GCD-SLAM: GEOMETRIC-CORRECTION DECOUPLING FOR RGB-D GAUSSIAN SPLATTING SLAM

Anonymous authors
Paper under double-blind review

ABSTRACT

Localization refinement in evolving scene representations can become inherently coupled with the representations used for optimization, when these representations are constructed from previously estimated states. Such coupling makes localization correction dependent on the underlying representation and its construction history, potentially limiting the effectiveness of state refinement. We identify this representation-dependent optimization issue in RGB-D Gaussian Splatting SLAM and propose GCD-SLAM, a Geometric-Correction Decoupling framework that introduces representation-independent localization correction beyond the native Gaussian tracking-mapping loop. Specifically, GCD-SLAM develops Independent Geometric Guidance to derive correction cues directly from temporal depth observations rather than trajectory-conditioned Gaussian representations. Reliability-aware Correction Selection identifies reliable correction conditions and enables selective localization refinement. Finally, Localization State Separation decouples corrected poses from the native tracking-mapping process, preserving Gaussian reconstruction consistency while providing corrected localization outputs. Experiments on challenging RGB-D SLAM scenarios demonstrate that GCD-SLAM achieves effective selective localization improvement while preserving the native Gaussian tracking-mapping process and maintaining robustness under challenging near-range conditions.

1 INTRODUCTION

Closed-loop state estimation increasingly refines poses against learned or evolving representations assembled from earlier states. If the reference used by the optimizer is itself constructed from previous estimates, the objective can inherit a trajectory-dependent preference. This representation-dependent optimization bias is not specific to Gaussian maps: learned depth and pose systems, recurrent SLAM, and neural implicit mapping also update a scene or feature representation while estimating state (Yang et al., 2020b; Teed & Deng, 2021; Zhu et al., 2022). Gaussian SLAM provides a concrete instance in which the reference, pose, and map update are explicit. It combines camera estimation with a Gaussian scene representation updated from earlier poses and rendered to guide subsequent tracking (Kerbl et al., 2023; Keetha et al., 2024; Matsuki et al., 2024; Yan et al., 2024; Yugay et al., 2023; Tran & Kosecka, 2026).

Existing Gaussian SLAM systems improve this loop with confidence weighting, geometric registration, uncertainty modeling, or loop closure (Hu et al., 2024; Ha et al., 2024; Tran & Kosecka, 2026; Zhu et al., 2025). These mechanisms strengthen tracking or map consistency, while the pose objective remains conditioned on a reference formed by earlier estimates. The resulting gap is a missing separation between current RGB-D evidence and the representation that defines the rendered objective. A controlled diagnosis therefore holds observations and candidate poses fixed while changing only the reference.

The fixed-probe diagnostic reveals a reference-dependent shift in pose preference: switching between maps built from estimated and ground-truth trajectories reverses the preferred pose under a common diagnostic objective. We call this tested dependency *Representation-Induced Localization Bias (RILB)*. The observation motivates a correction principle whose evidence does not depend on

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Gaussian-map parameters and whose output authority is kept separate from the native tracking–mapping recurrence.

