Videoconferencing System in Virtual Environment

Toleutay Elvira, Gregor Rozinaj
Institute of Applied Informatics Modeling and Simulation of Event Systems
xtoleutay@stuba.sk

Abstract - **This work presents an immersive videoconferencing system based on WebSocket technology and Splat model rendering.The proposed solution enables real-time video communication integrated within a dynamic virtual environment with first-person navigation.**

Unlike traditional platforms, it offers high responsiveness, low hardware requirements, and greater accessibility, bridging the gap between conventional videoconferencing and full VR immersion.

I. Introduction

In today's digitally interconnected world, effective communication is vital not only for personal interactions but also for the efficiency, productivity, and success of organizations globally. The rapid evolution of technology has significantly reshaped traditional means of communication, paving the way for innovative solutions such as videoconferencing systems. Particularly in recent times, the increasing demand for remote working and global collaboration has underscored the importance of robust, interactive, and immersive communication tools. One emerging area that holds immense promise for revolutionizing virtual interaction is the integration of videoconferencing within virtual environments. [1][2]

Videoconferencing systems embedded in virtual environments extend beyond conventional video calls, offering immersive experiences where participants can interact in multidimensional spaces. Such systems promise enhanced user engagement, increased attentiveness, and enriched collaborative experiences by providing realistic avatars, spatial audio, and interactive virtual objects that simulate face-to-face interactions[3]. By mimicking real-life scenarios, these platforms facilitate deeper interpersonal connections, reduce the limitations posed by geographic distances, and provide greater flexibility and inclusivity.[4]

In this bachelor's work, there is beeing developed a prototype of such a system, employing a realistic model of a conference office. Users can navigate this virtual office environment in a first-person perspective, allowing them to seamlessly move within the space and effortlessly connect to video calls. This practical implementation not only demonstrates the feasibility and functionality of immersive videoconferencing solutions but also provides valuable insights into user experience design and technological integration.

II. Existing Solutions

A. Comparison

Several notable existing systems illustrate the potential and ongoing evolution of videoconferencing within virtual environments. One prominent example is Meta Horizon Workrooms[6], developed by Meta (formerly Facebook), which offers users an immersive VR (virtual reality) workspace where they can collaborate as digital avatars, using spatial audio, interactive whiteboards, and intuitive gesture controls.

Another advanced platform, Microsoft Mesh[7], utilizes mixed reality (MR) technology, allowing individuals to interact as lifelike holograms in shared digital spaces, enhancing remote teamwork and collaboration capabilities.

Similarly, VRChat[8] represents a highly interactive social VR platform enabling users to engage socially, conduct meetings, and participate in various virtual events through customizable avatars and immersive worlds.

All three platforms represent different approaches to virtual collaboration and intercation. Meta Horizon Workrooms focuses primarily on professional remote collaboration, providing tools such as shared whiteboards, spatial audio, and seamless integration with traditional video conferencing. Microsoft Mesh pushes the boundaries further into MR by enabling users to appear as holographic avatars, blending physical and virtual spaces across a wide range of devices. In contrast, VRChat emphasizes social interaction and creative freedom, offering highly customizable avatars, community-driven world building, and extensive social event hosting.

While Horizon Workrooms and Microsoft Mesh prioritize productivity and enterprise collaboration, VRChat leans more towards informal, community-oriented interactions. These distinctions highlight the varied design goals: from structured business environments to open-ended social experiences.

Features	**Videocongerencing systems**		
	Meta Horizon Workrooms	***Microsoft mesh***	***VRChat***
Main focus	Remote collaboration	Mixes reality collaboration	Social intercation
Platform type	VR	MR	VR
Avatars	Semi-realistic	Holographic	Fully customizable
Device support	VR Headsets, PC	VR, HoloLens, PC, Mobile	VRHeadsets, PC
Target audience	Progessionals, businesses	Businesses, developers	General public, gamers
Customization level	Medium	Low to medium	Very high

B. Formulas

These existing systems underscore the viability and effectiveness of virtual-environment -integreted videocobferencing. To understand the scientific and visual aspects of these systems, we need to explore how VR is perceived.

