

G3Reg: Pyramid Graph-Based Global Registration Using Gaussian Ellipsoid Model

Zhijian Qiao^{ID}, Graduate Student Member, IEEE, Zehuan Yu^{ID}, Binqian Jiang^{ID}, Graduate Student Member, IEEE, Huan Yin^{ID}, Member, IEEE, and Shaojie Shen^{ID}

Abstract—This study introduces a novel framework, G3Reg, for fast and robust global registration of LiDAR point clouds. In contrast to conventional complex keypoints and descriptors, we extract fundamental geometric primitives, including planes, clusters, and lines (PCL) from the raw point cloud to obtain low-level semantic segments. Each segment is represented as a unified Gaussian Ellipsoid Model (GEM), using a probability ellipsoid to ensure the ground truth centers are encompassed with a certain degree of probability. Utilizing these GEMs, we present a distrust-and-verify scheme based on a Pyramid Compatibility Graph for Global Registration (PAGOR). Specifically, we establish an upper bound, which can be traversed based on the confidence level for compatibility testing to construct the pyramid graph. Then, we solve multiple maximum cliques (MAC) for each level of the pyramid graph, thus generating the corresponding transformation candidates. In the verification phase, we adopt a precise and efficient metric for point cloud alignment quality, founded on geometric primitives, to identify the optimal candidate. The algorithm’s performance is validated on three publicly available datasets and a self-collected multi-session dataset. Parameter settings remained unchanged during the experiment evaluations. The results exhibit superior robustness and real-time performance of the G3Reg framework compared to state-of-the-art methods. Furthermore, we demonstrate the potential for integrating individual GEM and PAGOR components into other registration frameworks to enhance their efficacy.

Note to Practitioners—Our proposed method aims to perform global registration for outdoor LiDAR point clouds. Our methodology, which extracts point cloud segments and utilizes their centers for registration, differs from conventional approaches that rely on keypoints and descriptors. We further propose GEM to model the uncertainty of the centers and embed it into our distrust-and-verify framework. In theory, our method can be applied to any registration task that involves primitives representable as sets of Gaussians or points. Additionally, practitioners should consider the following to enhance applicability. First, practitioners can fine-tune the parameters of the segmentation algorithm to generate more repeatable segmentation results. Second, although our default setting uses four compatibility test thresholds, fewer may suffice, especially when translations between point clouds are minor. Finally, for geometrically

Manuscript received 22 August 2023; revised 12 January 2024; accepted 3 April 2024. This article was recommended for publication by Associate Editor S. Wang and Editor D. Song upon evaluation of the reviewers’ comments. This work was supported in part by Hong Kong Center for Construction Robotics (InnoHK Center supported by Hong Kong ITC), in part by HKUST Postgraduate Studentship, and in part by HKUST-DJI Joint Innovation Laboratory. (Corresponding author: Huan Yin.)

The authors are with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, China (e-mail: zqiaoac@connect.ust.hk; zyuay@connect.ust.hk; bjiangah@connect.ust.hk; eehyin@ust.hk; eeshaojie@ust.hk).

Digital Object Identifier 10.1109/TASE.2024.3394519

uninformative segments such as vegetation, consider extracting descriptors within these segments to increase correspondences.

Index Terms—Global registration, point cloud, LiDAR, graph theory, robust estimation.

I. INTRODUCTION

ALIGNING two individual point clouds is a fundamental and indispensable operation in robotics and autonomous systems. Global registration, which estimates the transformation without prior information, has become crucial for tasks such as loop closure [2], [3], [4], multi-session SLAM [5], [6] and relocalization [7], [8]. The standard pipeline of global registration can typically be divided into front-end and back-end stages, which are responsible for establishing putative correspondences \mathcal{I} and estimating the transformation (\mathbf{R}, \mathbf{t}), respectively. The transformation estimation can be formulated as a least squares problem:

$$\min_{\mathbf{R} \in \text{SO}(3), \mathbf{t} \in \mathbb{R}^3} \sum_{(x_k, y_k) \in \mathcal{I}} \rho(r(y_k, f(x_k | \mathbf{R}, \mathbf{t}))) \quad (1)$$

where $\rho(\cdot)$ represents the robust kernel function, $r(\cdot)$ is the residual function, and $f(\cdot)$ denotes the rigid transformation.

Conventional approaches for establishing \mathcal{I} rely on complex methods of keypoint and descriptor extraction. These methods include handcrafted techniques [9] and deep learning-based local descriptors [10], [11], [12]. However, these approaches encounter challenges such as excessive processing time and reduced descriptor consistency, primarily due to varying viewpoints. Specifically, during the construction and matching of descriptors, frequent nearest neighbor searches are performed in both the spatial and descriptor space, which can be time-consuming. Additionally, changes in viewpoint result in occlusions and density variations, resulting in alterations in the neighborhood of each point. This further affects the computation of local features, such as normals [9], [10] and local reference frames [11], [12], which are essential for the consistency of descriptors relying on these local features. Consequently, an extremely high outlier ratio (e.g., 99%) is observed, significantly undermining the quality of the registration process.

To address this issue, recent advances [1], [13], [14], [15], [16], [17] suggest utilizing maximal clique inlier selection and coupling it with a robust estimator for transformation estimation. The inlier clique has been proven to belong to

at least one maximal clique in the compatibility graph [13]. To efficiently find this maximal clique, the maximum clique with the largest cardinality is computed. However, this strategy does not always work due to inappropriate threshold selection in the compatibility test [15], especially under extreme viewpoint changes and repetitive patterns, as discussed in [18].

In this study, we present a novel framework, named G3Reg, for the global registration of LiDAR point clouds that effectively addresses these issues. Our first key insight is to use segments and classify them into geometric primitives (planes, clusters, and lines) as a replacement for complex keypoints and descriptors. This approach has two advantages: firstly, using segments directly avoids the loss of geometric information caused by traditional keypoints; secondly, it eliminates the need for complex keypoint and descriptor extraction, thus reducing computational time. Our second key insight is a novel distrust-and-verify scheme that generates multiple transformation candidates and leverages compressed raw point clouds to verify and select the most appropriate candidate. Specifically, multiple maximum cliques are obtained by constructing a pyramid compatibility graph using a traversable confidence-related upper bound for the compatibility test. Each maximum clique is then used to estimate a transformation candidate.

This work is an extended version of our earlier conference paper [18]. In comparison, the journal version presents several notable enhancements, including a more efficient front-end approach that eliminates the need for the semantic segmentation network. Additionally, we have significantly improved our distrust-and-verify framework in terms of both efficiency and robustness. The contributions of this work are outlined as follows:

- We introduce a novel segment-based front-end approach to obtain putative correspondences. In this approach, each segment is parameterized by the proposed GEM, which lays the groundwork for the following multi-threshold compatibility test and enables distribution-to-distribution registration at the back end.
- We propose a distrust-and-verify scheme that generates multiple transformation candidates by distrusting the compatibility test and verifying their point cloud alignment quality. To achieve this, a pyramid compatibility graph is constructed using the multi-threshold compatibility test, and a graduated MAC solver is proposed to solve the maximum clique on each level. Finally, transformation candidates are estimated from these cliques, and an evaluation function is designed to select the optimal one.
- We conduct extensive evaluations of our proposed G3Reg on three publicly available datasets and a self-collected multi-session dataset in the real world. The experimental results demonstrate the superiority of our proposed method in terms of robustness and real-time performance compared to state-of-the-art handcrafted and deep learning methods.
- To enhance comprehension and encourage the adoption of our framework, we have released our source code¹ as

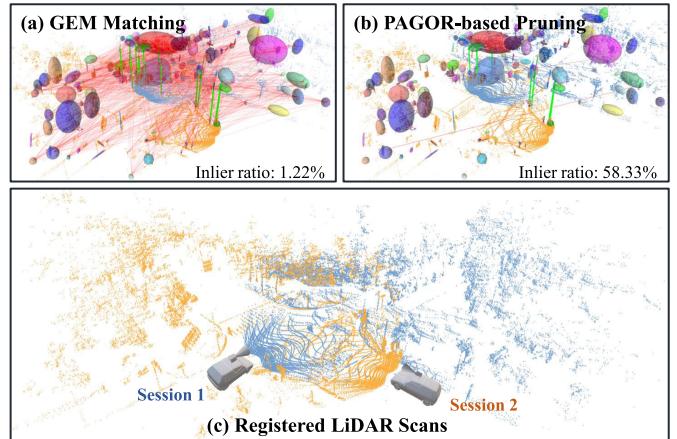


Fig. 1. Global registration procedure using our proposed G3Reg. (a) Front-end GEM matching for Gaussian ellipsoid modeling and correspondence building. (b) Back-end PAGOR for outlier pruning and pose estimation under a distrust-and-verify scheme. (c) A challenging global registration result at a road intersection, in which the two LiDAR point clouds are with low overlap and a large view difference.

open-source to the community. Our framework not only facilitates the replication of experiments but also offers the flexibility to modify individual components like detectors, matchers, maximum clique solvers, transformation estimators, and scalability to suit different datasets.