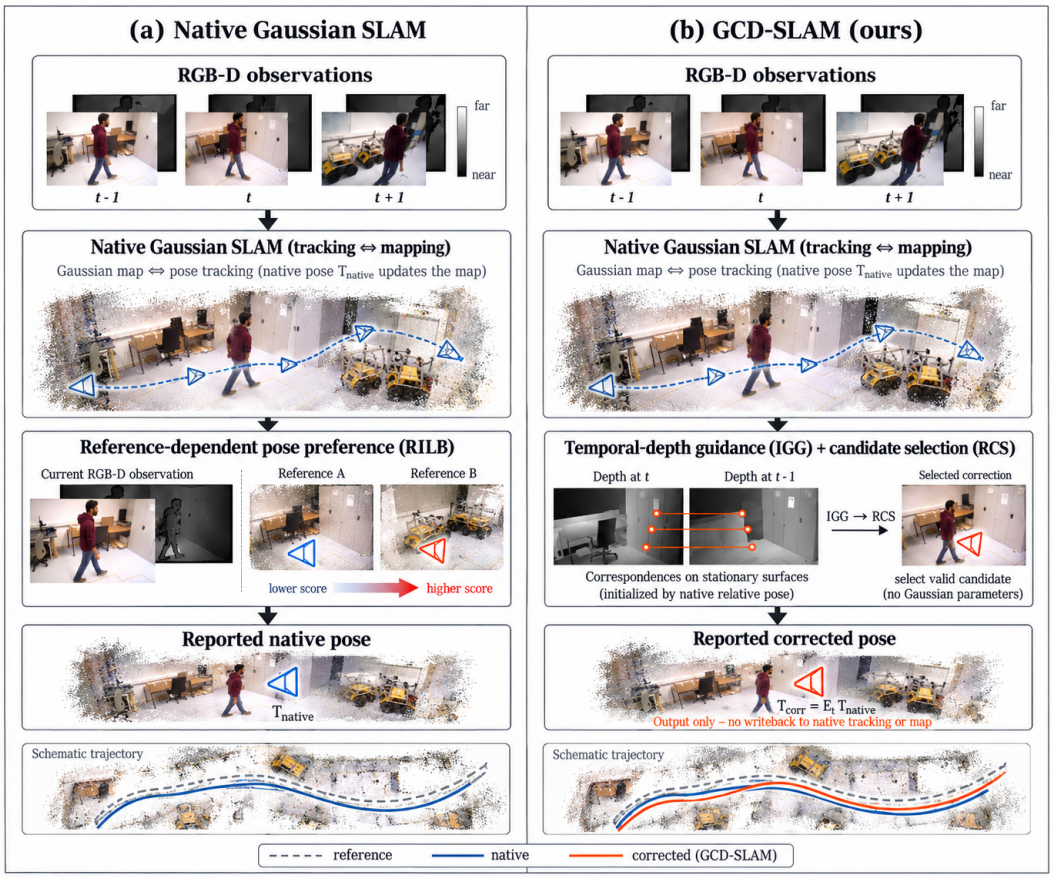

Figure 1: From tracking–mapping coupling to Geometric-Correction Decoupling. Native Gaussian SLAM uses a trajectory-conditioned representation as its tracking reference, introducing the representation-dependent optimization preference termed RILB. GCD-SLAM organizes an independent geometric correction pathway through guidance, selection, and state separation while preserving the native tracking–mapping loop.

We instantiate this principle in GCD-SLAM, a Geometric-Correction Decoupling (GCD) framework (Fig. 2). Independent Geometric Guidance extracts a temporal-depth candidate without reading Gaussian parameters. Reliability-aware Correction Selection restricts that candidate to the measured texture, range, support, and validity regime. Localization State Separation then composes the admitted correction into the reported pose while native poses retain tracking and Gaussian-update authority. The three stages form one authority chain: independent evidence, bounded admission, and zero-writeback state integration.

We evaluate the diagnosis, intervention, and authority boundary under matched within-backbone protocols. All 12 fixed pairs reverse preference between the two Gaussian references. On the selected 300-frame prefix, three-seed median ATE falls by 34.81%, whereas the median full-horizon reduction is 0.414% over 1,060 frames. The correction remains inactive on three public controls; room regresses by 0.31%, and one near-range seed fails in both Native and GCD. Fixed-raw replay attributes the prefix effect to geometric correction, while frozen-submap rendering measures the corrected output against the unchanged map.

The contributions of this work are summarized as follows:

- We formulate Representation-Induced Localization Bias as a representation-dependent optimization preference and isolate it by swapping the reference while holding observations and pose probes fixed.

- We introduce a representation-decoupled correction principle that separates geometric cues from the trajectory-conditioned reference and keeps correction output authority outside native reconstruction.
- We demonstrate selective localization improvement in the evaluated near-range regime while preserving the native Gaussian state and exposing the measured full-horizon and room boundaries.

2 Related Work

2.1 State Estimation with Evolving Representations

RGB-D systems track against geometric scene models and features (Newcombe et al., 2011; Campos et al., 2021); D3VO, DROID-SLAM, and NICE-SLAM use learned depth, recurrent updates, or implicit maps (Yang et al., 2020b; Teed & Deng, 2021; Zhu et al., 2022). Gaussian systems couple pose estimation to dense rendered maps (Keetha et al., 2024; Matsuki et al., 2024; Yan et al., 2024; Yugay et al., 2023). Photo-SLAM and LoopSplat use feature tracking and submap registration, respectively (Huang et al., 2024; Zhu et al., 2025); CG-SLAM, GS-ICP, and VarSplat incorporate confidence, geometry, or uncertainty (Hu et al., 2024; Ha et al., 2024; Tran & Kosecka, 2026). Our fixed-probe test targets reference-conditioned preference in the evolving Gaussian objective, without assuming the same behavior has been measured in other representations.

2.2 Geometric Localization and State Authority

Depth and point-to-plane residuals provide geometric pose evidence (Newcombe et al., 2011; Ha et al., 2024), while robust estimators limit outlier influence (Sunderhauf & Protzel, 2012; Agarwal et al., 2013; Yang et al., 2020a). Such constraints usually modify tracking directly. GCD derives a temporal-depth output correction, selects it with a fixed rule, and leaves the native tracking and mapping recurrence unchanged. Its distinction is the output-side placement of geometric evidence.

3 Method

3.1 Overview of Geometric-Correction Decoupling

Gaussian tracking uses a reference built from earlier poses. Under the fixed-probe diagnostic, changing this reference reverses pose preference even though the observations and probes stay fixed (Fig. 1). GCD responds by separating measured-depth correction evidence, candidate admission, and authority over the native tracking–mapping recurrence.

Figure 2 shows *Independent Geometric Guidance (IGG)*, *Reliability-aware Correction Selection (RCS)*, and *Localization State Separation (LSS)* in execution order. We instantiate the pathway on VarSplat (Tran & Kosecka, 2026). The symbols T_t^{native} and T_t^{corr} denote native and corrected camera-to-world poses; only the native pose drives recursive tracking and Gaussian reconstruction.

3.2 Independent Geometric Guidance

3.2.1 Motivation

IGG reads consecutive depth images rather than Gaussian map parameters. Native relative motion still initializes its correspondences, so the independent quantity is the measured-depth correction evidence, not the entire pose solve.

