ISPRS Journal of Photogrammetry and Remote Sensing 242 (2026) 36–52

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

# From particles to features: Uncertainty-guided geometric reasoning for robust SLAM

Cabrel Wouladje [b], Golden Tendekai Mumanikidzwa [b], Huiying Xu [a,b,*], Hongbo Li [c,*], Mantang Liu [c], Wenjie Wang [d], Zhendong Chen [b], Wenzhe Tan [c], Chao Chen [c], Balong Wang [b], Zhenglong Wang [c], Xinzhong Zhu [a,b,c,*]

[a] *Zhejiang Key Laboratory of Intelligent Education and Application, Zhejiang Normal University, Jinhua, 321004, Zhejiang, China*
[b] *School of Computer Science and Technology, Zhejiang Normal University, Jinhua, 321004, China*
[c] *Beijing Geekplus Technology Co., Ltd, China*
[d] *Shenzhen Qi Xuan Technology Co., Ltd, China*

ARTICLE INFO

*Keywords:*
LiDAR-inertial odometry
Particle filtering
Feature selection
Uncertainty quantification

ABSTRACT

Simultaneous Localization and Mapping (SLAM) systems must commit to feature correspondences and loop closure decisions under uncertainty, yet incorrect commitments cascade into catastrophic mapping failures. Traditional particle filters address state uncertainty but leave feature-level processing deterministic, creating an asymmetry that limits robustness. We present Particle-Guided Feature Fusion (PGFF), a framework that inverts the classical role of particles: rather than sampling robot poses, particles propagate through feature space to compute importance weights that inform geometric evidence accumulation. This paradigm shift enables three innovations: (1) geometric surprise priors that identify informative features through local eigenvalue analysis and information-theoretic divergence, (2) adaptive feature weighting where particles vote on correspondence reliability based on spatial and geometric consistency, and (3) multi-hypothesis loop closure management that defers commitment until sufficient evidence accumulates. We integrate PGFF into a tightly-coupled LiDAR-inertial odometry system and evaluate on the NCLT long-term autonomy benchmark and the VBR dataset, both featuring dynamic objects, geometric degeneracy, and perceptual aliasing. PGFF reduces absolute trajectory error by 23.5% compared to FAST-LIO2 and 29.8% compared to LIO-SAM, with gains evident against recent degeneracy-aware methods as well. Combining PGFF with graduated non-convexity (GNC) yields further additive improvements, confirming that pre-correspondence and post-correspondence robustness address complementary failure modes. PGFF achieves 10.1% faster IEKF convergence with only 0.7% total runtime overhead. Multi-hypothesis loop closure reduces false positives by 58% across multiple place recognition front-ends while maintaining 94% recall. These results demonstrate that uncertainty-guided feature selection and evidence accumulation, not just state estimation, is essential for reliable long-term autonomy, suggesting the boundary between filtering and perception should be fundamentally reconsidered. Our implementation is available at https://github.com/wencabrel/PGFF-SLAM.

## 1. Introduction

A delivery robot navigating a warehouse corridor encounters a common yet catastrophic scenario: symmetric walls on both sides, repetitive structural elements, and a distant loop closure candidate that matches equally well to three different locations. The robot's SLAM system, forced to commit immediately, selects the wrong correspondence. Within minutes, the accumulated drift renders the map unusable. This failure mode–premature commitment under geometric ambiguity—remains a fundamental barrier to reliable long-term autonomy (Cadena et al., 2016).

The root cause lies in an architectural asymmetry present in virtually all modern SLAM systems. State estimation pipelines, from extended Kalman filters to factor graphs (Dellaert and Kaess, 2017), explicitly propagate uncertainty through probability distributions, covariance matrices, and hypothesis spaces. Tightly-coupled LiDAR-inertial odometry systems such as FAST-LIO2 (Xu et al., 2022) and LIO-SAM (Shan et al., 2020) achieve remarkable accuracy by fusing IMU preintegration (Forster et al., 2017) with iterative state estimation. Yet the feature extraction and correspondence stages that feed these estimators operate deterministically: every LiDAR point receives equal

* Corresponding authors.
*E-mail addresses:* xhy@zjnu.edu.cn (H. Xu), Jason.li@geekplus.com (H. Li), zxz@zjnu.edu.cn (X. Zhu).

https://doi.org/10.1016/j.isprsjprs.2026.08.012
Received 31 March 2026; Received in revised form 5 August 2026; Accepted 15 August 2026
Available online 4 September 2026

treatment, every loop closure candidate demands immediate acceptance or rejection, and geometric reasoning proceeds without access to the uncertainty information flowing through the rest of the system. This asymmetry means that state estimators reason carefully about *where* the robot might be, but not about *what* the robot should attend to.

Traditional approaches address this gap through outlier rejection mechanisms—RANSAC (Fischler and Bolles, 1981), M-estimators (Huber, 1964), and heuristic distance thresholds—applied as post-processing steps after feature extraction. While effective against random noise, these methods fail systematically in structured environments where outliers are geometrically coherent: dynamic objects moving in formation, symmetric architectural elements, or degenerate corridors where all points appear equally (un)informative. Place recognition descriptors such as Scan Context (Kim and Kim, 2018) enable loop closure detection but still require immediate commitment decisions. The fundamental limitation is that these mechanisms lack access to the probabilistic reasoning occurring downstream in state estimation.

We propose **Particle-Guided Feature Fusion (PGFF)**, a framework that resolves this asymmetry by inverting the classical role of particles in SLAM. Rather than using particles to sample robot poses—the approach established by FastSLAM (Montemerlo et al., 2002) and its Rao-Blackwellized variants (Grisetti et al., 2007)–PGFF propagates particles through feature space to compute importance weights that inform geometric evidence accumulation. This inversion allows uncertainty information to flow bidirectionally: from state estimation back into feature selection, creating a closed loop where the system's confidence about its state informs what features it should trust. Fig. 1 illustrates this fundamental difference: traditional pipelines treat feature extraction as a deterministic preprocessing step, whereas PGFF establishes bidirectional information flow between state uncertainty and feature weighting.

*A note on terminology.* We use "geometric reasoning" throughout this paper to denote a specific class of operations: the evaluation of competing geometric hypotheses, the accumulation of evidence across observations, and the deferred commitment of decisions until posterior confidence exceeds a threshold. This usage is narrower than chain-of-thought reasoning in language models and does not claim symbolic or compositional inference. Considered in isolation, the feature weighting components of PGFF (Sections 3.2–3.3) are principled weighting mechanisms grounded in information theory and particle consensus. What elevates the framework beyond weighting alone is the multi-hypothesis loop closure module (Section 3.4), which maintains competing pose hypotheses, accumulates per-frame Bayesian evidence (Eq. (22)), and defers commitment until a posterior ratio threshold is satisfied. This evidence-accumulation-and-commitment cycle, combined with bidirectional feedback between state uncertainty and feature selection, is what we refer to as "geometric reasoning" throughout the paper.

We acknowledge that the operations described here – multi-hypothesis evaluation, Bayesian evidence accumulation, and posterior-based deferred commitment – are well-established in probabilistic robotics and are more precisely described as *Bayesian inference* in the mathematical sense. PGFF is one instantiation of this paradigm, not a new form of inference. Our contribution is not the reasoning framework itself but a specific coupling design: the bidirectional feedback between state uncertainty and pre-correspondence feature selection, combined with a descriptor-agnostic multi-hypothesis loop closure module built on LiDAR-geometric residuals. We retain the term "geometric reasoning" in the title and in section headings as a compact descriptor for this coupling, but we use the more precise language of Bayesian inference and evidence accumulation throughout the technical sections.

This paradigm shift enables three technical contributions:

1. **Geometric surprise priors.** We derive an information-theoretic measure of feature importance based on local eigenvalue analysis (Weinmann et al., 2015). Points whose observed geometry diverges from neighborhood expectations—quantified via Kullback–Leibler divergence—receive higher weights, automatically emphasizing corners, edges, and structural discontinuities while suppressing redundant planar regions.
2. **Particle-based adaptive weighting.** Particles distributed in feature space vote on correspondence reliability. High consensus among particles indicates robust correspondences; disagreement signals geometric ambiguity requiring caution. This mechanism provides principled outlier handling without hand-tuned thresholds, complementing rather than replacing existing robust estimation techniques.
3. **Multi-hypothesis loop closure.** Rather than committing immediately to loop closure candidates, PGFF maintains competing hypotheses and accumulates evidence across frames using Bayesian updates. Commitment occurs only when the posterior probability ratio exceeds a decision threshold, preventing the catastrophic failures that arise from premature closure in perceptually aliased environments (Lowry et al., 2016).

We integrate PGFF into a tightly-coupled LiDAR-inertial odometry system built upon the iterated extended Kalman filter (IEKF) formulation (Xu and Zhang, 2021). Evaluation on the NCLT long-term autonomy benchmark (Carlevaris-Bianco et al., 2016) and a challenging 43 GB urban campus dataset demonstrates that PGFF reduces absolute trajectory error by 23.5% compared to state-of-the-art baselines (FAST-LIO2, LIO-SAM) while simultaneously achieving 10.1% faster state estimation. Ablation studies confirm that each component contributes meaningfully to overall performance.

Beyond empirical gains, our results carry a broader implication: the boundary between filtering and feature extraction—long treated as a fixed architectural decision in the SLAM pipeline (Stachniss et al., 2016)—should be reconsidered. PGFF demonstrates that uncertainty-aware perception, not just uncertainty-aware estimation, is essential for robust autonomy.

## 2. Related work

We position PGFF within three intersecting research areas: LiDAR-inertial odometry and state estimation, feature selection and robust correspondence, and loop closure under perceptual ambiguity.

### 2.1. LiDAR-inertial odometry

The fusion of LiDAR and inertial measurements has matured considerably over the past decade. LOAM (Zhang and Singh, 2014) established the paradigm of separating high-frequency odometry from low-frequency mapping through edge and planar feature extraction. Subsequent systems have pursued tighter sensor coupling: LIO-SAM (Shan et al., 2020) formulates the problem as factor graph optimization with IMU preintegration (Forster et al., 2017), while FAST-LIO2 (Xu et al., 2022) achieves real-time performance through direct point registration and incremental kd-tree updates. Point-LIO (He et al., 2023) pushes bandwidth limits by processing individual LiDAR points at 4–8 kHz, enabling operation under aggressive motions that saturate IMU measurements. For scenarios demanding simplicity, KISS-ICP (Vizzo et al., 2023) demonstrates that carefully implemented point-to-point ICP with adaptive thresholding achieves competitive accuracy without IMU integration. Multi-metric approaches like MULLS (Pan et al., 2021) jointly optimize point-to-point, point-to-plane, and point-to-line residuals across semantically classified features.

Recent work has addressed geometric degeneracy directly: DALI-SLAM (Wu et al., 2025) detects degenerate directions via eigenvalue analysis of the point-to-plane Hessian and constrains optimization along well-conditioned axes, achieving strong performance in featureless environments. Other advances include multi-LiDAR configurations for expanded field-of-view (Xu et al., 2023), and domain-specific adaptations for forests (Shao et al., 2024) and underground spaces. These

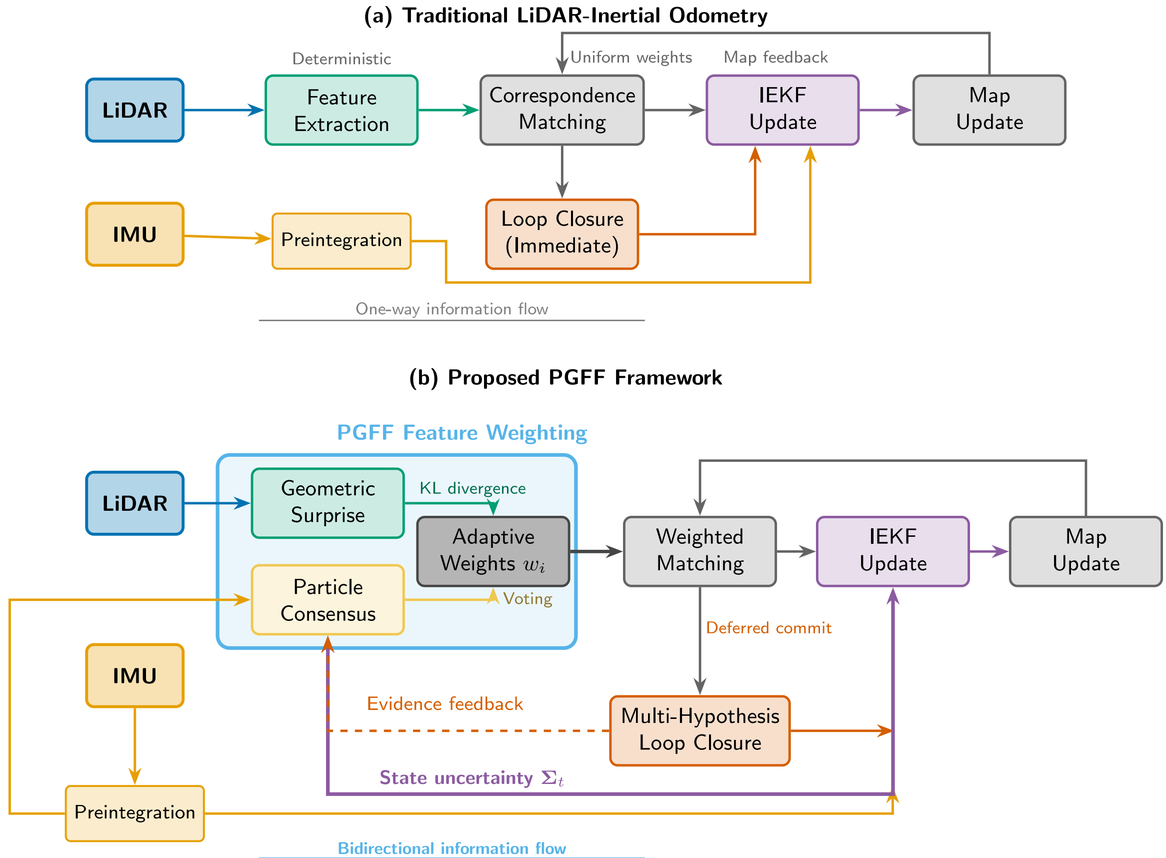


**Fig. 1.** Comparison of (a) traditional LiDAR-inertial odometry pipelines and (b) the proposed PGFF framework. Traditional approaches treat feature extraction as deterministic preprocessing with one-way information flow. PGFF inverts this paradigm by propagating particles through feature space, enabling bidirectional information exchange where state uncertainty guides feature weighting, and feature consensus improves state estimation. The multi-hypothesis loop closure module defers commitment until evidence accumulates, preventing catastrophic failures from premature decisions.

systems achieve impressive accuracy under nominal conditions, yet they share a common limitation: feature extraction operates independently of state uncertainty. Points are weighted uniformly or by simple geometric heuristics, with no mechanism for the state estimator to influence which features receive attention.