II. RELATED WORK

Global point cloud registration can be categorized into two primary models: correspondence-free and correspondence-based. The correspondence-free model achieves global registration using techniques like Fourier analysis [19], branch and bound strategies [20], and learned feature alignment techniques [21], [22]. However, correspondence-free models are generally time-consuming for global transformation search. This paper proposes a correspondence-based method, which involves various modules, including discriminative keypoint detection, descriptor extraction, efficient correspondence ranking for outlier pruning, and robust transformation estimation. In this section, we delve into related works that concentrate on these individual modules.

A. 3D Point Matching

1) *3D Keypoint Detection*: The process of 3D point matching involves two main steps: keypoint detection and descriptor designing. In the case of LiDAR point clouds in large-scale scenes, keypoint detection is crucial for reducing the number of point clouds that require registration, thus alleviating the burden on the registration algorithm.

Manual keypoint detection methods use local geometric properties to assess the saliency of a point. For example, ISS [23] uses the eigenvalue ratio of the neighborhood covariance matrix to evaluate point prominence. Similarly, KPQ [24] utilizes the largest eigenvalue ratio to identify candidate keypoints and calculates curvature for prominence estimation. On the other hand, the deep learning method D3Feat [25] predicts keypoint scores for each point in dense point clouds, assigning higher scores to matchable matches.

¹<https://github.com/HKUST-Aerial-Robotics/G3Reg>

In practical applications, LiDAR-based keypoint detection methods are often affected by factors like noise, occlusion, and point cloud density, resulting in poor repeatability. Additionally, keypoints may discard important geometric information, such as cluttered vegetation and smooth walls. Therefore, methods like Segmatch [4] and BoxGraph [3] have been developed, which extract segments to represent point clouds and use their centers for registration. Segment-based registration methods retain the structural information of the entire point cloud, showing robustness to noise and density variations. The common parameterization of segments could be 3D bounding box [3] or the ellipsoid [26], [27]. However, the repeatability of their center estimation is significantly influenced by viewpoints.

In our proposed method, we also utilize segments for global registration. The key difference lies in our introduction of a plane-assisted segmentation method to achieve segments with higher repeatability. Additionally, we employ GEM to model the uncertainty of segment centers, ensuring probabilistic coverage of the ground truth center. This approach enhances the robustness and accuracy of our registration process.

2) *3D Descriptor-Based Matching*: SOTA works focus on constructing descriptors that are rigid transformation invariant, discriminative, and highly generalizable. To achieve invariance, FPFH [9] and RPMNet [10] compute features such as distances and angles between points and normals that are invariant to transformations to construct their descriptors. Similarly, the local reference frame (LRF) is utilized in techniques like SHOT [28], SpinNet [11] and GeDi [12] to maintain the invariance to the rotation and translation. Furthermore, RoReg [29] introduces the icosahedral group and rotation-equivariant layers to achieve rotation equivariance, leading to rotation invariance through average pooling.

However, as the viewpoint changes, descriptor consistency can be compromised by occlusions and density variations, which means the descriptors of the same point from two different point clouds may not be similar. Furthermore, these factors can affect the local feature estimation, such as normals and LRFs, which in turn may degrade the descriptors' discriminative and generalization capabilities.

In contrast to point-based methods, our proposed GEM use segments for registration, which naturally reduces the impact of density changes since these do not significantly alter the segment's shape. Additionally, our method considers the uncertainty of segment center through the proposed probability coverage of the center (see Assumption 1). While GEM does not rely on strong descriptors, we use a weak descriptor coupled with a Mutual-K-Nearest Neighbors (MKNN) matching strategy to establish the initial correspondence set \mathcal{I} in Problem (1). This approach effectively prevents degeneracy, where inlier correspondences are fewer than three, thus ensuring a more robust registration.

B. Outlier Pruning

The robust loss function $\rho(\cdot)$ in Problem (1) is known for effectively rejecting outliers, but it may face challenges when dealing with problems involving high outlier ratios, such as up

to 99% outliers. To reduce the outlier ratio to a level where existing solvers (such as RANSAC and GNC [30]) perform well, a common approach in the computer vision community involves employing reciprocity checks and ratio tests [31]. Unlike the aforementioned approaches that heavily rely on descriptor performance, GORE [32] reformulates the problem, leveraging deterministic geometric properties to establish an upper bound on the inlier set size subproblem and a lower bound on the optimal solution, effectively eliminating incorrect matches.

In recent years, graph-theoretic methods have emerged as powerful tools in various registration works, demonstrating exceptional performance regardless of descriptor quality. Based on the compatibility graph, Enqvist et al. [33] propose using vertex cover to identify mutually consistent correspondences. MV (Mutual Voting) [34] introduces a mutual voting method for ranking 3D correspondences, enabling vertices and edges to refine each other in a mutually beneficial manner. Another widely used and stricter algorithm is the maximal clique inlier selection [1], [13], [14], [17], [35], [36], which requires every correspondence to be mutually compatible. However, finding the maximal clique is known to be an NP-hard problem, and its runtime often becomes impractical as the number of correspondences increases. To address this challenge, ROBIN [15] proposes a faster alternative by computing the maximum k -core of the graph.

In our proposed method, we also adopt the maximum clique approach but introduce a traversable confidence-related threshold to construct a pyramid compatibility graph and solve multiple maximum cliques in a graduated manner. This approach effectively increases the probability that inliers make up one of the solved maximum cliques.

C. Robust Transformation Estimation

Robust transformation estimation has been extensively studied in various research fields [37]. It can be formulated using different approaches, including least trimmed squares (LTS) in classical robust statistics [38], consensus maximization (MC) in computer vision [39], quadratic pose estimation problems [40], and truncated least squares (TLS) in robotics [13], [41]. Trimmed ICP [38] selects potential inlier correspondences and estimates the transformation using a predefined trimming percentage, which is often unknown in real applications. On the other hand, MC defines a consensus set that contains consistent correspondences with residuals below a threshold and aims to find the optimal (\mathbf{R}, \mathbf{t}) that maximizes the size of the consensus set. However, MC becomes NP-hard [42] when $\mathbf{R} \in \text{SO}(3)$, leading to the adoption of the suboptimal but efficient RANSAC algorithm to find the outlier-free set. Nevertheless, the runtime of RANSAC increases exponentially with the outlier rate.

In this study, we formulate the registration problem as a truncated least squares (TLS) problem. To address this problem, we leverage the Black-Rangarajan duality [43] to rewrite TLS using the corresponding outlier process. To solve the TLS problem, we adopt the fast heuristic Graduated Non-Convexity (GNC) algorithm based on alternating optimization

[30]. The GNC approach starts with a convex approximation of the cost function and gradually refines it to recover the original function. By employing graph-theoretic outlier pruning, GNC performs well in scenarios with moderate outlier rates, such as 80%. The primary difference between our formulation and conventional ones [13] lies in the usage of a distribution-to-distribution residual rather than a point-to-point residual, following a similar approach to Generalized ICP (GICP) [44].

III. PROBLEM FORMULATION

Given two 3D LiDAR point clouds, denoted as \mathcal{X} and \mathcal{Y} , our primary goal is to estimate the optimal rigid transformation, represented as $(\tilde{\mathbf{R}}, \tilde{\mathbf{t}})$, that best aligns the source point cloud \mathcal{X} with the target point cloud \mathcal{Y} . To achieve this, we first extract Gaussian Ellipsoid Models (GEM) from the segments within the point clouds during the front-end processing phase (Section IV). Consequently, we can denote \mathcal{X} and \mathcal{Y} as a series of GEMs, $\{x_1 \dots x_{N_x}\}$ and $\{y_1 \dots y_{N_y}\}$ respectively.

In the subsequent back-end phase of the framework, our strategy for estimating the optimal transformation parameters $(\tilde{\mathbf{R}}, \tilde{\mathbf{t}})$ employs a distrust-and-verify scheme. This scheme generates multiple transformation candidates using the extracted GEMs and then selects the most probable candidate that optimally aligns \mathcal{X} and \mathcal{Y} .

To hypothesize a candidate transformation $(\mathbf{R}^*, \mathbf{t}^*)$, we construct a compatibility graph (Section V-A) for the original putative correspondence set \mathcal{I} and derive a potential inlier set \mathcal{I}^* by identifying the maximum clique within the graph (Section V-C). We adapt the formulation presented in Equation (1) to express the registration problem as follows:

$$\mathbf{R}^*, \mathbf{t}^* = \min \sum_{(x_k, y_k) \in \mathcal{I}^*} \rho(r(y_k, f(x_k | \mathbf{R}, \mathbf{t}))) \quad (2)$$

In this equation, \mathcal{I} is replaced with \mathcal{I}^* compared to the formulation in Equation (1). The robust loss function $\rho(\cdot)$ and the residual function $r(\cdot)$ of two associated GEMs are defined in Section VI-A.

