Verifiable Reality: Contrasting Approaches to Photorealistic VR Using NeRF Streaming and Gaussian Splatting Technologies

Basile Simon
Starling Lab for Data Integrity
Stanford / USC
basile@stanford.edu

Mike Caronna
Starling Lab for Data Integrity
Stanford / USC
mike@starlinglab.org

Maciej Żemojcin
maciej@maciej-zemojcin.com

Paweł Szarzyński
Artem Ivanenko

Abstract

This paper presents two complementary approaches to immersive, photorealistic 3D environments in virtual reality, each utilizing different volumetric rendering techniques.

The first project implements a NeRF-based User Photo Selection System designed for remote streaming. This system leverages a render node with NVIDIA CUDA cores to process Neural Radiance Fields while using a lightweight low-poly proxy geometry for efficient interaction. The architecture streams rendered content to VR headsets via Unreal Engine's pixelstreaming, with testing across LAN and WiFi connections. This approach keeps computational demands off the headset while maintaining visual fidelity and interactive capabilities derived from verifiable source media.

The second project employs 3D Gaussian Splatting within Unity, offering both direct headset deployment and streaming options. This implementation features a complete pipeline from photo capture through alignment and training to real-time rendering on the Oculus Quest 2. The system integrates teleportation-based navigation to minimize motion sickness and includes a photo gallery showing source imagery for provenance and authentication. By utilizing the performance advantages of Gaussian Splatting over traditional NeRF, this approach can potentially run directly on VR hardware while also supporting Airlink/SteamVR streaming for flexibility.

Both projects address the challenge of creating verifiable and trustworthy interactive 3D reconstructions from photographs while taking different approaches to the deployment/streaming balance. This work **contributes to the emerging field of authenticated volumetric content with verifiable data integrity for VR applications, grounded in principles of content provenance, with particular relevance to documentation, education, and digital preservation of spatially complex environments.**

Problem Statements

Virtual reality offers unprecedented opportunities for immersive spatial experiences, but several fundamental challenges have limited the creation of photorealistic, interactive 3D environments with verifiable authenticity. That is, how can a user trust that what they are seeing is a faithful representation of a real place and time, and how can they inspect the history (or provenance) of the data? Our work addresses these limitations through two distinct technical approaches.

Challenges in NeRF-based VR Experiences

Neural Radiance Fields (NeRF) provide exceptional visual quality for reconstructing environments from photographs, but their computational demands make real-time rendering on standalone VR headsets prohibitive. Additionally, while NeRF excels at visual reproduction, it lacks inherent mechanisms for object selection and interaction that are essential for engaging VR experiences.

The first project tackles these challenges by asking: How can we create an interactive NeRF-based environment that maintains visual fidelity while enabling intuitive object selection? And further, can we deliver this experience to VR headsets of varying computational capabilities by utilizing the streaming protocols built into modern game engines?

In specific environments, such as the underground locations with challenging neon lighting mentioned in our documentation, traditional photogrammetry approaches proved inadequate. The capture application performed "ridiculously slow" in these conditions, limiting the number of usable photos, while the unusual lighting made resolving accurate photogrammetric models challenging.

Limitations in Current 3D Gaussian Splatting Implementation

3D Gaussian Splatting offers a more efficient alternative to NeRF, potentially enabling direct deployment on VR headsets. **However, this newer technology lacks established pipelines for end-to-end implementation from photo capture to VR interaction** and, similar to NeRF, it lacks inherent mechanisms for object selection and interaction.

Our second project addresses the question: Can we create a workflow for 3D Gaussian Splatting that enables both direct deployment on VR hardware and streaming options while maintaining source data authenticity?

Shared Concerns Across Both Approaches

Both projects confront several common challenges:

1. **Data Integrity and Authentication Verification**: How can users verify the authenticity of what they're seeing in VR and trace it back to original source imagery? More fundamentally, how can we cryptographically secure the entire "photon-to-polygon" (or "photon-to-splat") pipeline? This includes securing the initial capture (e.g., using C2PA-enabled hardware or mobile applications like ProofMode), ensuring the integrity of the model during training and storage (e.g., using content-addressing like CIDs), and presenting this chain of custody to the end-user in an intuitive way.
2. **User Comfort**: How can we enable fluid navigation within these environments while minimizing motion sickness and other VR-related discomforts?
3. **Performance Optimization**: What techniques can maintain visual quality while ensuring the frame rates necessary for comfortable VR experiences?
4. **Network Dependency**: How can these systems perform effectively across varying network conditions and standard internet connections?