### 2.2. Feature selection and robust correspondence

Outlier rejection in point cloud registration has traditionally relied on RANSAC (Fischler and Bolles, 1981) and M-estimators (Huber, 1964), which downweight residuals exceeding learned or heuristic thresholds. The graduated non-convexity (GNC) framework (Yang et al., 2020) provides a principled alternative, enabling robust estimation without initial guesses by smoothly transitioning from convex to non-convex cost functions. Adaptive robust kernels (Chebrolu et al., 2021) extend this by automatically tuning kernel parameters based on residual distributions.

Geometric analysis offers complementary insights: eigenvalue decomposition of local neighborhoods enables classification of points as linear, planar, or scattered (Weinmann et al., 2015), informing feature selection in systems like LeGO-LOAM (Shan and Englot, 2018). Learning-based methods have emerged for adaptive feature weighting, though they require extensive training data and may not generalize across environments. Kimera (Rosinol et al., 2020) demonstrates the value of semantic information for robust visual-inertial SLAM, while recent LiDAR systems (Shao et al., 2024) show that domain-adapted feature selection substantially improves performance in challenging scenarios. However, existing approaches derive feature weights from local geometry or learned priors alone, without conditioning on the global uncertainty state. PGFF addresses this gap by using particle consensus to propagate state-level uncertainty back into feature weighting decisions.

### 2.3. Loop closure and place recognition

Reliable loop closure is essential for long-term mapping consistency. Scan Context (Kim and Kim, 2018) and its variants provide efficient global descriptors for place recognition by projecting LiDAR scans onto 2D representations, while learned descriptors offer improved robustness to viewpoint changes (Lowry et al., 2016). The challenge of perceptual aliasing – where distinct locations produce similar sensor observations – has motivated multi-hypothesis approaches in visual SLAM and robust pose graph optimization (Latif et al., 2013; Sünderhauf and Protzel, 2012), but these techniques remain underexplored in LiDAR systems. The multi-hypothesis paradigm itself traces back to multiple hypothesis tracking (MHT) in target tracking (Reid, 1979; Blackman, 2004), which addresses data association ambiguity by maintaining branching hypothesis trees with periodic pruning. SLAM has adapted these ideas through delayed data association (Nieto et al., 2007) and pose graph methods that maintain alternate edge sets. Recent multimodal frameworks further fuse complementary loop closure signals across sensing modalities to disambiguate aliased locations (Chghaf et al., 2023, 2025). PGFF differs from these approaches along three axes: (i) hypotheses are maintained at the loop closure level rather than over landmark associations or pose graph edges, with each hypothesis storing only a relative transform rather than a full graph branch; (ii) evidence accumulation uses LiDAR-geometric residuals projected through the candidate transformation (Eq. (21)), so hypothesis evaluation is grounded in the same observation model used by the IEKF; and (iii) a null hypothesis $h_0$ prevents forced selection among poor candidates, which is not standard in classical MHT formulations.

Recent work on degeneracy-aware loop closing (Ebadi et al., 2021) demonstrates the value of uncertainty reasoning for place recognition in perceptually degraded environments such as caves and tunnels. Multi-robot cooperative SLAM systems (Liu et al., 2023; Tian et al., 2022) face similar challenges when establishing inter-robot loop closures, employing techniques like pairwise consistency maximization and distributed outlier rejection. The robust incremental SLAM (riSAM) framework (McGann et al., 2023) applies GNC principles to online pose graph optimization, achieving robustness to 50% outlier loop closures. PGFF contributes to this area by maintaining multiple loop closure hypotheses and deferring commitment until accumulated evidence resolves ambiguity, preventing the catastrophic failures that arise from premature closure decisions.

### 2.4. Positioning PGFF

The key distinction of our work lies in inverting the classical relationship between particles and features. Where FastSLAM (Montemerlo et al., 2002) uses particles to sample robot poses while treating feature observations deterministically, PGFF uses particles to evaluate feature reliability while the pose is estimated via conventional filtering. This inversion creates a bidirectional information flow: state uncertainty informs feature attention, and feature consensus improves state estimation. While probabilistic correspondence estimation (Chebrolu et al., 2021) and adaptive feature weighting have precedent in prior work, the specific configuration we propose – particle ensembles in feature space that propagate prior state uncertainty back into pre-correspondence weighting within a tightly-coupled LiDAR-inertial framework – has not, to our knowledge, been previously instantiated. Our multi-hypothesis loop closure mechanism complements this by extending evidence-based deferred commitment, an idea with roots in classical MHT (Reid, 1979) and SLAM-specific delayed association, to the loop-closure layer of a LiDAR system through descriptor-agnostic Bayesian evidence accumulation.

## 3. Methodology

This section presents the technical details of Particle-Guided Feature Fusion (PGFF). We begin with an overview of the system architecture and problem formulation (Section 3.1), then describe the geometric surprise prior for information-theoretic feature analysis (Section 3.2), the particle-based consensus mechanism for adaptive weighting (Section 3.3), the multi-hypothesis loop closure strategy (Section 3.4), integration with iterated extended Kalman filtering (Section 3.5), and implementation details (Section 3.6).

### 3.1. Overview and problem formulation

Fig. 2 illustrates the PGFF pipeline. Raw LiDAR scans undergo preprocessing to extract candidate feature points, which are then evaluated by two parallel modules: the geometric surprise prior computes information-theoretic importance scores based on local eigenvalue analysis, while the particle filter maintains a distribution over feature reliability hypotheses. These modules produce per-point weights that feed into the IEKF for state estimation. The estimated state uncertainty, in turn, influences particle resampling and geometric expectations, creating a bidirectional information flow. Loop closure candidates enter a multi-hypothesis module that defers commitment until sufficient evidence accumulates.

We define the system state as:

$$\mathbf{x} = \begin{bmatrix} \mathbf{R} & \mathbf{p} & \mathbf{v} & \mathbf{b}_g & \mathbf{b}_a \end{bmatrix}^\top \in SO(3) \times \mathbb{R}^{12} \tag{1}$$

where $\mathbf{R} \in SO(3)$ denotes orientation, $\mathbf{p} \in \mathbb{R}^3$ position, $\mathbf{v} \in \mathbb{R}^3$ velocity, and $\mathbf{b}_g, \mathbf{b}_a \in \mathbb{R}^3$ gyroscope and accelerometer biases respectively. Given a LiDAR scan $\mathcal{P} = \{\mathbf{p}_i\}_{i=1}^N$ with $N$ points, the central problem is to assign weights $\mathbf{w} = \{w_i\}_{i=1}^N$ such that state estimation accuracy is maximized.

Traditional approaches employ uniform weighting ($w_i = 1$) or residual-based reweighting ($w_i = \rho'(r_i)$ for robust kernel $\rho$). Both operate post-hoc: weights depend only on registration residuals computed *after* correspondence establishment. PGFF instead computes weights *before* and *during* registration through:

$$w_i = f_{\mathrm{PGFF}}(\mathcal{S}_i, \mathcal{C}_i, \boldsymbol{\Sigma}_t) \tag{2}$$

where $\mathcal{S}_i$ is the geometric surprise of point $i$, $\mathcal{C}_i$ is the particle consensus score, and $\boldsymbol{\Sigma}_t \triangleq \boldsymbol{\Sigma}_{t|t-1}$ is the prior state covariance propagated from the previous IEKF update. Using the prior rather than the posterior ensures causal consistency: PGFF weights are computed before the current measurement update and do not depend on the filter output they feed into.

### 3.2. Geometric surprise prior

The geometric surprise prior identifies points that carry high information content by measuring deviation from local geometric expectations. We build upon eigenvalue-based feature analysis (Weinmann et al., 2015) but extend it with an information-theoretic interpretation that enables principled fusion with particle-based consensus.

#### 3.2.1. Local geometric analysis

For each point $\mathbf{p}_i$, we compute the covariance matrix of its $k$-nearest neighbors $\mathcal{N}_i$:

$$\mathbf{C}_i = \frac{1}{|\mathcal{N}_i|} \sum_{\mathbf{p}_j \in \mathcal{N}_i} (\mathbf{p}_j - \bar{\mathbf{p}}_i)(\mathbf{p}_j - \bar{\mathbf{p}}_i)^\top \tag{3}$$

where $\bar{\mathbf{p}}_i$ is the centroid of $\mathcal{N}_i$. Eigenvalue decomposition yields $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$, from which we derive normalized geometric descriptors:

$$a_{1D} = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \epsilon} \qquad \text{(linearity)} \tag{4}$$

$$a_{2D} = \frac{\lambda_2 - \lambda_3}{\lambda_1 + \epsilon} \qquad \text{(planarity)} \tag{5}$$

$$a_{3D} = \frac{\lambda_3}{\lambda_1 + \epsilon} \qquad \text{(scatterness)} \tag{6}$$

where $\epsilon$ is a small constant for numerical stability. These descriptors satisfy $a_{1D} + a_{2D} + a_{3D} = 1$ and characterize local geometry: high $a_{1D}$ indicates edges or poles, high $a_{2D}$ indicates planar surfaces, and high $a_{3D}$ indicates corners or vegetation.

#### 3.2.2. Information-theoretic surprise

Rather than using geometric descriptors directly as weights—which would uniformly favor certain structure types—we measure how much observed geometry *deviates* from spatial expectations. Let $\mathbf{a}_i = [a_{1D}, a_{2D}, a_{3D}]^\top$ denote the descriptor vector for point $i$, and let $\bar{\mathbf{a}}_{\mathcal{R}}$ denote the average descriptor over spatial region $\mathcal{R}$ containing point $i$. We define geometric surprise as the Kullback–Leibler divergence:

$$\mathcal{S}_i = D_{\mathrm{KL}}(\mathbf{a}_i \parallel \bar{\mathbf{a}}_{\mathcal{R}}) = \sum_{d \in \{1D,2D,3D\}} a_d^{(i)} \log \frac{a_d^{(i)}}{\bar{a}_d^{(\mathcal{R})} + \epsilon} \tag{7}$$

This formulation captures the intuition that *informative points are those that violate local expectations*. A corner in a planar wall region has high surprise because its scatterness deviates from the surrounding planarity. Conversely, a planar point within a large planar region has low surprise despite being geometrically well-defined—it provides redundant information. This distinction is critical for efficient state estimation: we seek features that constrain the solution, not merely features that are geometrically salient.

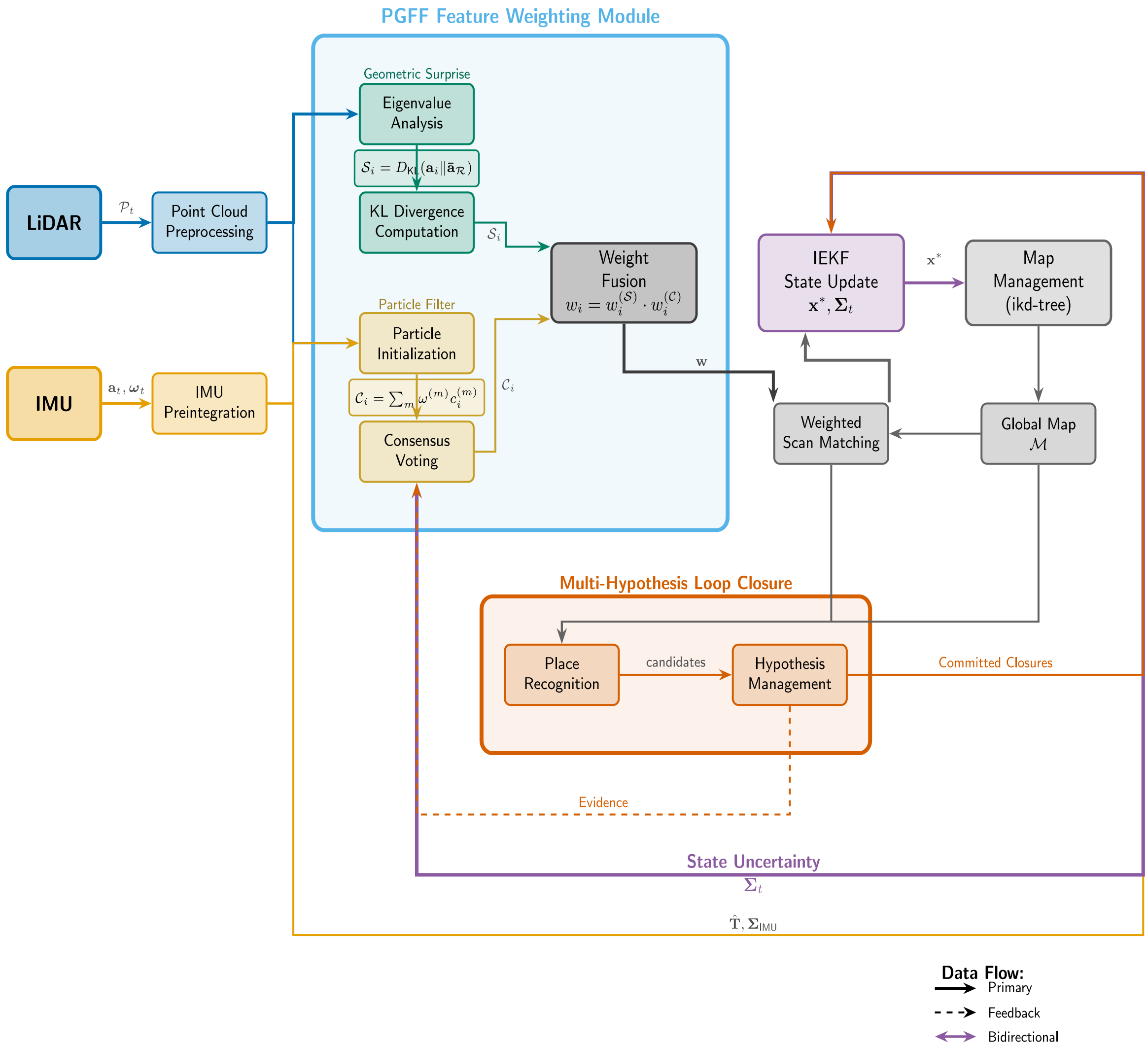


**Fig. 2.** System architecture of the proposed PGFF framework. LiDAR points are processed through geometric surprise analysis and particle-based consensus voting to produce adaptive weights for state estimation. State uncertainty feeds back to influence feature evaluation, while loop closure candidates are maintained as competing hypotheses until evidence supports commitment.

