Towards Robust Gaussian Splatting SLAM: Lessons from Benchmarking and Pipeline Analysis

Björn Ellensohn, Peyman Kahrizi, Christian Rauch, Elmar Rueckert

Abstract

Gaussian Splatting (GS) [3] has recently emerged as a promising representation for real-time 3D reconstruction. While several GS-based SLAM systems have been proposed, their robustness in practical RGB-D scenarios remains insufficiently understood. In this work, we present a compact evaluation of representative GS-SLAM pipelines and analyze the role of tracking components within a GS-based system. This study is motivated by the need to identify reliable geometric foundations for future semantic scene representations built on top of GS-based SLAM. Our results indicate that tracking accuracy remains the dominant limitation, while reconstruction quality can be partially decoupled from trajectory correctness. We further observe that preprocessing strategies, such as downsampling, significantly impact both runtime and accuracy. Based on these findings, we identify key failure modes and discuss implications for improving robustness in GS-based SLAM systems.

1 Introduction

High-fidelity and geometrically consistent 3D scene reconstruction is a fundamental requirement for many robotics and vision applications. In particular, emerging approaches to semantic scene representations and language-conditioned reasoning rely on dense and spatially accurate geometry to associate structure with semantic or embedding-based features.

Recent advances in Gaussian Splatting (GS) [3] enable real-time rendering of complex environments with high visual fidelity, making them a promising foundation for such representations. However, extending GS to SLAM settings introduces additional challenges. While several GS-based SLAM systems have been proposed [5, 8, 2, 1], their robustness in real-world RGB-D scenarios remains limited. In practice, these systems depend strongly on accurate camera tracking, and pose estimation errors directly affect the consistency of the reconstructed scene.

This study focuses on systematically evaluating existing GS-SLAM pipelines as a necessary step towards reliable semantic extensions. Futhermore, the goal is to identify practical limitations and design constraints that influence the robustness of GS-based scene representations in real-world settings.

2 Related Work

Gaussian Splatting (GS) was introduced as an efficient representation for high-quality real-time rendering [3] and has recently been extended to SLAM settings, including MonoGS [5], Uni-SLAM [8], SplaTAM [2], and GS-ICP-SLAM [1]. These approaches differ primarily in their tracking strategies and system design, but all rely on accurate pose estimation to maintain consistent scene representations. Related work in neural implicit SLAM, such as iMAP [7] and NICE-SLAM [9], similarly demonstrates dense reconstruction using learned representations while emphasizing the importance of tracking stability. Classical systems such as ORB-SLAM2 [6] further highlight that robust pose

The third Austrian Symposium on AI and Vision (AIROV26).

estimation remains a fundamental requirement across different SLAM paradigms, with ICP-based methods and variants such as FastGICP [4] that are widely used for geometric alignment.

In contrast to prior work that primarily proposes new SLAM methods or representations, this work focuses on a systematic evaluation of existing GS-SLAM pipelines. Our goal is to analyze their robustness and identify practical limitations, particularly in the context of downstream semantic scene representations that depend on consistent geometric structure.

3 Method

We design a controlled evaluation framework to assess representative GS-SLAM systems, including MonoGS [5], Uni-SLAM [8], SplaTAM [2], and GS-ICP-SLAM [1], on RGB-D datasets (Replica, TUM). The evaluation focuses on reconstruction quality, trajectory accuracy, and runtime performance, measured using PSNR, SSIM, and LPIPS for rendering quality, absolute trajectory error (ATE RMSE) for pose estimation, and frames per second (FPS) for computational efficiency.

To analyze the influence of the tracking component, we focus on GS-ICP-SLAM, which relies on ICP-based registration using FastGICP [4] as its backend. Here we compare two preprocessing strategies within the GS-ICP-SLAM pipeline. External downsampling refers to reducing the input image resolution prior to depth estimation and point cloud generation, whereas internal voxelization performs downsampling directly on dense point clouds using a voxel grid. This setup allows us to compare early versus late reduction of geometric information. We hypothesize that tracking performance depends more on correspondence quality than on raw point density.