Our work explored two distinct technical approaches to tackle these challenges and uncovered valuable insights into the trade-offs and potential of different volumetric rendering technologies for creating verifiable, interactive VR environments.

Methodology

The systems were primarily tested on Oculus Quest 2 headsets, with the NeRF approach streaming content via a dedicated render node equipped with NVIDIA GPUs supporting CUDA cores. The Gaussian Splatting implementation tried out both direct deployment and streaming via Oculus AirLink for comparison testing.

Provenance integrity verification was implemented using a multi-ledger approach. This registration of content hashes (or CIDs) creates immutable timestamps and proof-of-existence. We implemented references to AVALANCHE, NUMBERS, ISCN, FILECOIN (for decentralized, content-addressed storage), and emerging standards like C2PA providing a robust chain of verification (or “provenance chain”). This allowed users to trace any point in the environment back to its original source imagery and verify capture metadata including location, time, and equipment used. (Figure 7)

NeRF-based Selection System

The first approach centers on streaming NeRF content from a powerful render node to VR headsets. Key elements include:

1. **Low-Poly Proxy Geometry**: We developed a simplified collision mesh that mirrors the visible

environment but with drastically reduced polygon counts. This geometry is divided into smaller parts based on surface curvature or clustering, with each section assigned a Unity box collider component. The system utilizes a specific file hierarchy, with transforms.json serving as the core data structure linking spatial coordinates with image data. Selectable objects are exported as nested FBX files with metadata stored as attributes, enabling seamless reference back to the original source imagery with verifiable provenance.

2. **Selectable Objects System**: Using transforms.json data from the NeRF generation process, we create interactive objects that represent photos in 3D space. When a ray intersects the proxy geometry, the system retains a reference object containing file name, image resolution, and camera data.

3. **Selection Mechanism**: The system captures reference points in space when rays intersect the proxy geometry, then performs near-point detection using a defined radius around this intersection. Detected objects are stored in a list and sorted by distance.

4. **Streaming Architecture**: The render node processes NeRF data using NVIDIA CUDA cores and streams the output via Unreal Engine's pixelstreaming feature. This approach was tested across LAN and WiFi connections to ensure viable performance. The implementation relies on Unreal Engine's .nvol filetype for importing NeRF data structures. Performance across both LAN and WiFi was assessed qualitatively by the team for each scenario.

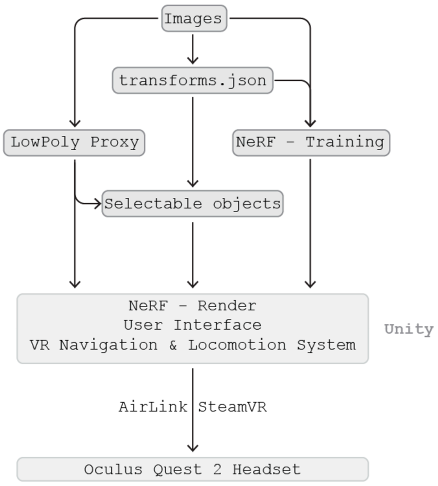

Fig 1: Data flow, NeRF system

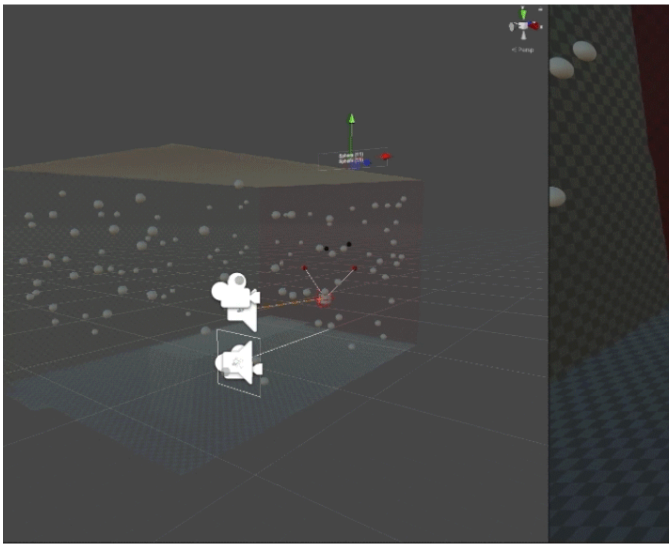

Fig 2: Selection system

Gaussian Splatting Approach

The second approach uses 3D Gaussian Splatting with the following key components:

1. **Data Acquisition Pipeline**: Photos are captured with at least 70% overlap between consecutive images. Ideally, this capture process itself would be authenticated at the source using standards like C2PA or tools like ProofMode to create an initial, verifiable record of provenance (time, location, device) before any processing begins. After filtering out blurry photos, orientation parameters are assigned, and real-world scale is established using known-sized elements in the scene. Our Unity implementation requires a specific directory structure with a 'model' directory containing 'cameras.json' and 'point_cloud/iteration_7000/point_cloud.ply' for

proper functionality. The system optimizes performance by defaulting to DX12 on Windows to leverage a more efficient GPU sorting routine.