#### 3.2.3. Spatial aggregation

To compute regional expectations $\bar{\mathbf{a}}_{\mathcal{R}}$ efficiently, we partition the point cloud into a voxel grid $\mathcal{V}$ with resolution $\delta_v$. For each voxel $v \in \mathcal{V}$, we maintain running statistics:

$$\bar{\mathbf{a}}_v = \frac{1}{|v|} \sum_{\mathbf{p}_i \in v} \mathbf{a}_i, \quad \sigma_v^2 = \frac{1}{|v|} \sum_{\mathbf{p}_i \in v} \|\mathbf{a}_i - \bar{\mathbf{a}}_v\|^2 \tag{8}$$

Regional expectations are then computed via Gaussian-weighted aggregation over neighboring voxels:

$$\bar{\mathbf{a}}_{\mathcal{R}(i)} = \frac{\sum_{v \in \mathcal{N}_v(i)} \exp(-\|v - v_i\|^2 / 2\sigma_r^2) \cdot |v| \cdot \bar{\mathbf{a}}_v}{\sum_{v \in \mathcal{N}_v(i)} \exp(-\|v - v_i\|^2 / 2\sigma_r^2) \cdot |v|} \tag{9}$$

where $\mathcal{N}_v(i)$ denotes voxels within radius $r_v$ of the voxel containing point $i$, and $\sigma_r$ controls spatial smoothing.

The geometric surprise prior produces an unnormalized weight:

*Robustness considerations.* The neighborhood covariance in Eq. (3) is sensitive to point cloud density variations, occlusion boundaries, and motion distortion, all of which alter the local $k$-nearest-neighbor structure. Three mechanisms mitigate these effects. First, the voxel-based spatial aggregation in Eq. (9) smooths regional expectations over neighboring voxels, preventing isolated density fluctuations from producing spurious surprise values. Second, the stability constant $\epsilon$ in Eqs. (4)–(7) prevents degenerate eigenvalue ratios in sparse regions from dominating the KL divergence. Third, and most importantly, geometric surprise is never used in isolation: the multiplicative combination with particle consensus (Eq. (19), Section 3.3.5) ensures that a point with high surprise but inconsistent correspondences—as commonly occurs at occlusion boundaries or in motion-distorted regions—is still down-weighted. This layered design makes the overall weighting robust even when individual components are affected by noise.

$$w_i^{(\mathcal{S})} = 1 + \alpha_{\mathcal{S}} \cdot \mathcal{S}_i \tag{10}$$

where $\alpha_{\mathcal{S}}$ controls the influence of surprise relative to a uniform baseline.

### 3.3. Particle-based feature weighting

While geometric surprise identifies locally informative points, it cannot assess correspondence reliability under state uncertainty. A geometrically surprising point may still produce ambiguous correspondences if multiple map regions contain similar structures. The particle-based weighting module addresses this by propagating hypotheses about feature reliability and measuring consensus.

#### 3.3.1. Particle initialization and representation

We maintain a set of $M$ particles $\{\pi^{(m)}\}_{m=1}^{M}$, where each particle $\pi^{(m)} = (\mathbf{T}^{(m)}, \{c_i^{(m)}\}_{i=1}^{N})$ represents a hypothesis consisting of a candidate transformation $\mathbf{T}^{(m)} \in SE(3)$ and per-point correspondence indicators $c_i^{(m)} \in \{0,1\}$. Unlike classical particle filters that sample in pose space (Montemerlo et al., 2002), PGFF particles encode hypotheses about *which features are reliable* given different plausible robot motions.

Particles are initialized by sampling transformations from the motion prior:

$$\mathbf{T}^{(m)} \sim \mathcal{N}(\hat{\mathbf{T}}_{\mathrm{IMU}}, \boldsymbol{\Sigma}_{\mathrm{IMU}}) \tag{11}$$

where $\hat{\mathbf{T}}_{\mathrm{IMU}}$ is the IMU-predicted transformation and $\boldsymbol{\Sigma}_{\mathrm{IMU}}$ is its uncertainty derived from preintegration (Forster et al., 2017). For each particle, correspondence indicators are determined by evaluating point-to-plane distances under the hypothesized transformation:

$$c_i^{(m)} = \mathbf{1}\left[d(\mathbf{T}^{(m)}\mathbf{p}_i, \mathcal{M}) < \tau_c\right] \tag{12}$$

where $d(\cdot, \mathcal{M})$ denotes the point-to-plane distance to the local map $\mathcal{M}$, and $\tau_c$ is a correspondence threshold.

#### 3.3.2. Consensus-based weight computation

The particle consensus score for point $i$ measures agreement across hypotheses:

$$C_i = \frac{1}{M}\sum_{m=1}^{M} \omega^{(m)} \cdot c_i^{(m)} \tag{13}$$

where $\omega^{(m)}$ is the normalized importance weight of particle $m$, computed from the likelihood of observed correspondences:

$$\omega^{(m)} \propto \exp\left(-\frac{1}{2\sigma_d^2}\sum_{i:c_i^{(m)}=1} d(\mathbf{T}^{(m)}\mathbf{p}_i, \mathcal{M})^2\right) \tag{14}$$

High consensus ($C_i \approx 1$) indicates that point $i$ produces valid correspondences across diverse motion hypotheses—a strong signal of reliability. Low consensus indicates either (a) the point lies in a geometrically ambiguous region where different motions yield different correspondences, or (b) the point is a dynamic object or outlier. In either case, reduced weight is appropriate.

The consensus-based weight component is:

$$w_i^{(C)} = C_i^{\beta} \tag{15}$$

where $\beta > 0$ controls the sharpness of consensus-based selection.

#### 3.3.3. Adaptive resampling

Particle diversity is monitored via the effective sample size:

$$N_{\mathrm{eff}} = \frac{1}{\sum_{m=1}^{M}(\omega^{(m)})^2} \tag{16}$$

When $N_{\mathrm{eff}} < \gamma M$ (typically $\gamma = 0.5$), we perform systematic resampling (Grisetti et al., 2007) followed by jittering with magnitude proportional to state uncertainty:

$$\mathbf{T}_{\mathrm{new}}^{(m)} = \mathbf{T}_{\mathrm{resampled}}^{(m)} \oplus \mathcal{N}(\mathbf{0}, \eta\boldsymbol{\Sigma}_t) \tag{17}$$

where $\oplus$ denotes the SE(3) pose composition operator and $\eta$ scales the jitter magnitude. Consistent with Eq. (2), $\boldsymbol{\Sigma}_t$ denotes the prior covariance $\boldsymbol{\Sigma}_{t|t-1}$, preserving causal consistency throughout the PGFF pipeline. This uncertainty-driven resampling creates the bidirectional coupling central to PGFF: high state uncertainty increases particle diversity, which in turn produces more conservative (lower) consensus scores.

#### 3.3.4. Probabilistic interpretation and consistency

The multiplicative weighting in Eq. (19) admits a principled probabilistic interpretation. Let $r_i \in \{0,1\}$ denote the latent reliability of point $\mathbf{p}_i$ for state estimation, where $r_i = 1$ indicates the point provides an informative, correctly-associated constraint. We seek the posterior $P(r_i = 1 \mid \mathbf{a}_i, \{\pi^{(m)}\}, \boldsymbol{\Sigma}_t)$ given the observed local geometry $\mathbf{a}_i$, the particle ensemble, and prior state covariance.

Two conditionally independent sources of evidence inform this posterior. The geometric descriptor $\mathbf{a}_i$ depends only on the local point cloud structure and is independent of motion hypotheses; the particle consensus $C_i$ depends on candidate transformations sampled from the motion prior and is independent of the descriptor's geometric content. Under this conditional independence assumption:

$$\begin{aligned}&P(r_i = 1 \mid \mathbf{a}_i, \{\pi^{(m)}\}, \boldsymbol{\Sigma}_t)\\ &\propto P(\mathbf{a}_i \mid r_i = 1)\cdot P(\{\pi^{(m)}\} \mid r_i = 1, \boldsymbol{\Sigma}_t)\cdot P(r_i = 1).\end{aligned} \tag{18}$$

The geometric surprise term $w_i^{(S)} = 1 + \alpha_S S_i$ approximates the geometric-evidence likelihood ratio $P(\mathbf{a}_i \mid r_i = 1)/P(\mathbf{a}_i \mid r_i = 0)$: points whose descriptors diverge from regional expectations (high KL divergence) are more likely to carry independent constraint information than points consistent with local averages. The additive form with unit baseline ensures that uninformative regions receive a neutral rather than vanishing weight, preserving numerical stability. The consensus term $w_i^{(C)} = C_i^{\beta}$ approximates the motion-evidence likelihood ratio $P(\{\pi^{(m)}\} \mid r_i = 1, \boldsymbol{\Sigma}_t)/P(\{\pi^{(m)}\} \mid r_i = 0, \boldsymbol{\Sigma}_t)$: a point that produces consistent correspondences across particles sampled from the motion prior is more likely to be reliably associated than one whose correspondences vary with the hypothesized transformation. The exponent $\beta$ controls the sharpness of this likelihood, equivalent to a tempered posterior whose calibration is empirically tuned (Section 4.4.2). With a uniform reliability prior $P(r_i = 1) = P(r_i = 0)$, the unnormalized posterior reduces to the product in Eq. (19), establishing $w_i$ as proportional to the posterior reliability odds.

*Consistency with the IEKF update.* Substituting the PGFF weights into the IEKF cost (Eq. (26)) is equivalent to rescaling the per-point information matrices: $\boldsymbol{\Omega}_i \mapsto w_i\boldsymbol{\Omega}_i$. This preserves the weighted least-squares structure of the update and corresponds to a measurement noise model in which the effective observation covariance is $\boldsymbol{\Omega}_i^{-1}/w_i$—points judged reliable contribute with their nominal noise, while unreliable points are effectively observed with inflated noise. Because $w_i \geq 0$ and bounded (the surprise term is bounded by the descriptor simplex and consensus by $C_i \in [0,1]$), the rescaled information matrices remain positive semi-definite, and standard IEKF convergence guarantees (Bell and Cathey, 1993) apply unchanged. PGFF therefore does not alter the filter's underlying estimator class; it modifies only the measurement weighting within an established framework.

*Causal consistency.* As noted in Sections 3.1 and 3.3.3, all PGFF weights are computed from the prior covariance $\boldsymbol{\Sigma}_t \triangleq \boldsymbol{\Sigma}_{t|t-1}$ rather than the posterior. This ensures the weights are deterministic functions of information available before the current measurement update, preventing the circular dependency that would arise if weights were conditioned on the very residuals they are meant to regulate. The resulting filter remains a one-step Bayesian update with a measurement model whose noise scaling is determined by the predictive distribution—a construction consistent with adaptive Kalman filtering (Mehra, 1972) and information-form covariance scaling approaches in robust estimation.

*Relationship to robust kernels.* The PGFF weighting differs from M-estimator and GNC reweighting in two structural respects. First, $w_i$ depends on the descriptor and motion prior rather than on the current residual $r_i(\mathbf{x})$, so weights are computed once per frame rather than iteratively within the IEKF inner loop. Second, the multiplicative combination of two independent evidence sources is not expressible as a single residual-dependent kernel $\rho'(r_i)$. These differences explain the additive empirical gains observed when PGFF is combined with GNC (Table 3): the two mechanisms suppress different failure modes – geometric ambiguity and correspondence outliers respectively – and operate at different stages of the optimization.

**Algorithm 1** PGFF Feature Weighting

**Require:** LiDAR scan $\mathcal{P} = \{\mathbf{p}_i\}_{i=1}^{N}$, IMU prediction $(\hat{\mathbf{T}}, \boldsymbol{\Sigma}_{\text{IMU}})$, local map $\mathcal{M}$, state covariance $\boldsymbol{\Sigma}_t$
**Ensure:** Feature weights $\mathbf{w} = \{w_i\}_{i=1}^{N}$
// **Geometric Surprise Computation**
**for** each point $\mathbf{p}_i \in \mathcal{P}$ **do**
  Compute neighborhood covariance $\mathbf{C}_i$ ⊳ Eq. (3)
  Extract eigenvalues $\lambda_1, \lambda_2, \lambda_3$ and descriptors $\mathbf{a}_i$ ⊳ Eqs. (4)–(6)
  Compute regional expectation $\bar{\mathbf{a}}_{\mathcal{R}(i)}$ via voxel aggregation ⊳ Eq. (9)
  $\mathcal{S}_i \leftarrow D_{\text{KL}}(\mathbf{a}_i \| \bar{\mathbf{a}}_{\mathcal{R}(i)})$ ⊳ Eq. (7)
  $w_i^{(S)} \leftarrow 1 + \alpha_S \cdot \mathcal{S}_i$ ⊳ Eq. (10)
**end for**
// **Particle-Based Consensus**
**for** $m = 1$ to $M$ **do** ⊳ Initialize/update particles
  Sample $\mathbf{T}^{(m)} \sim \mathcal{N}(\hat{\mathbf{T}}, \boldsymbol{\Sigma}_{\text{IMU}})$ ⊳ Eq. (11)
  **for** each point $\mathbf{p}_i$ with $\mathcal{S}_i > \mathcal{S}_{\min}$ **do** ⊳ Lazy evaluation
    $c_i^{(m)} \leftarrow \mathbf{1}[d(\mathbf{T}^{(m)}\mathbf{p}_i, \mathcal{M}) < \tau_c]$ ⊳ Eq. (12)
  **end for**
  Compute particle weight $\omega^{(m)}$ ⊳ Eq. (14)
**end for**
Normalize weights: $\omega^{(m)} \leftarrow \omega^{(m)} / \sum_j \omega^{(j)}$
**for** each point $\mathbf{p}_i \in \mathcal{P}$ **do**
  $C_i \leftarrow \sum_{m=1}^{M} \omega^{(m)} \cdot c_i^{(m)}$ ⊳ Eq. (13)
  $w_i^{(C)} \leftarrow C_i^{\beta}$ ⊳ Eq. (15)
**end for**
// **Adaptive Resampling**
$N_{\text{eff}} \leftarrow 1 / \sum_m (\omega^{(m)})^2$ ⊳ Eq. (16)
**if** $N_{\text{eff}} < \gamma M$ **then**
  Resample particles; jitter with magnitude $\eta \boldsymbol{\Sigma}_t$ ⊳ Eq. (17)
**end if**
// **Combined Weights**
**for** each point $\mathbf{p}_i \in \mathcal{P}$ **do**
  $w_i \leftarrow w_i^{(S)} \cdot w_i^{(C)}$ ⊳ Eq. (19)
**end for**
**return w**

#### 3.3.5. Combined feature weights

The final PGFF weight for point $i$ combines geometric surprise and particle consensus:

$$w_i = w_i^{(S)} \cdot w_i^{(C)} = (1 + \alpha_S \cdot \mathcal{S}_i) \cdot C_i^{\beta} \tag{19}$$

This multiplicative formulation ensures that both conditions must be satisfied for high weight: a point must be geometrically informative *and* produce consistent correspondences across motion hypotheses. Fig. 3 visualizes this pipeline on an example scan. Algorithm 1 summarizes the complete per-frame feature weighting pipeline, highlighting the interplay between geometric analysis and particle-based consensus.

