An Efficient Neural Scene Generation Pipeline for High-Quality XR Content

Hye Sun Kim
Content Research Division
ETRI
Daejeon, Korea
hsukim@etri.re.kr

Sung Jin Hong
Content Research Division
ETRI
Daejeon, Korea
sjhong0117@etri.re.kr

Cho Rong Yu
Content Research Division
ETRI
Daejeon, Korea
crryu@etri.re.kr

Youn Hee Gil
Content Research Division
ETRI
Daejeon, Korea
yhgil@etri.re.kr

***Abstract*— Gaussian Splatting has shown strong potential for 3D scene representation and view synthesis, but its direct use in real-time RGB-D SLAM is limited by pose drift and redundant gaussian insertions, restricting its applicability to XR content generation. We propose an efficient post-processing pipeline that integrates learning-based pose refinement with quadtree-based adaptive gaussian allocation. Pose refinement minimizes photometric and depth consistency losses to reduce drift, while the quadtree allocation distributes gaussians adaptively based on color and depth variance, suppressing redundancy. On handheld RGB-D sequences, our method reduced the number of gaussians by 47% with negligible PSNR loss, and further training improved quality beyond uniform allocation. These results demonstrate that the proposed pipeline enables efficient and accurate neural scene generation, supporting high-quality VR/XR content creation.**

Keywords*— *3D Scene Representation, RGB-D SLAM, 3D Gaussian Splatting

I. INTRODUCTION

Gaussian Splatting, which represents 3D scenes using translucent ellipsoids, has demonstrated outstanding performance in scene representation and real-time novel view synthesis [1]. Building on this success, a wide range of follow-up studies have been actively conducted, particularly in the context of RGB-D based SLAM systems, where it has been applied to camera tracking, 3D map representation, and optimization [2,3]. However, when extended to content generation, two critical limitations remain. First, due to real-time constraints, camera tracking accuracy often lags behind that of traditional Visual SLAM methods. Second, dense RGB-D inputs frequently lead to inefficient scene representations with excessive redundancy.

To overcome these limitations, this work proposes a pipeline that prioritizes high-quality offline scene reconstruction, making it especially suitable for VR/XR content creation. Specifically, we employ a real-time Visual SLAM system as the frontend and augment it with two post-processing modules: (1) a learning-based camera pose refinement and (2) a quadtree-based adaptive gaussian mapping. The proposed method leverages photometric and depth consistency losses to refine camera extrinsic parameters (RT), thereby mitigating cumulative drift and instability. Furthermore, gaussians are adaptively allocated according to scene complexity, reducing redundant insertions and enabling a lightweight yet high-fidelity scene representation.

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (RS-2023-00224358, Multi-view wide sensing based XR high-DoF full body motion interface development)

II. METHOD

The proposed system generates a gaussian-based 3D scene representation from an input RGB-D sequence through three main stages.

A. Real-Time RGB-D SLAM

We employ RTAB-Map to estimate the initial camera trajectory[4]. Due to the real-time nature of the process, however, cumulative drift and local pose inaccuracies inevitably occur.

B. Learning-Based Camera Pose Refinement

The estimated extrinsic parameters (RT) are refined through learning-based optimization to mitigate both cumulative drift and local pose errors. In this process, the reconstructed point cloud is reprojected onto the image plane, and photometric as well as depth consistency losses are minimized to ensure accurate frame-to-frame alignment. Beyond local refinement, optimization is also performed along the edges of the pose graph, enforcing global consistency and stability across the entire sequence. Consequently, the corrected camera trajectories are more reliable and enhance the overall fidelity of scene reconstruction.

C. Quadtree-Based Adaptive Gaussian Placement

To reduce redundancy while preserving structural details, we adopt a quadtree-based adaptive gaussian allocation. The image domain is recursively subdivided according to RGB/depth variance: rich textured regions receive more gaussians, while homogeneous regions receive fewer.

The procedure is as follows:

- **Root Node Initialization**: A visible mask is obtained by comparing the rendered depth buffer of the current gaussian map with the input depth map. By initializing the quadtree with the visible mask area, redundant gaussian generation is suppressed, ensuring efficiency in representation.
- **Recursive Subdivision**: At each node, RGB and depth variances are evaluated; if below thresholds, subdivision stops, otherwise the node splits into four quadrants. Recursion yields deeper quadtrees in textured or complex regions, enabling multi-scale scene representation.
- **Leaf Node Assignment**: For each leaf node, a single gaussian is instantiated with its centroid and color set to the mean of the corresponding point samples, and its scale proportional to the node's spatial extent. Point normals from the depth map are used to orient and flatten the gaussians, improving accuracy near surfaces and object boundaries.