In virtual reality systems, the field of view (FOV) is a crucial parameter that defines how much of the virtual world the user can see at once. The horizontal field of view (FOVh) can be calculated by the following formula:

$$Fov_h = 2\,arctan\left(\frac{w}{2f}\right) \tag{1}$$

Where *w*- the width of the display, *f*- the focal length (distance from the lens to the eye). A larger FOV generally increases the sense of immersion, making the virtual environment feel more natural and realistic. VR headset designers often aim to maximize FOV while maintaining visual clarity and minimizing distortions.

Spatial audio is essential for creating a convincing sense of presence in VR and MR environments. The perceived loudness of a sound source based on its distance and angle relative to the listener can be modeled by the following formula:

$$L = \frac{1}{r^2}cos(\theta) \tag{2}$$

Where *L* -the perceived loudness, *r*- the distance between listener and sound source, *θ*- the angle between the listener's forward direction and the source. This model illustrates that sound intensity decreases with the square of the distance and varies based on directional alignment. Accurate spatial audio significantly enhances the realism of interactions in virtual environments, enabling users to detect the direction and distance of sounds naturally.

III. Proposed Solution

In proposed solution, there has been utilized WebSocket connections to enable real-time video calls within an immersive virtual environment, aiming to blend seamless interactive communication with a realistic virtual space. The core of the video conferencing component relies on WebSockets, which facilitate real-time, bidirectional data transfer. This technology allows for minimal latency and a smooth communication experience, crucial for maintaining natural conversational dynamics during virtual meetings.

A. Splat

The virtual environment is constructed using a novel approach known as the Splat model. Unlike conventional polygonal rendering techniques, which rely on meshes and rigid geometric structures, the Splat model represents the environment as a dense cloud of oriented surfels (surface elements) or Gaussian splats. Each splat carries information about position, normal, scale, and color, allowing for a continuous, flexible representation of complex surfaces.

In this system, users navigate freely within the virtual office space through a first-person perspective, enabled by dynamically adjusting the viewpoint over the dense splat cloud. The Splat model provides smoother and more responsive movement by eliminating the traditional constraints of polygonal topology, resulting in continuous surfaces without visible tessellation artifacts.

Additionally, the user experience benefits from intuitive looking around and direct interaction with virtual objects and participants, enhancing both immersion and realism. This method inherently supports level-of-detail (LOD) optimizations, where nearby splats are rendered in high detail and distant splats are approximated with fewer resources.

Thus, the Splat model significantly advances the sense of presence in virtual environments while maintaining computational efficiency, making it particularly suitable for real-time WebSocket-integrated video conferencing systems.

Each splat can be mathematically represented as a 3D anisotropic Gaussian:

$$G(x) = a\,exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \tag{3}$$

Where *x*- 3D position in space, *μ*- center position of the splat, Σ- covariance matrix (describes size, orientation, and elongation of the splat), *a*- amplitude (intensity/color weight).

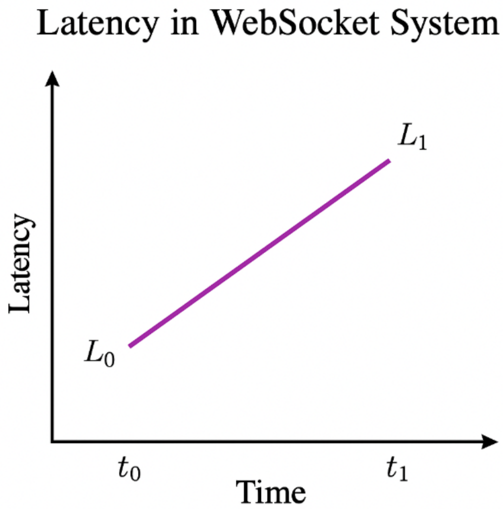
B. WebSocket

WebSocket technology provides a full-duplex communication channel over a single TCP connection, allowing continuous bidirectional data transfer without the overhead of repeatedly opening and closing HTTP requests. This method ensures low latency, high throughput, and a persistent connection, which are critical for maintaining smooth video conferencing experiences and seamless interactions in immersive virtual spaces.

