HOPE: Holographic Optimized Processing Engine

V. E. Kotsis-Panakakis[2], I. Evangelou[2], F. Bistas[2],

A. Gkaravelis[1], N. Vitsas[1], A. A. Vasilakis[1], G. Papaioannou[2]

Phasmatic, Greece[1]

Athens University of Economics & Business, Greece[2]

Corresponding author: iordanise@aueb.gr

Keywords: adaptive sampling, 3D compression, holographic communication, telepresence

Introduction

Real-time holographic communication is rapidly transforming the way people interact remotely by enabling immersive, lifelike 3D representations of individuals in motion. This technology goes beyond conventional video calls by capturing and streaming dynamic human features in three dimensions, delivering a strong sense of physical presence. These capabilities open new possibilities for a wide range of applications, from remote collaboration to social interaction, education, and virtual training experiences. As volumetric video technologies continue to evolve, several key challenges have emerged, driving ongoing research and innovation. A central requirement is the efficient delivery of high-bandwidth, data-rich content suitable for real-time telepresence applications based on fluctuating network conditions and the processing capabilities of end-user devices. These challenges lie at the core of making holographic communication scalable, accessible, and seamless across platforms and use cases. In this work, we present HOPE (Holographic Optimized Processing Engine), an ongoing research effort in advancing real-time holographic communication through a suite of context-aware technologies. HOPE employs importance-driven, semantic-aware filtering to preserve critical details while degrading less relevant data. In this scenario, gesture recognition is utilized to dynamically highlight points of interest enabling more intuitive interaction and enhancing the telepresence realism. By integrating smart tuning of efficient mesh compression methods, HOPE aims to offer high-quality 3D streaming under bandwidth constraints. A dynamic adaptation mechanism will fine-tune processing parameters in real time based on device capabilities and network conditions. Together, these components position HOPE as a promising enabler of scalable, bandwidth-efficient holographic experiences with high visual fidelity and responsive performance, particularly in interactive use cases. This poster/demo presents ongoing work and early insights into the system's design and potential.

Methodology

The HOPE architecture is structured around a producer-consumer pipeline, where the end-to-end process begins with the capture of a human subject's 3D shape and motion using a depth camera. The choice of camera setup depends on the positional requirements of the application. The captured raw data is then processed into a point cloud-based volumetric video. The data generated undergoes preprocessing steps, including background removal, tracking, segmentation and compression, to reduce data size while retaining visual fidelity. Rather than relying on uniform data strategies, HOPE project focuses on context-aware sampling and adaptive compression, preserving crucial details (such as user suggestion areas) while discarding redundant or less important data. The processed 3D content is then packaged and transmitted in real time to the consumer side. Upon reception, the data are decoded, reconstructed into 3D form, and rendered for visualization on the user's XR device. Currently, the pipeline supports unidirectional (asymmetric) communication, where one participant is viewed by many, enabling real-time, one-to-many holographic streaming conferencing. Figure 1 illustrates the overall HOPE pipeline.

Depth Camera Capturing. Our system captures 3D data using a single Intel RealSense Depth Camera D435i, positioned to provide a frontal view of the user. However, the generation of a fused point cloud from multiple depth cameras is inherently supported. The depth stream is used to generate a point cloud, where the spatial position of each point is calculated based on depth measurements. Simultaneously, the color stream is utilized to map RGB values to the corresponding points in the cloud, resulting in a colorized 3D representation of the subject.

Preprocessing. To manage the large volume of 3D data produced by depth sensing, HOPE incorporates a multi-step preprocessing pipeline designed to adapt to varying client conditions and prioritize meaningful visual content. The first step applies a distance-based filtering to remove points located beyond a depth threshold from the camera [3]. This early filtering significantly reduces the raw point cloud size by excluding irrelevant background data, allowing subsequent processes to focus on the subject. The main approach to reducing data complexity is a flexible, context-aware segmentation method. The point cloud is divided into N independent descriptions corresponding to objects and regions detected by neural image analysis [1,2]. This segmentation is further refined by incorporating gesture recognition, enabling users to intuitively select or highlight important objects and areas (see Figure 1, second column). These segmented descriptions can then be processed into different quality presets and selectively streamed, prioritizing regions of interest for enhanced visual fidelity. Optionally, when required due to bandwidth constraints, uniform sampling may be applied as a simple downsampling step [3]. This combination of distance filtering, semantic segmentation, tracking and gesture-driven selection enables adaptive, bandwidth-efficient streaming that preserves critical details while responding to user intent.