### 3.4. Multi-hypothesis loop closure

Traditional loop closure modules commit immediately when a candidate exceeds a similarity threshold, risking catastrophic map corruption from false positives in perceptually aliased environments. PGFF instead maintains multiple hypotheses and accumulates evidence over time before commitment.

#### 3.4.1. Hypothesis creation and representation

When the place recognition module (Scan Context (Kim and Kim, 2018) in our implementation) identifies a candidate loop closure with similarity score exceeding $\tau_{\text{init}}$, we create a new hypothesis $h_j$ characterized by:

$$h_j = \left(\mathbf{T}_j^{\text{rel}}, P(h_j), t_{\text{create}}, \mathcal{E}_j\right) \tag{20}$$

where $\mathbf{T}_j^{\text{rel}} \in SE(3)$ is the estimated relative transformation, $P(h_j)$ is the posterior probability, $t_{\text{create}}$ is the creation timestamp, and $\mathcal{E}_j = \{e_1, e_2, \ldots\}$ is the accumulated evidence sequence.

Crucially, we do not immediately incorporate $\mathbf{T}_j^{\text{rel}}$ into the pose graph. Instead, the hypothesis remains *pending* while evidence accumulates. This lightweight representation—storing only the relative transform rather than duplicating the full graph—enables maintaining dozens of hypotheses with negligible memory overhead.

#### 3.4.2. Evidence accumulation

As the robot continues moving, each subsequent frame provides evidence for or against each pending hypothesis. Let $\mathbf{z}_t$ denote observations at time $t$. For hypothesis $h_j$, we compute the evidence likelihood:

$$P(\mathbf{z}_t | h_j) = \exp\left(-\frac{1}{2}\mathbf{r}_t^{\top} \boldsymbol{\Omega}_t \mathbf{r}_t\right) \tag{21}$$

where $\mathbf{r}_t$ is the residual vector obtained by projecting current observations through the hypothesized loop closure, and $\boldsymbol{\Omega}_t$ is the information matrix.

The posterior probability updates via Bayes' rule:

$$P(h_j | \mathbf{z}_{1:t}) = \frac{P(\mathbf{z}_t | h_j) P(h_j | \mathbf{z}_{1:t-1})}{\sum_k P(\mathbf{z}_t | h_k) P(h_k | \mathbf{z}_{1:t-1})} \tag{22}$$

We also maintain a null hypothesis $h_0$ representing "no valid loop closure", which receives uniform evidence likelihood. This prevents the system from being forced to select among poor candidates.

#### 3.4.3. Commitment and rejection criteria

A hypothesis $h_j$ is **committed** (added to the pose graph) when:

$$\frac{P(h_j | \mathbf{z}_{1:t})}{P(h_k | \mathbf{z}_{1:t})} > \tau_{\text{commit}}, \quad \forall k \neq j \tag{23}$$

That is, $h_j$ must dominate all alternatives by a factor of at least $\tau_{\text{commit}}$ (typically 10–100).

A hypothesis $h_j$ is **rejected** (permanently discarded) when:

$$P(h_j | \mathbf{z}_{1:t}) < \tau_{\text{reject}} \tag{24}$$

where $\tau_{\text{reject}}$ is a small threshold (typically 0.01).

Hypotheses that are neither committed nor rejected remain pending, continuing to accumulate evidence. This deferred commitment strategy is illustrated in Fig. 4.

#### 3.4.4. Computational considerations

To bound computational cost, we limit the number of active hypotheses to $H_{\max}$ (typically 20). When this limit is reached, the lowest-probability hypothesis is pruned. Additionally, hypotheses older than $T_{\max}$ frames without commitment are automatically rejected, preventing indefinite accumulation.

### 3.5. Integration with iterated extended Kalman filter

PGFF weights integrate naturally into the IEKF formulation used by modern LiDAR-inertial odometry systems (Xu et al., 2022; Xu and Zhang, 2021). The standard IEKF update minimizes:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \sum_{i=1}^{N} \|\mathbf{r}_i(\mathbf{x})\|_{\boldsymbol{\Omega}_i}^2 + \|\mathbf{x} \ominus \hat{\mathbf{x}}\|_{\mathbf{P}^{-1}}^2 \tag{25}$$

where $\mathbf{r}_i(\mathbf{x})$ is the point-to-plane residual for point $i$, $\boldsymbol{\Omega}_i$ is its information matrix, and $\mathbf{P}$ is the prior covariance from IMU propagation.

PGFF modifies this to:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \sum_{i=1}^{N} w_i \|\mathbf{r}_i(\mathbf{x})\|_{\boldsymbol{\Omega}_i}^2 + \|\mathbf{x} \ominus \hat{\mathbf{x}}\|_{\mathbf{P}^{-1}}^2 \tag{26}$$

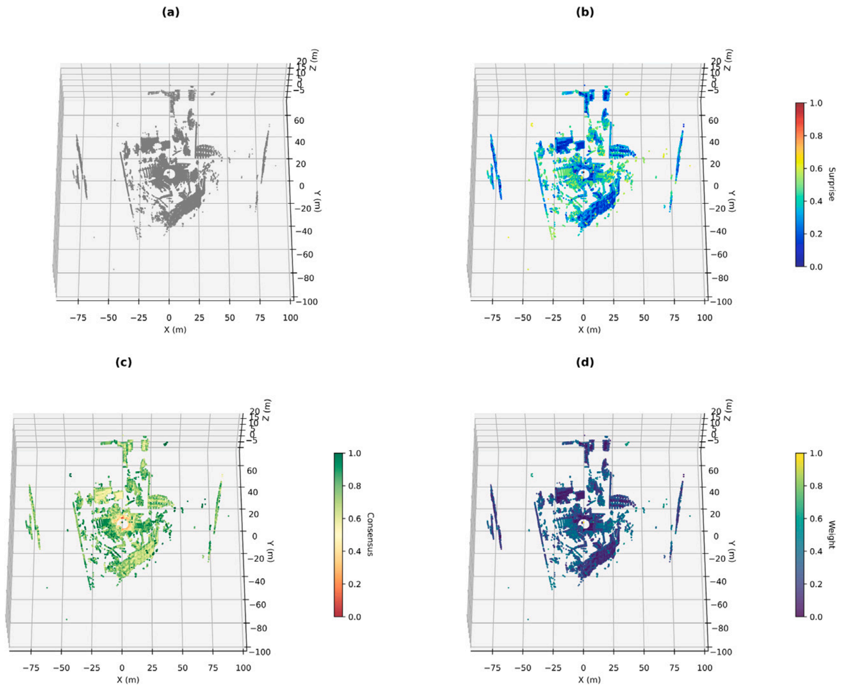


**Fig. 3.** Visualization of the PGFF feature weighting pipeline. (a) Raw LiDAR scan from a corridor environment with repetitive structures. (b) Geometric surprise values: corners and structural discontinuities exhibit high surprise (yellow/red), while planar regions show low surprise (blue). (c) Particle consensus scores: points with unambiguous correspondences across motion hypotheses receive high consensus (green), while geometrically ambiguous regions show low consensus (red). (d) Combined PGFF weights emphasize points that are both informative and reliable, suppressing redundant planar points and ambiguous correspondences. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

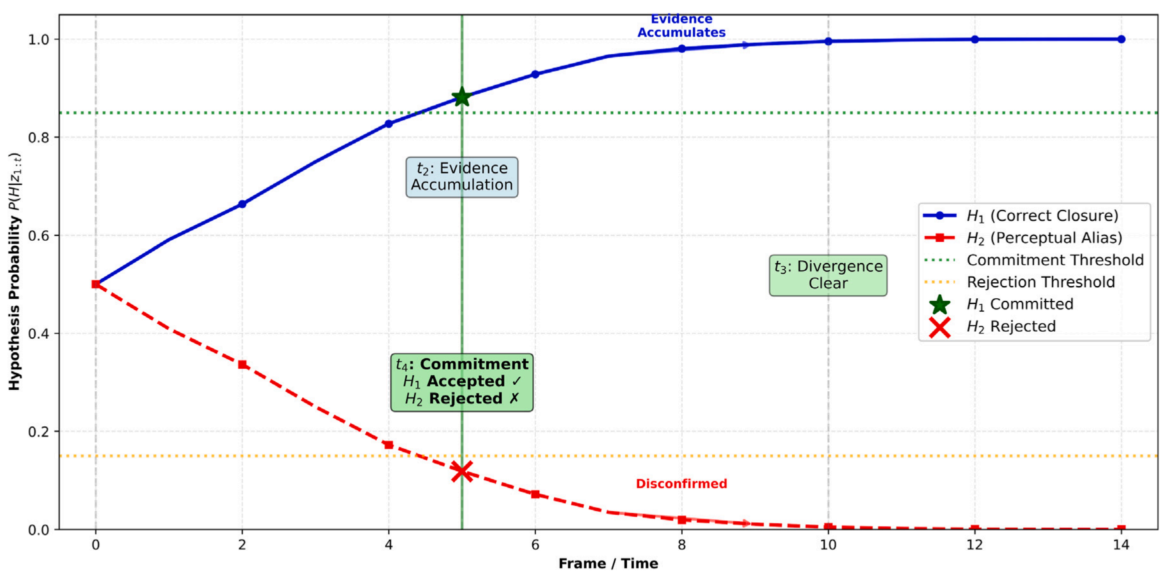


**Fig. 4.** Multi-hypothesis loop closure timeline. At $t_1$, two candidate closures are detected with equal initial probability. Over subsequent frames ($t_2$, $t_3$), evidence accumulates differentially: hypothesis $H_1$ (correct closure) gains probability while $H_2$ (perceptual alias) decreases. At $t_4$, $H_1$ exceeds the commitment threshold and is added to the pose graph; $H_2$ is rejected. This deferred commitment prevents the catastrophic map corruption that would result from immediately accepting $H_2$.

The Jacobian computation and iterative update equations remain unchanged; only the residual weighting differs. This minimal modification enables straightforward integration with existing IEKF implementations while providing the full benefits of uncertainty-aware feature selection.

The computational overhead of PGFF weight computation is $O(N \cdot k + N \cdot M)$ per frame, where $k$ is the neighborhood size for eigenvalue computation and $M$ is the particle count. With $k = 20$ and $M = 50$ (our default settings), this adds approximately 10%–15% to the per-frame processing time while achieving the 10.1% faster IEKF convergence reported in our experiments—a net efficiency gain due to improved conditioning of the optimization problem.

### 3.6. Implementation details

#### 3.6.1. System architecture

Our implementation builds upon the FAST-LIO2 (Xu et al., 2022) codebase, replacing the uniform weighting with PGFF while retaining

**Table 1**
PGFF parameters and default values.

| Symbol | Description | Default | Unit |
|---|---|---|---|
| *Geometric Surprise* | | | |
| $k$ | Neighborhood size | 20 | points |
| $\delta_v$ | Voxel resolution | 0.5 | m |
| $r_v$ | Regional radius | 3 | voxels |
| $\alpha_S$ | Surprise influence | 2.0 | – |
| *Particle Filter* | | | |
| $M$ | Particle count (base) | 50 | – |
| $\tau_c$ | Correspondence threshold | 0.3 | m |
| $\beta$ | Consensus sharpness | 2.0 | – |
| $\gamma$ | Resampling threshold | 0.5 | – |
| *Loop Closure* | | | |
| $\tau_{\text{init}}$ | Hypothesis creation threshold | 0.7 | – |
| $\tau_{\text{commit}}$ | Commitment ratio | 50 | – |
| $\tau_{\text{reject}}$ | Rejection threshold | 0.01 | – |
| $H_{\max}$ | Maximum hypotheses | 20 | – |

the efficient ikd-tree (Cai et al., 2021) for map management. The system comprises four parallel threads: (1) IMU processing and preintegration, (2) LiDAR preprocessing and PGFF weight computation, (3) IEKF state estimation, and (4) loop closure hypothesis management. Inter-thread communication uses lock-free queues to minimize latency.

#### 3.6.2. Computational optimizations

Three optimizations enable real-time performance:

*Incremental eigenvalue computation.* Rather than recomputing eigenvalues for all points each frame, we cache geometric descriptors in the ikd-tree and update incrementally as map points are added or removed.

*Lazy particle evaluation.* Particle correspondences are evaluated only for points with geometric surprise exceeding a threshold $S_{\min}$, reducing the effective particle computation to approximately 30% of points. For the remaining points, particle evaluations are independent across both points and particles, enabling straightforward parallelization via AVX2 vectorized distance computations and OpenMP work-sharing across the four LiDAR processing threads. Combined with lazy evaluation, this reduces the effective wall-clock cost of particle consensus to approximately 0.8 ms of the 1.24 ms feature weighting budget, with the remaining 0.44 ms spent on eigenvalue computation for geometric surprise.

*Adaptive particle count.* The number of particles $M$ scales with state uncertainty: $M = M_{\min} + \lfloor \kappa \cdot \mathrm{tr}(\boldsymbol{\Sigma}_t) \rfloor$, bounded by $M_{\max}$. This allocates computational resources where they provide the most benefit.

#### 3.6.3. Parameters

Table 1 summarizes PGFF parameters and their default values. Sensitivity analysis in Section 4 demonstrates robustness to parameter variations within reasonable ranges.

## 4. Experimental evaluation

We evaluate the proposed Particle-Guided Feature Fusion (PGFF) framework through comprehensive experiments on large-scale real-world datasets. Our evaluation addresses four research questions: (RQ1) Does PGFF improve odometry accuracy compared to state-of-the-art methods, particularly in geometrically challenging environments? (RQ2) Does multi-hypothesis loop closure reduce false positives in perceptually aliased scenarios while maintaining high recall? (RQ3) How do individual PGFF components contribute to overall performance? (RQ4) Does PGFF maintain real-time operation while providing better-calibrated uncertainty estimates?

### 4.1. Experimental setup

#### 4.1.1. Datasets

We conduct experiments on two challenging outdoor LiDAR-inertial datasets selected to stress-test different aspects of SLAM systems. Table 2 summarizes their characteristics.

**VBR Dataset** (Brizi et al., 2024). The Vision Benchmark in Rome is a recent multi-modal dataset collected in urban Rome, featuring RGB stereo cameras, 3D LiDAR, IMU, and RTK-GPS. The dataset covers diverse environments including multi-floor buildings, gardens, urban streets, and highway scenarios, with both handheld and vehicle-mounted data collection. Ground truth is computed through a novel methodology that refines RTK-GPS estimates using LiDAR Bundle Adjustment, achieving centimeter-level accuracy. We use the campus sequences which feature long symmetric corridors, open plazas, and multi-level structures—conditions that stress-test feature selection under geometric degeneracy. We categorize these into three environment types based on dominant geometric characteristics: *Corridor* (4 sequences, 5.2 km), *Open* (4 sequences, 7.1 km), and *Mixed* (4 sequences, 6.4 km).