The total latency L of a WebSocket-based transmission can be approximated as:

$L = L_1 + L_2 + L_3$	(4)

Where $L1$- pure network transmission time, $L2$- server-side processing delay, $L3$- any queuing delay at intermediate points.



Higher throughput enables smoother video rendering and more responsive user interactions within the virtual environment.

Compared to traditional HTTP-based communication, WebSocket technology offers significant advantages for real-time virtual environments. While HTTP relies on a request-response model introducing additional latency and overhead for each interaction, WebSocket maintains a persistent, low-latency connection, enabling continuous streaming of data without repetitive handshakes

Although WebRTC is another common protocol for real-time media exchange, it often requires complex session negotiation and is optimized for peer-to-peer scenarios, sometimes increasing setup complexity in multi-user environments. By contrast, WebSocket offers a simpler, lightweight solution that perfectly aligns with the needs of immersive VR systems, where latency minimization and responsiveness are crucial to preserving the user's sense of presence.

Thus, the integration of WebSocket technology provides a scalable, responsive, and efficient communication backbone, making it ideal for real-time video conferencing within immersive virtual spaces.

C. Progress

At the current stage of development, bachelor project features a functional local setup in which a Splat-based model of an office environment is rendered, and first-person navigation (FPS-style) within the virtual space is fully operational.

Presently, the work is at the state on implementing multi-user connectivity for real-time conferencing, aiming to allow participants to join the environment and have their live video streams displayed dynamically on the virtual monitors positioned within the office.

This integration will enhance the realism and interactivity of the environment, supporting seamless communication within the immersive virtual workspace.

D. Some Common Mistakes

- **HTML**- The project includes a main HTML page that serves as the base of the application. This page features a simple CSS styling block to format the loading indicator and interactive buttons, and it connects to the main TypeScript script via a module import.
The HTML also contains hidden screen buttons, which are later dynamically positioned on top of the 3D scene according to the user's viewpoint.
- **Scene Initialization**: Creating a 3D scene and WebGL renderer using the SPLAT library.
- **Camera Setup**: Defining an initial camera position and rotation, and implementing FPS-style (first-person) controls for navigation within the virtual space.
- **Environment Loading**: Loading a 3D environment model (.splat file) to create an immersive office space.
- **Button Projection**: Mapping predefined 3D points to 2D screen coordinates to dynamically position HTML buttons over the corresponding virtual monitors.
- **Animation Loop**: Continuously updating the user's movement, rendering the scene, and adjusting button positions to maintain correct alignment with the virtual world.
- **Real-Time Interactivity**: Detecting when objects come into the user's field of view and enabling corresponding screen interactions.
- For this project, GSplat library[5] is beeing used, a lightweight and efficient 3D engine designed for rendering environments based on Splat models (dense point cloud representations).
GSplat facilitates smooth navigation, flexible camera control, and efficient environment rendering, making it particularly suitable for real-time virtual conferencing scenarios.

IV. Comparison with Existing Solutions and Advantages of the Proposed System

Current virtual conferencing platforms, such as Meta Horizon Workrooms, Microsoft Mesh, and VRChat, have made significant strides in merging communication technologies with immersive environments.
However, these solutions typically focus on either high-end mixed reality (MR) systems requiring specialized hardware (such as HoloLens or Meta Quest headsets) or socially oriented platforms with limited direct integration of real-time video conferencing inside dynamic virtual spaces.