4 Key Results

4.1 Benchmark Comparison

We compare representative GS-SLAM systems across datasets in terms of reconstruction quality, trajectory accuracy, and runtime performance. Table 1 reports the full per-sequence evaluation results.

GS-ICP-SLAM achieves comparable reconstruction quality to other methods while providing significantly higher runtime performance. However, trajectory accuracy varies substantially across datasets, particularly in real-world scenarios.

These results reveal a consistent performance gap between synthetic and real-world data, indicating sensitivity to sensor noise, calibration inaccuracies, and scene complexity.

4.2 Preprocessing Impact

To further analyze the influence of data processing on tracking performance, we compare different preprocessing strategies within the GS-ICP-SLAM pipeline.

Table 2 shows that external downsampling improves both runtime and trajectory accuracy, while internal voxel-based downsampling degrades performance. These observations strengthen the hypothesis that tracking performance depends more on correspondence quality than on raw point density.

In particular, early reduction of input resolution leads to more stable tracking, whereas voxel-based downsampling introduces additional noise and correspondence ambiguity in dense point clouds.

This suggests that the structure and quality of the generated point cloud, rather than its raw density, plays a critical role in ICP-based alignment.

4.3 Failure Modes and Limitations

Figure 1 illustrates typical failure modes caused by inaccurate pose estimation in GS-based SLAM systems.

Based on these observations, we identify several recurring failure modes:

- Degenerate geometry (e.g., planar scenes)

Table 1: Full GS-SLAM evaluation across Replica and TUM datasets (all sequences).

Method	Metric	Replica r0	r1	r2	of0	of1	of2	of3	of4	TUM f1	f2	f3
MonoGS	PSNR	35.8	38.4	39.1	42.7	43.5	37.2	37.2	-	23.4	24.4	24.9
	SSIM	0.96	0.96	0.97	0.98	0.98	0.96	0.96	-	0.78	0.79	0.83
	LPIPS	0.05	0.04	0.05	0.03	0.03	0.04	0.04	-	0.25	0.22	0.20
	RMSE	0.40	0.03	0.35	0.53	0.56	0.20	0.19	-	1.43	1.44	1.50
	FPS	0.90	1.03	1.02	1.14	1.19	0.95	0.92	-	1.83	3.53	2.83
Uni-SLAM	PSNR	27.9	29.9	29.9	36.2	35.9	29.1	30.0	32.7	18.1	22.5	19.3
	SSIM	0.93	0.93	0.94	0.97	0.97	0.95	0.96	0.97	0.73	0.85	0.77
	LPIPS	0.27	0.23	0.20	0.15	0.17	0.21	0.17	0.18	0.55	0.39	0.44
	RMSE	0.47	0.46	0.47	0.37	0.44	0.57	0.59	0.42	2.72	1.53	3.04
	FPS	0.79	0.58	0.65	1.04	1.33	1.23	1.04	0.85	0.23	0.44	0.93
GS-ICP-SLAM	PSNR	35.7	37.9	38.6	43.0	43.3	36.9	37.0	39.0	18.1	23.6	20.9
	SSIM	0.96	0.97	0.97	0.98	0.98	0.97	0.96	0.97	0.71	0.83	0.76
	LPIPS	0.04	0.04	0.04	0.026	0.03	0.05	0.04	0.04	0.28	0.13	0.21
	RMSE	0.14	0.16	0.11	0.18	0.12	0.17	0.19	0.21	2.44	1.81	2.53
	FPS	12.5	12.9	12.3	12.9	12.8	12.5	12.9	13.2	15.3	14.8	15.0
SplaTAM	PSNR	32.0	33.6	34.8	38.1	39.0	32.0	29.8	36.0	21.6	24.5	21.8
	SSIM	0.97	0.96	0.98	0.98	0.98	0.96	0.95	0.94	0.84	0.94	0.87
	LPIPS	0.07	0.09	0.73	0.08	0.09	0.09	0.11	0.15	0.24	0.09	0.20
	RMSE	0.34	0.39	0.30	0.47	0.28	0.26	0.33	0.58	3.35	1.32	5.23
	FPS	0.23	0.19	0.19	0.19	0.21	0.25	0.24	0.23	0.41	0.09	0.41