3.2.2 Geometric Constraint Extraction

Given consecutive depth observations D_{t-1}, D_t, IGG constructs temporal constraints and obtains a geometric candidate:

$$\begin{aligned} \mathcal{C}_{\text{geo},t} &= f_{\text{geo}}(D_{t-1}, D_t; K, \overline{T}_{t-1\leftarrow t}), \\ \Delta T_{\text{geo},t} &= g(\mathcal{C}_{\text{geo},t}). \end{aligned} \tag{1}$$

Here K denotes intrinsics and $\overline{T}_{t-1\leftarrow t}$ the native relative pose initializing correspondences and linearization. Thus independence concerns the measured-depth evidence, which reads no Gaussian

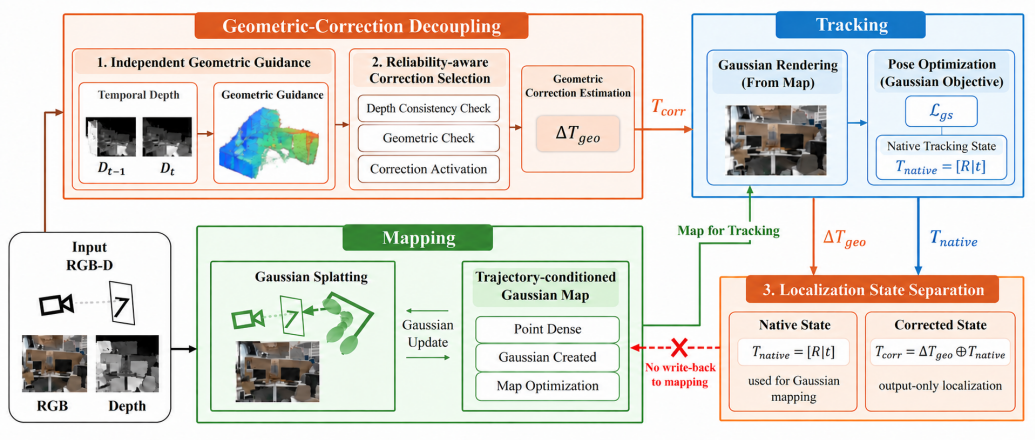

Figure 2: GCD-SLAM preserves the native RGB-D Gaussian tracking–mapping backbone and adds a decoupled correction pathway. Independent Geometric Guidance extracts temporal-depth evidence, Reliability-aware Correction Selection controls activation, and Localization State Separation produces an output pose while the native pose retains tracking and mapping authority.

parameters. Back-projected current points p_i are associated with reference vertices q_i and normals n_i, giving

$$r_i(\xi) = n_i^\top \left(T_{r\leftarrow t}(\xi)p_i - q_i\right). \tag{2}$$

where $r = t-1$ and $T_{r\leftarrow t}(\xi)$ transforms the current camera into the reference under perturbation ξ. We sample every four pixels, retain depths in $[0.10, 8.0]$ m, reject innovations larger than $\max(0.05\text{ m}, 0.05D)$, and require at least 256 correspondences. We linearize $r_i \simeq r_i^0 + J_i\eta$ and form

$$H_{\text{raw}} = \frac{1}{N}\sum_i w_i J_i^\top J_i, \qquad b_{\text{raw}} = \frac{1}{N}\sum_i w_i J_i^\top r_i^0. \tag{3}$$

The regularized solve uses GNC-TLS (Yang et al., 2020a) weights at scale 0.005 m. Translation is normalized by median factor depth z_{ref} and restored after solving. The twist ξ_t, bounded to 2 cm and 1°, left-perturbs the native world-to-camera pose. Its camera-to-world correction is

$$\begin{aligned} \widetilde{T}_t &= \left[\text{Exp}(\xi_t)(T_t^{\text{native}})^{-1}\right]^{-1}, \\ \Delta T_{\text{geo},t} &= \widetilde{T}_t(T_t^{\text{native}})^{-1}. \end{aligned} \tag{4}$$

The correction therefore acts in the world frame used by LSS.

3.3 Reliability-aware Correction Selection

3.3.1 Motivation

Incomplete surfaces, ambiguous geometry, and noisy depth can weaken a temporal-depth candidate. RCS applies fixed gates to admit corrections within the selected near-range, low-texture observation regime.

3.3.2 Reliability Estimation and Activation

Let s_t be the mean central-difference grayscale gradient and v_t indicate a valid finite IGG candidate. The joint support, range, and persistent-texture rule is

$$\begin{aligned} \alpha_t = \mathbf{1}\left[\max_{k\in\{0,1,2\}} s_{t-k} < 0.015\right] \\ \cdot\, \mathbf{1}[z_{\text{ref}} \le 1.05\text{ m}] \cdot \mathbf{1}[N \ge 256] \cdot \mathbf{1}[v_t]. \end{aligned} \tag{5}$$

All three frames must be consecutive and available; depth must be finite. These hand-selected thresholds are fixed across experiments; RCS is not a learned reliability predictor. The admitted

