Z-3PO: Seamless Switching Between 3 Representation Perspectives for Real-Time Immersive Teleconferencing

Simon N.B. Gunkel, Tessa Klunder, Galit Rahim, and Gianluca Cernigliaro

TNO (Netherlands Organization for Applied Scientific Research)

Corresponding author: Simon.Gunkel@tno.nl

Keywords: Holographic Communication, Extended Reality (XR), Social XR, RGBD Capture, Photorealistic 3D Representation, Remote Training, Immersive Teleconferencing, Education

Abstract

In this (demo) paper we present Z-3PO a capture solution to support seamless switching between 3 real-time representations for humans and objects in immersive 3D applications. We integrated this capture application into a unity demonstrator connecting end devices via a commercial WebRTC pipeline (openrainbow.com). Our prototype allows a teacher and student to collaboratively engage in a Lego block building task while seamlessly explore the different hybrid representation perspectives (3D stereo, 2D billboard and 3D Mesh) in Mixed Reality.

Introduction

The accelerating pace of technological advancement across industrial sectors has exposed a critical global challenge: a growing shortage of skilled experts (Gunkel et al., 2025a). As industries evolve, the demand for specialized knowledge and technical proficiency increasingly outpaces the capacity of traditional training systems (Islam et al., 2025). This imbalance is further exacerbated by limited access to training facilities, geographic constraints, and the complexity of modern industrial tasks. Consequently, organizations face mounting difficulties in maintaining operational efficiency, ensuring safety, and fostering innovation, all of which are contingent on a well-trained workforce.

To address this skills gap, immersive technologies such as Extended Reality (XR)—encompassing virtual, augmented, and mixed reality—offer a compelling solution. XR enables realistic, interactive training environments that simulate complex industrial scenarios, allowing for hands-on learning without the constraints of physical presence (Gunkel et al., 2025b). Photorealistic, real-time representations of humans and objects enhance the fidelity and effectiveness of remote instruction, quality control, and collaboration. These capabilities not only mitigate the impact of expert scarcity but also pave the way for scalable, accessible, and high-quality industrial training solutions.

The main goal of Z-3PO is to enable the next paradigm for collaborative business meetings and remote training experiences in XR, with a specific focus on natural communications between humans based on the following three limitations.

Limitation 1, **Human presence**: bringing real people in XR is key for realistic interactions. Remote XR experiences are normally based on virtual avatars not able to convey the true feeling of "being there together". Z-3PO enables real humans remotely located to join shared XR experiences, specifically considering the use cases: i) business meetings, where users communicate with each other in a many-to-many scenario and ii) remote training where a trainer and a trainee interact in a one-to-one scenario.

Limitation 2, **Real-life experience**: XR applications are usually based on computer generated assets that, in case of collaborative experiences, can be controlled by multiple users. However, the integration of virtual assets in real-world scenarios often lacks the tactile authenticity of object manipulation. Thanks to the multiple holographic representation formats available, Z-3PO allows augmentation of reality with different virtual representations of users and objects.

Limitation 3, **Real-time and multi-modal communication in XR**: Holographic communications encounter challenges due to high volumes of data to be processed and solutions vulnerable to real-time constraints. Furthermore, even though multiple representation modalities were introduced and tested in the past (Gunkel, 2024). It is not always clear how suitable these formats are under different conditions and use cases and direct comparisons are cumbersome. With Z-3PO we introduce a demonstrator that provides 3 photorealistic real-time representation formats and allows seamless switching between them without the need of complex reconfiguration or restarting of components or end-devices.