In contrast, the system proposed in this project introduces several key innovations that distinguish it from existing platforms:

a) *Direct real-time video* streaming integration into a freely explorable 3D environment without reliance on heavy MR infrastructure.

b) *Use of WebSocket technology* to maintain a low-latency, bidirectional connection, ensuring a smooth, scalable experience even for multiple participants.

c) *Implementation of the Splat model* for rendering the environment, which provides a lightweight yet immersive spatial experience, optimizing performance across a wide range of devices.

d) *Intuitive FPS-style navigation*, allowing users to move naturally within the virtual space rather than teleporting or being limited to predefined paths, as seen in many VR systems.

e) *Accessibility:* Unlike traditional MR solutions that require expensive hardware, the proposed system can operate on standard VR setups or even conventional computers with minor adjustments, making it more accessible to a broader audience.

This project represents one of the first attempts to combine real-time WebSocket-based video conferencing directly within a dynamic Splat-rendered virtual environment. Such an approach merges the benefits of real-time audiovisual communication with the spatial advantages of virtual reality, creating a hybrid system that improves both presence and interactivity without demanding excessive computational resources.

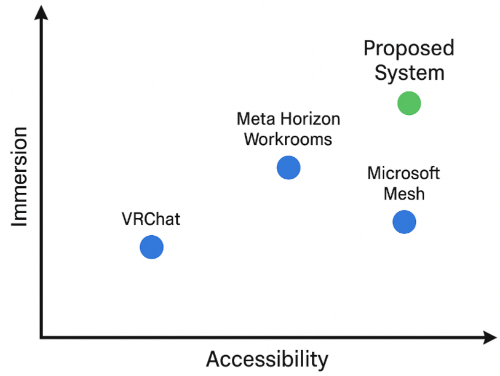
The potential benefits of this system are numerous:

f) *Educational applications:* Virtual classrooms where students can interact through both video and spatial movements.

g) *Corporate collaboration:* Virtual offices where team members maintain face-to-face interaction while navigating shared workspaces.

h) *Remote socialization:* Events, meetups, and collaborative projects where participants feel genuinely "present" together.

i) *Scalability:* The lightweight nature of Splat rendering combined with WebSocket architecture supports expansion to large numbers of users without degradation in performance.



By integrating real-time video streams seamlessly into an actively navigable environment, the proposed system addresses the critical limitations of many existing platforms: the lack of fluid movement, the dependence on costly hardware, and the separation between communication and navigation within virtual spaces.

Thus, this work pioneers a new class of virtual conferencing systems, offering an accessible, highly interactive, and immersive experience that can evolve alongside technological advances in Web communication and 3D rendering.

ACKNOWLEDGMENT

This paper was supported by NEXT (ERASMUS-EDU-2023-CBHE-STRAND-2), EULiST (ERASMUS), and (VEGA 1/0605/23).

REFERENCES

[1] Bailenson, J. N., & Yee, N. (2021). Infinite Reality: Avatars, Eternal Life, New Worlds, and the Dawn of the Virtual Revolution. HarperCollins Publishers.

[2] Schroeder, R. (2010). Being There Together: Social Interaction in Shared Virtual Environments. Oxford University Press.

[3] Biocca, F., & Levy, M. R. (Eds.). (2013). *Communication in the Age of Virtual Reality*. Routledge.

[4] Slater, M., & Sanchez-Vives, M. V. (2016). "Enhancing our Lives with Immersive Virtual Reality." *Frontiers in Robotics and AI*, 3, 74.

[5] Gsplat library URL: https://github.com/huggingface/gsplat.js

[6] Meta Horizon Workrooms – *Meta (Facebook)* **URL**: https://www.meta.com/workrooms/

[7] Microsoft Mesh – *Microsoft* **URL**: https://www.microsoft.com/en-us/mesh

[8] VRChat – *VRChat Inc.* **URL**: https://hello.vrchat.com/