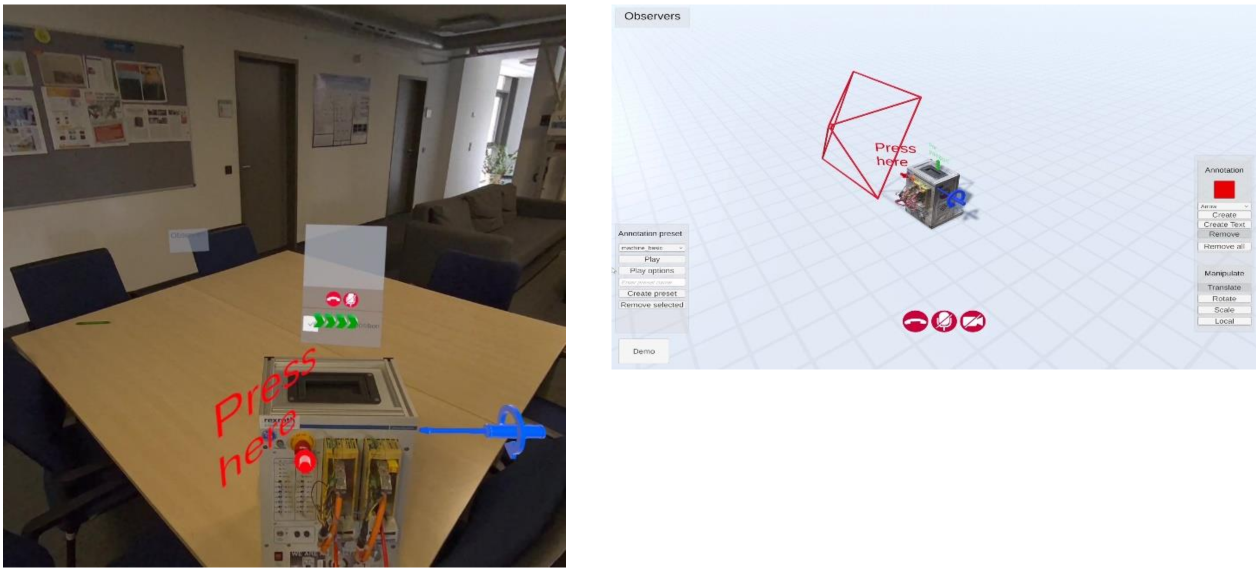

Figure 1: An overview of our setup. Left image: The technician view annotated with moveable augmentations. Right image: The remote export view with the view of the technician visualized as frame and an interface capable of creating 3D annotations and add them to the scene as augmentation.

While existing systems such as HoloLens and mixed reality platforms focus on immersive experiences, our approach is tailored to the industrial environment. It addresses the current issue of heterogeneous hardware availability and usage, allowing flexibility through multi-device integration and adapting to the hardware on hand.

Technical description

Our application is based on Rainbow, a cloud-based communication platform developed by Alcatel-Lucent Enterprise, which offers APIs for video/audio sharing and data transfer. Rainbow's base communication layer operates independently of the device, which means interoperability across different platforms. It supports short-lived sessions, managing dynamic drop-in/out participation. We developed our application with Unity by utilizing the Rainbow SDK. This application supports common communication features such as video/audio transfer and an experimental low-level data channel for message transfer in a call between clients. It also employs distributed state storage, enabling participants to rejoin ongoing conversations, regardless of their device type, maintaining session continuity.

Our implementation currently runs on desktop and XR headsets. The desktop version employs mouse and keyboard as interaction means and uses the object tracker developed by Rambach et al. [4] for providing the camera pose. The 3D data available from the 3D scan allowed us to implement effects such as occlusion and shadows, which makes the experience more realistic.

The headset version is based on OpenXR which makes it simpler to adapt and port to different headsets. The interaction is based on hand/controller tracking which allows the user to select the most appropriate input. As the pose tracking is embedded in most headsets, the technician places the virtual object on top of the real one at the start of the session. After the placement has been done, the virtual object is hidden, but still used to provide occlusion.

Limitations

Currently OpenXR is not in a mature or stable state, which consequently introduces several bugs in official packages and vendor plugins. In terms of XR devices, they also come with several limitations. These include not being able to access sensor data, such as RGB and depth, or missing services such as surface reconstruction. While the API is the same across devices, different headsets might have small discrepancies, which makes it difficult to use the same code (e.g. Passthrough projection doesn't align correctly with rendering).

Summary and Future work

In this paper we presented an XR telecommunication application prototype targeting industrial use cases, with the potential of generalization. Our prototype operates on desktop PCs and multiple XR headsets, allowing participation and interaction in 3D. For the future, we will extend our application to be capable of quickly scanning the environment and add it to the scene. Moreover, we will extend the roles, allowing multiple technicians to work simultaneously. Additionally, we will test external RGB-based interaction methods for users of desktop PCs, enabling them to have hand and gesture interactions. To make augmentations more immersive, we will use depth-cameras, either external or provided by the headset vendors themselves.

Acknowledgements

This research has been funded by the EU project CORTEX[2] (GA: 101070192).

References

[1] Chen H, Lee AS, Swift M, Tang JC. 3D Collaboration Method over HoloLens™ and Skype™ End Points. Proceedings of the 3rd International Workshop on Immersive Media Experiences, New York, NY, USA: Association for Computing Machinery; 2015, p. 27–30. https://doi.org/10.1145/2814347.2814350.

[2] Mahmood T, Fulmer W, Mungoli N, Huang J, Lu A. Improving Information Sharing and Collaborative Analysis for Remote GeoSpatial Visualization Using Mixed Reality. 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Beijing, China: IEEE; 2019, p. 236–47. https://doi.org/10.1109/ISMAR.2019.00021.

[3] Drey T, Albus P, Der Kinderen S, Milo M, Segschneider T, Chanzab L, et al. Towards Collaborative Learning in Virtual Reality: A Comparison of Co-Located Symmetric and Asymmetric Pair-Learning. CHI Conference on Human Factors in Computing Systems, New Orleans LA USA: ACM; 2022, p. 1–19. https://doi.org/10.1145/3491102.3517641.

[4] Rambach J, Pagani A, Stricker D. [POSTER] Augmented Things: Enhancing AR Applications leveraging the Internet of Things and Universal 3D Object Tracking. 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), 2017, p. 103–8. https://doi.org/10/ggr6rz.