SPIRIT

VistaXR

EUROXR Winterthur 2025

VISTA-XR | Virtual Immersive Solutions for Tourism and Architecture in Extended Reality

Introduction & Project Goal

VISTA-XR is an immersive telepresence solution transforming cultural tourism and education through holographic storytelling. Traditional virtual tours often lack true interactivity and realism. VISTA-XR addresses this by combining volumetric capture, XR technology, and optimized WebSocket streaming for realistic, pre-recorded holographic guided tours.

Our goal:

Deliver immersive holographic experiences using low-latency (<200ms), bandwidth-efficient XR streaming.

Enable inclusive, interactive virtual exploration of global cultural heritage sites.

Showcase edge computing (Jetson Orin NX) and advanced XR workflows.

Promote cross-cultural collaboration through innovative digital presence.

Real-time Streaming (WebSockets)

Server Component:

Handles fast, efficient transmission of pre-recorded holographic textures from the server to multiple XR headsets using WebSocket technology.

Latency Goal:

Optimized workflow ensures smooth and responsive XR experiences, maintaining end-to-end streaming latency below 200 ms.

Multi-Client Handling:

Supports multiple simultaneous XR headset connections, allowing multiple users to experience immersive, synchronized holographic content concurrently.

Technical Setup

Technical Setup (Hardware & Software)

Camera: StereoLabs ZED2i for volumetric capture

GPU Processing: NVIDIA GPU (CUDA-dependent)

Edge Device: Jetson Orin NX

Development Platform: Unity 6, ZED SDK (v5.0), CUDA toolkit

Output: Standalone Unity server (.exe)

Fig 1: 2D Holographic Guide

Fig 2: 3D Point Cloud Holographic Guide

Streaming Performance Comparison

To support bandwidth-efficient XR streaming while maintaining visual fidelity, we implemented a series of pipeline-level optimizations for 3D point cloud content. These enhancements significantly reduced total frame size and processing overhead. By adjusting resolution and applying tuned compression strategies, we achieved meaningful reductions in data volume with minimal perceptual loss. Unlike the 2D workflow, which transmits only a single texture, the 3D configuration encodes and streams two synchronized channels. The table below summarizes key differences in data size and processing time across the 2D and 3D streaming paths

Metric	2D Video	3D Point Cloud
Average Frame Size	20KB/frame	45KB/frame
Encoding Latency	11ms/frame	25ms/frame

Tab.: 2D–3D Comparison (VISTA-XR)

Future Work

Implement synchronized spatial audio streaming.

Further optimize adaptive streaming for varying network conditions.

Expand scalability to support larger concurrent user groups.

Evaluate performance and compatibility on additional hardware.

Contact

Isida Karpuzi & Nikhil Mandrekar
The Carbon Games
play@thecarbongames.com

LinkedIn: https://www.linkedin.com/company/thecarbongames/

the carbongames

References (Selection)

Stereolabs. (2025). ZED SDK Documentation. Stereolabs Developer Portal. https://www.stereolabs.com/developers

Co-funded by the European Union

Project funded by

UK Research and Innovation

www.spirit-project.eu