XROI-GS: Real-time XR Interactive Inspection of High-quality Objects of Interest in a 3D Gaussian Splats Scene

Quoc-Anh Bui[1,2], Serhane Lim[1], Tom Boudard[1], Gilles Rougeron[1], Simone Gasparini[2], and Géraldine Morin[2]

[1] Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France
[2] Université de Toulouse, Toulouse INP – IRIT, France

Abstract. We present XROI-GS, an immersive XR framework enabling real-time exploration and interaction with complex real-world scenes. The system addresses the challenge of delivering photorealistic detail for user-specified objects within large-scale scenes while maintaining real-time XR performance. Our demonstration showcases a cultural heritage scene from the Hôtel de la Marine museum, where users can freely explore the scene space and naturally interact with historical artifacts using controller-free virtual hands. Our contribution is an automatic mechanism that significantly improves the visual quality of objects: when a user approaches or interacts with an artifact, the system enhances its resolution to increase rendering quality. In addition, we provide physically based manipulation for realistic and intuitive object handling. This framework democratizes access to fragile cultural treasures by enabling close inspection and interaction impossible in traditional museum settings, opening new possibilities for digital heritage preservation and public engagement.

Keywords: eXtended Reality (XR) · Cultural Heritage · 3D Gaussian Splatting · Level of Details · Scene Composition · Real-time Interaction.

1 Introduction and Context

The transformative potential of eXtended reality (XR) technologies in cultural heritage preservation and presentation has been demonstrated across numerous applications, from digital conservation [7] and virtual exploration [3] to interactive museum experiences [8]. Unlike traditional multimedia formats — static images, videos, or basic 3D reconstructions such as textured meshes and colored point clouds with limited photorealism — XR technologies enable truly immersive, interactive, and engaging experiences.

Recent neural rendering advances, particularly 3D Gaussian Splatting (3DGS) [6], offer photorealistic 3D reconstructions for XR. Deploying these high-fidelity models in resource-constrained XR remains challenging due to their huge memory footprint, especially for detailed inspection of specific artifacts. Several works have extended 3DGS for interactive and object-focused applications [10]. VR-GS [5] addresses physics-aware interaction in VR by integrating XPBD simulation with 3DGS, but its object-level reconstruction supports only manipulation, not enhancing object visual fidelity. Object-Centric 2D Gaussian Splatting [12] achieves accurate surface reconstruction via background removal and occlusion-aware pruning but depends on high-quality segmentation masks. 3DGS-DET [2] improves object-background distribution for detection tasks, while COB-GS [14] and SAGD [4] resolve segmentation boundary ambiguity. Despite the advances, these existing approaches either encode entire scenes without considering user-defined Regions Of Interest (ROIs), require consistent high-quality object masks, or lack adaptive view selection focused on ROIs. Moreover, apart from VR-GS, most of the existing 3DGS methods lack support for direct XR deployment.

To overcome these limitations, we introduce XROI-GS, which combines automatic object-focused camera selection for targeted ROI training with efficient scene composition to enhance user-specified objects while keeping moderate quality elsewhere. We also provide an integrated XR framework for real-time immersion and interaction through VR headsets. The rest of the paper is structured as follows: Section 2 presents the demonstration, Section 3 details the XROI-GS architecture and system, and Section 4 concludes.

2 Demonstration Experience

The demonstration showcases an immersive eXtended Reality experience that enables users to explore an 18th-century dining room at the Hôtel de la Marine, a prestigious museum in Paris, France.

Leveraging our previous work ROI-GS [1], this application provides detailed exploration and interaction with historical artifacts in a cultural heritage setting. The scene is reconstructed from 916 wide-angle images (8192×5464) and 222 close-ups (3024×4032). For this demonstration, six objects of interest — ornate bowls, vases, plates, and statuettes — are designated as ROIs and reconstructed using object-aware camera selection and targeted training. This strategy yields high-fidelity 3D Gaussian representations that preserve fine surface details of the objects while maintaining real-time rendering performance on the whole scene, suitable for XR applications.