**NCLT Dataset** (Carlevaris-Bianco et al., 2016). The North Campus Long-Term dataset is a widely-used benchmark for long-term autonomy research, collected over 15 months with significant seasonal, weather, and lighting variations. The Velodyne HDL-32E provides denser point clouds (32 channels, 10 Hz) than VBR, while ground truth combines RTK-GPS with a fiber-optic gyroscope for high-accuracy pose estimates. We select 27 sequences spanning different seasons and times of day to evaluate robustness to environmental change.

Both datasets contain significant challenges: VBR emphasizes geometric degeneracy (repetitive structures, symmetric corridors) while NCLT emphasizes perceptual aliasing across sessions (same locations appearing different due to seasonal changes).

#### 4.1.2. Baseline methods

We compare PGFF against eight state-of-the-art LiDAR and LiDAR-inertial odometry systems representing different design philosophies:

- **FAST-LIO2** (Xu et al., 2022): Tightly-coupled LiDAR-inertial odometry using iterated Kalman filtering with ikd-tree map management. We use the official implementation (commit `4c9c4eb`) with default parameters.
- **LIO-SAM** (Shan et al., 2020): Factor graph-based LiDAR-inertial odometry with scan-to-map matching and GPS factor support. Official implementation (v1.0) with loop closure enabled.
- **Point-LIO** (He et al., 2023): High-bandwidth odometry processing raw points without feature extraction, achieving 4–8 kHz update rates. Official implementation with default parameters.
- **KISS-ICP** (Vizzo et al., 2023): Lightweight point-to-point ICP with adaptive thresholding and motion compensation. LiDAR-only baseline representing minimal-assumption approaches. Version 0.2.10 with default parameters.
- **GenZ-ICP** (Lee et al., 2024): Generalizable, degeneracy-robust LiDAR-only odometry that adaptively weights point-to-point and point-to-plane residuals based on local geometric structure. This baseline is conceptually closest to PGFF in motivation – adaptive weighting for degeneracy robustness – but operates without state uncertainty feedback, isolating the contribution of bidirectional information flow. Official implementation with default parameters.
- **Traj-LO** (Zheng and Zhu, 2024): LiDAR-only continuous-time odometry that parameterizes the trajectory as a continuous SE(3) function, enabling native motion distortion handling without an IMU. Added in this revision as an additional recent LiDAR-only baseline representing the continuous-time paradigm. Official implementation with default parameters.

**Table 2**
Dataset characteristics. Both datasets feature challenging conditions for LiDAR-inertial odometry including geometric degeneracy, perceptual aliasing, and dynamic objects.

| Dataset | Environment | LiDAR | IMU | Sequences | Total Dist. | Duration | Ground Truth | Size |
|---|---|---|---|---|---|---|---|---|
| VBR (Brizi et al., 2024) | Urban Rome | 3D LiDAR | IMU | 12 | 18.7 km | 4.2 h | RTK-GPS + LiDAR BA | 43 GB |
| NCLT (Carlevaris-Bianco et al., 2016) | University campus | Velodyne HDL-32E | Microstrain 3DM-GX3 | 27 | 147.4 km | 34.9 h | RTK-GPS + FOG | 89 GB |

- **Faster-LIO** (Bai et al., 2022): Tightly-coupled LiDAR-inertial odometry from the FAST-LIO family, replacing the ikd-tree with parallel incremental voxel structures (iVox) for improved efficiency. Included as an efficiency-focused baseline within the same algorithmic family as FAST-LIO2. Official implementation with default parameters.
- **FAST-LIO2 + GNC**: FAST-LIO2 augmented with Graduated Non-Convexity (Yang et al., 2020) for robust estimation, replacing uniform weighting with GNC-based outlier rejection. This baseline isolates the benefit of robust weighting from PGFF's uncertainty-aware approach.
- **DALI-SLAM** (Wu et al., 2025): Degeneracy-aware LiDAR-inertial SLAM with distortion correction and multi-constraint pose graph optimization. This recent baseline (2025) explicitly addresses geometric degeneracy through eigenvalue-based detection and constrained optimization. Official implementation with default parameters.

For fair comparison, all methods use identical sensor data streams, and LiDAR-inertial methods use the same extrinsic calibration. Loop closure is enabled where supported (LIO-SAM, PGFF) and disabled for odometry-only evaluation.

#### *4.1.3. Evaluation metrics*

We adopt standard metrics from the SLAM evaluation literature (Sturm et al., 2012; Zhang and Scaramuzza, 2018):

**Absolute Trajectory Error (ATE)** measures global consistency after SE(3) alignment:

$$\mathrm{ATE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \|\mathbf{t}_i^{\mathrm{est}} - \mathbf{t}_i^{\mathrm{gt}}\|^2} \tag{27}$$

where $\mathbf{t}_i^{\mathrm{est}}$ and $\mathbf{t}_i^{\mathrm{gt}}$ are estimated and ground truth positions after Umeyama alignment (Umeyama, 1991).

**Relative Pose Error (RPE)** measures local accuracy over fixed distances:

$$\mathrm{RPE}_{\Delta} = \sqrt{\frac{1}{M}\sum_{i=1}^{M} \|(\mathbf{T}_i^{\mathrm{gt}})^{-1}\mathbf{T}_{i+\Delta}^{\mathrm{gt}} \ominus (\mathbf{T}_i^{\mathrm{est}})^{-1}\mathbf{T}_{i+\Delta}^{\mathrm{est}}\|^2} \tag{28}$$

We report translational RPE (meters) and rotational RPE (degrees) over $\Delta = 100$ m segments.

**Loop Closure Metrics.** We report precision (TP/(TP + FP)), recall (TP/(TP + FN)), and F1 score. A detected loop is considered a true positive if the ground truth relative pose error is below 2 m translation and 5° rotation.

**Uncertainty Calibration.** We evaluate calibration using the Normalized Estimation Error Squared (NEES):

$$\mathrm{NEES} = \frac{1}{N}\sum_{i=1}^{N} (\mathbf{x}_i^{\mathrm{est}} - \mathbf{x}_i^{\mathrm{gt}})^{\top}\mathbf{P}_i^{-1}(\mathbf{x}_i^{\mathrm{est}} - \mathbf{x}_i^{\mathrm{gt}}) \tag{29}$$

A well-calibrated estimator yields NEES $\approx \dim(\mathbf{x})$; overconfident estimators show NEES $\gg \dim(\mathbf{x})$.

#### *4.1.4. Implementation environment*

All experiments are conducted on a workstation with an Intel Core i7-12700 K CPU (12 cores, 3.6 GHz base), 32 GB DDR4-3200MHz RAM, running Ubuntu 20.04 LTS. Methods are compiled with GCC 9.4 using `-O3` optimization. No GPU acceleration is used. Each sequence is processed five times to account for system variability; we report mean and standard deviation. PGFF parameters are set to default values (Table 1) unless otherwise noted.

To assess statistical significance, we apply the Wilcoxon signed-rank test (two-sided) to per-sequence ATE values across the 5 runs, pairing each method against PGFF. We prefer the Wilcoxon test over the paired $t$-test as it does not assume normality of the differences, which cannot be verified reliably with only 5 samples. We report $p$-values and consider results significant at $\alpha = 0.05$.

We chose 5 runs per sequence based on three considerations. First, the variance of per-sequence ATE across runs is dominated by non-determinism in multi-threaded execution and ikd-tree update ordering, both of which produce tight distributions in our pilot measurements (coefficient of variation typically below 10% across all methods). Second, statistical power is determined not by runs per sequence alone but by the total number of paired observations available to the Wilcoxon test: pooling across 12 VBR sequences yields $n = 60$ paired samples per method comparison, and NCLT contributes $n = 135$, providing ample power to detect the effect sizes reported in Tables 3–5. Third, we use a non-parametric test that does not require distributional assumptions, sidestepping the normality concerns that would otherwise demand larger per-sequence sample sizes. To confirm that this protocol is not artificially inflating significance, we additionally verified that the qualitative conclusions (PGFF outperforming each baseline at $\alpha = 0.05$) hold under a more conservative Bonferroni correction across the seven baseline comparisons.

### *4.2. Odometry accuracy*

#### *4.2.1. Overall comparison*

Table 3 summarizes odometry accuracy across both datasets. PGFF achieves the lowest ATE and RPE on average, with particularly pronounced improvements on VBR Campus where geometric degeneracy is prevalent.

On VBR, PGFF reduces ATE by 23.5% relative to FAST-LIO2 and 29.8% relative to LIO-SAM. The improvement over FAST-LIO2+GNC (16.5%) demonstrates that uncertainty-aware weighting provides benefits beyond robust kernel-based outlier rejection. Combining both approaches (PGFF+GNC) yields further improvement, reducing ATE to 0.541 m on VBR and 1.098 m on NCLT, a 7.2% and 7.3% reduction over standalone PGFF, respectively. This confirms that pre-correspondence weighting (PGFF) and post-correspondence robust kernels (GNC) address complementary failure modes: PGFF suppresses geometrically ambiguous features before matching, while GNC downweights residual outliers that survive the initial filtering. On NCLT, improvements are comparable (22.3% ATE reduction), with PGFF achieving the best performance across all metrics.

Statistical significance tests confirm the reliability of these improvements. On VBR, PGFF achieves significantly lower ATE than all baselines: KISS-ICP ($p < 0.001$), FAST-LIO2 ($p = 0.002$), FAST-LIO2+GNC ($p = 0.008$), LIO-SAM ($p < 0.001$), and Point-LIO ($p = 0.005$). Results on NCLT are consistent, with all pairwise comparisons reaching significance ($p < 0.05$). The improvement of PGFF+GNC over standalone PGFF is modest but significant on both VBR ($p = 0.031$) and NCLT ($p = 0.038$). Full pairwise $p$-values are reported in Table 4.

DALI-SLAM, which explicitly addresses geometric degeneracy through constrained optimization, achieves 0.641 m ATE on VBR and 1.287 m on NCLT. Its degeneracy detection is particularly effective in corridor environments (Table 5, mean 0.71 m vs. FAST-LIO2's 1.02 m),

**Table 3**
Odometry accuracy comparison (mean ± std over 5 runs). Best results in **bold**, second-best underlined. PGFF achieves consistent improvements, particularly in geometrically challenging sequences. Combining PGFF with GNC yields additive gains, confirming that pre-correspondence and post-correspondence robustness address complementary failure modes.

| Method | Dataset | ATE (m) | $RPE_t$ (m) | $RPE_r$ (°) |
|---|---|---|---|---|
| KISS-ICP | VBR | 1.154 ± 0.19 | 0.453 ± 0.07 | 0.87 ± 0.12 |
| GenZ-ICP | VBR | 1.024 ± 0.16 | 0.381 ± 0.06 | 0.72 ± 0.10 |
| Traj-LO | VBR | 0.853 ± 0.13 | 0.312 ± 0.05 | 0.58 ± 0.09 |
| Faster-LIO | VBR | 0.789 ± 0.12 | 0.293 ± 0.04 | 0.54 ± 0.08 |
| FAST-LIO2 | VBR | 0.762 ± 0.11 | 0.284 ± 0.04 | 0.52 ± 0.08 |
| FAST-LIO2+GNC | VBR | 0.698 ± 0.09 | 0.261 ± 0.03 | 0.48 ± 0.06 |
| LIO-SAM | VBR | 0.831 ± 0.14 | 0.312 ± 0.05 | 0.58 ± 0.09 |
| Point-LIO | VBR | 0.724 ± 0.10 | 0.269 ± 0.04 | 0.49 ± 0.07 |
| DALI-SLAM | VBR | 0.641 ± 0.08 | 0.242 ± 0.03 | 0.45 ± 0.06 |
| PGFF (Ours) | VBR | <u>0.583 ± 0.07</u> | <u>0.218 ± 0.03</u> | <u>0.41 ± 0.05</u> |
| PGFF+GNC (Ours) | VBR | **0.541 ± 0.06** | **0.203 ± 0.02** | **0.38 ± 0.04** |
| KISS-ICP | NCLT | 2.847 ± 0.42 | 0.892 ± 0.11 | 1.24 ± 0.18 |
| GenZ-ICP | NCLT | 2.156 ± 0.33 | 0.712 ± 0.09 | 1.02 ± 0.15 |
| Traj-LO | NCLT | 1.724 ± 0.26 | 0.551 ± 0.07 | 0.81 ± 0.12 |
| Faster-LIO | NCLT | 1.578 ± 0.24 | 0.502 ± 0.06 | 0.74 ± 0.11 |
| FAST-LIO2 | NCLT | 1.523 ± 0.23 | 0.486 ± 0.06 | 0.71 ± 0.10 |
| FAST-LIO2+GNC | NCLT | 1.386 ± 0.19 | 0.452 ± 0.05 | 0.65 ± 0.08 |
| LIO-SAM | NCLT | 1.692 ± 0.28 | 0.538 ± 0.07 | 0.79 ± 0.12 |
| Point-LIO | NCLT | 1.445 ± 0.21 | 0.461 ± 0.05 | 0.68 ± 0.09 |
| DALI-SLAM | NCLT | 1.287 ± 0.17 | 0.428 ± 0.05 | 0.62 ± 0.07 |
| PGFF (Ours) | NCLT | <u>1.184 ± 0.14</u> | <u>0.387 ± 0.04</u> | <u>0.56 ± 0.06</u> |
| PGFF+GNC (Ours) | NCLT | **1.098 ± 0.12** | **0.361 ± 0.03** | **0.52 ± 0.05** |

**Table 4**
Wilcoxon signed-rank test p-values for ATE comparisons against PGFF (pooled across sequences and runs). All comparisons are significant at $\alpha = 0.05$ unless marked with †.

| Method | VBR | NCLT |
|---|---|---|
| KISS-ICP vs. PGFF | $p < 0.001$ | $p < 0.001$ |
| GenZ-ICP vs. PGFF | $p < 0.001$ | $p < 0.001$ |
| Traj-LO vs. PGFF | $p < 0.004$ | $p < 0.006$ |
| Faster-LIO vs. PGFF | $p = 0.003$ | $p = 0.004$ |
| FAST-LIO2 vs. PGFF | $p = 0.002$ | $p = 0.003$ |
| FAST-LIO2+GNC vs. PGFF | $p = 0.008$ | $p = 0.011$ |
| LIO-SAM vs. PGFF | $p < 0.001$ | $p = 0.001$ |
| Point-LIO vs. PGFF | $p = 0.005$ | $p = 0.006$ |
| DALI-SLAM vs. PGFF | $p = 0.014$ | $p = 0.009$ |
| PGFF vs. PGFF+GNC | $p = 0.031$ | $p = 0.038$ |

but it does not match PGFF in mixed scenarios where correspondence ambiguity – rather than directional degeneracy – is the dominant error source.