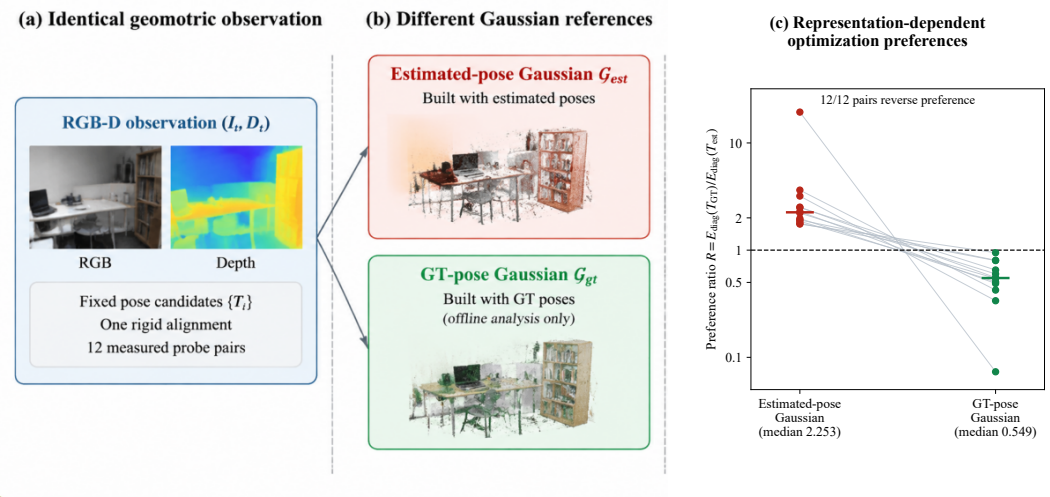

Figure 3: Quantitative analysis of Representation-Induced Localization Bias. **(a)** Each probe fixes an identical RGB-D observation and a set of aligned pose candidates under the same geometric condition. **(b)** Two Gaussian references are constructed with matched evaluation settings, where the estimated-pose Gaussian $\mathcal{G}_{\text{est}}$ is generated from estimated poses and the GT-pose Gaussian $\mathcal{G}_{\text{gt}}$ is generated from ground-truth poses for offline diagnosis. **(c)** Under the same localization objective and candidate poses, the preference ratio changes across different Gaussian references. The consistent preference reversal shows that the tested objective's pose ranking depends on the Gaussian reference under fixed observations and probes. Panels (a,b) are schematic illustrations; (c) plots all 12 measured pairs under the common uniform-weight diagnostic objective.

correction is

$$C_t = \begin{cases} \Delta T_{\text{geo},t}, & \alpha_t = 1, \\ I, & \alpha_t = 0. \end{cases} \tag{6}$$

Identity denotes no new correction. Hessian spectra are logged but have no selection authority in the final policy. Sec. 4.4 evaluates the activation boundary.

3.4 Localization State Separation

3.4.1 Motivation

Feeding corrected poses into native tracking or mapping would reintroduce the original coupling. LSS assigns separate state ownership to reconstruction and localization output.

3.4.2 Native and Corrected Localization States

Let G_{t-1} denote the Gaussian representation available before processing frame t. The native RGB-D tracking and mapping recurrence is

$$\begin{aligned} T_t^{\text{native}} &= \text{Track}_{\text{G}}(T_t^{\text{init}}; G_{t-1}, I_t, D_t), \\ G_t &= \mathcal{M}(G_{t-1}, T_t^{\text{native}}, I_t, D_t), \end{aligned} \tag{7}$$

where T_t^{init} is initialized from previous native poses. The corrected output composes a separate correction state $E_t \in \text{SE}(3)$ with the native pose:

$$T_t^{\text{corr}} = E_t \oplus T_t^{\text{native}} = E_t T_t^{\text{native}}. \tag{8}$$

Here $\oplus$ denotes pose composition. Only T_t^{native} enters mapping and subsequent tracking initialization. The output T_t^{corr} has zero writeback into Gaussian parameters, keyframes, or native tracking.

3.5 Decoupled Localization Optimization

Combining source, admission, and state ownership under GCD gives the temporal integration

$$E_t = \mathrm{Exp}(\beta \,\mathrm{Log}(E_{t-1}) + \mathrm{Log}(C_t))\,,$$
$$T_t^{\mathrm{corr}} = E_t T_t^{\mathrm{native}}, \qquad E_0 = I, \quad \beta = 0.95. \tag{9}$$

For valid candidates, $\mathrm{Log}(C_t) = \alpha_t \,\mathrm{Log}(\Delta T_{\mathrm{geo},t})$ expresses activation in the tangent space. Inactive selection admits identity while the previous error state decays through β. A wholly inactive sequence reproduces Native; nonconsecutive frames reset the error state and texture history. Thus T_t^{native} continues to update G_t, while T_t^{corr} is reported only as localization output.

4 Experiments

4.1 Experimental Setup

We test whether (i) the Gaussian reference changes fixed-probe preference, (ii) geometric output correction changes localization under matched protocols, and (iii) state separation leaves the native map recurrence intact. The tests use frozen VarSplat (Tran & Kosecka, 2026) artifacts. TUM RGB-D (Sturm et al., 2012) protocols include public-sequence intersections, the original `fr3/structure_notexture_near` at 300/1,060 frames starting from frame 0, texture-near and notexture-far controls, and low-texture `fr1/xyz`, `room`, and `desk2` derivatives. The derivatives deterministically transform RGB while preserving depth, GT, timestamps, and intrinsics. They form controlled stress tests that preserve the original geometric signal.

ATE RMSE uses global rigid $\mathrm{SE}(3)$ alignment without scale. Near-prefix absolute values use three paired seeds; the displayed full-horizon absolute values use seed 0, while full-horizon reductions use the median of three seed-wise relative reductions. Other within-backbone rows use seed 0. Table 1 contains only locally controlled correction variants; Appendix Table 3 separately records available external results as protocol context. Native and GCD share the map-pose stream and Gaussian map; independent raw-factor and trajectory replays verify the reported outputs.