By traversing the compatibility test threshold across various confidence levels (Section V-B) to construct a pyramid compatibility graph, we can derive multiple potential \mathcal{I}_m^* from their maximum cliques. We can then solve the corresponding hypothetical transformation $(\mathbf{R}_m^*, \mathbf{t}_m^*)$ using Equation (2). Subsequently, we introduce an evaluation function g (Section VI-B) to determine the most suitable transformation based on the geometric information of \mathcal{X} and \mathcal{Y} , as follows:

$$\tilde{\mathbf{R}}, \tilde{\mathbf{t}} = \arg \min_{\tilde{\mathbf{R}} \in \{\mathbf{R}_m^*\}, \tilde{\mathbf{t}} \in \{\mathbf{t}_m^*\}} g(\mathbf{R}_m^*, \mathbf{t}_m^* | \mathcal{X}, \mathcal{Y}) \quad (3)$$

In summary, our proposed G3Reg method firstly introduces the novel concept of GEM and utilizes it to obtain the putative correspondence set. It then follows a carefully designed distrust-and-verify framework that facilitates fast and robust registration.

IV. GEM-BASED DATA ASSOCIATION

In this section, we present our method for extracting Gaussian Ellipsoid Models (GEMs) from two given point clouds

and acquiring the initial putative correspondences, \mathcal{I} . This entire procedure is composed of three steps. First, we design a plane-aided segmentation algorithm to process the input LiDAR scans, resulting in a series of plane, line and cluster (PCL) segments (Section IV-A). Second, for each of these segments, we model it as a GEM. This is achieved by estimating its statistical Gaussian distribution parameters along with an additional pseudo-Gaussian parameter (Section IV-B). Lastly, based on the GEM parameters, we select a subset of these GEMs that are deemed significant and implement a soft association strategy. This strategy is facilitated through the use of mutual K nearest neighbors (Section IV-C).

A. Plane-Aided Segmentation

The initiation of GEM extraction involves identifying repeatable segments within a LiDAR point cloud from varying viewpoints. We segment point clouds with reference to TRAVEL [45]; however, a significant distinction lies in our categorization of the resulting segments into three types: planes, lines, and clusters (PCL). This methodology is underpinned by three considerations. First, planes and lines, being distinct geometric structures, are advantageous for precise point cloud registration. Second, plane segmentation helps prevent extensive portions of the point cloud segments from being incorrectly merged, resulting in under-segmentation. This strengthens the repeatability of the derived clusters. Lastly, these low-level semantics assist in data association. Figure 3 presents a visual comparison of outcomes obtained with and without plane-assisted segmentation, and also the GEM-based matching proposed in our approach.

More specifically, we first voxelise the point cloud with a voxel size of s_v , and determine whether each voxel forms a plane based on the ratio of the smallest eigenvalue λ_3 to the second smallest eigenvalue λ_2 and its threshold σ_p . We then apply region growing to merge plane voxels into plane segments. The decision to merge plane voxels is predicated on several factors, including point-to-plane distance, the similarity of normal vectors n_1 and n_2 , and adjacency relations. These conditions are elaborated as follows:

- 1) Distance condition: $abs(n_1^T(p_1 - p_2))$ and $abs(n_2^T(p_1 - p_2))$ are smaller than a threshold σ_{p2n} . Here, p represents the center of plane points.
- 2) Normal condition: $abs(n_1^T n_2)$ is smaller than a threshold σ_n .
- 3) Adjacency condition: two planes should contain at least one voxel that is adjacent to each other.

Subsequently, we segment the remaining point cloud into clusters using the TRAVEL algorithm [45]. This step is the most time-consuming in the GEM extraction process, with a time complexity of $O(n \log(n))$ where n is the total number of nodes in a range image row. Furthermore, for each cluster, we employ RANSAC to fit a line, employing a point-to-line distance threshold σ_{p2l} . Clusters that exhibit an inlier rate exceeding a predefined threshold σ_{line} are identified as lines. Ultimately, every voxel is re-labeled based on the segmented PCL segments and is used in the verification stage for evaluating alignment quality (Section VI-B).

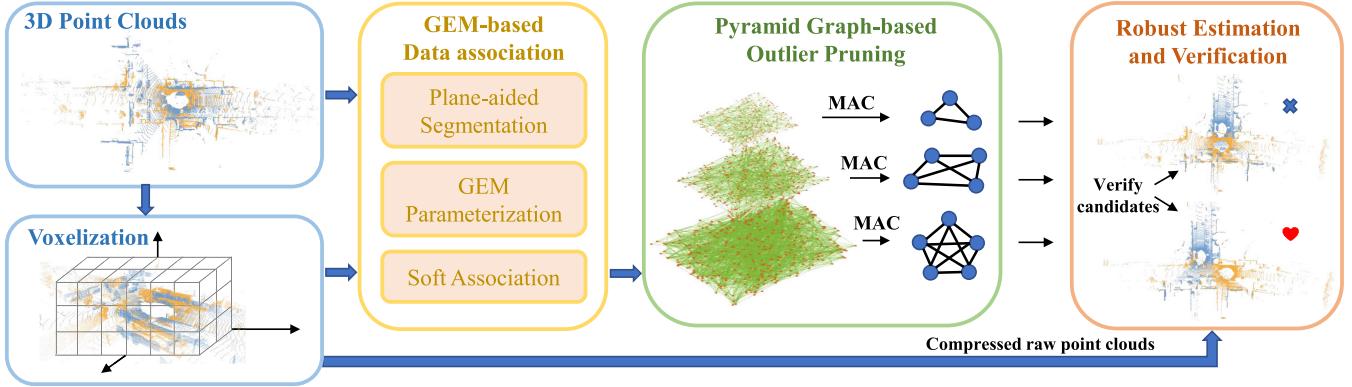


Fig. 2. The proposed global registration method G3Reg adopts a distrust-and-verify framework, consisting of three components. First, putative correspondences between input point clouds are established via GEM-based data association (block in yellow). Then a pyramid compatibility graph is constructed to generate maximal cliques to obtain transformation candidates (block in green). Finally, the original point cloud information is re-utilized to select the optimal candidate (block in brownish-orange).

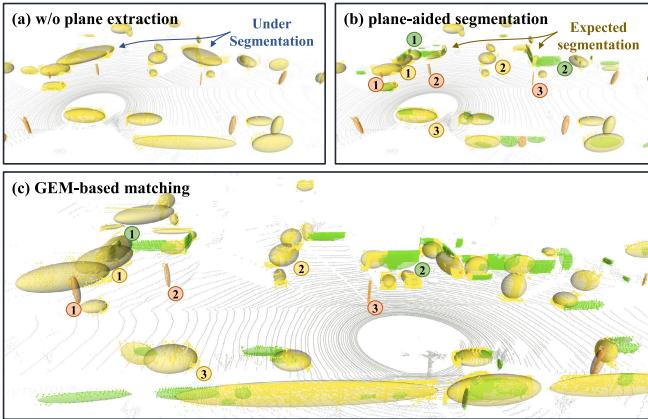


Fig. 3. Visualization of GEM-based data association. Green, yellow, and orange represent planes, clusters, and lines respectively. (a) and (b) show the difference in segmentation results without or with plane assistance. As the arrow indicates, the former leads to under-segmentation. (c) is a different frame from (b). We use different colors and indices to illustrate the matching between GEMs of these two frames.

Data Structure 1: Gaussian Ellipsoid Model

```

1 Struct GEM:
2   PCL type; % Plane, cluster or line
3   Vector3 μ; % Statistical center
4   Matrix3 Σ; % Statistical covariance
5   Matrix3 Σ̂; % Pseudo covariance
6   Vector3 λ̂; % Eigenvalues of Σ̂
7 end

```

B. Parameter Estimation

In this section, we represent the low-level semantic segments obtained in the previous step via a standardized parameterization mentioned as the Gaussian Ellipsoid Model (GEM). The attributes of the GEM are outlined in **Data Structure 1**. Each GEM's type is associated with its specific geometric primitive. We directly derive the statistical Gaussian distribution parameters from the parameters of the contained

voxels, thereby eliminating the need for recomputation:

$$\mu = \frac{1}{\sum_{k=1}^{N_v} N_k} \sum_{k=1}^{N_v} N_k \mu_k$$

$$\Sigma = \frac{1}{\sum_{k=1}^{N_v} N_k} \sum_{k=1}^{N_v} (N_k (\Sigma_k + \mu_k \mu_k^T)) - uu^T \quad (4)$$

where N_v is the number of contained voxels, and N_v and (μ_k, Σ_k) denote the point number, statistical center and covariance matrix of the k -th voxel, respectively.