System: RTX4090

Table 2: Comparison of External vs Internal Downsampling

Mode A: external downsampling		**Mode B: internal downsampling**	
Frames	2506	Frames	2506
Runtime [sec]	12.027	Runtime [sec]	122.326
Avg input points/frame	3819.4	Avg input points/frame	243944.1
Min input points/frame	2981	Min input points/frame	191149
Max input points/frame	4262	Max input points/frame	271328
ATE RMSE [m]	0.183453	ATE RMSE [m]	0.371779
Avg effective points/frame	3819.4	Avg effective points/frame	4153.4
Min effective points/frame	2981	Min effective points/frame	931
Max effective points/frame	4262	Max effective points/frame	7664

Data: TUM scene freiburg3_long_office_household

- Drift accumulation over time
- Sensitivity to noise and input quality
- Mismatch between visual reconstruction quality and geometric accuracy

Additionally, observations shown in Figure 2 reinforce that reconstruction quality in GS-based SLAM is highly dependent on tracking accuracy.

5 Discussion and Outlook

The results indicate that tracking remains the primary limitation in current GS-based SLAM systems. While Gaussian-based representations provide efficient and high-quality rendering, they rely on accurate pose estimation to maintain geometric consistency.

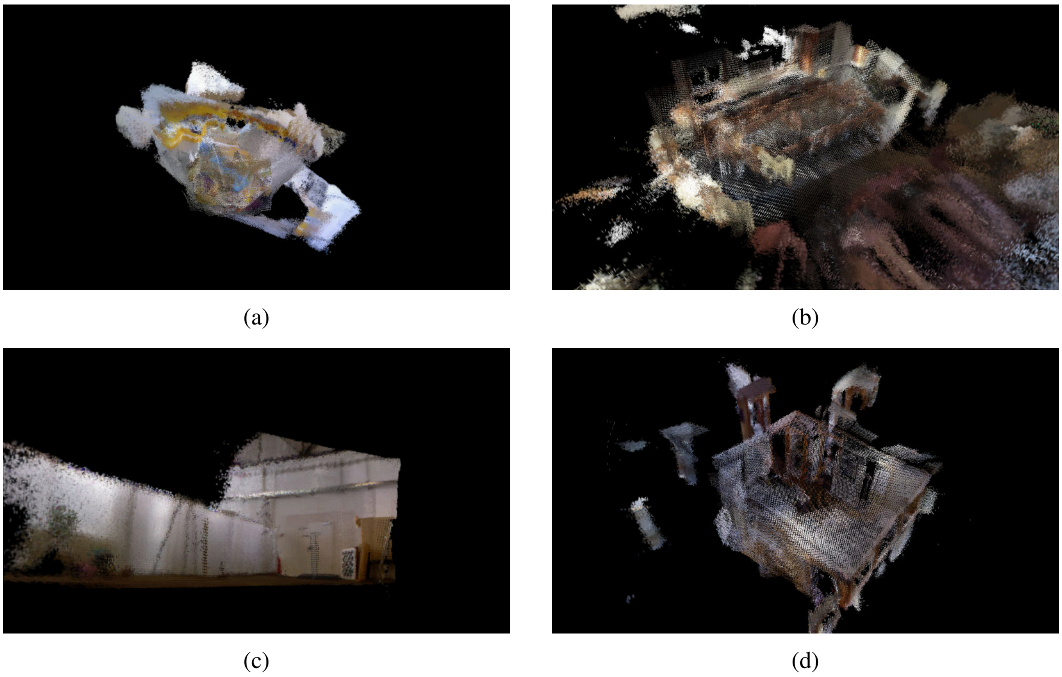

Figure 1: Qualitative examples of reconstruction artifacts caused by tracking errors in GS-based SLAM: (a) structural collapse due to severe pose inconsistency, (b) duplication artifacts from systematic pose offsets, (c) incomplete reconstruction under partial tracking failure, and (d) geometric distortions caused by small pose inaccuracies.