4.2 Analysis of Representation-Induced Localization Bias

4.2.1 Experimental Design

We evaluate identical RGB-D observations and estimated/GT pose pairs on Gaussian references constructed from estimated and GT trajectories. A global rigid alignment fixes the gauge. Mapping budget, realized submap/keyframe schedule, raw-valid-depth pixels, and uniform weights are matched. Define

$$R(\mathcal{M}) = \frac{E_{\mathrm{diag}}(T^{\mathrm{GT}}; \mathcal{M})}{E_{\mathrm{diag}}(T^{\mathrm{est}}; \mathcal{M})}. \tag{10}$$

where $R > 1$ favors the estimated probe and $R < 1$ favors GT. The common E_{diag} differs from the native variance-weighted objective. All 12 fixed pairs reverse preference between references, with median ratios 2.253 and 0.549 (Fig. 3). This measures reference-conditioned objective preference on the tested texture-poor near-range protocol; it does not isolate the unique cause of downstream ATE.

4.2.2 Discussion

Reference-dependent preferences motivate GCD's independent correction source. We assess that source against GT relative motion anchored to the same native reference pose. For $r = t - 1$, define

$$T_t^{\star} = T_r^{\mathrm{native}} (T_r^{\mathrm{GT}})^{-1} T_t^{\mathrm{GT}},$$
$$\xi_{\mathrm{GT}} = \mathrm{Log}\big[(T_t^{\star})^{-1} T_t^{\mathrm{native}}\big]\,. \tag{11}$$

Cosine agreement with ξ_{GT} is measured for native descent and IGG in the same left-world-to-camera tangent. Finite differences freeze native masks and weights. Median cosines are -0.060 and $+0.452$, with positive agreement in 4/12 and 8/12 cases. Temporal depth provides a complementary direction in this local diagnostic. GT is confined to offline analysis, and the measured cosines characterize local direction agreement.

Table 1: Locally controlled localization correction. ATE RMSE is in cm (lower is better). Public sequences use exact full-trajectory frame intersections. Near-300 absolute values are three-seed medians. The displayed Near-1060 absolute values are seed-0 values; the reported Near-1060 reduction is the median of three seed-wise relative reductions.

	Public sequences: common frame intersections			Targeted near-range sequence	
Method	fr1/desk	fr2/xyz	fr3/office	Near-300	Near-1060
Matched frames	592	3,397	2,515	300	1,060
Native VarSplat	2.070	1.526	3.641	1.743	16.052
GCD-SLAM (Ours)	2.070	1.526	3.641	1.137	15.806
Output-only ICP control	–	–	–	2.974	17.921

Public columns use one frozen local run. Dashes denote unavailable matched controls. GCD's three-seed median paired ATE reduction is **34.81%** at 300 frames and **0.414%** at 1,060 frames. The output-only ICP control's corresponding reductions are -71.02% and -9.90%.

Table 2: Near-300 correction controls. The fixed-raw rows share a candidate log and differ only in RCS admission; ICP and feedback are cross-configuration controls. Positive ATE change denotes improvement over Native.

Variant	Factor	ATE change (%)	Admitted	Interpretation
Native Gaussian SLAM	Reference	0.000	0	Native reference
Raw candidate, RCS-off	Admission	+34.961	298	All valid
GCD-SLAM, RCS-on	Admission	+34.810	286	Recorded RCS
Output-only Open3D ICP	Solver/policy	-71.020	–	Cross-configuration
Tracking-init. feedback (V2.2-H)	State ownership	-8.660	286	Native-pose feedback

4.3 Localization Performance Comparison

4.3.1 Protocol-Separated Tracking Comparison

Table 1 reports locally controlled Native, GCD, and raw-depth trajectories. On three public sequences RCS remains inactive, so GCD reproduces Native. The paired three-seed median ATE reduction is 34.81% on the targeted 300-frame near prefix and 0.414% over all 1,060 frames. A cross-configuration ICP variant worsens both horizons. These results concern the reported localization stream; native tracking and mapping stay fixed.

4.3.2 Horizon and Scene-Wise Behavior

Across seeds 0/1/2, near-prefix ATE reductions are $33.715/40.605/34.810\%$, versus $1.536/0.006/0.414\%$ over 1,060 frames; seed 1 fails under both Native and GCD. On low-texture XYZ/desk2, ATE falls from $8.135/6.035$ to $7.540/5.888$ cm, while room increases from 4.162 to 4.175 cm (0.31%). Texture-near/far remain at $2.535/0.530$ cm. Near-prefix rotational RPE-1 also rises from 0.316° to 0.356° (three-run medians). Appendix B gives the remaining RPE values; external tracking numbers are protocol-separated in Table 3.

4.4 Validation of Geometric-Correction Decoupling Principles

Table 2 tests correction source, admission, and state ownership. Without IGG the output is Native; the direction diagnostic in Sec. 4.2.2 tests its geometric evidence. ICP and feedback are cross-configuration controls.

4.4.1 Effect of Reliability Selection

RCS activates on 95.97% of near-prefix candidates, 98.87% over the full near sequence, and 0% on texture/far controls (Fig. 4c). Fixed-raw replay changes only admission while holding candidates, native initialization, recurrence, frame order, and evaluator fixed. Across three near-prefix seeds, admitting all valid candidates yields a 34.961% median ATE reduction versus 34.810% with RCS. RCS shapes activation boundaries but does not account for the main prefix gain. The Open3D (Zhou et al., 2018) control changes solver and policy, so it is cross-configuration.