Generally, the distribution of an object in the real world adheres to a uniform distribution. Therefore, the use of statistical Gaussian covariance as a descriptor may result in distortions and limitations, particularly when confronted with variations in point density. To address this issue, we introduce pseudo-Gaussian parameters to model the uncertainty associated with a segment's center point μ , based on Assumption 1, as illustrated in Figure 4.

Assumption 1 (Probability Coverage of Center): Given an observed segment, the observation's incompleteness implies that the statistically derived μ cannot represent its true center $\hat{\mu}$. However, we expect $\hat{\mu}$ to converge within the minimal 3D Oriented Bounding Box (OBB) that encloses the segment, with a high degree of probability. Moreover, the difference $\delta\mu$ between μ and $\hat{\mu}$ is constrained within this 3D OBB, as both μ and $\hat{\mu}$ converge within it.

Following this assumption, we estimate pseudo-Gaussian parameters in two stages. Initially, we regress the 3D OBB by projecting the points onto a 2D plane along a specified axis (normal for plane and cluster, direction for line). We then derive the 2D OBB from the projected points as per [46]. The 3D OBB is finally attained by merging the 2D OBB with its projection axis.

In the subsequent stage, we derive the 3D Oriented Bounding Ellipsoid (OBE) that is tangential to the 3D OBB and designate it as the probability ellipsoid for the center point μ of the GEM. Assuming that $\mu \in \mathcal{N}(\hat{\mu}, \hat{\Sigma})$, this probability ellipsoid can be represented as follows:

$$(\mu - \hat{\mu})^T \hat{\Sigma}^{-1} (\mu - \hat{\mu}) \leq \chi^2_{(p)} \quad (5)$$

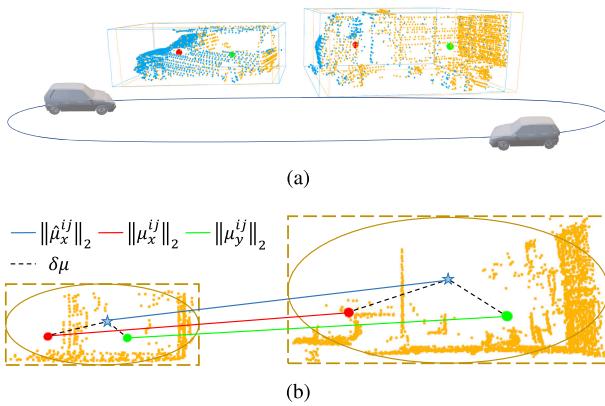


Fig. 4. (a) The blue and orange point clouds are from a pair of loop closure frames. The figure shows two clusters and their centers (the red and green spheres represent the centers of the blue and yellow clusters, respectively) with 3D OBBs. (b) The red and green spheres are defined the same as in (a), and the blue pentagons are the ground truth centers of the two clusters. The deviations of the observed centers from the ground truth are constrained within OBEs tangent to the OBBs. In addition, the translation and rotation invariant measurements (TRIMs) $\|\mu_x^{ij}\|_2$ and $\|\mu_y^{ij}\|_2$ and their ground truths $\|\hat{\mu}_x^{ij}\|_2$ are denoted by the red, green and blue lines, respectively.

where $\chi_{(p)}^2$ is the 3-DoF χ^2 value with a probability $(1 - p)$ to reject the possible center outside of the ellipsoid. Based on the estimated size s and orientation \mathbf{R}_o of OBE, we obtain

$$\hat{\lambda}_i = \left(\frac{s_i}{2\sqrt{\chi_{(p)}^2}} \right)^2 \quad (6)$$

$$\hat{\Sigma} = \mathbf{R}_o \text{diag}(\hat{\lambda}) \mathbf{R}_o^T \quad (7)$$

We set $\chi_{(p)}^2 = \chi_{(0.05)}^2 = 7.815$ to ensure a 95% probability that the ground truth center converges. The derived pseudo-gaussian parameters offer an account of the uncertainty in the segment's center and will be utilized in the compatibility test in Section V-B. Conversely, the statistical Gaussian parameters, which capture the geometric features of the segments, will be used in Section IV-C to build putative correspondences and Section VI-A to estimate the transformation.

C. Soft Association

Starting with a collection of GEMs extracted from the point cloud, we initially select the top-J GEMs from each type of PCL segment. This selection process is based on respective measures of area, volume, and length, as larger segments generally demonstrate superior stability. Following this, we employ a mutual-K-nearest neighbor (MKNN) matching strategy for GEMs that share labels, generating the initial set of putative correspondences represented as \mathcal{I} . This MKNN strategy aids in enhancing the likelihood of including true correspondences. The distance between a pair of GEMs (x, y) utilized in the MKNN is determined by the Wasserstein distance of their statistical covariance matrices,

$$\begin{aligned} \mu'_x &= \mathbf{R}^* \mu_x + \mathbf{t}^* \\ \Sigma'_x &= \mathbf{R}^* \Sigma_x \mathbf{R}^{*T} \\ W(f(x | \mathbf{R}^*, \mathbf{t}^*), y) &= \|\mu'_x - \mu_y\|_2^2 \end{aligned} \quad (8)$$

$$+ \text{Tr} \left(\Sigma'_x + \Sigma_y - 2 \left((\Sigma'_x)^{\frac{1}{2}} \Sigma_y (\Sigma'_x)^{\frac{1}{2}} \right)^{\frac{1}{2}} \right) \quad (9)$$

In an ideal scenario, the optimal rigid transformation $(\mathbf{R}^*, \mathbf{t}^*)$ is ascertained via Equation (2). However, because it is infeasible during the data association step, we instead derive its substitute $(R_j^o R_i^{oT}, \mu_y - R_j^o R_i^{oT} \mu_x)$ by aligning the two GEMs.

Our proposed MKNN-based association method not only eliminates the need to devise complex descriptors—which are typically employed for one-to-one matching—but also avoids the quadratic growth associated with all-to-all matching. The effectiveness of the Wasserstein distance is validated through comparisons with other metrics in Section VII-B.

V. GRAPH-THEORETIC OUTLIER PRUNING

In this section, we introduce an outlier pruning method based on a pyramid compatibility graph. This approach leverages the compatibility graph (Section V-A), which captures the inter-compatibility among putative correspondences. We assume that inliers will form the maximum clique within this graph. Our framework, operating on a distrust-and-verify scheme, utilizes a multi-threshold compatibility test (Section V-B) to construct a pyramid compatibility graph. Subsequently, a graduated maximum clique solver is introduced (Section V-C) to identify the maximum cliques (MACs) at each level and construct multiple inlier sets. These sets are then employed to generate an array of transformation candidates (Section VI-A).

A. Compatibility Graph Construction

Given a set of GEM correspondences $\mathcal{I} = \{(x_k, y_k)\}$, the centers (μ_x^k, μ_y^k) of each GEM correspondence follows the following model,

$$\begin{aligned} \mu_y^k &= \mathbf{R}^* \mu_x^k + \mathbf{t}^* + \mathbf{o}_k + \epsilon_k \\ \mu_x^k &\sim \mathcal{N}(\hat{\mu}_x^k, \hat{\Sigma}_x^k), \mu_y^k \sim \mathcal{N}(\hat{\mu}_y^k, \hat{\Sigma}_y^k) \\ \epsilon_k &\sim \mathcal{N}(0, \mathbf{R}^* \hat{\Sigma}_x^k \mathbf{R}^{*T} + \hat{\Sigma}_y^k) \end{aligned} \quad (10)$$

where μ , $\hat{\mu}$ and $\hat{\Sigma}$ represent the center, true center and pseudo covariance matrix of the GEM, respectively. For inlier correspondences, \mathbf{o}_k is a zero vector, while for outlier correspondences, it is an arbitrary vector. ϵ_k captures the uncertainty of the GEM center resulting from viewpoint changes and occlusion.

For any pair of correspondences (x_i, y_i) and (x_j, y_j) , we can construct a pairwise invariant using two translation and rotation invariant measurements (TRIMs [13]) for the compatibility test. TRIMs are defined as $\|\mu_x^i - \mu_x^j\|_2$ and $\|\mu_y^i - \mu_y^j\|_2$, exploiting the length-preserving property of rigid transformation, as shown in Figure 4. The two correspondences pass the compatibility test if their pairwise invariant satisfies the condition:

$$\left| \|\mu_x^i - \mu_x^j\|_2 - \|\mu_y^i - \mu_y^j\|_2 \right| < \delta_{ij} \quad (11)$$

where δ_{ij} is a threshold used to reject potential outliers in this correspondence pair. It is determined based on the pseudo

covariance of the four involved GEMs and will be analyzed in detail in Section V-B.