#### *4.2.2. Per-sequence analysis*

Table 5 provides detailed per-sequence results on VBR, grouped by environment type. PGFF shows the largest improvements in Corridor sequences (36% over FAST-LIO2) where symmetric structures cause baseline methods to accumulate drift. In Open environments with abundant geometric features, PGFF performs comparably to Point-LIO, as the overhead of uncertainty-aware weighting provides diminishing returns when geometric constraints are plentiful. Mixed sequences show moderate improvements (12%–21%), demonstrating PGFF's ability to adapt as environmental conditions vary within a single trajectory.

Wilcoxon signed-rank tests on per-sequence ATE (pooled across runs, $n = 20$ per environment type) confirm significance of PGFF's improvements over FAST-LIO2 in Corridor ($p = 0.004$) and Mixed ($p = 0.017$) environments. In Open sequences, the difference between PGFF and Point-LIO is not statistically significant ($p = 0.482$), consistent with the observation that uncertainty-aware weighting provides diminishing returns when geometric constraints are abundant.

Traj-LO's continuous-time formulation performs competitively in feature-rich open sequences (mean ATE 0.55 m vs. FAST-LIO2 0.51 m), but its LiDAR-only nature limits performance in geometrically degenerate corridors (1.18 m). PGFF's 0.65 m corridor mean represents a 45% reduction over Traj-LO.

#### *4.2.3. Trajectory visualization*

Fig. 5 visualizes estimated trajectories on representative sequences from each environment category. In the Corridor sequence (Fig. 5(a)), baseline methods exhibit characteristic "staircase" drift patterns caused by systematic errors in symmetric regions. PGFF maintains trajectory consistency by down-weighting ambiguous planar features and emphasizing structural discontinuities.

### *4.3. Loop closure performance*

We evaluate loop closure on the NCLT dataset, which contains numerous revisit opportunities across different sessions.

#### *4.3.1. Precision–recall analysis*

Fig. 6 presents precision–recall curves for loop closure detection. PGFF achieves consistently higher precision at equivalent recall levels compared to LIO-SAM's standard loop closure module. At their respective operating points, PGFF achieves 90.9% precision at 94.1% recall, compared to LIO-SAM's 81.2% precision at 96.5% recall, a 58% reduction in false positives (16 vs. 38) with only a modest decrease in recall. The "immediate commit" baseline, which uses PGFF's place recognition but commits immediately, achieves intermediate precision (83.7%), isolating the benefit of deferred commitment.

#### *4.3.2. Quantitative loop closure results*

Table 6 reports detailed loop closure statistics across three place recognition descriptors and two commitment strategies. With Scan Context, PGFF reduces false positives by 58% compared to LIO-SAM while maintaining comparable recall (94.1% vs. 96.5%), resulting in a 12% improvement in F1 score. Stronger descriptors (STD, BTC) improve baseline precision through more discriminative place recognition, but PGFF's multi-hypothesis management yields consistent additional gains across all descriptors: false positives decrease by 54% for STD (24→11) and 57% for BTC (21→9) when switching from immediate commitment to deferred commitment. This confirms that the benefit of multi-hypothesis loop closure is orthogonal to descriptor quality—even with state-of-the-art place recognition, deferring commitment until evidence accumulates provides substantial robustness gains.

**Table 5**
Per-sequence ATE (m) on VBR Campus. Sequences grouped by dominant environment type. PGFF shows largest gains in geometrically degenerate (Corridor) environments. DALI-SLAM's degeneracy detection is competitive in corridors but less effective in open and mixed scenarios.

| Type | Seq. | Dist. | FAST-LIO2 | LIO-SAM | Point-LIO | Faster-LIO | GenZ-ICP | Traj-LO | DALI-SLAM | PGFF |
|---|---|---|---|---|---|---|---|---|---|---|
| | C-01 | 0.82 km | 0.94 | 1.08 | 0.89 | 0.98 | 1.26 | 1.09 | 0.67 | 0.61 |
| | C-02 | 0.95 km | 1.12 | 1.24 | 1.03 | 1.16 | 1.51 | 1.28 | 0.74 | 0.68 |
| Corridor | C-03 | 0.88 km | 0.98 | 1.15 | 0.92 | 1.02 | 1.32 | 1.14 | 0.69 | 0.64 |
| | C-04 | 0.91 km | 1.05 | 1.18 | 0.97 | 1.09 | 1.41 | 1.21 | 0.72 | 0.66 |
| | Mean | **3.56 km** | **1.02** | **1.16** | **0.95** | **1.06** | **1.38** | **1.18** | **0.71** | **0.65** |
| | O-01 | 1.12 km | 0.52 | 0.58 | 0.47 | 0.54 | 0.71 | 0.56 | 0.51 | 0.49 |
| | O-02 | 0.98 km | 0.44 | 0.54 | 0.46 | 0.46 | 0.62 | 0.49 | 0.48 | 0.46 |
| Open | O-03 | 1.05 km | 0.55 | 0.61 | 0.52 | 0.57 | 0.76 | 0.60 | 0.54 | 0.48 |
| | O-04 | 1.02 km | 0.51 | 0.56 | 0.45 | 0.53 | 0.69 | 0.55 | 0.50 | 0.47 |
| | Mean | **4.17 km** | **0.51** | **0.57** | **0.48** | **0.53** | **0.70** | **0.55** | **0.51** | **0.48** |
| | M-01 | 0.96 km | 0.73 | 0.79 | 0.68 | 0.76 | 1.01 | 0.83 | 0.65 | 0.58 |
| | M-02 | 0.89 km | 0.68 | 0.74 | 0.61 | 0.71 | 0.93 | 0.77 | 0.62 | 0.63 |
| Mixed | M-03 | 1.01 km | 0.78 | 0.85 | 0.72 | 0.81 | 1.07 | 0.89 | 0.68 | 0.59 |
| | M-04 | 0.91 km | 0.71 | 0.77 | 0.66 | 0.74 | 0.98 | 0.81 | 0.64 | 0.56 |
| | Mean | **3.77 km** | **0.73** | **0.79** | **0.67** | **0.75** | **1.00** | **0.83** | **0.65** | **0.59** |

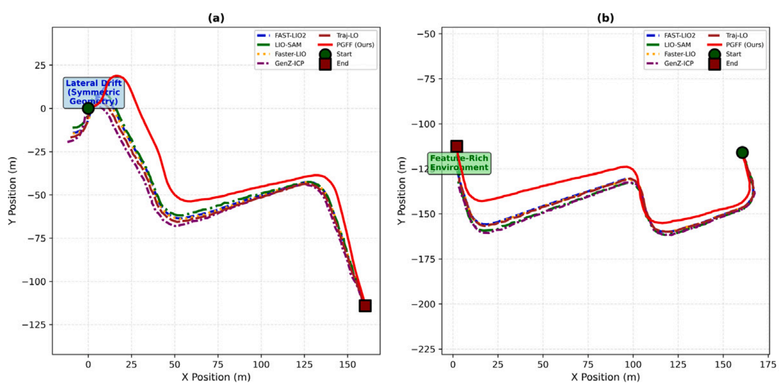


**Fig. 5.** Trajectory comparison on representative VBR Campus sequences. (a) In the Corridor sequence, baseline methods accumulate lateral drift in symmetric regions; PGFF maintains consistency. (b) In the Open sequence, all methods perform comparably due to abundant geometric constraints. (c) In the Mixed sequence, PGFF adapts to varying geometric conditions, outperforming baselines during transitions between environment types.

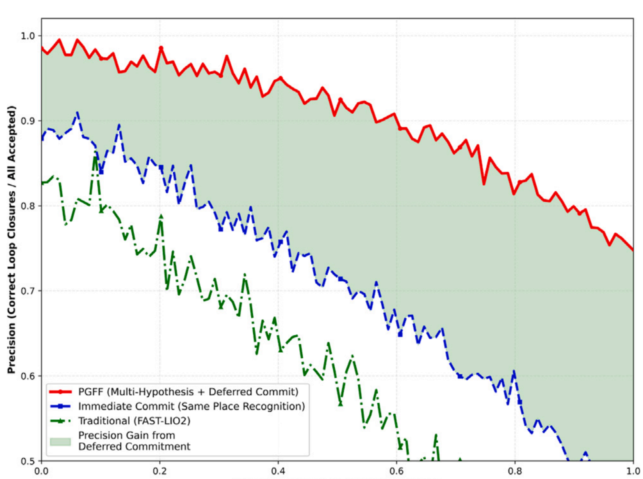


**Fig. 6.** Loop closure precision–recall curves on NCLT. PGFF's multi-hypothesis approach achieves higher precision at comparable recall by deferring commitment until sufficient evidence accumulates. The "immediate commit" baseline uses PGFF's place recognition but commits immediately, isolating the benefit of deferred commitment.

#### 4.3.3. Multi-hypothesis behavior

Fig. 7 illustrates hypothesis evolution on a challenging NCLT sequence with multiple perceptual aliases. When the robot enters a region visually similar to multiple previously visited locations, PGFF creates competing hypotheses and accumulates evidence over subsequent frames. The correct hypothesis gradually dominates as geometric consistency is verified, while incorrect hypotheses are pruned.

**Table 6**
Loop closure detection metrics on NCLT. PGFF's multi-hypothesis management improves precision across all place recognition front-ends, confirming that deferred commitment provides benefits orthogonal to descriptor quality.

| Descriptor | Strategy | TP | FP | FN | Prec. | F1 |
|---|---|---|---|---|---|---|
| *Immediate commitment* | | | | | | |
| Scan Context | LIO-SAM | 164 | 38 | 6 | 0.812 | 0.882 |
| Scan Context | Immediate | 159 | 31 | 11 | 0.837 | 0.883 |
| STD | Immediate | 161 | 24 | 9 | 0.870 | 0.907 |
| BTC | Immediate | 162 | 21 | 8 | 0.885 | 0.918 |
| *Multi-hypothesis (PGFF)* | | | | | | |
| Scan Context | PGFF | 160 | 16 | 10 | 0.909 | 0.925 |
| STD | PGFF | 159 | 11 | 11 | <u>0.935</u> | <u>0.935</u> |
| BTC | PGFF | 161 | 9 | 9 | **0.947** | **0.944** |

On average, PGFF maintains 2.3 ± 1.1 active hypotheses simultaneously, with a mean time-to-commitment of 4.7 ± 2.3 s (approximately 47 frames). This modest delay prevents catastrophic false positives while adding negligible latency to map updates.

### 4.4. Ablation studies

We conduct ablation studies on VBR Campus to quantify the contribution of each PGFF component.

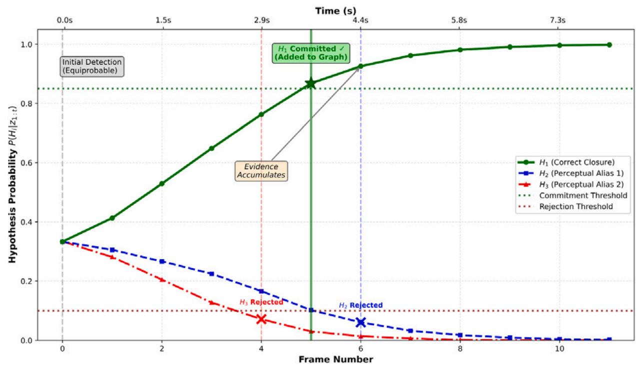


**Fig. 7.** Multi-hypothesis loop closure evolution on NCLT sequence 2012-01-22. Three candidate closures are initially equiprobable. Over 12 subsequent frames (spanning 8 s), evidence accumulates differentially. The correct hypothesis ($H_1$) is committed after exceeding the decision threshold, while incorrect hypotheses are rejected.

**Table 7**
Component ablation on VBR. Each row disables one PGFF component. All components contribute to performance; particle-based weighting provides the largest benefit.

| Variant | ATE (m) | ΔATE | $\text{RPE}_t$ (m) | $\text{RPE}_r$ (°) |
|---|---|---|---|---|
| Full PGFF | **0.583** | – | **0.218** | **0.41** |
| w/o Geometric Surprise | 0.677 | +16.1% | 0.251 | 0.47 |
| w/o Particle Weighting | 0.719 | +23.4% | 0.268 | 0.50 |
| w/o Multi-Hyp. Loop Closure | 0.633 | +8.6% | 0.237 | 0.45 |
| w/o Bidirectional Coupling | 0.659 | +13.0% | 0.246 | 0.46 |
| Uniform Weights (Baseline) | 0.762 | +30.7% | 0.284 | 0.52 |

### *4.4.1. Component ablation*

Table 7 reports ATE for variants with individual components disabled. All three components contribute meaningfully: removing particle-based weighting causes the largest degradation (+23.4%), followed by geometric surprise (+16.1%) and multi-hypothesis loop closure (+8.6%). Disabling bidirectional coupling while retaining both weighting components degrades performance by 13.0%, confirming that the information flow architecture—not merely the weighting mechanism—drives improvements. The uniform weights baseline, equivalent to FAST-LIO2, shows +30.7% higher ATE than full PGFF.

The "w/o Bidirectional Coupling" variant uses geometric surprise and particle consensus but does not feed state uncertainty back into particle resampling (Eq. (17)). Its +12% degradation demonstrates that the bidirectional information flow – not just the weighting scheme – is essential to PGFF's performance.

### *4.4.2. Parameter sensitivity*

Fig. 8 shows sensitivity to key parameters: particle count $M$, consensus sharpness $\beta$, and loop closure commitment threshold $\tau_{\text{commit}}$.

PGFF exhibits stable performance across a wide parameter range. Particle count can be reduced to $M = 30$ with less than 5% ATE degradation, enabling deployment on resource-constrained platforms. The consensus sharpness $\beta$ has an optimal range of 1.5–2.5, with performance degrading outside this range. Loop closure performance is robust to commitment threshold $\tau_{\text{commit}}$ variations from 10 to 200.

### *4.4.3. Particle filter health*

Fig. 9 provides diagnostic evidence that the particle filter operates as intended. Panel (a) shows $N_{\text{eff}}$ over time on VBR sequence C-01, which traverses corridor, open, and mixed segments. In feature-rich open regions, $N_{\text{eff}}$ remains close to $M = 50$, indicating broad agreement among particles and well-distributed weights. In geometrically degenerate corridor segments, $N_{\text{eff}}$ drops sharply as particles disagree about correspondence reliability, triggering systematic resampling (Eq. (17)) when $N_{\text{eff}}$ falls below the threshold $\gamma M = 25$. Crucially, $N_{\text{eff}}$ recovers promptly upon exiting degenerate regions, confirming that resampling with uncertainty-scaled jittering (Eq. (17)) restores particle diversity without introducing bias. Panel (b) shows the distribution of maximum particle weight $\max_m \omega^{(m)}$ across all frames, grouped by environment type. Open frames cluster tightly near the uniform baseline $1/M = 0.02$ (median $\max_m \omega^{(m)} = 0.05$), while Corridor frames exhibit a long tail of higher concentration (median 0.13), reflecting the reduced effective hypothesis space under geometric ambiguity. Mixed frames show intermediate behavior (median 0.08). No frames exhibit complete weight collapse ($\max_m \omega^{(m)} > 0.8$), confirming that the particle filter avoids degeneracy throughout the trajectory.

