Metaverse Communication System

Jakub Tury, Marek Vančo
Faculty of Electrical Engineering and Information Technology, Slovak Univesity of Technology Bratislava, Slovakia
xtury@stuba.sk, marek_vanco@stuba.sk

Abstract - **This bachelor thesis focuses on the development of virtual meeting rooms and collaborative educational spaces using the FrameVR platform. FrameVR enables web-based creation, navigation, and interaction in immersive 3D environments. Unlike static 2D platforms, FrameVR supports live communication, media presentation, and modeling capabilities in real-time. A key part of this project was evaluating the potential of integrating interactive 3D modeling directly within FrameVR, alongside leveraging Gaussian Splatting for advanced visualization. The result is a set of interactive virtual classrooms and meeting rooms where students can communicate, collaborate, and co-create content inside a shared virtual world.**

Keywords - FrameVR; Virtual Classrooms;Educational Videoconferencing; Gaussian Splatting;

I. Introduction

Technology-enhanced education has played a pivotal role in student learning and development [1]. Following this trend, many educational systems globally have begun exploring emerging technologies to improve engagement and learning outcomes. One notable initiative is South Korea's 'SMART' classroom project, launched in 2011, which aimed to embed Self-directed, Motivated, Adaptive, Resource Free, and Technology-enhanced practices into the education system [2]. This project involved significant investment into digital textbooks and online learning environments, with the goal of modernizing the K–12 curriculum and making content more interactive and accessible.

Although digital textbooks have improved interactivity and multimodal engagement through multimedia sources such as videos, animations, and quizzes, their flat, screen-based interface continues to limit the depth of student interaction [3,4]. Research shows that while they enhance digital literacy and flexibility, students often do not experience a significant difference in learning engagement compared to traditional materials [5,6,7]. To address this, newer methods—such as Augmented and Virtual Reality (AR/VR)—have been introduced to enhance presence and immersion. Despite positive early feedback, VR content has only been implemented in limited subjects like Science and Social Studies, and there is limited research on its broader classroom integration .

This bachelor thesis builds on previous experiments with Mozilla Hubs and Unity, which were initially explored as platforms for hosting scalable virtual student events. While Mozilla Hubs offered early ease-of-use and browser compatibility, it lacked the visual fidelity and scalability needed for immersive academic environments. Unity, though powerful, posed significant accessibility and performance barriers for non-developers. These limitations led to the adoption of FrameVR—a platform combining the accessibility of browser-based tools with advanced support for real-time collaboration and high-performance 3D visualization.

This project presents a practical advancement within this ongoing evolution: implementing a fully immersive educational space in FrameVR and applying the Gaussian Splatting technique to enhance visual fidelity and engagement. Unlike traditional virtual rooms, Gaussian Splatting allows for highly realistic rendering of classroom environments based on real-world image data. My environment within FrameVR is already developed and functional, and this current phase of the project focuses on the implementation and integration of the Gaussian Splatting method into that space. This work aims to explore whether this combination can overcome previous limitations of digital textbooks and VR content by offering a more immersive, flexible, and scalable solution for digital learning environments.

II. Accessibillity and User Experience

FrameVR operates entirely through a browser, making it accessible without any dedicated installation. Room creation is based on visual interfaces, asset uploads, and a scriptable logic system. The scene editor allows for drag-and-drop placement, adjustment of scale, and interactive linking of zones. These features allow students and organizers to co-design immersive experiences collaboratively. Automatic streaming optimization and zone loading make even complex environments performant and stable. Furthermore, voice zones, customizable avatars, real-time language translation, and intuitive interaction controls lower the technical barrier for participation, even for non-technical users.

III. Workflow Pipeline

A. Data collecting

High-quality visual and audio assets are crucial for building an immersive FrameVR environment. The foundation of each virtual room begins with thoughtful and structured data acquisition. Depending on the objective—be it a keynote hall, laboratory simulation, or informal lounge—real-world spatial references are recorded using either high-resolution cameras or 360° video devices. This ensures that physical dimensions, object placement, and perspective are faithfully recreated in the virtual scene.