Users explore the dining room at 1:1 scale within a free-walk area, enabling natural spatial perception and physical movement inside the scene without the need of teleportation. This preserves the sense of presence and spatial continuity essential for an authentic examination. The Meta Quest headset provides controller-free interaction through real-time hand tracking via onboard camera-based image analysis, allowing users to reach and grasp objects using natural hand gestures. This intuitive interaction paradigm requires no prior skill, knowledge, or training, making the experience accessible to diverse audiences. To enhance realism, we integrated the XDE physics engine [9], simulating weight, collisions, and momentum as objects are picked up, rotated, or repositioned. Visual affordances guide interaction through a subtle rhythmic pulsing glow highlight-

Fig. 1. An example interaction with an object in the scene using the virtual hand.

ing manipulable artifacts and nearby pop-up labels providing historical context. A key feature of our demonstration is the dynamic adaptive Level Of Detail (LOD) rendering: distant objects remain at a lower quality for performance optimization but are enhanced to high fidelity when approached or interacted with, thus revealing the fine surface details.

This demonstration directly addresses critical challenges in cultural heritage preservation and public engagement. By combining the detail-focused reconstruction capabilities of XROI-GS with immersive XR interaction, fragile and valuable artifacts can be examined closely and safely by the general public without risk of physical damage or handling restrictions. The realistic virtual setting preserves the prestige and context of the original museum environment while democratizing access, enabling visitors to engage with cultural treasures through tactile interaction and close inspection — experiences that would be impossible or prohibited in traditional museum contexts.

3 XROI-GS

Our XROI-GS framework comprises two distinct phases: an offline scene learning phase and an online immersion phase. The first phase, the scene reconstruction, is largely based on our previous work ROI-GS [1]: the resulting scene assets are then fed into the XR rendering pipeline to perform high-quality, real-time immersive experiences.

3.1 ROI-GS Architecture

The offline component ROI-GS is built on the standard 3DGS method [6]. This method takes as input a set of 2D color images along with camera poses, intrinsic parameters, a sparse point cloud, and models the scene as multiple sets of anisotropic 3D Gaussians to produce highly realistic renderings. The method employs user-defined axis-aligned bounding boxes (AABBs) enclosing objects of

Fig. 2. Comparison of Object isolation method: AABB Boxes (1) vs. Boundary-aware Gaussian Splitting (2).

interest to guide an advanced camera selection technique that selects optimal object-focused viewpoints for corresponding ROI training. Scene components consisting of a global scene model (Scene-GS) and individual object models (Object-GS) are trained independently, then composed by replacing Scene Gaussians within object bounding boxes with the more detailed Object Gaussians.

During preprocessing, we used the Colmap Structure-from-Motion (SfM) [13] algorithm to perform image calibration on the input 2D images, obtaining camera poses for each view, a sparse 3D keypoint cloud, but also a global triangle mesh of the entire scene. As shown in [1], training with all available scene views to reconstruct a given object both degrades its level of detail because of low resolution rendering from distant cameras and reduces computational efficiency due to numerous images: we thus apply the proposed view selection workflow [1].

Improving upon the manual object AABB annotation in ROI-GS, we developed an intuitive interface tool that enables users to select objects of interest with just a few clicks in some overview scene content images. This tool leverages the SAM 2 [11] segmentation model to obtain consistent object masks, which are projected onto the 3D keypoint cloud to automatically compute object AABBs defining our ROIs. Furthermore, to achieve precise object separation, we additionally adapt the Boundary-aware Gaussian Splitting technique inspired from SAGD [4]. This approach splits edge Gaussians using object masks described above to reduce boundary spike artifacts, producing clean object representations that accurately conform to surface geometry. Therefore, we obtain Scene-GS in which the ROIs have been extracted mand Object-GSs that offer an accurate representation of the object surfaces. As the results in Fig. 2 show, this approach is well-suited for seamless object compositing into the scene and interaction.

3.2 XR Framework

Rendering Performance Considerations. XR applications impose quite high performance requirements compared to traditional computer graphics standards.