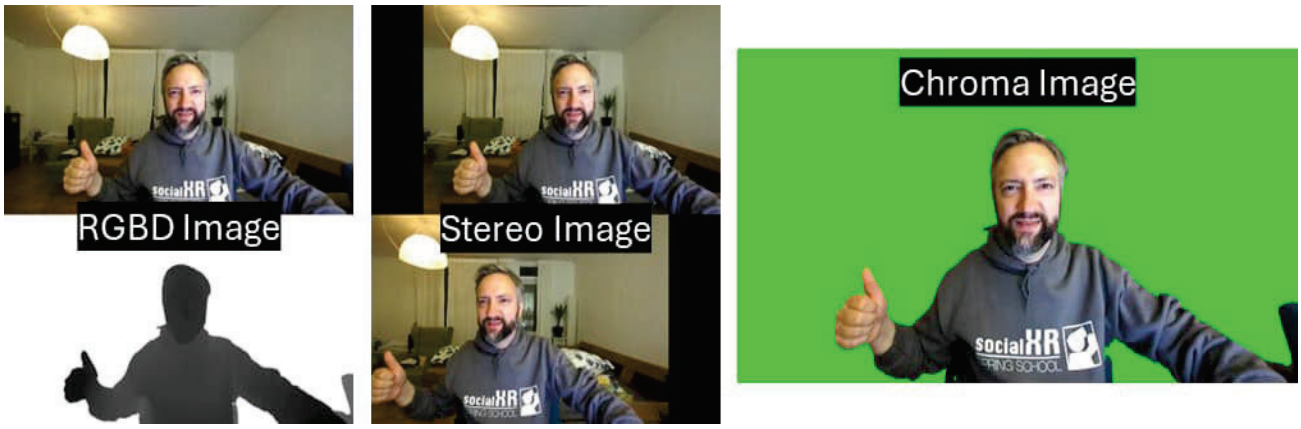

Figure D 1. Three capture and render perspectives for industrial trainings (3D mesh, 3D stereo, 2D Chroma).

Solution

The technical basis for Z-3PO is the RGBD 3D user representation format, and modular capture component presented in Gunkel et al. (2023), and the stereo format presented in the "remote training" prototype (Gunkel et al., 2025b). Ultimately the Z-3PO prototype integrates the following 3 representations (see Figure 1) into a flexible end-to-end (capture, networking, and rendering) pipeline:

- **2D billboard (chroma):** detailed representation of humans or objects with greenscreen cutout effect and thus blended into the virtual environment
- **3D Stereo:** using stereo sensors for a highly detailed 3D view, this for example allows a view of objects or a workbench in complex (technical) training tasks
- **3D RGBD:** photorealistic 3D representation of humans as point cloud or mesh texture

All 3 formats are widely compatible with existing video compression and transmission technologies and allow real-time photorealistic capture and rendering. Figure D 2 shows the full end-to-end architecture of Z-3PO. A capture module (served by a stereo and depth camera, Zed 2i) is responsible for capturing the 3 formats, creating all necessary metadata and allowing switching between the formats. The capture module is connected to a unity client via virtual webcam interface (for video data) and a local webserver (for metadata). The unity client (running on a Laptop with dedicated graphics card) sends and receives video, audio and metadata via WebRTC (Rainbow SDK)[25] and renders the combined local and remote view (users and objects) into a mixed reality head mounted display (HMD). The rendering is changed based on the formats and metadata, to reproduce a photorealistic natural representation of the users and objects. The main benefits of the Z-3PO are:

A. 3 representation formats integrated into a commercial (operational) WebRTC pipeline
B. Easy way to compare different formats (real-time switch between formats)

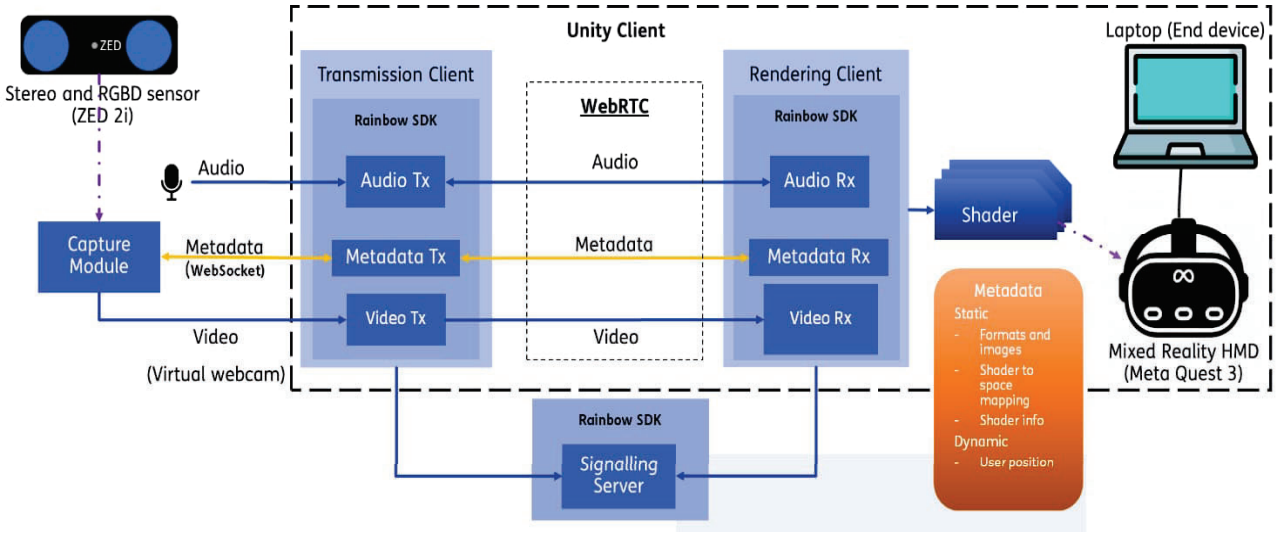

Figure D 2. System Architecture.