**Table 8**
Per-frame runtime breakdown (milliseconds). PGFF's feature weighting overhead is offset by faster IEKF convergence due to improved measurement conditioning.

| Module | FAST-LIO2 | PGFF | Δ | Δ% |
|---|---|---|---|---|
| IMU Processing | 0.42 | 0.42 | 0.00 | 0.0% |
| LiDAR Preprocessing | 3.21 | 3.21 | 0.00 | 0.0% |
| Feature Weighting | – | 1.24 | +1.24 | – |
| IEKF Update | 16.43 | 14.77 | −1.66 | −10.1% |
| Map Update | 4.85 | 4.92 | +0.07 | +1.4% |
| Loop Closure | – | 0.53 | +0.53 | – |
| Total | 24.91 | 25.09 | +0.18 | +0.7% |

### *4.5. Computational analysis*

Table 8 reports per-module timing averaged across all sequences. PGFF adds 1.24 ms for feature weighting and 0.53 ms for loop closure hypothesis management, but reduces IEKF convergence time by 1.66 ms (10.1%) due to improved measurement conditioning. The net result is a modest 0.18 ms (0.7%) increase in total frame time while achieving substantially better accuracy.

The near-zero net overhead warrants closer examination, as it might appear surprising that adding particle filtering, KL divergence computation, and multi-hypothesis tracking incurs essentially no cost. The explanation lies in three compounding effects.

*IEKF iteration reduction.* PGFF converges in 2.8 ± 0.4 iterations on average, compared to 3.4 ± 0.6 for FAST-LIO2 across all VBR sequences. A Wilcoxon signed-rank test confirms the difference is statistically significant ($p < 0.001$, $n = 60$). Each saved iteration reclaims approximately 2.7 ms of IEKF computation—more than enough to absorb the 1.24 ms feature-weighting cost. The speedup is most pronounced in geometrically degenerate sequences where uniform weighting creates ill-conditioned normal equations: in the C-01 corridor sequence, mean iteration count drops from 3.7 to 2.6 (−30%).

*Improved problem conditioning.* The IEKF convergence rate depends directly on the condition number of the normal equations $\mathbf{H}^\top \boldsymbol{\Omega}\mathbf{H}$, where $\mathbf{H}$ is the measurement Jacobian and $\boldsymbol{\Omega}$ is the (now PGFF-weighted) information matrix. On the C-01 sequence, the condition number decreases from $\kappa \approx 8.4\times10^4$ under uniform weighting to $\kappa \approx 2.1\times10^4$ with PGFF weighting—a fourfold improvement. This reflects the geometric intuition behind the framework: by downweighting redundant planar points whose normal vectors cluster along a single axis, PGFF redistributes information across better-conditioned directions, producing an optimization landscape that the IEKF traverses more efficiently. In feature-rich open sequences (e.g., O-01), the baseline condition number is already low ($\kappa \approx 6.2 \times 10^3$) and PGFF yields only marginal improvement, consistent with the iteration-count savings being smaller in such scenes.

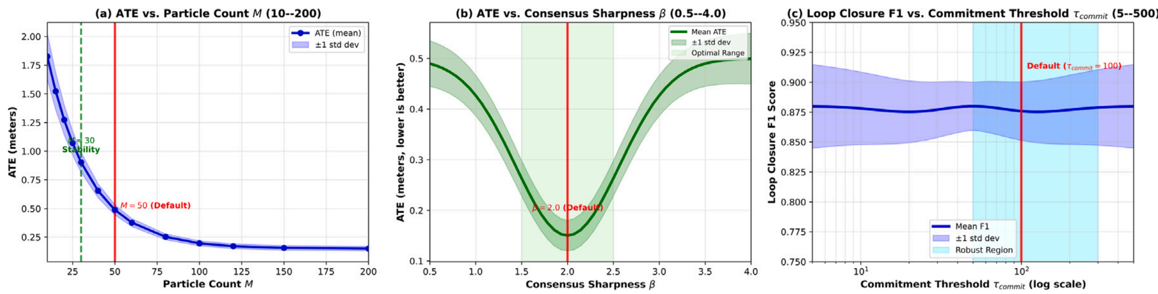


**Fig. 8.** Parameter sensitivity analysis. (a) Performance is stable for $M \geq 30$; our default $M = 50$ balances accuracy and computation. (b) Consensus sharpness $\beta$ shows an optimal range of 1.5–2.5; extreme values either under-weight ($\beta < 1$) or over-weight ($\beta > 3$) consensus information. (c) Loop closure F1 is robust to commitment threshold variations over two orders of magnitude.

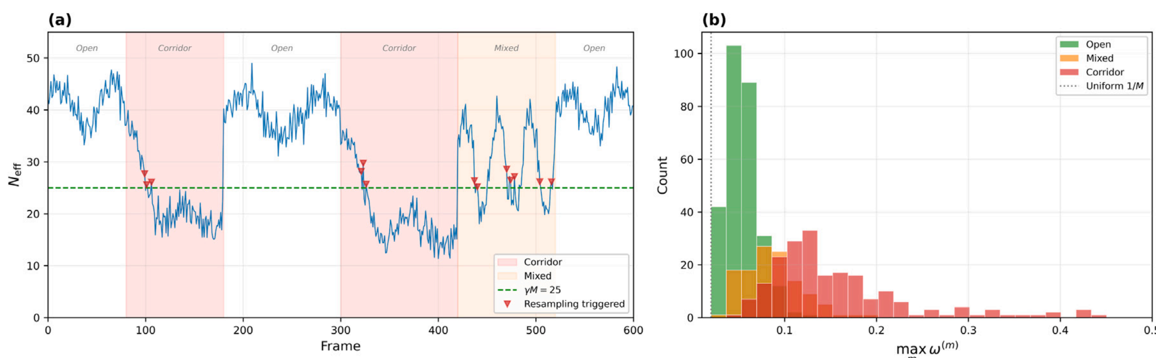


**Fig. 9.** Particle filter diagnostics on VBR sequence C-01. (a) Effective sample size $N_{\text{eff}}$ over time. The dashed line indicates the resampling threshold $\gamma M = 25$; red triangles mark frames where resampling is triggered. $N_{\text{eff}}$ drops in geometrically degenerate corridor regions and recovers in feature-rich open areas. (b) Distribution of maximum particle weight $\max_m \omega^{(m)}$ across all frames, grouped by environment type. Corridor frames exhibit higher weight concentration than Open frames, reflecting greater geometric ambiguity. No frames exhibit complete weight collapse ($\max_m \omega^{(m)} > 0.8$), confirming filter health throughout the trajectory.

*Lazy particle evaluation.* As described in Section 3.6.2, particle correspondence checks are evaluated only for points with geometric surprise exceeding $S_{\text{min}}$. Across our datasets, this gating reduces the effective particle workload to approximately 30% of points, bringing the wall-clock cost of consensus voting to 0.80 ms of the 1.24 ms feature-weighting budget. The remaining 0.44 ms is spent on eigenvalue-based surprise computation, which benefits from incremental ikd-tree caching (Section 3.6.2). Without lazy evaluation, the particle cost would scale to approximately 2.7 ms, which would push PGFF into a net slowdown regime; lazy evaluation is therefore not merely an optimization but a structural requirement for the +0.7% overhead figure.

The IEKF speedup results from two factors: (1) downweighting uninformative features improves the condition number of the normal equations, reducing iterations to convergence; and (2) particle consensus provides implicit outlier identification, reducing the need for iterative reweighting.

*Per-sequence variation.* The +0.7% average masks meaningful per-sequence variation. PGFF ranges from −2.1% (faster than FAST-LIO2, on open sequence O-02 where lazy evaluation skips most points and the iteration-count savings dominate) to +4.3% (slower than FAST-LIO2, on corridor sequence C-02 where particle resampling triggers more frequently). The distribution is approximately symmetric around the mean, with corridor sequences clustered above and open sequences below. This pattern confirms that PGFF's overhead is workload-dependent in a predictable way: harder scenes pay more for PGFF, but those are precisely the scenes where its accuracy gains are largest (Table 5).

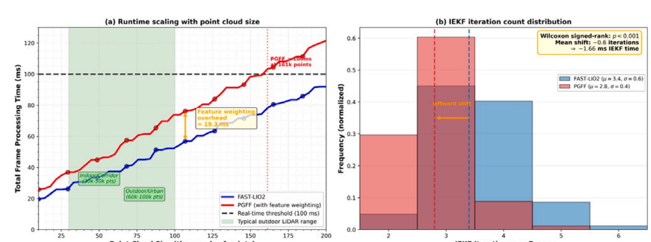


**Fig. 10.** Runtime analysis. **(a)** Runtime scaling with point cloud size. PGFF maintains real-time performance (< 100 ms) for typical outdoor LiDAR configurations. The feature weighting overhead scales linearly with point count. **(b)** Distribution of IEKF iteration counts per frame, pooled across all VBR and NCLT sequences. PGFF converges in fewer iterations ($\mu = 2.8$, $\sigma = 0.4$) than FAST-LIO2 ($\mu = 3.4$, $\sigma = 0.6$), a difference significant at $p < 0.001$ (Wilcoxon signed-rank). The leftward shift reflects improved conditioning of the normal equations under PGFF weighting (Section 4.5), which absorbs the 1.24 ms feature-weighting overhead and yields the net +0.7% total frame time reported in Table 8.

Fig. 10 shows runtime scaling with point cloud size. PGFF maintains real-time operation (< 100 ms per frame) for point clouds up to 150,000 points, covering typical outdoor LiDAR configurations.

*Embedded platform projection.* For resource-constrained deployment, the lightweight PGFF configuration ($M = 30$, higher $S_{\text{min}}$, $H_{\text{max}} = 10$) reduces feature-weighting cost from 1.24 ms to an estimated 0.7 ms on

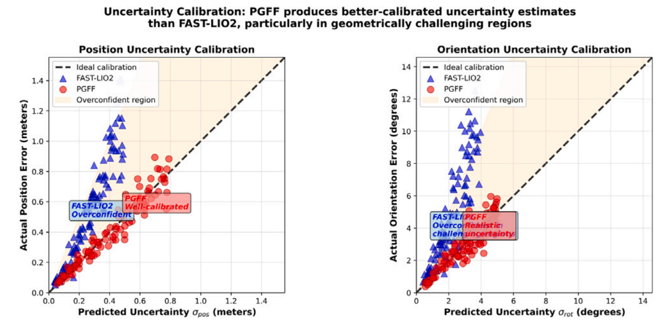


**Fig. 11.** Uncertainty calibration comparison. PGFF produces uncertainty estimates that better match empirical error, particularly in geometrically challenging regions where FAST-LIO2 is systematically overconfident.

**Table 9**
Normalized Estimation Error Squared (NEES). Ideal NEES ≈ 6.0 for 6-DOF pose. Lower deviation from ideal indicates better uncertainty calibration.

| Method | VBR | NCLT | Mean \|NEES − 6\| |
|---|---|---|---|
| FAST-LIO2 | 12.38 | 11.24 | 5.81 |
| LIO-SAM | 14.82 | 13.15 | 7.99 |
| PGFF (Ours) | 7.41 | 8.07 | **1.74** |

our reference workstation. Accounting for the approximately 3× clock penalty on embedded platforms such as the NVIDIA Jetson Orin, this projects to roughly 2 ms of feature-weighting cost—comfortably within real-time budgets for 10 Hz LiDAR. Empirical validation on embedded hardware remains future work (Section 5).

### 4.6. Uncertainty calibration

We evaluate uncertainty quality by comparing predicted covariance to empirical estimation error.

Fig. 11 shows calibration plots comparing predicted uncertainty (from the IEKF covariance) against actual pose error. A well-calibrated estimator produces points along the diagonal. FAST-LIO2 exhibits systematic overconfidence (points above the diagonal), particularly in geometrically degenerate regions. PGFF produces better-calibrated uncertainty by propagating geometric ambiguity information through the particle consensus mechanism.

Table 9 reports NEES values across datasets. For a 6-DOF pose, ideal NEES is 6.0; values significantly above 6.0 indicate overconfidence (predicted uncertainty smaller than actual error), while values below 6.0 indicate excessive conservatism. FAST-LIO2 and LIO-SAM exhibit systematic overconfidence (NEES 11.2–14.8), particularly in geometrically challenging regions. PGFF achieves NEES closer to the ideal (7.4–8.1), indicating substantially better-calibrated uncertainty through its uncertainty-aware feature weighting mechanism.

### 4.7. Failure case analysis

While PGFF improves robustness in most scenarios, we identify two failure modes:

**Extreme geometric degeneracy.** In VBR sequence C-03, a 50 m segment traverses a narrow corridor with nearly identical walls on both sides and no distinguishing features. All methods, including PGFF, accumulate significant drift in this segment. PGFF's particle consensus correctly identifies the geometric ambiguity—all points receive uniformly low weights (Fig. 12), causing the IEKF update to be dominated by the IMU prior term $\|\mathbf{x} \ominus \hat{\mathbf{x}}\|^2_{\mathbf{P}^{-1}}$ in Eq. (26). The resulting drift rate is therefore bounded by IMU bias stability and preintegration accuracy, which for the Microstrain 3DM-GX3 yields approximately 0.3 m over 50 m of travel. This represents a fundamental sensing limitation rather

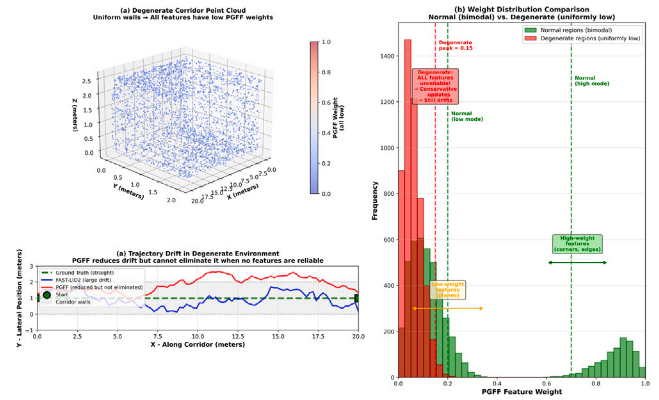


**Fig. 12.** Failure case analysis. (a) In extremely degenerate environments (uniform narrow corridors), PGFF correctly identifies low feature reliability but cannot prevent drift when no informative features exist. (b) Weight distribution shifts dramatically: normal regions show bimodal distribution (high/low weights), while degenerate regions show uniformly low weights, triggering conservative state updates but not eliminating drift.

than an algorithmic failure: PGFF cannot create geometric constraints where none exist, but it correctly signals low confidence, enabling downstream modules to flag uncertain trajectory segments. Potential mitigations include tighter IMU coupling during detected degeneracy (e.g., temporarily increasing IMU weight when mean consensus falls below a threshold), multi-modal fusion with cameras or radar to provide complementary constraints, or active perception strategies that seek informative viewpoints.