Finally, we construct a compatibility graph by representing each correspondence as a vertex and connecting an edge between two vertices if they pass the compatibility test.

Algorithm 1 Compatibility Graph Construction

Params: Pre-defined $p_1 \geq p_2 \geq \dots \geq p_M$

Input : Initial putative GEM correspondences

$$\mathcal{I} = \{(x_k, y_k) | k = 1 : N\}$$

Output : Pyramid graph $\{\mathcal{G}_m | m = 1 : M\}$

1 Function ConstructGraphs (\mathcal{I})

2 % parallel running

3 **for** (i, j) in $\binom{N}{2}$ **do**

$$4 \quad d_{ij} = \|\mu_x^i - \mu_x^j\|_2 - \|\mu_y^i - \mu_y^j\|_2; \\ 5 \quad \lambda_{x,1}^{ij} = \text{UpperEigenvalue}(\hat{\Sigma}_x^i, \hat{\Sigma}_x^j); \\ 6 \quad \lambda_{y,1}^{ij} = \text{UpperEigenvalue}(\hat{\Sigma}_y^i, \hat{\Sigma}_y^j);$$

7 **for** $m = 1 : M$ **do**

$$8 \quad \delta_{ij} = \sqrt{\chi_{(p_m)}^2 \lambda_{x,1}^{ij}} + \sqrt{\chi_{(p_m)}^2 \lambda_{y,1}^{ij}}; \\ 9 \quad \text{if } d_{ij} \leq \delta_{ij} \text{ then}$$

10 **for** $k = m : M$ **do**

11 | $\mathcal{G}_k.\text{addEdge}(i, j);$

12 **end**

13 **break;**

14 **end**

15 **end**

16 **end**

17 **End Function**

18 **Function** UpperEigenvalue (Σ_1, Σ_2)

19 $\Sigma = \Sigma_1 + \Sigma_2;$

20 %Perron Frobenius theorem

$$21 \quad \text{ub}_1 = \max\left(\sum_{i=1}^3 |\Sigma_{i,j}|\right) \quad \forall j \in \{1, 2, 3\}; \\ 22 \quad \%Wolkowicz's method$$

$$23 \quad m \leftarrow \frac{\text{trace}(\Sigma)}{3}; s^2 \leftarrow \frac{\text{trace}(\Sigma^2)}{3} - m^2;$$

$$24 \quad \text{ub}_2 = m + \sqrt{2s^2};$$

25 %Weyl's inequality

$$26 \quad \text{ub}_3 = \lambda_1(\Sigma_1) + \lambda_1(\Sigma_2);$$

27 **return** ($\min(\text{ub}_1, \text{ub}_2, \text{ub}_3)$)

28 **End Function**

B. Multi-Threshold Compatibility Test

In this section, we delve into the methodology for selecting the appropriate δ_{ij} , an essential parameter that helps ascertain whether a pair of correspondences meet the criteria of the compatibility test. As previously detailed in Section IV-B, we have utilized pseudo Gaussian distribution parameters to effectively model the uncertainty inherent in the center μ . This leads us to the following relationships:

$$\begin{aligned} \mu_x^i &\sim \mathcal{N}(\hat{\mu}_x^i, \hat{\Sigma}_x^i), \mu_x^j \sim \mathcal{N}(\hat{\mu}_x^j, \hat{\Sigma}_x^j) \\ \mu_y^i &\sim \mathcal{N}(\hat{\mu}_y^i, \hat{\Sigma}_y^i), \mu_y^j \sim \mathcal{N}(\hat{\mu}_y^j, \hat{\Sigma}_y^j) \\ \mu_x^i - \mu_x^j &\sim \mathcal{N}(\hat{\mu}_x^i - \hat{\mu}_x^j, \hat{\Sigma}_x^i + \hat{\Sigma}_x^j) = \mathcal{N}(\hat{\mu}_x^{ij}, \hat{\Sigma}_x^{ij}) \\ \mu_y^i - \mu_y^j &\sim \mathcal{N}(\hat{\mu}_y^i - \hat{\mu}_y^j, \hat{\Sigma}_y^i + \hat{\Sigma}_y^j) = \mathcal{N}(\hat{\mu}_y^{ij}, \hat{\Sigma}_y^{ij}) \end{aligned} \quad (12)$$

Let $\mu_x^{ij} = \mu_x^i - \mu_x^j = \hat{\mu}_x^{ij} + \epsilon_x^{ij}$, where $\epsilon_x^{ij} \sim \mathcal{N}(0, \hat{\Sigma}_x^{ij})$, and $\mu_y^{ij} = \mu_y^i - \mu_y^j = \hat{\mu}_y^{ij} + \epsilon_y^{ij}$, where $\epsilon_y^{ij} \sim \mathcal{N}(0, \hat{\Sigma}_y^{ij})$. We consider the two correspondences be both inliers, thus $\hat{\mu}_y^{ij} = \mathbf{R}^* \hat{\mu}_x^{ij}$. According to the triangle inequality, we have

$$\|\mu_x^{ij}\|_2 \leq \|\hat{\mu}_x^{ij}\|_2 + \|\epsilon_x^{ij}\|_2 \quad (13)$$

in which ϵ_x^{ij} characterizes the joint uncertainty between the GEMs μ_x^i and μ_x^j . This joint uncertainty can also be similarly confined within its corresponding uncertainty ellipsoid. To simplify matters and maintain generality, we adopt the probability p as a means to model this uncertainty, as detailed below:

$$(\epsilon_x^{ij})^T (\hat{\Sigma}_x^{ij})^{-1} \epsilon_x^{ij} \leq \chi_{(p)}^2 \quad (14)$$

where $\hat{\Sigma}_x^{ij}$ is a symmetric and positive-definite matrix that can be expressed in a diagonalized form, as illustrated below:

$$\hat{\Sigma}_x^{ij} = U \begin{bmatrix} \lambda_{x,1}^{ij} & 0 & 0 \\ 0 & \lambda_{x,2}^{ij} & 0 \\ 0 & 0 & \lambda_{x,3}^{ij} \end{bmatrix} U^T \quad (15)$$

where U is an orthogonal matrix and $\lambda_{x,1}^{ij} \geq \lambda_{x,2}^{ij} \geq \lambda_{x,3}^{ij}$. Furthermore, we can get a lower bound on $(\epsilon_x^{ij})^T (\hat{\Sigma}_x^{ij})^{-1} \epsilon_x^{ij}$,

$$\begin{aligned} &(\epsilon_x^{ij})^T (\hat{\Sigma}_x^{ij})^{-1} \epsilon_x^{ij} \\ &= (U^T \epsilon_x^{ij})^T \begin{bmatrix} \lambda_{x,1}^{ij} & 0 & 0 \\ 0 & \lambda_{x,2}^{ij} & 0 \\ 0 & 0 & \lambda_{x,3}^{ij} \end{bmatrix}^{-1} (U^T \epsilon_x^{ij}) \\ &= \sum_k^3 \frac{1}{\lambda_{x,k}^{ij}} (U^T \epsilon_x^{ij})_k^2 \geq \frac{1}{\lambda_{x,1}^{ij}} \sum_k^3 (U^T \epsilon_x^{ij})_k^2 \\ &= \frac{1}{\lambda_{x,1}^{ij}} (\epsilon_x^{ij})^T \epsilon_x^{ij} = \frac{1}{\lambda_{x,1}^{ij}} \|\epsilon_x^{ij}\|_2^2 \end{aligned} \quad (16)$$

where $\lambda_{x,1}^{ij}$ is the largest eigenvalue of $\hat{\Sigma}_x^{ij}$. By consolidating equations (13), (14) and (16), we can derive an upper bound for $\|\mu_x^{ij}\|_2$, as follows:

$$\begin{aligned} \|\mu_x^{ij}\|_2 &\leq \|\hat{\mu}_x^{ij}\|_2 + \|\epsilon_x^{ij}\|_2 \\ &\leq \|\hat{\mu}_x^{ij}\|_2 + \sqrt{\chi_{(p)}^2 \lambda_{x,1}^{ij}} \end{aligned} \quad (17)$$

Following a similar derivation, we can establish the upper bound of $\|\mu_y^{ij}\|_2$ as follows,

$$\begin{aligned} \|\mu_y^{ij}\|_2 &\leq \|\hat{\mu}_y^{ij}\|_2 + \|\epsilon_y^{ij}\|_2 \\ &\leq \|\mathbf{R}^* \hat{\mu}_x^{ij}\|_2 + \sqrt{\chi_{(p)}^2 \lambda_{y,1}^{ij}} \\ &= \|\hat{\mu}_x^{ij}\|_2 + \sqrt{\chi_{(p)}^2 \lambda_{y,1}^{ij}} \end{aligned} \quad (18)$$

Finally, by integrating equations (17) and (18), we can determine δ_{ij} by assigning a confidence probability p (e.g., 90%) to accept the observation. It is noteworthy that a larger value of p implies a reduction in uncertainty.

$$|\|\mu_x^{ij}\|_2 - \|\mu_y^{ij}\|_2| \leq \sqrt{\chi_{(p)}^2 \lambda_{x,1}^{ij}} + \sqrt{\chi_{(p)}^2 \lambda_{y,1}^{ij}} = \delta_{ij} \quad (19)$$

The comprehensive process of constructing the pyramid graph using a multi-threshold compatibility test is encapsulated in Algorithm 2. For N correspondences, it is necessary to compute the TRIMs and their upper bounds for $N(N - 1)/2$ pairs, as indicated in lines 4–6 of Algorithm 2. In the worst case, the time complexity reaches $O(M \cdot N^2)$, accounting for the majority of the computational effort in the PAGOR. To reduce the computational burden associated with eigenvalue calculations, we employ the Perron-Frobenius theorem [47], [48], Wolkowicz's method [49], and Weyl's inequality [50]. These techniques aid in computing the upper bound of the largest eigenvalue λ_1 , as presented in line 18, thereby accelerating the computation process.

