

HOPE: Holographic Optimized Processing Engine

V. E. Kotsis-Panakakis², I. Evangelou², F. Bistas²,
A. Gkaravelis¹, N. Vitsas¹, A. A. Vasilakis¹, G. Papaioannou²

Phasmatic, Greece¹

Athens University of Economics & Business, Greece²

Introduction

Holographic telepresence is recently revolutionizing remote interaction by streaming lifelike 3D representations of people in motion, going far beyond traditional video calls. This immersive technology has wide-ranging potential across collaboration, education, and virtual training, but delivering high-quality, real-time 3D content remains challenging due to bandwidth and device limitations.

In this work, we introduce **HOPE: Holographic Optimized Processing Engine**, a context-aware system designed to enable efficient, scalable holographic communication. HOPE uses gesture recognition & semantic-aware filtering to prioritize key details, improving interactivity and realism. Smart mesh compression and adaptive tuning allow responsive performance.

This poster presents the **design, components, and early results** of HOPE, aimed at advancing real-time, bandwidth-efficient 3D telepresence.

HOPE Methodology

Depth Capture. An Intel RealSense D435i depth camera is positioned frontally of the user. The depth stream generates a *spatial colorized 3D point cloud*. The system may also support multi-camera fusion for higher accuracy.

Preprocessing: To handle high data volumes, HOPE employs a multi-step preprocessing pipeline:

- **Semantic segmentation** splits the point cloud into regions based on neural image analysis [1,2].
- Refinement via **gesture recognition**, allowing users to highlight important areas or points of interest.
- Distance-based **filtering** removes *background points* [3].
- Point **downsampling** may also be used to ensure efficiency under network constraints.
- Segmented regions are *assigned to predefined presets* and **prioritized** for streaming.

Compression. Each segmented region is compressed using advanced techniques (e.g., *Draco*, *MeshOpt*) with different quality profiles.

Communication. Real-time mesh data is streamed via WebSockets in a one-way channel. Future updates will enable client feedback to dynamically adjust stream quality based on available bandwidth.

VR Rendering. On the consumer side, 3D data is decoded, reconstructed & rendered in a virtual environment using WebXR and Three.js.

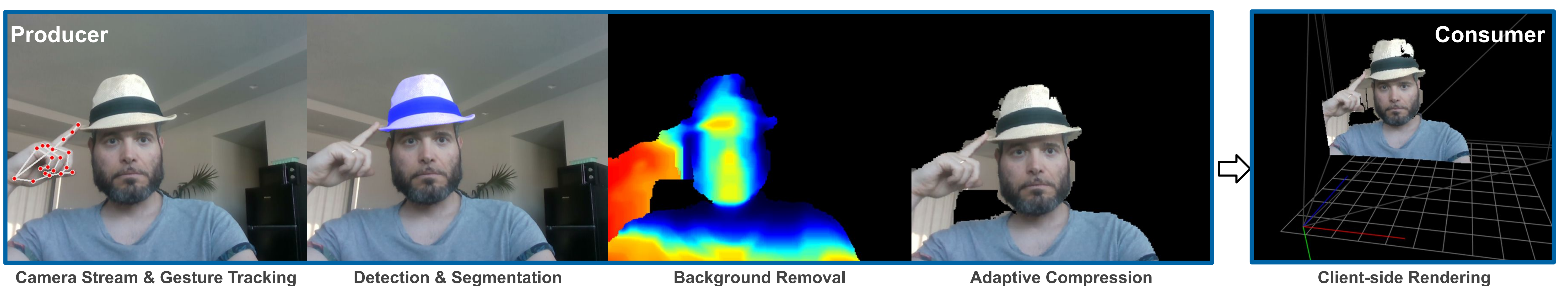


Fig. 1: Overview of the HOPE pipeline.

References

- [1] Wang, et al. (2025) YOLOE: Real-Time Seeing Anything. arXiv preprint arXiv:2503.07465.
- [2] Ravi, et al. (2024) SAM 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714.
- [3] De Fré, et al. (2024) Scalable MDC-based volumetric video delivery for real-time one-to-many WebRTC conferencing. Proceedings of the 15th ACM multimedia systems conference, 121-131.
- [4] Gourlay & Held (2017) Head-Mounted-Display tracking for augmented and virtual reality. Information Display, 33(1), 6-10.