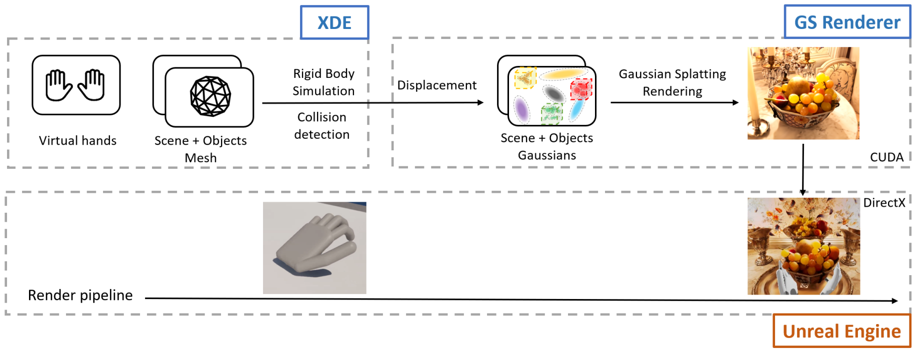

Fig. 3. Our XROI-GS pipeline integrates physics simulation XDE for object manipulation, then GS renderer for the LOD of the scene components, with final XR output handled by Unreal Engine.

Our system employs a Meta Quest 3 VR headset that renders stereoscopic images at 2000 $\times$ 2000 pixels per eye, representing a higher resolution than most 3DGS VR studies. Furthermore, due to the sensitivity of human visual perception to temporal artifacts, VR headsets must maintain a minimum refresh rate of 72 FPS, or approximately 13.8 ms per frame. These demanding constraints make real-time performance critical.

Our GS renderer, derived from INRIA's original implementation [6], uses a tile-based CUDA architecture that takes scene component representations as input and generates both RGB images and estimated depth maps as output. The renderer handles multiple sets of Gaussians simultaneously, thus enabling the seamless integration of several dynamic ROIs within the global scene representation. We precompute spatial partitioning structures for each Gaussian set to optimize runtime performance, allowing efficient frustum culling of entire ROIs and portions of the global scene before frame rendering. This preprocessing strategy adds minimal computational overhead, ensuring rendering cost does not scale with the total number of objects in the scene, but rather with the total count and distribution of rendered visible gaussians. However, the fixed-size tile architecture slows down rendering when many Gaussians cluster in small screen regions, reducing parallel efficiency, particularly when detailed objects are viewed from distant viewpoints. This limitation underscores the necessity of LOD mechanisms in XR. In this work, we used an NVIDIA RTX 4090 graphics card for GS rendering and achieved around 72 FPS performance for stereoscopic rendering on the Meta Quest 3 headset, demonstrating that the system meets real-time requirements.

Physics Simulation and Interaction. Beyond photorealistic rendering, our XR framework incorporates a physics engine that provides users with visual pseudo-haptic feedback, significantly enhancing immersion. Real-time hand tracking enables natural interactions - such as grabbing, manipulating, and releasing actions

- while objects respond realistically to collisions and gravity. Physical properties like mass and material are manually estimated to ensure plausible dynamic behavior during user interactions.

The framework uses the XDE engine to power physics computations, operating on surface meshes rather than volumetric or point-based geometries. Consequently, we must obtain collision meshes for each set of Gaussians representing interactive objects. The global scene mesh is computed through triangulation of the point cloud generated during COLMAP preprocessing. For individual objects, we employ the geometry-aware separation technique described in Section 3.1 (excluding the Gaussian splitting step), which utilizes segmentation masks to precisely isolate object meshes without corrupting neighboring geometry. These separated meshes are then processed in Blender to produce watertight manifold representations suitable for robust physics simulation.

Unreal Engine Integration. To construct our demonstration, we leverage the comprehensive VR capabilities provided by Unreal Engine 5 (UE 5), as shown in Fig. 3. The integration of our CUDA-based GS renderer into UE is achieved through a custom plugin utilizing CUDA-DirectX interoperability, enabling efficient GPU memory sharing between the renderer and the engine's graphics pipeline. The Gaussian render pass is incorporated as a separate stage in the Unreal rendering pipeline, with composition performed based on the depth map generated by the GS renderer. The plugin architecture maintains real-time performance while enabling artists and developers to use familiar UE tools for scene composition, lighting setup, and user interface design.