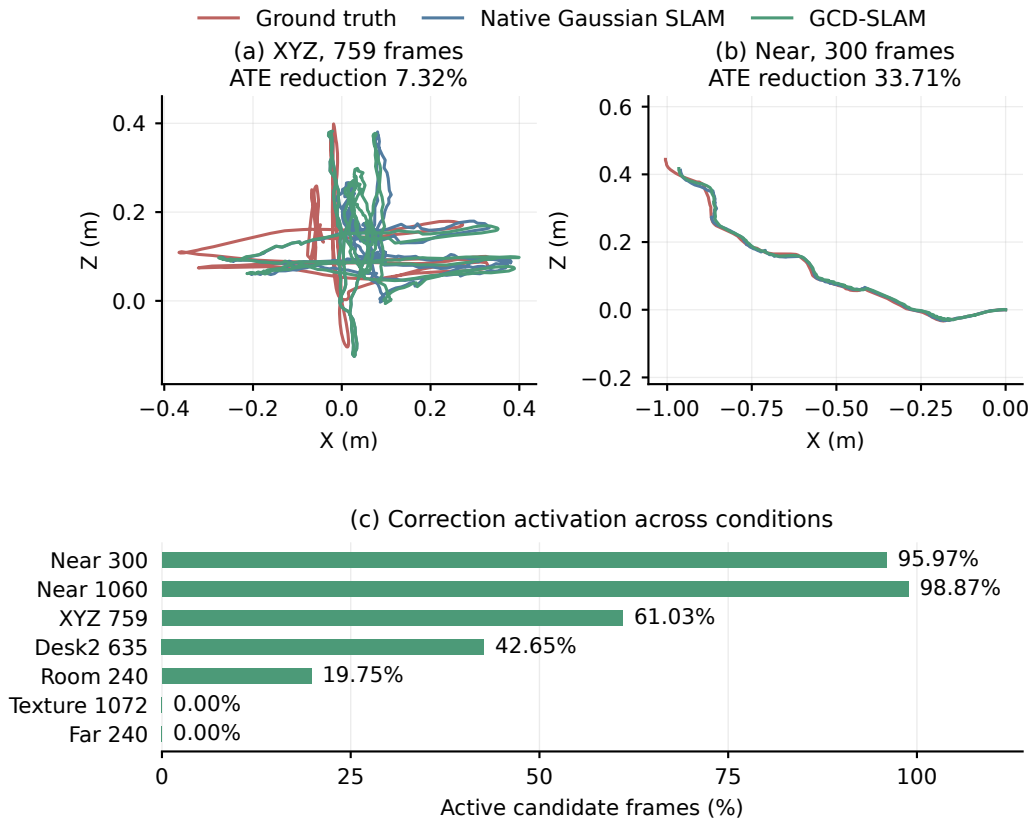

Figure 4: Localization trajectories and correction activation. **(a,b)** Ground truth, Native Gaussian SLAM, and GCD-SLAM on full XYZ and near-300 (seed 0), rigidly aligned at the first pose for visualization. The displayed ATE reductions use trajectory-wide rigid alignment; near-300's 33.71% is a single-run value, versus the three-seed median 34.81% in Table 1. **(c)** Measured correction activation fractions across near horizons, low-texture stresses, and inactive controls. No correction is admitted on texture-near or notexture-far; room still exhibits a 0.31% ATE regression.

4.4.2 Controlled RILB-to-GCD Bridge

Four matched 300-frame arms compare Native, all valid raw corrections, RCS output correction, and feedback to native tracking and mapping. Across 12 probes, RCS improves on all-valid correction in only 2 cases; the RILB ratio correlates -0.322 with that improvement. The mixed result supports RILB as an objective diagnosis, not the sole cause of ATE reduction. Full paired comparisons appear in Appendix B.

4.5 Validation of Localization State Separation

Historical V2.2-H feeds a reoptimized native pose to subsequent tracking and mapping, not the output-only corrected state. Its paired near-300 ATE rises from 1.725 to 1.875 cm (8.66%), despite lower translational RPE-30. Final V2.3 GCD retains native state ownership with zero writebacks.

4.6 Corrected Pose Consistency with the Frozen Gaussian Map

Against the same finalized submap and observations, both pose streams retain mean depth support 1.000 across 900 near-prefix frames (three seeds). Median seed-level mean absolute depth residual rises from 0.003442 m (Native) to 0.004006 m (GCD); depth RMSE rises from 0.012079 to 0.013376 m. Frame-wise map snapshots remain a stricter online test.

4.7 Gaussian Reconstruction Preservation

Across 2,454 stored views from three active protocols, Native and GCD share the same map and evaluation views, yielding identical reconstruction metrics ($\Delta = 0$) with zero corrected-pose write-backs (Appendix Table 5).

5 Discussion and Limitations

The gain concentrates on the targeted near-range prefix of the VarSplat backbone: the full-horizon median reduction is 0.414%, room regresses by 0.31%, and one seed fails under both methods. The frozen-map audit finds full depth support at a higher corrected-pose residual.

RCS uses hand-selected thresholds and shapes activation, but fixed-raw replay shows little contribution to the main prefix gain. Held-out reliability learning, frame-wise map audits, and additional backbones remain untested.

6 Conclusion

All 12 fixed-probe pairs reverse pose preference with the Gaussian reference. GCD-SLAM separates temporal-depth correction, admission, and output authority. Three-seed median ATE falls by 34.81% on the selected 300-frame prefix but only 0.414% over 1,060 frames; room regresses by 0.31%. Fixed-raw replay attributes the prefix gain primarily to geometry, while the unchanged map carries a small corrected-pose depth-residual cost.