C. Graduated MAC Solver

Given M p -values $p_1 \geq p_2 \geq \dots \geq p_M$, we correspondingly have χ^2 -values $\chi_{(p_1)}^2 \leq \chi_{(p_2)}^2 \leq \dots \leq \chi_{(p_M)}^2$. From these, we derive M compatibility graphs \mathcal{G} along with their associated maximum cliques \mathcal{G}^{MC} . Theoretically, as the value of $\chi_{(p)}^2$ increases, δ_{ij} follows the change, which results in more pairwise correspondences passing the compatibility test, thereby leading to denser compatibility graphs \mathcal{G} . We arrange all the consistency graphs \mathcal{G} in the order of their sparsity, placing the sparsest at the top and the densest at the bottom. Our objective is to compute these M maximum cliques \mathcal{G}^{MC} , where their vertices \mathcal{I}_m^* are potential inlier sets of the initial correspondences \mathcal{I} . At each level of the pyramid graph, our goal is to solve the following optimization problem:

$$\begin{aligned} & \text{maximize}_{\mathcal{I}_m^* \subseteq \{1, \dots, N\}} |\mathcal{I}_m^*| \\ & \text{s.t. } \left| \|\mu_x^i - \mu_x^j\|_2 - \|\mu_y^i - \mu_y^j\|_2 \right| < \delta_{ij}^m, \quad \forall i, j \in \mathcal{I}_m^* \end{aligned} \quad (20)$$

Proposition 1 (Lower Bound of Clique Cardinality):

Given $\chi_{(p_{m+1})}^2 > \chi_{(p_m)}^2$, it follows that $|\mathcal{G}_{m+1}^{MC}| \geq |\mathcal{G}_m^{MC}|$.

Proof: Proceeding by contradiction, let us assume the contrary: $\chi_{(p_{m+1})}^2 > \chi_{(p_m)}^2$, but $|\mathcal{G}_{m+1}^{MC}| < |\mathcal{G}_m^{MC}|$. Let us denote the consistency thresholds corresponding to $\chi_{(p_m)}^2$ and $\chi_{(p_{m+1})}^2$ as δ_{ij}^m and δ_{ij}^{m+1} respectively. Given that $\chi_{(p_{m+1})}^2 > \chi_{(p_m)}^2$, we can infer that $\delta_{ij}^{m+1} > \delta_{ij}^m$ for all i, j .

Consider the maximum cliques at levels m and $m + 1$:

$$\begin{aligned} |\mathcal{G}_m^{MC}| &= |\mathcal{I}_m^*| \\ \text{s.t. } & \left| \|\mu_x^i - \mu_x^j\|_2 - \|\mu_y^i - \mu_y^j\|_2 \right| < \delta_{ij}^m, \quad \forall i, j \in \mathcal{I}_m^* \end{aligned} \quad (21)$$

$$\begin{aligned} |\mathcal{G}_{m+1}^{MC}| &= |\mathcal{I}_{m+1}^*| \\ \text{s.t. } & \left| \|\mu_x^i - \mu_x^j\|_2 - \|\mu_y^i - \mu_y^j\|_2 \right| < \delta_{ij}^{m+1}, \quad \forall i, j \in \mathcal{I}_{m+1}^* \end{aligned} \quad (22)$$

Given that $\delta_{ij}^{m+1} > \delta_{ij}^m$, any feasible set for \mathcal{I}_m^* must also be a feasible set for \mathcal{I}_{m+1}^* . Hence, the cardinality of $|\mathcal{G}_{m+1}^{MC}|$ must be greater than or equal to that of $|\mathcal{G}_m^{MC}|$. This contradicts our earlier assumption that $|\mathcal{G}_{m+1}^{MC}| < |\mathcal{G}_m^{MC}|$, and consequently, our proof is concluded. ■

Proposition 1 suggests that we can leverage the maximum clique deduced from a sparse clique to produce a lower bound on the cardinality of MAC for the denser compatibility graph of the subsequent level. This insight works well with an efficient parallel maximum clique finder algorithm, PMC [51],

enabling us to devise a graduated PMC algorithm that calculates the maximum clique from the top to the bottom of the pyramid compatibility graph. Specifically, the original PMC algorithm is based on a branch-and-bound method and prunes using the core numbers of the vertices. In the graduated PMC, once we have determined the maximum clique from the sparser graph of the preceding layer, we can exclude all nodes in this layer whose degree is less than the cardinality of the clique. This process substantially reduces the search space, as these nodes cannot be included in the final MAC of the current layer, according to Proposition 1.

Although finding the maximum clique typically has exponential computational complexity in the worst-case scenario, this is mitigated by the sparsity of the compatibility graph under high outlier ratios. Combined with parallel computing techniques, this sparsity ensures that the maximum clique problem can be solved efficiently in practice. Moreover, as we demonstrate empirically in Section VII-D, the integration of an additional compatibility graph does not significantly increase computation time when using the graduated MAC solver, due to the prior lower bound provided by the preceding graph.

VI. ROBUST TRANSFORMATION ESTIMATION

In this section, we illustrate how to derive multiple transformation candidates from the M hypothetical inlier sets \mathcal{I}^* utilizing robust estimation methods (Section VI-A). We then examine the selection of the evaluation function g in Eq. 3 from a robust statistics viewpoint, with the goal of identifying the most appropriate transformation from the set of candidates that can optimally align \mathcal{X} and \mathcal{Y} (Section VI-B).

A. Distribution-to-Distribution Registration

Given a hypothetic inlier set $\mathcal{I}^* = (x_k, y_k)$, we reformulate the Equation (2) as a distribution-to-distribution registration problem using the statiscal Gaussian parameters of GEMs. Let $\mu_x^k \in \mathcal{N}(\hat{\mu}_x^k, \Sigma_x^k)$ and $\mu_y^k \in \mathcal{N}(\hat{\mu}_y^k, \Sigma_y^k)$, Equation (2) can be written as,

$$\begin{aligned} \mathbf{R}^*, \mathbf{t}^* &= \arg \min_{\mathbf{R} \in \text{SO}(3), \mathbf{t} \in \mathbb{R}^3} \sum_{(\mu_x^k, \mu_y^k) \in \mathcal{I}^*} \min(r(\mu_y^k, \mu_x^k), \bar{c}^2) \\ r(\mu_y^k, \mu_x^k) &= d_k^T (\Sigma_y^k + \mathbf{R} \Sigma_x^k \mathbf{R}^T)^{-1} d_k \\ d_k &= \mu_y^k - (\mathbf{R} \mu_x^k + \mathbf{t}) \end{aligned} \quad (23)$$

where $d_k \sim \mathcal{N}(\hat{\mu}_y^k - (\mathbf{R} \hat{\mu}_x^k + \mathbf{t}), \Sigma_y^k + \mathbf{R} \Sigma_x^k \mathbf{R}^T)$. $r(\cdot)$ is now defined as the Truncated Least Squares (TLS) cost, with \bar{c} representing the inlier cost threshold which is initialized by $\chi_{(0.01)}^2$. The minimization of the residual function $r(\cdot)$ is equivalent to maximizing the log-likelihood of the new Gaussian model d_k .

To solve Equation (23), we use graduated non-convexity [30] implemented in GTSAM [52] where the non-minimal global solver is replaced by Levenberg-Marquardt optimization for the reason that there is no closed-form solution for distribution-to-distribution registration problem. In addition, for plane segments, the covariance matrix is regularized by replacing its eigenvalue with $(1, 1, 10^{-3})$ to work as a plane-to-plane registration. GNC acts as an alternative optimization

with inner and outer iteration for L-M optimization and weight update, respectively.