Compression. After segmentation, each cluster is smartly tuned and compressed individually using advanced mesh compression techniques such as Draco and MeshOpt. Different compression profiles (low, medium, and high priority) have been identified, with selected parameters optimized for each case. This targeted approach allows the system to preserve essential details in critical regions while minimizing data size for less important areas. By

applying this compression strategy selectively, HOPE aims to achieve efficient bandwidth usage without sacrificing visual quality.

Communication. To achieve low-latency, real-time transmission between the capturing and rendering components, the system employs WebSockets. In the current implementation, WebSockets is used for one-way communication, where 3D mesh data flows from the producer (capture side) to the consumer (rendering side) without feedback. In future iterations, the client will be able to transmit feedback messages to the server, enabling adaptive quality control based on the estimated available bandwidth.

VR Rendering. The 3D data are finally transmitted to the consumer-side, where it is decoded and reconstructed into a full 3D mesh representation of the subject. This reconstructed point cloud is then rendered in a virtual environment for immersive visualization on the user's XR device. The rendering has been implemented using WebXR in combination with Three.js, enabling cross-platform compatibility and seamless deployment within web-based XR experiences. The solution has been successfully tested on a Meta Quest 3.

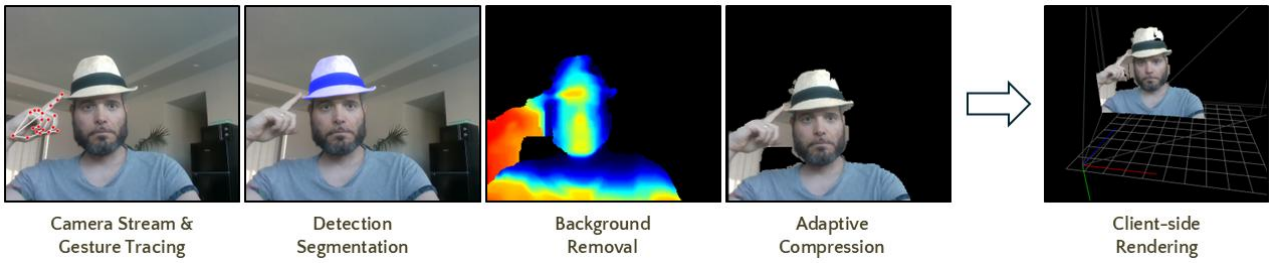

Figure 1. HOPE pipeline in detail.

Conclusion & Future Work

HOPE represents a step toward realizing the vision of the "holographic internet," where real-time 3D communication becomes a natural part of digital interaction. This ongoing work lays the foundation for a technical framework that supports immersive, scalable, and bandwidth-efficient holographic experiences. Looking ahead, we aim to enhance HOPE's capabilities through several key development areas: integrating multi-camera capture setups to improve 3D reconstruction accuracy and depth perception, optimizing client-side rendering for resource-constrained devices by reducing latency (for example supporting adaptive rendering strategies [4]), and expanding device compatibility across a wider range of VR headsets. These directions will further strengthen HOPE's role as a versatile solution for next-generation interactive holographic communication.

Acknowledgements. This work was funded by the European Union under the SPIRIT project (Grant Agreement no. 101070672).

References

[1] Wang, Ao, et al. "YOLOE: Real-Time Seeing Anything." arXiv preprint arXiv:2503.07465 (2025).

[2] Ravi, Nikhila, et al. "Sam 2: Segment anything in images and videos." arXiv preprint arXiv:2408.00714 (2024).

[3] De Fré, Matthias, et al. "Scalable MDC-based volumetric video delivery for real-time one-to-many WebRTC conferencing." Proceedings of the 15th ACM multimedia systems conference. 2024 (pp. 121-131).

[4] Gourlay, M. J., & Held, R. T. Head-Mounted-Display tracking for augmented and virtual reality. Information Display, 33(1), 6-10. (2017)