When capturing reference media, it is important to avoid using wide-angle modes on phones, as these can distort geometry and edges, making them unsuitable for accurate modeling or texture references. If 360° video is used, it is recommended to extract images using cubic projection—

splitting the spherical video into six faces of a cube—to preserve perspective and reduce distortion.

The most effective technique for collecting spatial reference images is to record the scene from multiple heights and angles: one pass at a low angle slightly tilted upward, a second at chest level aimed horizontally, and a final pass above head height tilted slightly downward. This approach ensures full spatial coverage and helps when modeling room layout or generating realistic lighting and interaction zones. These materials serve as a foundation for generating accurate room reconstructions and background environments, which are later enhanced or replaced with interactive digital equivalents. Photographs of whiteboards, posters, and interfaces can also be embedded directly into the space, adding familiarity and practical context for users engaging in lessons or conferences.

B. Image selection and Calibration

The process of preparing content for FrameVR begins with careful selection of source imagery. For virtual conference and educational rooms, quality and consistency are essential. Low-resolution or blurred images are discarded to maintain clarity during environment reconstruction. When working with videos, FrameVR-compatible tools such as Postshot or Blender extract keyframes—ideally 4 to 5 per second—to capture sufficient environmental variance. This ensures broad spatial awareness in the final room layout.

Next, camera poses are calculated to properly align and stitch content in FrameVR. Using a Structure-from-Motion (SfM) algorithm, keypoints within the images are triangulated to determine the spatial position of each frame. Rooms with minimal features, such as blank walls or repeating patterns, present challenges during this step. To mitigate this, objects like signage, posters, or architectural details are often added or emphasized to improve frame alignment. This results in better orientation accuracy when placing the content inside the virtual room.

The actual layout design begins in the FrameVR editor. Rooms are assigned specific roles based on event type—keynote halls include projection screens, stages, and seating rows; breakout zones include circular tables, collaborative canvases, and interactive whiteboards; informal networking areas may include sound zones and casual seating. Asset placement is done using FrameVR's drag-and-drop tools, with fine adjustments made through scripting for custom behaviors. Linking these rooms through teleporters and portals ensures fluid movement through the event.

C. Point Cloud Construction and Virtual Layout

Once spatial orientation has been resolved, the next phase involves converting the calibrated image data into a rough 3D representation. This point cloud, generated from the Structure-from-Motion results, acts as a digital scaffold of the target environment. It provides a visual approximation of walls, objects, seating areas, and overall structure. Within FrameVR, this cloud is not directly visualized but is used as a template to begin placing static and interactive content.

Virtual props—desks, chairs, banners, and screens—are then arranged in accordance with the cloud to replicate real-world positions. At this stage, scale and proportion are critical, especially when matching virtual walking paths with real-world counterparts. This layout becomes the interactive base for FrameVR rooms, with navigation boundaries, interaction hotspots, and visual zones layered on top.

To bring life into these rooms, developers can insert multimedia panels for streamed content, linked web resources, and embedded forms or polls. Session content can be controlled using FrameVR scripting, allowing objects like countdown timers, light triggers, or group-specific visual assets to react to the presence of users. Collaboration tools, such as shared documents or real-time drawing surfaces, are tied to virtual surfaces or 3D props, reinforcing the feeling of physical interaction inside a virtual context. Through FrameVR's API hooks, LMS and CRM tools record metrics such as attendance, quiz responses, or individual movement data across the virtual campus.

Figure 1. Representation of a conference meeting room

D. Creating a point cloud

To create a point cloud, Structure-from-Motion is also used, which tries to find details from the frames it has correctly positioned. Therefore, assigns a point in space to the details and color changes in the frames. With all the points assigned we can usually see the outline of the object we are trying to model. With this point cloud we can tell if the modeling will be of sufficient quality, more points equal more detail in a given location which equals to better quality.