AI use statement

ChatGPT/Codex assisted with language polishing, literature organization, LaTeX preparation, code development, and illustrative figures. The authors verified all methods, experiments, references, code, figures, and claims.

Reproducibility statement

The method specifies geometric and admission thresholds; the experiments report seeds, horizons, alignment, and paired controls. Appendix C describes the frozen-map audit and shared-map reconstruction checks.

References

Pratik Agarwal, Gian Diego Tipaldi, Luciano Spinello, Cyrill Stachniss, and Wolfram Burgard. Robust map optimization using dynamic covariance scaling. In *IEEE International Conference on Robotics and Automation*, pp. 62–69, 2013. doi: 10.1109/ICRA.2013.6630557.

Carlos Campos, Richard Elvira, Juan J. Gomez Rodriguez, Jose M. M. Montiel, and Juan D. Tardos. ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multi-map SLAM. *IEEE Transactions on Robotics*, 37(6):1874–1890, 2021. doi: 10.1109/TRO.2021.3075644.

Seongbo Ha, Jiung Yeon, and Hyeonwoo Yu. RGBD GS-ICP SLAM. In *European Conference on Computer Vision*, pp. 180–197, 2024. doi: 10.1007/978-3-031-72764-1_11.

Jiarui Hu, Xianhao Chen, Boyin Feng, Guanglin Li, Liangjing Yang, Hujun Bao, Guofeng Zhang, and Zhaopeng Cui. CG-SLAM: Efficient dense RGB-D SLAM in a consistent uncertainty-aware 3D gaussian field. In *European Conference on Computer Vision*, pp. 93–112, 2024. doi: 10.1007/978-3-031-72698-9_6.

Huajian Huang, Longwei Li, Hui Cheng, and Sai-Kit Yeung. Photo-slam: Real-time simultaneous localization and photorealistic mapping for monocular, stereo, and RGB-D cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21584–21593, 2024. doi: 10.1109/CVPR52733.2024.02039.

Nikhil Keetha, Jay Karhade, Krishna Murthy Jatavallabhula, Gengshan Yang, Sebastian Scherer, Deva Ramanan, and Jonathon Luiten. SplaTAM: Splat, track and map 3D gaussians for dense RGB-D SLAM. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21357–21366, 2024.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuehler, and George Drettakis. 3D gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023. doi: 10.1145/3592433.

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*, pp. 18039–18048, 2024.

Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohli, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. KinectFusion: Real-time dense surface mapping and tracking. In *IEEE International Symposium on Mixed and Augmented Reality*, 2011.

Juergen Sturm, Nikolas Engelhard, Felix Endres, Wolfram Burgard, and Daniel Cremers. A benchmark for the evaluation of RGB-D SLAM systems. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 573–580, 2012. doi: 10.1109/IROS.2012.6385773.

Niko Sunderhauf and Peter Protzel. Switchable constraints for robust pose graph SLAM. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1879–1884, 2012. doi: 10.1109/IROS.2012.6385590.

Zachary Teed and Jia Deng. DROID-SLAM: Deep visual SLAM for monocular, stereo, and RGB-D cameras. In *Advances in Neural Information Processing Systems*, 2021.

Anh Thuan Tran and Jana Kosecka. VarSplat: Uncertainty-aware 3D gaussian splatting for robust RGB-D SLAM. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 26072–26082, 2026.

Chi Yan, Delin Qu, Dan Xu, Bin Zhao, Zhigang Wang, Dong Wang, and Xuelong Li. GS-SLAM: Dense visual SLAM with 3D gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19595–19604, 2024.

Heng Yang, Pasquale Antonante, Vasileios Tzoumas, and Luca Carlone. Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection. *IEEE Robotics and Automation Letters*, 5(2):1127–1134, 2020a.

Nan Yang, Lukas von Stumberg, Rui Wang, and Daniel Cremers. D3VO: Deep depth, deep pose and deep uncertainty for monocular visual odometry. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1281–1292, 2020b.

Vladimir Yugay, Yue Li, Theo Gevers, and Martin R. Oswald. Gaussian-slam: Photo-realistic dense SLAM with gaussian splatting. *arXiv preprint arXiv:2312.10070*, 2023.

Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3D: A modern library for 3d data processing. *arXiv preprint arXiv:1801.09847*, 2018.

Liyuan Zhu, Yue Li, Erik Sandström, Shengyu Huang, Konrad Schindler, and Iro Armeni. Loop-Splat: Loop closure by registering 3D gaussian splats. In *International Conference on 3D Vision*, pp. 156–167, 2025. doi: 10.1109/3DV66043.2025.00020.

Zihan Zhu, Songyou Peng, Viktor Larsson, Weiwei Xu, Hujun Bao, Zhaopeng Cui, Martin R. Oswald, and Marc Pollefeys. NICE-SLAM: Neural implicit scalable encoding for SLAM. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12786–12796, 2022.

A External Tracking Context

Table 3: Available external tracking results retained as protocol-separated context. ATE RMSE is in cm (lower is better). Values come from frozen available artifacts with method-specific execution and aggregation; they do not constitute a controlled cross-system ranking against Table 1.