The time complexity of the optimization process is denoted as $O(D^2 N \cdot L)$, where $D = 6$ represents the dimension of the optimization variable, N denotes the number of inliers denoted by $|\mathcal{I}^*|$, and L indicates the number of outer iterations. In practical applications, the GEM-based matching algorithm tends to yield a limited number of inliers, which keeps N relatively small. Additionally, the use of maximum clique inlier selection effectively eliminates the majority of outliers, resulting in a robust initial guess. This, in turn, reduces L , the count of outer iterations required for convergence. Consequently, both N and L contribute to a lower practical time complexity, enhancing the overall efficiency of the optimization process, which is demonstrated in Figure 8.

B. Transformation Verification

Given M hypothetical transformations $(\mathbf{R}_m^*, \mathbf{t}_m^*)$, our objective is to construct the evaluation function $g(\cdot)$ to identify the most suitable transformation by resolving Equation (3). Whereas existing approaches endeavor to certify the global optimality [13], [53], [54] of minimizing the objective function, we take a step back to re-think how to design an objective function that accurately reflects the alignment quality of the point clouds. We argue the objective in Equation (23) may be inadequate for two reasons. First, although GEMs recover geometries of \mathcal{X} and \mathcal{Y} maximally, information loss persists since \mathcal{I}^* contains only a subset of all potential correspondences. Second, \mathcal{I}^* unavoidably contains adversarial outliers that the TLS cannot accommodate.

To address this, we leverage the semantic voxel map V from Section IV to verify candidates, as it retains most raw geometric information. Assuming every $(\mathbf{R}_m^*, \mathbf{t}_m^*)$ is optimal, correspondences producing large residuals are considered outliers. Thus, we formulate $g(\cdot)$ based on the Chamfer distance with a new robust kernel $\hat{\rho}(\cdot)$:

$$\begin{aligned} g(\mathbf{R}_m^*, \mathbf{t}_m^* | \mathcal{X}, \mathcal{Y}) &= \frac{1}{|\mathcal{X}|} \sum_{v_x \in V_{\mathcal{X}}} \hat{\rho}(r(v'_x, v'_y)) \\ v'_x &= \mathbf{R}_m^* v_x + \mathbf{t}_m^* \\ v'_y &= \arg \min_{v_y \in V_{\mathcal{Y}}} \|v'_x - v_y\|_2 \end{aligned} \quad (24)$$

where $V_{\mathcal{X}}$ and $V_{\mathcal{Y}}$ represent the voxel maps of \mathcal{X} and \mathcal{Y} .

To expedite the computational process, we compress the raw point clouds by retaining only the sampled five points within each voxel $V_{\mathcal{X}}$ and the centroid of each voxel $V_{\mathcal{Y}}$. For each sampled point v_x in $V_{\mathcal{X}}$, we determine its nearest counterpart v'_y in $V_{\mathcal{Y}}$ after transforming it with the optimal rotation \mathbf{R}_m^* and translation \mathbf{t}_m^* . We then calculate the residual $r(v'_x, v'_y)$ as follows:

$$\begin{cases} |n_y^T(v'_x - v'_y)| & \text{if type}(v'_y) = \text{plane} \\ \left\| (\mathbf{I} - d_y^T d_y)(v'_x - v'_y) \right\|_2 & \text{if type}(v'_y) = \text{line} \\ \|v'_x - v'_y\|_2 & \text{if type}(v'_y) = \text{cluster} \end{cases} \quad (25)$$

In this equation, n_y denotes the normal to the plane, and d_y represents the direction of the line. The nearest neighbor search leverages a KD-Tree structure, where building the

TABLE I
DATASETS FOR PERFORMANCE EVALUATION

	KITTI-10m	KITTI-loop	KITTI-360	Apollo-SouthBay	Self-collected
Num. of Pairs	556	3325	18469	55118	4707
Testing Strategy	Loop Close	Loop Close	Loop Close	Loop Close	Multi Sessions
Sensor Type	Velodyne	Velodyne	Velodyne	Velodyne	Livox

tree requires a time complexity of $O(n \log(n))$ and querying the nearest point has a complexity of $O(m \log(n))$, with m and n being the numbers of sampled points in $V_{\mathcal{X}}$ and $V_{\mathcal{Y}}$, respectively. Furthermore, the verification of $(\mathbf{R}_m^*, \mathbf{t}_m^*)$ will be omitted if it is too close to an already evaluated one.

For the selection of the evaluation function $\hat{\rho}(\cdot)$, we employ Dynamic Covariance Scaling (DCS) [55] and compare its performance to other robust kernels, including Tukey, Cauchy, Huber [56], and TLS. This comparative analysis is presented in Section VII-D.

VII. EXPERIMENTS

This section conducts a comprehensive comparative analysis between our proposed method and existing SOTA techniques, employing various datasets and sensor types. Our method achieves better performance and delivers sub-optimal results with a diminished computational overhead. Additionally, the integration of our novel sub-modules, namely GEM and PAGOR, into any registration approach demonstrates a substantial enhancement in performance. Finally, a review of selected failure cases is also presented to provide practitioners with an in-depth understanding of our methodology.

A. Experimental Setup

1) *Dataset*: As presented in Table I, we conduct experiments on three public and one self-collected multi-session dataset, namely Campus-MS. The KITTI-10m registration set, widely utilized for the evaluation of LiDAR-based global registration techniques [17], [57], [58], is formed by sampling consecutive frames at 10-meter intervals from sequences 08, 09, and 10 of the KITTI odometry dataset. However, this dataset's concentrated distribution of relative poses between the two point cloud frames may not comprehensively assess the true performance of global registration.

To address this constraint, we adopt a strategy of sampling registration pairs from revisited locations, a testing approach commonly utilized in re-localization and loop closing scenarios. This approach yields a broader distribution of poses and more pronounced disparities in point cloud data as shown in Figure 5. The translation-based down-sampling is also employed to decrease the test set size, while ICP is utilized to refine the ground truth.

The strategy outlined above is implemented to process three publicly accessible datasets, including KITTI-LC [59], KITTI360-LC [60], and Apollo-LC [61] (LC means loop closing). The datasets cover a wide range of scenarios pertinent to autonomous navigation, including urban centers, highways, overpasses, and suburban areas. We have ensured the public availability of the code used to generate this benchmark

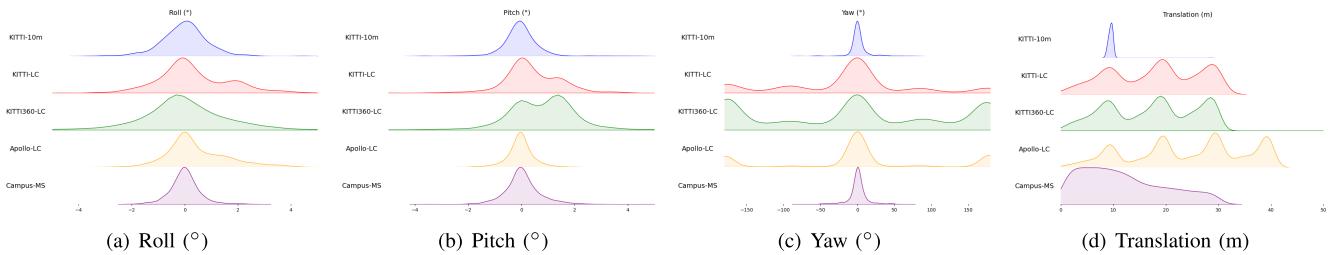


Fig. 5. The distributions of ground truth rotations and translations in different datasets. This figure shows broader distributions that can effectively validate global registration performances.

TABLE II
ALL PARAMETERS OF OUR APPROACH

Parameters			Value
Plane	Extraction	Voxel size s_v λ_2/λ_3 threshold σ_p	1 m 30
	Merge	Normal threshold σ_n Distance threshold σ_{p2n}	0.95 0.2
Line extraction		Distance threshold σ_{p2l} Inlier ratio threshold σ_{line}	0.5 m 0.5
Association		Segment number J K in MKNN strategy	50 20
MAC solver		p-values	[0.99, 0.95, 0.9, 0.8]
Geometric verification		Robust kernel function	DCS

dataset, with the intention of benefiting the wider community.² The self-collected Campus-MS dataset will be introduced in detail in Section VII-F.

2) *Comparisons*: Concerning the front end, we employ two approaches: the handcrafted descriptor FPFH [9] and the deep learning-based descriptor FCGF [62]. As for the back end, our selection encompasses SOTA methods based on different frameworks: (1) graph-theoretic methods: TEASER++ [13], Quattro [16], and 3DMAC [17]; (2) deep learning-based methods: DGR [57] and PointDSC [58]; (3) enhanced RANSAC; (4) advanced loop closing methods that integrate global registration, like STD [63], Cont2 [64].

