SPIRIT
PHASMATIC
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS
EUROXR Winterthur 2025

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]

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 System Architecture

The HOPE pipeline follows a **producer-consumer** model for real-time holographic streaming (Fig. 1). It begins with 3D capture of a subject using a depth camera setup.

Captured data is transformed into point cloud-based volumetric video and undergoes key **preprocessing steps** to reduce bandwidth while maintaining fidelity, including:

- Segmentation & tracking
- Background removal
- 3D point cloud compression

A core innovation in HOPE is its **context-aware processing**: instead of treating all data equally, it selectively prioritizes important visual regions (see Fig. 1 second column) while discarding redundant or low-value information.

The resulting data stream is efficiently encoded, transmitted in real time, decoded and visualized on the client's XR device.

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.

Future directions include:

- **Multi-camera capture** for more accurate 3D reconstruction.
- **Optimizing client-side rendering** to reduce latency [4].
- **Broadening VR** platform support.

These developments aim to position HOPE as a robust, next-generation solution for interactive holographic communication.

Contact

Iordanis Evangelou
Athens University of Economics & Business

iordanise@aueb.gr
info@phasmatic.com

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 ensures 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.

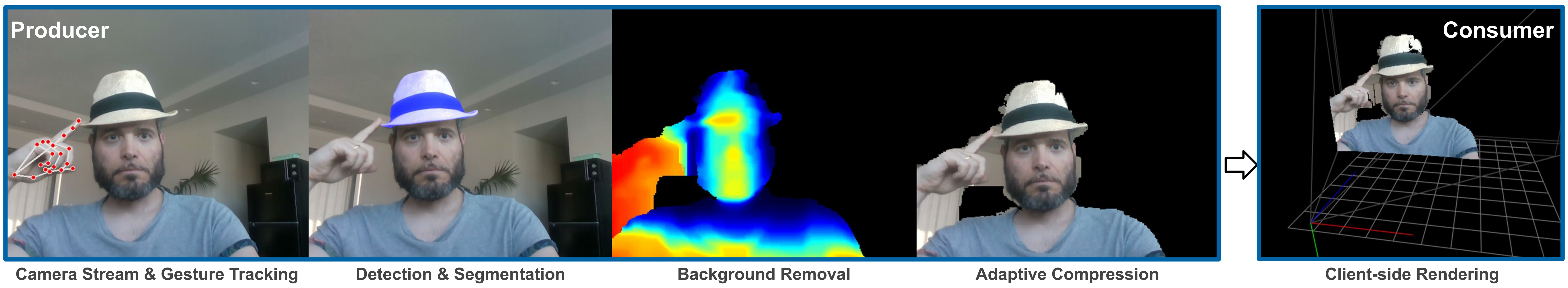

Fig. 1: Overview of the HOPE pipeline.

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.

Co-funded by the European Union
Project funded by
UK Research and Innovation
www.phasmatic.com