4 Conclusion, perspective

We presented XROI-GS, a framework demonstrating the practical application of ROI-GS methodology in XR environments, enabling exploration of complex 3D scenes with interactive and high resolution objects of interest while maintaining real-time performance for immersive experiences. The system opens interesting opportunities across multiple domains, including cultural heritage preservation, industrial supervision, and virtual tourism. Despite its effectiveness, challenges warrant future investigation. The depth-guided scene composition can produce visual artifacts when depth estimates are inaccurate. Moreover, the not-fully-tight coupling between UE and our GS renderer restricts advanced lighting effects: manipulated objects retain independently learned lighting and static shadows, preventing lighting interaction with additional 3D models, or coherent relighting of 3DGS objects and scene during manipulations. Additionally, future work includes implementing inpainting techniques to fill holes left when moving objects, improving GS representation for higher-quality reconstruction in both geometry and appearance, and support for collaborative multi-user experiences.

Acknowledgments. We would like to thank the Centre des Monuments Nationaux and the team of the the Hôtel de la Marine for providing access to this magnificent place. The dataset will be publicly released upon publication of the paper.

References

1. Bui, Q.A., Rougeron, G., Morin, G., Gasparini, S.: Roi-gs: Interest-based local quality 3d gaussian splatting. In: Accepted for International Conference on Visual Communications and Image Processing (VCIP) December, 2025 (2025), https://arxiv.org/abs/2510.01978
2. Cao, Y., Ju, Y., Xu, D.: 3dgs-det: Empower 3d gaussian splatting with boundary guidance and box-focused sampling for 3d object detection. arXiv preprint arXiv:2410.01647 (2024)
3. Haibt, M.: End-to-end digital twin creation of the archaeological landscape in uruk-warka (iraq). International Journal of Digital Earth **17**(1) (Mar 2024). https://doi.org/10.1080/17538947.2024.2324964
4. Hu, X., Wang, Y., Fan, L., Fan, J., Peng, J., Lei, Z., Li, Q., Zhang, Z.: Semantic anything in 3d gaussians. arXiv preprint arXiv:2401.17857 (2024)
5. Jiang, Y., Yu, C., Xie, T., Li, X., Feng, Y., Wang, H., Li, M., Lau, H., Gao, F., Yang, Y., Jiang, C.: Vr-gs: A physical dynamics-aware interactive gaussian splatting system in virtual reality. In: Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers. p. 1–1. SIGGRAPH '24, ACM (Jul 2024). https://doi.org/10.1145/3641519.3657448
6. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3D Gaussian Splatting for Real-Time Radiance Field Rendering. ACM Transactions on Graphics **42**(4) (07 2023)
7. Kong, X., Hucks, R.G.: Preserving our heritage: A photogrammetry-based digital twin framework for monitoring deteriorations of historic structures. Automation in Construction **152**, 104928 (Aug 2023). https://doi.org/10.1016/j.autcon.2023.104928
8. Liu, Z., Chang, S.: A study of digital exhibition visual design led by digital twin and vr technology. Measurement: Sensors **31**, 100970 (Feb 2024). https://doi.org/10.1016/j.measen.2023.100970
9. Merlhiot, X., Le Garrec, J., Saupin, G., Andriot, C.: The xde mechanical kernel: Efficient and robust simulation of multibody dynamics with intermittent nonsmooth contacts. In: The 2nd joint international conference on multibody system dynamics. vol. 13, p. 84 (2012)
10. Qiu, S., Xie, B., Liu, Q., Heng, P.A.: Advancing extended reality with 3d gaussian splatting: Innovations and prospects. In: 2025 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR). p. 203–208. IEEE (Jan 2025). https://doi.org/10.1109/aixvr63409.2025.00039
11. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K.V., Carion, N., Wu, C.Y., Girshick, R., Dollár, P., Feichtenhofer, C.: Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714 (2024), https://arxiv.org/abs/2408.00714
12. Rogge, M., Stricker, D.: Object-centric 2d gaussian splatting: Background removal and occlusion-aware pruning for compact object models. In: Proceedings of the 14th International Conference on Pattern Recognition Applications and Methods. p. 519–530. SCITEPRESS - Science and Technology Publications (2025). https://doi.org/10.5220/0013305500003905
13. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (Jun 2016). https://doi.org/10.1109/cvpr.2016.445

14. Zhang, J., Jiang, J., Chen, Y., Jiang, K., Liu, X.: Cob-gs: Clear object boundaries in 3dgs segmentation based on boundary-adaptive gaussian splitting. In: CVPR (2025)