Figure 2. Modified enviroment

E. *Assigning Gaussians to each point*

And very large number of gaussians are used to create the model, which can be read as an ellipsoidal object in space. This gaussian has its own position, dimension, rotation, color and opacity. Thanks to the large number of these elements, we can create our model. For larger and more detailed models, we can use up to several million of these elements. The disadvantage of these models is the poor quality in places, where there were not enough frames or poor quality of the frames. In this case, the models will show artifacts or floaters with poor quality.

F. *Correction of errors on the created model*

Gaussian Embellishment and Optimization To supplement the realism of these digital spaces, Gaussian splatting may be used as a final rendering overlay, particularly for environmental details like walls, backgrounds, or less interactive props. Each Gaussian point represents a smooth 3D ellipsoid with its own orientation, color, and opacity. When layered together, millions of these points form a visually rich texture without the polygonal overhead of traditional mesh models.

FrameVR does not yet support Gaussian models natively, so they are preprocessed externally and converted into compatible visuals. This process ensures optimized rendering for conference zones that prioritize performance. Gaussian overlays can be used to replicate scenic architecture, museum walls, or static classroom backgrounds where high realism is needed but no physical interaction is required.

Because Gaussian splats don't produce mesh-based geometry, they are best reserved for passive decoration. However, when strategically integrated alongside fully interactive glTF assets, the result is a visually rich and lightweight environment. These elements also help bridge the visual gap between scanned real-world locations and procedural virtual content. In the context of a virtual university, such visuals can recreate iconic hallways, branded backdrops, or cityscapes in the background, creating immersion while preserving system performance. The balance of Gaussian visuals and interactable 3D models creates a seamless blend of high performance and visual fidelity suitable for a scalable virtual university campus.

Figure 3. Illustration of the locked space for students for personal lessons

IV. DISADVANTAGES OF GAUSSIAN SPLATTING

While Gaussian Splatting provides a high level of visual fidelity with minimal computational load, it also comes with notable limitations that affect usability in fully interactive FrameVR environments.

Because the model is purely graphical and lacks traditional mesh geometry, it cannot be used for direct user interaction such as collisions, walk paths, or physics-based actions. For users navigating virtual classrooms or conference booths, this absence of physical presence can result in a disjointed experience if not mitigated.

For immersive realism, photogrammetry-based visuals or pre-recorded panoramic content can be embedded. In specific zones, Gaussian Splatting may be used for visual context (e.g., scanning real campus scenes). However, most modeling is done via FrameVR's built-in systems or imported glTF assets.

Figure 4. Gaussian splatting first models

A. *Importing and aligning the mesh*

Once the mesh is created, it must be carefully imported into FrameVR and synchronized in scale and position with the visual model. Alignment is crucial, especially when both Gaussian renderings and interactable props exist in the same space. This dual-layer approach—visual splats for aesthetics, mesh for interactivity—ensures that the scene remains immersive while still functionally navigable.

Advanced tools within FrameVR's object editor support alignment through XYZ adjustments, bounding box references, and snapping to anchor points defined during the mesh creation process. Once aligned, this hybrid structure can be further enhanced with spatial audio and logic-based triggers to simulate realistic environments.

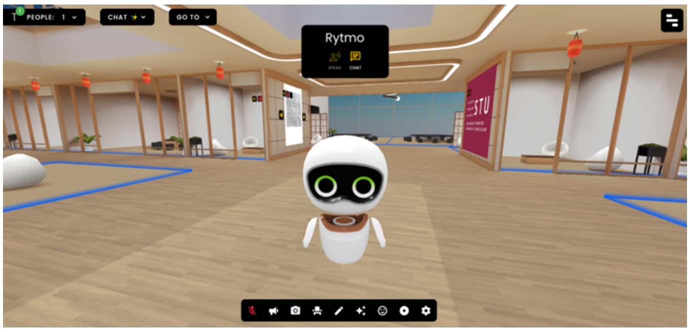

Figure 5. Example of the hallway ai robot which can help in topics from lessons

V. Results

The project yielded a modular, scalable, and highly interactive virtual conference environment designed in FrameVR. It demonstrated the ability to support over 200 users per session through the use of breakout rooms and linked Frames, each tailored for different conference functions such as keynote halls, collaboration zones, and informal networking areas.