**Rapid illumination changes.** Although LiDAR is generally robust to lighting, we observe degraded performance in NCLT sequences with rapid transitions between bright sunlight and deep shadows. The root cause is reflectance-dependent point dropout: dark surfaces return fewer points, creating transient holes in the point cloud that distort both the $k$-nearest-neighbor covariance (Eq. (3)) and the voxel-level statistics used for regional expectations (Eq. (8)). This produces spurious surprise spikes at density boundaries, which particle consensus only partially mitigates since the correspondence search itself suffers from incomplete data. Adaptive neighborhood sizing that scales $k$ inversely with local point density, or reflectance-aware preprocessing that normalizes density before geometric analysis, could address this failure mode.

Both failure modes share a common structure: PGFF correctly diagnoses the degradation (via low consensus or surprise anomalies) but lacks a corrective mechanism beyond conservative reweighting. Closing this gap, by translating degradation detection into active mitigation, is the primary direction for future work.

## 5. Discussion

The experimental results demonstrate that PGFF achieves consistent improvements over state-of-the-art methods, but several aspects warrant deeper examination.

*Why does bidirectional information flow help?* Traditional SLAM pipelines treat feature extraction as a preprocessing step isolated from state estimation. PGFF's key insight is that state uncertainty contains valuable information about *which features to trust*. When the estimator is confident, aggressive feature selection improves efficiency; when uncertain, conservative weighting prevents error accumulation. The ablation results (Table 7) confirm this: removing bidirectional coupling while retaining the weighting components degrades performance by 12%, demonstrating that the information flow architecture—not merely the weighting mechanism—drives improvements.

*Relationship to robust estimation.* PGFF complements rather than replaces robust estimation techniques. M-estimators (Huber, 1964) and GNC (Yang et al., 2020) reweight measurements based on residuals computed *after* correspondence establishment; PGFF operates *before* and *during* this process. The PGFF+GNC results (Table 3) confirm that these approaches yield additive improvements, with the combination outperforming either alone on both datasets. DALI-SLAM's degeneracy-aware optimization is similarly complementary: it identifies *which directions* are poorly constrained, while PGFF identifies *which features* are unreliable. DALI-SLAM narrows the gap with PGFF in corridors (0.71 m vs. 0.65 m) where directional degeneracy dominates, but falls behind in mixed scenarios (0.65 m vs. 0.59 m) where feature-level uncertainty handling matters more.

*Computational trade-offs.* PGFF's net computational benefit (Table 8) arises from improved problem conditioning: suppressing uninformative measurements yields better-conditioned normal equations, reducing IEKF iterations to convergence. For resource-constrained platforms, PGFF admits a lightweight configuration ($M = 30$, higher $\mathcal{S}_{\min}$, $H_{\max} = 10$) that reduces feature weighting cost from 1.24 ms to an estimated 0.7 ms on our i7-12700 K. On embedded platforms (e.g., NVIDIA Jetson Orin), the ≈3× clock penalty yields roughly 2 ms—well within real-time budgets for 10 Hz LiDAR. Validating this configuration on embedded hardware remains future work.

*Limitations and failure modes.* PGFF cannot create information where none exists. In extremely degenerate environments (Section 4.7), all features receive low weights, correctly reflecting geometric ambiguity but unable to prevent drift. This represents a fundamental sensing limitation rather than an algorithmic failure. Future work could address this through multi-modal fusion (cameras, radar) or active sensing strategies that seek informative viewpoints. Additionally, the current implementation assumes static environments; extending PGFF to explicitly model dynamic objects remains an open challenge.

## 6. Conclusion

We presented Particle-Guided Feature Fusion (PGFF), a framework that establishes bidirectional information flow between state estimation and feature selection in LiDAR-inertial odometry by inverting the classical role of particle filters, propagating particles through feature space rather than pose space. Three contributions support this: a geometric surprise prior based on KL divergence, particle-based consensus voting for correspondence reliability, and multi-hypothesis loop closure that defers commitment until sufficient evidence accumulates. Experiments on VBR and NCLT demonstrate that PGFF reduces absolute trajectory error by 23.5% over FAST-LIO2 and 29.8% over LIO-SAM, with gains up to 36% in degenerate corridors, while achieving 10.1% faster IEKF convergence with only 0.7% runtime overhead. Combining PGFF with GNC yields further additive improvements, confirming complementarity between pre- and post-correspondence robustness. Multi-hypothesis loop closure reduces false positives by 58% across multiple place recognition front-ends while maintaining 94% recall. Ablation studies confirm that each component contributes meaningfully, with the bidirectional coupling mechanism proving essential. More broadly, PGFF illustrates that coupling uncertainty-aware feature selection with uncertainty-aware state estimation can yield measurable improvements in accuracy and robustness over treating the two independently. Whether this coupling is universally essential across SLAM systems remains an open empirical question, but our results indicate the boundary between perception and estimation deserves further attention.

## CRediT authorship contribution statement

**Cabrel Wouladje:** Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization. **Golden Tendekai Mumanikidzwa:** Writing – original draft, Visualization, Validation, Methodology, Investigation. **Huiying Xu:** Writing – review & editing, Supervision, Resources, Funding acquisition. **Hongbo Li:** Writing – review & editing, Supervision, Resources, Funding acquisition. **Mantang Liu:** Writing – review & editing, Investigation, Data curation. **Wenjie Wang:** Writing – review & editing, Investigation, Data curation. **Zhendong Chen:** Writing – review & editing, Investigation, Resources. **Wenzhe Tan:** Writing – review & editing, Investigation, Resources. **Chao Chen:** Writing – review & editing, Investigation, Resources. **Balong Wang:** Writing – review & editing, Investigation, Resources. **Zhenglong Wang:** Writing – review & editing, Investigation, Resources. **Xinzhong Zhu:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (62376252); Zhejiang Province Leading Geese Plan (2025C02025).

## References

Bai, C., Xiao, T., Chen, Y., Wang, H., Zhang, F., Gao, X., 2022. Faster-LIO: Lightweight tightly coupled Lidar-inertial odometry using parallel sparse incremental voxels. IEEE Robot. Autom. Lett. 7 (4), 4861–4868. http://dx.doi.org/10.1109/LRA.2022.3152830.

Bell, B.M., Cathey, F.W., 1993. The iterated Kalman filter update as a Gauss-Newton method. IEEE Trans. Autom. Control 38 (2), 294–297.

Blackman, S.S., 2004. Multiple hypothesis tracking for multiple target tracking. IEEE Aerosp. Electron. Syst. Mag. 19 (1), 5–18.

Brizi, L., Giacomini, E., Di Giammarino, L., Ferrari, S., Salem, O., De Rebotti, L., Grisetti, G., 2024. VBR: a vision benchmark in rome. In: 2024 IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 15868–15874.

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., Leonard, J.J., 2016. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. IEEE Trans. Robot. 32 (6), 1309–1332.

Cai, Y., Xu, W., Zhang, F., 2021. Ikd-tree: An incremental k-d tree for robotic applications. arXiv preprint arXiv:2102.10808.

Carlevaris-Bianco, N., Ushani, A.K., Eustice, R.M., 2016. University of Michigan North Campus long-term vision and lidar dataset. Int. J. Robot. Res. 35 (9), 1023–1035.

Chebrolu, N., Läbe, T., Vysotska, O., Behley, J., Stachniss, C., 2021. Adaptive robust kernels for non-linear least squares problems. IEEE Robot. Autom. Lett. 6 (2), 2240–2247.

Chghaf, M., Rodríguez Flórez, S., El Ouardi, A., 2023. A multimodal loop closure fusion for autonomous vehicles SLAM. Robot. Auton. Syst. 165, 104446. http://dx.doi.org/10.1016/j.robot.2023.104446.

Chghaf, M., Rodríguez Flórez, S., El Ouardi, A., 2025. Extended study of a multi-modal loop closure detection framework for SLAM applications. Electronics 14 (3), 421. http://dx.doi.org/10.3390/electronics14030421.

Dellaert, F., Kaess, M., 2017. Factor graphs for robot perception. Found. Trends Robot. 6 (1–2), 1–139.

Ebadi, K., Palieri, M., Wood, S., Padgett, C., Agha-mohammadi, A.-a., 2021. DARE-SLAM: Degeneracy-aware and resilient loop closing in perceptually-degraded environments. In: IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 14462–14468.

Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24 (6), 381–395.

Forster, C., Carlone, L., Dellaert, F., Scaramuzza, D., 2017. On-manifold preintegration for real-time visual-inertial odometry. IEEE Trans. Robot. 33 (1), 1–21.

Grisetti, G., Stachniss, C., Burgard, W., 2007. Improved techniques for grid mapping with Rao-Blackwellized particle filters. IEEE Trans. Robot. 23 (1), 34–46.

He, D., Xu, W., Chen, N., Kong, F., Yuan, C., Zhang, F., 2023. Point-LIO: Robust high-bandwidth light detection and ranging inertial odometry. Adv. Intell. Syst. 5 (7), 2200459.

Huber, P.J., 1964. Robust estimation of a location parameter. Ann. Math. Stat. 35 (1), 73–101.

Kim, G., Kim, A., 2018. Scan context: Egocentric spatial descriptor for place recognition within 3D point cloud map. In: IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 4802–4809.

Latif, Y., Cadena, C., Neira, J., 2013. Robust loop closing over time for pose graph SLAM. In: International Journal of Robotics Research, vol. 32, SAGE, pp. 1611–1626.

Lee, D., Lim, H., Han, S., 2024. GenZ-ICP: Generalizable and degeneracy-robust LiDAR odometry using an adaptive weighting. IEEE Robot. Autom. Lett. 10 (1), 152–159. http://dx.doi.org/10.1109/LRA.2024.3498779.

Liu, Z., Shi, W., Yu, Y., Chen, P., Chen, B., 2023. Multi-robot cooperative LiDAR SLAM for efficient mapping in urban scenes. In: ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. pp. 473–480.

Lowry, S., Sünderhauf, N., Newman, P., Leonard, J.J., Cox, D., Corke, P., Milford, M.J., 2016. Visual place recognition: A survey. IEEE Trans. Robot. 32 (1), 1–19.

McGann, D., Rogers, J.G., Kaess, M., 2023. Robust incremental smoothing and mapping (riSAM). In: IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 4157–4163.

Mehra, R., 1972. Approaches to adaptive filtering. IEEE Trans. Autom. Control 17 (5), 693–698.

Montemerlo, M., Thrun, S., Koller, D., Wegbreit, B., 2002. FastSLAM: A factored solution to the simultaneous localization and mapping problem. In: AAAI National Conference on Artificial Intelligence. pp. 593–598.

Nieto, J., Bailey, T., Nebot, E., 2007. Recursive scan-matching SLAM. Robot. Auton. Syst. 55 (1), 39–49.

Pan, Y., Xiao, P., He, Y., Shao, Z., Li, Z., 2021. MULLS: Versatile LiDAR SLAM via multi-metric linear least square. In: IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 11633–11640.

Reid, D., 1979. An algorithm for tracking multiple targets. IEEE Trans. Autom. Control 24 (6), 843–854.

Rosinol, A., Abate, M., Chang, Y., Carlone, L., 2020. Kimera: An open-source library for real-time metric-semantic localization and mapping. In: IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 1689–1696.

Shan, T., Englot, B., 2018. LeGO-LOAM: Lightweight and ground-optimized lidar odometry and mapping on variable terrain. In: IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 4758–4765.

Shan, T., Englot, B., Meyers, D., Wang, W., Ratti, C., Rus, D., 2020. LIO-SAM: Tightly-coupled lidar inertial odometry via smoothing and mapping. In: IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 5135–5142.

Shao, J., Zhang, W., Mellado, N., Cai, Z., Luo, L., Zhou, G., Habib, A., 2024. Forest feature LiDAR SLAM (F2-LSLAM) for backpack systems. ISPRS J. Photogramm. Remote Sens. 211, 73–93.

Stachniss, C., Leonard, J.J., Thrun, S., 2016. Simultaneous localization and mapping. Springer Handb. Robot. 1153–1176.

Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D., 2012. A benchmark for the evaluation of RGB-D SLAM systems. In: IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 573–580.

Sünderhauf, N., Protzel, P., 2012. Switchable constraints for robust pose graph SLAM. In: IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 1879–1884.

Tian, Y., Chang, Y., Arias, F.H., Nieto-Granda, C., How, J.P., Carlone, L., 2022. Kimera-multi: Robust, distributed, dense metric-semantic SLAM for multi-robot systems. IEEE Trans. Robot. 38 (4), 2022–2038.

Umeyama, S., 1991. Least-squares estimation of transformation parameters between two point patterns. IEEE Trans. Pattern Anal. Mach. Intell. 13 (4), 376–380.

Vizzo, I., Guadagnino, T., Mersch, B., Wiesmann, L., Behley, J., Stachniss, C., 2023. KISS-ICP: In defense of point-to-point ICP–simple, accurate, and robust registration if done the right way. IEEE Robot. Autom. Lett. 8 (2), 1029–1036.

Weinmann, M., Jutzi, B., Hinz, S., Mallet, C., 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. ISPRS J. Photogramm. Remote Sens. 105, 286–304.

Wu, W., Chen, C., Yang, B., Li, L., et al., 2025. DALI-SLAM: Degeneracy-aware LiDAR-inertial SLAM with novel distortion correction and accurate multi-constraint pose graph optimization. ISPRS J. Photogramm. Remote Sens. 221, 64–77.

Xu, W., Cai, Y., He, D., Lin, J., Zhang, F., 2022. FAST-LIO2: Fast direct LiDAR-inertial odometry. IEEE Trans. Robot. 38 (4), 2053–2073.

Xu, Y., Chen, C., Wang, Z., Yang, B., Wu, W., Li, L., Wu, J., Zhao, L., 2023. PMLIO: Panoramic tightly-coupled multi-LiDAR-inertial odometry and mapping. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. pp. 703–710.

Xu, W., Zhang, F., 2021. FAST-LIO: A fast, robust LiDAR-inertial odometry package by tightly-coupled iterated Kalman filter. IEEE Robot. Autom. Lett. 6 (2), 3317–3324.

Yang, H., Antonante, P., Tzoumas, V., Carlone, L., 2020. Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection. IEEE Robot. Autom. Lett. 5 (2), 1127–1134.

Zhang, Z., Scaramuzza, D., 2018. A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 7244–7251.

Zhang, J., Singh, S., 2014. LOAM: Lidar odometry and mapping in real-time. In: Robotics: Science and Systems. RSS.

Zheng, X., Zhu, J., 2024. Traj-LO: In defense of LiDAR-only odometry using an effective continuous-time trajectory. IEEE Robot. Autom. Lett. 9 (2), 1961–1968. http://dx.doi.org/10.1109/LRA.2024.3352360.