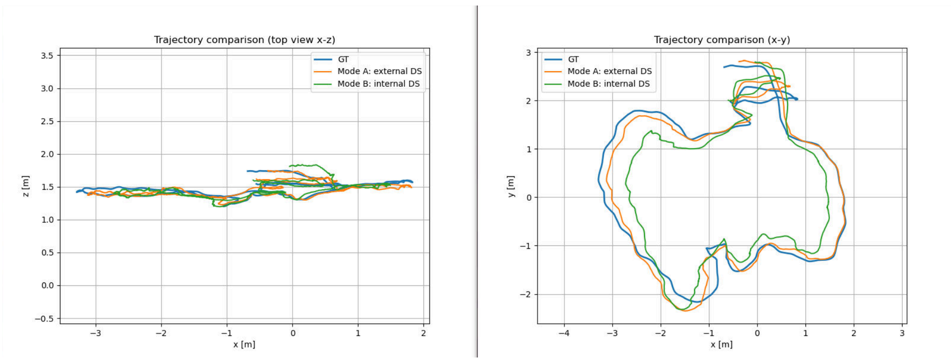

Figure 2: Trajectory comparison showing divergence between estimated and ground truth trajectories. Internal voxelization introduces drift and misalignment, while external downsampling yields more stable tracking.

This has direct implications for semantic scene representations, where inaccuracies in geometry may propagate to higher-level representations and affect their reliability. Future work should therefore focus on improving tracking robustness and integrating geometric and semantic representations more tightly.

6 Conclusion

We presented a compact evaluation of GS-based SLAM systems and analyzed the role of tracking and preprocessing. The results highlight that tracking robustness is a key limiting factor and that preprocessing choices significantly influence performance.

7 Acknowledgement

This work was created as part of the research projects 'KIRAMET' (FO999899661), MUTAVIA (FO999922732) and NNATT (FO999907606), which are funded by the Federal Ministry Republic of Austria Climate Action, Environment, Energy, Mobility, Innovation and Technology.

References

[1] Seongbo Ha, Jiung Yeon, and Hyeonwoo Yu. Rgbd gs-icp slam. In Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol, editors, *Computer Vision – ECCV 2024*, pages 180–197, Cham, 2025. Springer Nature Switzerland.

[2] Nikhil Keetha, Jay Karhade, Krishna Murthy Jatavallabhula, Gengshan Yang, Sebastian Scherer, Deva Ramanan, and Jonathon Luiten. Splatam: Splat, track & map 3d gaussians for dense rgb-d slam, 2024.

[3] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 2023.

[4] Kenji Koide. FastGICP: Fast and accurate point cloud registration. `https://github.com/koide3/fast_gicp`, 2021. Accessed: 2026-03-23.

[5] Hidenobu Matsuki, Riku Murai, Paul H.J. Kelly, and Andrew J. Davison. Gaussian splatting slam. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18039–18048, June 2024.

[6] Raul Mur-Artal and Juan Tardos. Orb-slam2: An open-source slam system. *IEEE Transactions on Robotics*, 2017.

[7] Edgar Sucar et al. imap: Implicit mapping and positioning in real-time. In *ICCV*, 2021.

[8] Shaoxiang Wang, Yaxu Xie, Chun-Peng Chang, Christen Millerdurai, Alain Pagani, and Didier Stricker. Uni-slam: Uncertainty-aware neural implicit slam for real-time dense indoor scene reconstruction. In *Proceedings of the Winter Conference on Applications of Computer Vision (WACV)*, pages 2228–2239, February 2025.

[9] Zihan Zhu et al. Nice-slam: Neural implicit scalable encoding for slam. In *CVPR*, 2022.