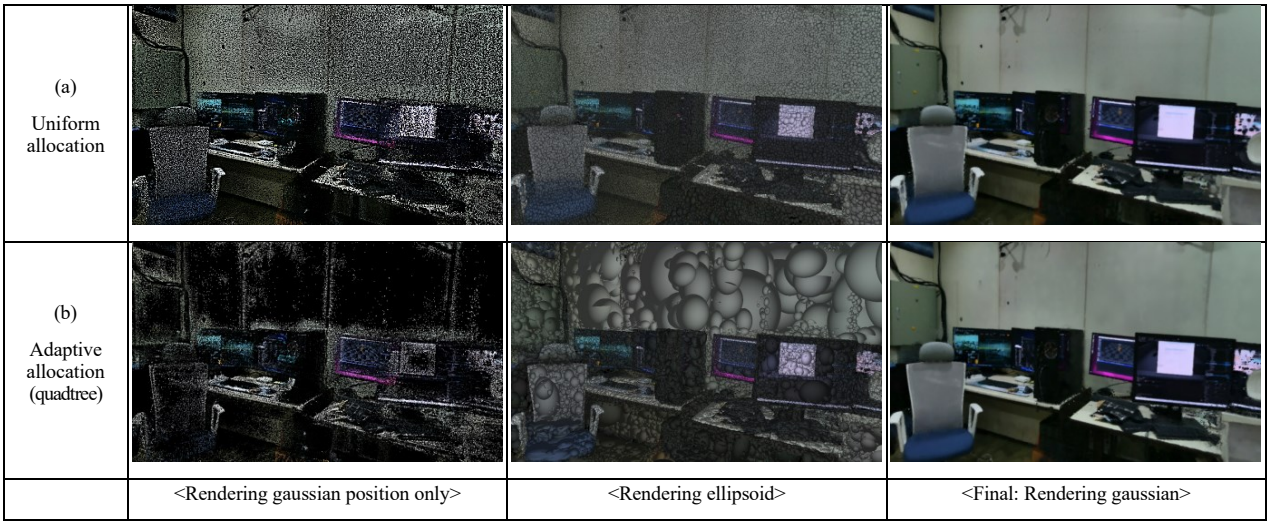

Fig 1. Comparison of gaussian maps: (a) uniform allocation, (b) quadtree-based adaptive allocation.

III. EXPERIMENTS

We evaluated the proposed method using RGB-D sequences captured via a visual SLAM pipeline. To mitigate pose drift and trajectory errors, a learning-based pose refinement was first applied. Subsequently, two gaussian placement strategies were compared: (a) a uniform allocation baseline and (b) the proposed quadtree-based adaptive allocation.

In the uniform allocation baseline, the input resolution was reduced by half due to memory constraints. gaussians of identical size were placed only within the visible mask. In contrast, the quadtree-based allocation could be performed at the original resolution.

The dataset consisted of a 7 m × 5 m indoor scene captured at a resolution of 1180 × 620, comprising 265 pairs of {RGB, depth} frames. The data was captured in a handheld manner, leading to motion blur and noticeable variations in auto-exposure.

Figure 1 presents a comparison of the initial gaussian maps. The uniform allocation (a) densely inserts gaussians across the entire image domain, resulting in significant redundancy. By contrast, the quadtree allocation (b) concentrates gaussians around edges, boundaries, and texture-rich regions, thereby achieving a structurally more efficient representation.

TABLE I. QUANTITATIVE RESULTS

	(a) Regular allocation	***(b) Adaptive allocation (quadtree)***
Number of gaussians	4,377,162	**2,334,956** (47% reduction)
Quality metrics (before optimization)	SSIM: 0.5808791 PSNR: 17.9501629 LPIPS: 0.4611209	SSIM: 0.5964603 PSNR: 17.8341503 LPIPS: 0.4764062
Qualti metrics (after optimization) (iteration=1000)	SSIM: 0.8382929 PSNR: 24.6341171 LPIPS: 0.2490266	**SSIM: 0.8415418** **PSNR: 24.8280029** LPIPS: 0.2561470
Processing time (insertion + optimization)	196 sec	145 sec

Table 1 summarizes the quantitative results. At the initial stage, the quadtree-based allocation reduced the number of gaussians by approximately 47% compared to uniform allocation, with only a minor drop in PSNR. After 1,000 training iterations, however, the quadtree allocation not only converged faster but also surpassed the baseline, yielding a PSNR improvement of +0.194. These results demonstrate that the proposed method achieves a more compact representation while ultimately producing higher-quality scene reconstructions.

The experiments were conducted using a system equipped with an Intel® Core™ i9-12900K processor running at 3.20 GHz, 128 GB of RAM, an NVIDIA A6000 ADA GPU, and the Windows 11 (64-bit) operating system.

IV. CONCLUSION

In this work, we presented a post-processing pipeline for RGB-D based visual SLAM that combines learning-based camera pose refinement with quadtree-based gaussian allocation. The proposed approach enables (1) precise pose correction through photometric and depth consistency optimization, and (2) efficient gaussian placement via quadtree subdivision adaptive to scene complexity, thereby improving both memory efficiency and convergence speed over uniform allocation strategies. Experimental results demonstrate that our method reduces gaussian redundancy by nearly half without degrading quality, while further optimization enhances reconstruction fidelity. These findings confirm that the quadtree-based allocation scheme provides a more compact and effective representation for 3D scene reconstruction.

REFERENCES

[1] Kerbl, B., Kopanas, G., Leimkuehler, T., & Drettakis, G. (2023). 3D Gaussian Splatting for Real-Time Radiance Field Rendering. ACM Transactions on Graphics, 42(4), 1–14.

[2] Keetha, N., Karhade, J., Jatavallabhula, K. M., Yang, G., Scherer, S., Ramanan, D., & Luiten, J. (2024). *SplaTAM: Splat, Track & Map 3D Gaussians for Dense RGB-D SLAM*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 21495–21504.

[3] Yan, C., Qu, D., Xu, D., Zhao, B., Wang, Z., Wang, D., & Li, X. (2024). *GS-SLAM: Dense Visual SLAM with 3D Gaussian Splatting*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024.

[4] RTAB-Map, Real-Time Appearance-Based Mapping, https://introlab.github.io/rtabmap/