2. **Format Conversion**: Camera orientation parameters are exported using the bundler exchange format (v0.4) and subsequently converted to COLMAP format using the Kapture application (github.com/naver/kapture). This standardized format ensures compatibility with the Gaussian Splatting training pipeline.
3. **Training Process**: The Gaussian Splatting method generates models at two quality levels: an initial version after 7,000 iterations and a higher-quality model after 30,000 iterations. The training process generates two key model versions: an initial representation after 7,000 iterations, serving as a draft model, and a higher-fidelity version after 30,000 iterations. Model quality is assessed using the SIBR viewer before integration into the Unity environment.
4. **VR Implementation**: Unity XR setup with a 6DOF camera rig enables precise movement. Teleportation is used for navigation to minimize motion sickness. The system includes a photo gallery that displays the original source images.
5. **Deployment Options**: The system supports both direct deployment on Oculus Quest 2 via side-loading and streaming options via Airlink/SteamVR.

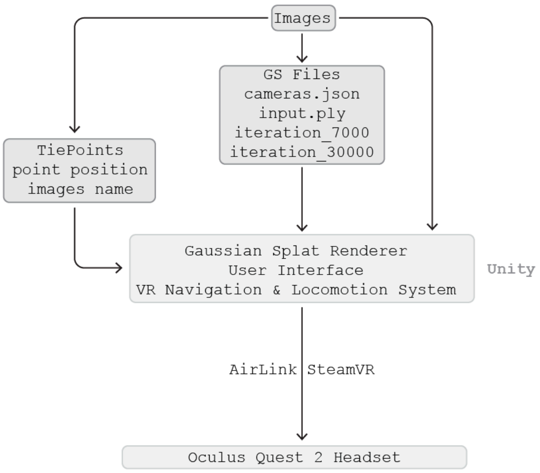

Fig 3: Data flow, GS system

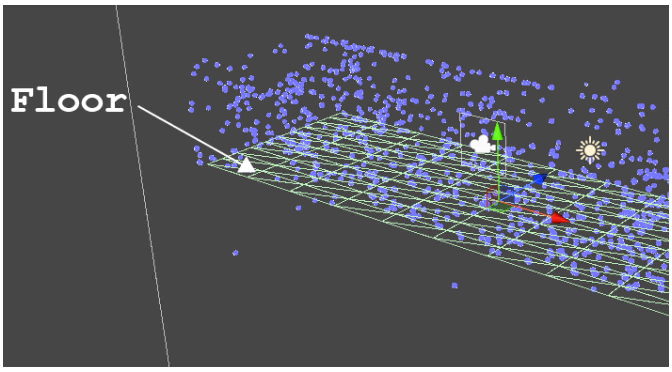

Fig 4: Surfacing original photos

Results

Network performance was evaluated for both LAN and WiFi configurations, with particular attention to latency thresholds that might induce VR motion sickness. Testing protocols established minimum bandwidth requirements for acceptable quality streaming with the NeRF implementation.

NeRF-based System Performance

The NeRF-based system demonstrated reliable performance across various network configurations.

Key findings include:

1. **Streaming Viability**: Pixelstreaming provided acceptable quality over LAN and WiFi connections, with some degradation over standard internet connections.
2. **Interaction Accuracy**: The low-poly proxy geometry provided sufficiently accurate collision detection for intuitive selection of objects within the NeRF environment.
3. **Authentication and Provenance Functionality**: The system successfully maintained connections between points in the 3D environment and their source photos, allowing users to verify

content authenticity and inspect its provenance. (Figure 5)

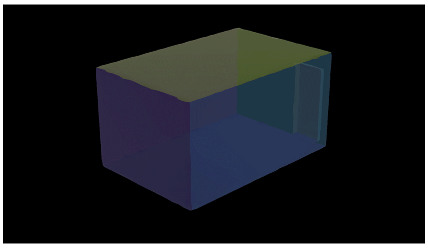

Fig 5: NeRF projection of anchors

[A screen recording](#) of the NeRF render is available

Gaussian Splatting System Evaluation

The Gaussian Splatting approach showed several advantages over the NeRF-based system:

1. **Performance Efficiency**: The optimized Gaussian Splatting renderer achieved higher frame rates than the NeRF implementation, making it more suitable for direct VR headset deployment.
2. **Visual Quality**: While both approaches produced high-quality visuals, the Gaussian Splatting method after 30,000 iterations showed superior detail in complex lighting situations.
3. **User Experience**: Teleportation-based navigation proved effective at minimizing motion sickness while maintaining presence within the environment (Figure 6)
4. **Dual Deployment Flexibility**: The ability to deploy either directly on headsets or via streaming provided valuable flexibility depending on available hardware.

[A screen recording](#) of the 3DGS render is available.

Fig 6: Teleportation in 3DGS

Discussion

Comparative Analysis

When comparing the two approaches, several trade-offs emerge:

1. **Computational Requirements**: The NeRF-based system requires more powerful hardware but can potentially deliver higher visual fidelity. The Gaussian Splatting approach sacrifices some visual quality for improved performance on less powerful hardware.
2. **Deployment Flexibility**: The Gaussian Splatting approach offers more deployment options, including direct headset installation, while the NeRF system is fundamentally streaming-dependent.
3. **Development Complexity**: The NeRF system's separation of visual rendering and interaction geometry adds complexity but provides more optimization opportunities. The Gaussian Splatting approach offers a more straightforward implementation pipeline.
4. **Integrity and Provenance Integration**: Both systems were designed to connect 3D points back to source images, which is the foundation of verification. The Gaussian Splatting approach, with its explicit photo gallery (Fig 7), provided a more direct user-facing mechanic for provenance inspection. However, the underlying challenge for both is formally documenting the entire pipeline – from C2PA-stamped source images to the processing steps and the final CIDs of the 3D models – in a standardized, verifiable manifest. Neither NeRF nor GS has an inherent advantage here.

Limitations

Both NeRF streaming and Gaussian Splatting technologies encounter limitations, particularly concerning environmental complexities. While complex lighting environments pose challenges for both, Gaussian Splatting demonstrates greater robustness in difficult lighting scenarios.

Scaling is another significant concern for both approaches. As virtual environments expand in size, maintaining consistent performance becomes increasingly difficult. Future research could focus on dynamic loading techniques to mitigate these performance challenges in larger-scale applications.

The user experience for both systems is heavily reliant on network conditions. Although both NeRF-based and Gaussian Splatting systems can operate across various network configurations, the quality of the experience, especially for the NeRF-based approach, is directly dependent on the stability and bandwidth of the internet connection.

Finally, a key limitation in the current implementation is the user experience of verification. While the data backend (multi-ledger, CIDs) is robust, presenting this complex provenance information to a user in an intuitive, non-obtrusive way within a VR headset remains a significant design challenge. Figure 7 shows a 2D panel, but future work must explore more deeply integrated and immersive ways to convey trust and authenticity.

AUTHENTICATION CERTIFICATE

AUTHOR

Captured by: Hala Systems
Captured Using: Starling Capture app

CAPTURE INFORMATION

Data: 21.06.2023
Time: 13:52:00
Location: Hala Systems (Бородянка) Ukraine

Verifited by Starling Lab

Fig 7

Conclusion

This paper presented two complementary approaches to creating immersive, photorealistic 3D environments in VR with a focus on content authenticity, data integrity, and verification. The NeRF-based streaming system provides high visual quality but requires more powerful hardware and stable network connections. The Gaussian Splatting approach offers greater deployment flexibility and potentially better performance on standalone headsets.

Both methods successfully address the challenge of creating verifiable, interactive 3D reconstructions from photographs. The pipelines developed for each approach provide valuable frameworks for future work in authenticated volumetric content and verifiable digital provenance for VR applications.

Future research could explore hybrid approaches that combine the strengths of both methods, further optimize performance for larger environments, and develop more sophisticated interaction mechanisms within these photorealistic spaces. **Ultimately, as content generated by these technologies become indistinguishable from reality, building them on a foundation of verifiable provenance is not just a feature, but an ethical necessity.**

References

1. Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3D Gaussian Splatting for Real-Time Radiance Field Rendering.
2. Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis.
3. Epic Games. (2023). Unreal Engine Pixelstreaming Documentation.
4. Coalition for Content Provenance and Authenticity (C2PA). (2024). C2PA Technical Specification