Demonstrator Outline

Participants can engage in a remote Lego building task with a "student" building an Lego object locally assisted by a remote teacher (similar as presented in Gunkel et.al., 2025b) [26]. Users can see their local environment, including (Lego) objects and interaction, in a Meta Quest 3 Mixed Reality glasses (with pass-through activated) and the possibility to toggle between the 3 representation perspectives (stereo, chroma and 3D). Thus, the demo is similar to the "Ready Expert One" experience presented in Gunkel et al[26]., (2025b) with the key difference of a different rendering HMD (the Meta Quest 3 offering higher detail in reading with a higher resolution and field of view) and the ability to directly compare different representation formats in a fluent experience (no restart needed during switching).

[25] openrainbow.com
[26] https://youtu.be/5XFPV2Zzosg

Initial Measurements

To evaluate the technical feasibility of our demonstrator we performed some initial measurements deployed in local setup connected via the Rainbow infrastructure. The test was executed on a laptop (Acer Predator Helios 300, PH315-55s-917f) with internet over Wi-Fi. We measured the CPU / GPU usage of the Z-3PO capture module and latency for the 3 representation formats. Latency was measured with the DelAyrUco tool[27] as capture to render (local) and full end-to-end latency (remote). For the latency measurements we utilized the Z-3PO capture module and the browser version of Rainbow[28]. The results of our measurements can be seen in Table D 1. The performance (CPU / GPU) of the capture module is stable and similar across a representation modality. The remote delays measures might indicate some inconsistencies and need further investigation but overall show technical readiness of our approach.

Table D 1. Initial technical measurements of Z-3PO.

	3D-RGBD (MESH)	3D-Stereo	2D-Chroma
Capture CPU in %	4.22 (std. 0.21)	4.13 (std. 0.19)	4.96 (std. 0.26)
Capture GPU in %	35.76 (std. 2.13)	35.34 (std. 2.5)	37.54 (std. 1.0)
Capture Latency in ms	170 (std. 27)	148 (std. 21)	143 (std. 18)
Full Latency in ms	514 (std. 35)	498 (std. 27)	641 (std. 17)

Conclusion and Future work

We presented Z-3PO, a real-time XR teleconferencing solution that enables seamless switching between three photorealistic representation formats—2D chroma, 3D stereo, and 3D RGBD. Integrated into a commercial WebRTC pipeline, Z-3PO supports immersive, natural communication for remote training and collaboration. Initial technical tests confirm stable performance and acceptable latency across formats. Future work includes expert user evaluations to assess usability and representation preferences. We also plan to explore adaptive streaming (via Artificial Intelligence) to enhance performance and user experience.

Acknowledgments

This work was partially funded by the European Union: CORTEX2[29] OC#2 (101070192), R3in3D (RGBD Real-Time Representations of Humans and Objects in 3D).

References

Gunkel, S.N.B., Dijkstra-Soudarissanane, S., Stokking, H.M., & Niamut, O.A. 2023. From 2D to 3D video conferencing: Modular RGB-D capture and reconstruction for interactive natural user representations in immersive extended reality (XR) communication. Frontiers in Signal Processing, 3, 1139897. https://doi.org/10.3389/frsip.2023.1139897

[27] https://github.com/TNO/DelAyrUco
[28] https://web.openrainbow.com/
[29] https://cortex2.eu/

Gunkel, S.N.B. 2024. From 2D Video Conferencing to Photorealistic Immersive 3D Communication.

Gunkel, S.N.B., Klunder, T., & Cernigliaro, G. 2025a. R3-D3: (Photo-) Realistic Real-time Representations of Dynamic 3D-Driven Industrial Trainings. In ACM International Conference on Interactive Media Experiences Workshops (IMXw) (pp. 140–144). SBC.

Gunkel, S.N.B., Klunder, T., & Cernigliaro, G. 2025b. Ready Expert One: Universal 3D Workbench for Remote Industrial Training. In Proceedings of the 2025 ACM International Conference on Interactive Media Experiences (pp. 359–361).

Islam, M.T., Sepanloo, K., Woo, S., Woo, S.H., & Son, Y.-J. 2025. A review of the industry 4.0 to 5.0 transition: Exploring the intersection, challenges, and opportunities of technology and human–machine collaboration. Machines, 13(4), 267. https://doi.org/10.3390/machines13040267