Key communication features such as real-time chat, multi-screen sharing, embedded polls, and live video feeds were seamlessly integrated. The environment also included adaptive voice zones with configurable ranges, moderation controls, and language translation powered by AI captioning and real-time interpreters.

Navigation throughout the virtual campus was intuitive and flexible, leveraging portals and linkable teleportation points to create a coherent user journey between various rooms. The FrameVR editor allowed for real-time customization of content and assets during live events, enabling both organizers and participants to adapt sessions dynamically based on the audience or context.

Analytics and behavior tracking tools were implemented to monitor user engagement and traffic flow. Heatmaps and usage logs provided organizers with insight into the most visited zones, session durations, and interaction frequency.

User testing indicated a high level of satisfaction across several metrics. Participants noted minimal latency during sessions, stable media performance across devices, and ease of use even for those with limited technical backgrounds. The system also supported hybrid event structures, allowing physical sessions to be streamed directly into the virtual platform and enabling remote attendees to interact with real-time content.

Organizers benefited from robust management features including access control, user role assignment, embedded branding opportunities, and automatic archiving of session materials and interactions for later review. Overall, the setup has proven capable of delivering professional, engaging, and technically reliable academic events that extend beyond the limitations of traditional physical spaces.

VI. Conclusion

This thesis confirms that FrameVR is a viable platform for hosting large-scale academic or educational events. With added support for dynamic content, translation, and real-time conferencing tools, FrameVR can replicate and extend the functionality of real-world venues. The combination of scalable environments, real-time interaction, and web accessibility supports a new generation of hybrid and international academic collaboration. Future work can explore integration with AR wearables, more sophisticated gamification mechanics, and tighter links to academic management systems.

Figure 6. Example of the presentation and interacting

Acknowledgment

This paper was supported by DISIC (09I05-03-V2), NEXT (ERASMUS-EDU-2023-CBHE-STRAND-2), EULiST (ERASMUS), CYB-FUT (ERASMUS+), InteRViR (VEGA 1/0605/23).

References

[1] FrameVR Inc. "Getting Started with FrameVR: Platform Overview and Best Practices." https://learn.framevr.io/start.

[2] FrameVR Inc. "Advanced Features for Education and Events." https://learn.framevr.io/features.

[3] FrameVR Community Blog. "Case Study: Hosting International Virtual Conferences with FrameVR." https://blog.framevr.io.

[4] Postshot Documentation. "Optimizing Image-Based Modeling for Web VR." https://www.jawset.com/docs.

[5] Blender Foundation. "Exporting 3D Assets for Web-Based Platforms." https://docs.blender.org.

[6] Google Developers. "Integrating Cloud Translation API into Virtual Spaces." https://cloud.google.com/translate.

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[8] Joo, Y.J.; Park, S.; Shin, E.K. Students′ expectation, satisfaction, and continuance intention to use digital textbooks. Comput. Hum. Behav. **2017**, 69, 83–90

[9] Pande, P.; Thit, A.; Sørensen, A.E.; Mojsoska, B.; Moeller, M.E.; Jepsen, P.M. Long-term effectiveness of immersive VR simulations in undergraduate science learning: Lessons from a media-comparison study. Res. Learn. Technol. 2021, 29, EJ1293535.

[10] Schittek, M., Mattheos, N., Lyon, H.C., and Attstrom, R. (2001) Computer Assisted Learning: A Review.European Journal of Dental Education 5:93-100