Method	fr1/desk	fr2/xyz	fr3/office	Near-300	Near-1060
Matched frames	592	3,397	2,515	300	1,060
SplaTAM	1.885	1.554	2.862	1.437	2.241
LoopSplat	1.989	1.496	4.106	–	–
GS-ICP	3.387	2.299	3.196	3.683	2.206
Gaussian-SLAM	2.964	2.180	5.085	–	–

Public-sequence values use one frozen run per method. Near-300 uses one SplaTAM run and a three-run GS-ICP median; Near-1060 uses one SplaTAM run and a three-run GS-ICP median. Dashes denote unavailable results under these horizons.

B Additional Localization Diagnostics

On the near prefix, translational RPE-1/30 changes from $0.368/1.744$ cm to $0.324/1.159$ cm, while rotational RPE-30 changes from 1.005° to 0.761° (three-run medians). RCS activates on 61.03%, 42.65%, and 19.75% of candidates on low-texture XYZ, desk2, and room, respectively. RCS-on fixed-raw replay reproduces the logged GCD trajectory to 8.9×10^{-16} maximum absolute error.

The four-arm bridge uses A0 Native, A1 all-valid raw correction, A2 recorded RCS in the output-only state, and A3 the same RCS-admitted candidate in native tracking and mapping. All arms share observations, Native initialization, probe schedule, fallback, and the 300-frame horizon. Across 12 probes, A2 improves over A1 in 2 cases (median A1–A2 difference -0.002361 m), while A3 improves over A2 in 7 cases (median A2–A3 difference $+0.000085$ m).

Table 4: Runtime of the decoupled correction pathway.

Sequence	Total (min)	GPU (GB)	Correction (ms/frame)	Share (%)
Near-1060	46.19	1.63	49.95	1.91
XYZ-759	24.43	3.00	59.58	3.08
Desk2-635	24.93	1.36	63.87	2.70

Table 5: Gaussian reconstruction preservation under localization correction.

(a) Native/GCD preservation on identical maps and evaluation views

Sequence	Output	Views	PSNR ↑ (dB)	SSIM ↑	LPIPS ↓	Depth L1 ↓ (m)	Δ	Writes
fr3/structure-notexture-near	Native	1,060	32.089	0.9431	0.2495	0.1017	–	–
	GCD-SLAM	1,060	32.089	0.9431	0.2495	0.1017	0	0
fr1/xyz-low-texture	Native	759	28.231	0.8873	0.2431	0.1294	–	–
	GCD-SLAM	759	28.231	0.8873	0.2431	0.1294	0	0
fr1/desk2-low-texture	Native	635	30.351	0.9403	0.1198	0.1926	–	–
	GCD-SLAM	635	30.351	0.9403	0.1198	0.1926	0	0

Native is VarSplat. Each pair evaluates the same physical map, not independent reconstructions. All four metric differences Δ (GCD minus Native) are zero by map/view identity; Writes counts corrected-pose mapper writebacks.

(b) Evaluation compatibility of available reconstruction artifacts

Method	Available evidence	Evaluation views	Use in this paper
Native / GCD	Shared checkpoints and exhaustive renders	Same 635–1,060 stored views per run	Preservation pair
SplaTAM	Saved parameters and rendering metrics	2–7 selected views per public scene	Qualitative context
Photo-SLAM	Saved maps, renders, frozen replay	Method-specific saved views	Qualitative context
Gaussian-SLAM	Completed runs and saved renders	Method-specific saved views	Qualitative context

External methods lack a common-view evaluation matched to GCD. Their saved renders in Fig. 5 are not used for a quantitative mapping ranking.

C ADDITIONAL MAPPING VISUALIZATION

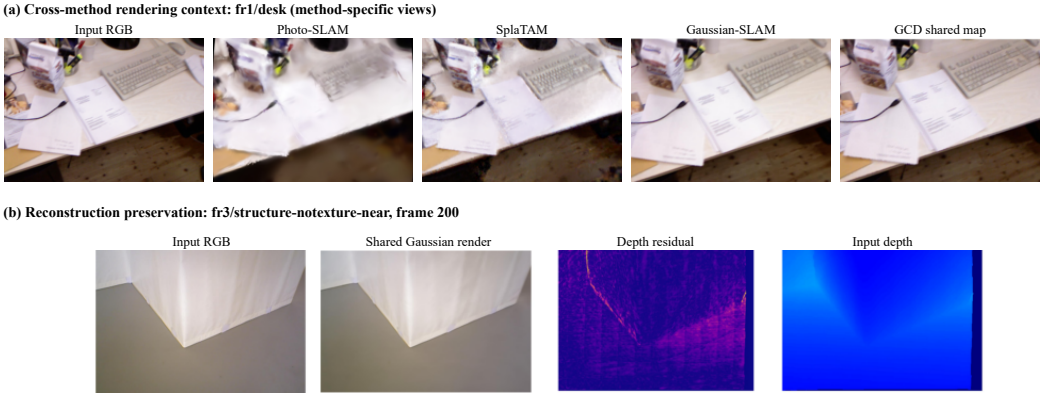

Figure 5: Mapping context and reconstruction preservation. **(a)** Real saved renders on TUM `fr1/desk`: Photo-SLAM uses saved view 93, SplaTAM/Gaussian-SLAM frame 499, and GCD frame 500. These are method-specific views, not a synchronized quality ranking. **(b)** Near-range frame 200: input RGB, shared Native/GCD rendering, depth residual, and input depth. Corrected localization has zero map writeback.