3) *Evaluation Metric*: To conduct a thorough evaluation of performance across different levels of difficulty and variations in point cloud distribution, we categorize the test set into three subsets. These subsets are defined by the translational distance between pairs: [0, 10] for easy scenarios, [10, 20] for medium complexity, and [20, 30] for hard cases. These intervals are particularly relevant for robotic applications such as loop closure [2], [65].

For feature matching evaluation in Section VII-B, the inlier ratio (IR), total number of correspondences, correspondence recall, and time consumption are used. A correspondence (x_i, y_i) is classified as an inlier if it adheres to the following condition:

$$\|\mathbf{R}_{gt}x_i + \mathbf{t}_{gt} - y_i\|_2 < 0.5 \quad (26)$$

where $(\mathbf{R}_{gt}, \mathbf{t}_{gt})$ denote the ground truth rotation and translation, respectively. If the count of inliers exceeds 3, that pair is considered successfully recalled. Subsequently, we compute the recall rate for all pairs in the test set.

²<https://github.com/HKUST-Aerial-Robotics/LiDAR-Registration-Benchmark>

For the global registration evaluation in Section VII-C, we use registration recall for evaluation. Given the estimated transformation $(\mathbf{R}_{est}, \mathbf{t}_{est})$, the registration is classified as successfully recalled if

$$\arccos\left(\frac{\text{tr}(\mathbf{R}_{est}^T \mathbf{R}_{gt}) - 1}{2}\right) < 5 \text{ deg} \\ \|\mathbf{R}_{est}^T (\mathbf{t}_{gt} - \mathbf{t}_{est})\|_2 < 2 \text{ m} \quad (27)$$

The rotation and translation thresholds of 5 degrees and 2 meters, respectively, are consistent with the configurations used by other state-of-the-art methods [1], [17], [57]. This metric setting is aligned with the fact that global registration methods typically prioritize robustness, while local registration methods like iterative closest point (ICP) [66] and normal distribution transformation (NDT) [67], emphasize the accuracy of point cloud registration.

4) *Implementation Details*: All the experiments are evaluated on an Intel i7-13700K CPU and an Nvidia RTX4080 graphics card. All parameters in GEM and PAGOR are detailed in Table II. Additionally, we maintain uniformity in parameter across all experiments, even when the LiDAR type and scenarios vary. For ground removal and clustering, we adopt the default parameters as outlined in their respective original papers [45], [68].

We optimized FPFH-based matching for both efficiency and effectiveness by performing ground removal, voxel down-sampling at a 0.5m resolution, and adjusting the search radii for normals and FPFH to 1m and 2.5m, respectively. For FCGF matching, we used pre-trained models from the KITTI dataset's open-source code. Correspondences of both FPFH and FCGF are filtered by the mutual nearest neighbors (MNN) strategy. In graph-theoretical methods, we fine-tuned key parameters to optimize performance: noise bounds are set at 0.4 m for TEASER++ and Quattro, and the compatibility threshold is 0.9 for 3DMAC. For RANSAC, we use a default of 1,000,000 iterations and employ TRIMs with a high threshold to efficiently filter out incorrect correspondences. Multi-processing acceleration is used for TEASER++, Quattro, 3DMAC and RANSAC.

B. Feature Matching Evaluation

We assess the quality of putative correspondences from our GEM-based front end against FPFH [9] and FCGF [62], using the KITTI-LC dataset. Our evaluation explores various aspects of our method, considering segment types, matching strategies, and clustering algorithms as follows:

TABLE III
FEATURE MATCHING EVALUATION ON KITTI-LC DATASET

	Method	0-10m				10-20m				20-30m			
		IR ¹ (%)↑	Num	Recall (%)↑	Time (ms)↓	IR (%)↑	Num	Recall (%)↑	Time (ms)↓	IR (%)↑	Num	Recall (%)↑	Time (ms)↓
Point-based	FPFH	9.198	642	99.89	170.69	2.575	590	94.52	177.76	0.923	557	64.44	175.16
	FCGF	29.08	2762	96.72	96.21	13.12	2586	87.84	86.52	5.678	2403	71.59	84.12
Segment-based	Plane only	1.109	450	73.08	21.69	0.580	447	42.83	21.87	0.337	441	18.73	22.11
	Line only	2.678	180	56.45	20.65	1.454	185	34.23	21.14	0.741	175	17.38	21.25
	Cluster only	1.677	713	98.03	22.76	0.744	714	81.49	22.77	0.356	712	43.65	23.87
	Ours (Random)	0.618	1737	95.18	24.43	0.342	1753	80.19	24.57	0.203	1738	53.25	25.02
	Ours (EigenVal)	0.418	1379	81.62	24.97	0.270	1381	61.16	24.83	0.175	1361	36.19	25.47
	Ours (FPFH+T)	1.330	1196	98.14	45.38	0.617	1203	89.23	42.58	0.323	1189	60.24	43.34
	Ours (All-to-all)	0.692	3598	99.02	24.41	0.360	3626	96.35	24.50	0.199	3581	85.08	25.77
	Ours (Wass+D)	1.376	1292	96.94	25.12	0.687	1291	92.61	25.10	0.376	1273	71.75	25.27
	Ours (Wass+T)	1.526	1336	98.69	25.02	0.737	1339	94.53	24.80	0.391	1320	74.52	25.17

¹Inlier ratio.

- Use of individual segment types (plane, cluster, line).
- Replacement of the covariance matrix's Wasserstein distance with alternative matchings, such as IoU3D, the eigenvalue-based descriptor [69] (EigenVal), matching randomly (Random), and enumerating all possible correspondences (All-to-all). Furthermore, we try to follow FPFH to design a GEM descriptor for encoding the local information of the neighborhood (20m) using GEM's centers and normals (FPFH+T).
- Use of different clustering methods: DCVC [68] (Wass+D) and TRAVEL [45] (Wass+T).

Table III summarizes our findings. Segment-based matching tends to produce more outliers than point-based due to its less robust descriptors and difficulties in estimating segment centers. However, our PAGOR algorithm mitigates this issue, effectively handling high outlier ratios, which indicates that the recall rate is more important than inlier ratio. Additionally, as the translation increases, the segment-based method exhibits superior recall compared to the descriptor-based approach like FPFH and FCGF, since the change of neighborhood damage the descriptor performance as discussed in Section II-A2. For efficiency, we employ the MKNN strategy over the all-to-all strategy, which balances performance with computational practicality. Table III also presents the contribution of any individual geometry type. Moreover, Wasserstein distance-based matching proves to be the most effective, aided by our MKNN strategy. Finally, TRAVEL clustering slightly outperforms DCVC, but the latter's fewer parameters may be preferable depending on the application context. We use DCVC for solid-state LiDAR and TRAVEL for mechanical LiDAR.

C. Registration Evaluation

1) *KITTI-10m Dataset*: Although the KITTI-10m dataset has limitations in registration performance evaluation, we include it in Table IV for consistency with prior studies. We use the more accurate Semantic KITTI [70] ground truth for benchmarking, addressing inaccuracies in the official KITTI ground truth. Table IV shows that deep learning-based methods using new ground truth pose perform comparably to their original reported results. By adjusting FPFH and TEASER++ parameters, we found traditional methods outperform those in earlier research [17]. It should be noted that a discrepancy in 3DMAC performance arises when using MNN for correspondence filtering, resulting in fewer corre-

TABLE IV
GLOBAL REGISTRATION EVALUATION ON KITTI-10M DATASET

Front-end	Back-end	Recall (%) on KITTI-10m ↑				Average Time (ms)
		08	09	10	Total	
FCGF	DGR	96.74	99.38	97.70	97.66	573.34
FCGF	PointDSC	98.37	99.38	98.85	98.74	333.78
FPFH	RANSAC	97.39	100.0	95.40	97.84	264.31
FPFH	TEASER	98.04	98.76	93.10	97.48	236.35
FPFH	3DMAC	95.44	98.15	90.80	95.50	258.29
FPFH	PAGOR	99.67	100.0	100.0	99.82	282.38
GEM	RANSAC	83.39	86.42	70.11	82.19	64.68
GEM	TEASER	99.35	99.38	94.25	98.56	45.55
GEM	3DMAC	99.02	100.0	93.10	98.38	55.05
GEM	PAGOR	99.34	100.0	98.85	99.46	61.88

spondences (5000 in the original paper [17]) and decreased results.

2) *Results on Loop Closure Datasets*: In this study, we assess our registration method on loop closure point clouds with distances up to 30 ~ 40 meters, meeting most robotics application requirements. Results on KITTI-LC, KITT360-LC, and Apollo-LC datasets are detailed in Tables V, VI, and VII, respectively. Methods are categorized into deep learning-based, RANSAC-based, and graph-theoretical at the backend, with our GEM and PAGOR performances evaluated when integrated with these methods. We also benchmark against state-of-the-art loop closure techniques capable of global registration. Additional GEM matching and registration results are visualized in Figure 6.

Deep learning methods show a performance drop in loop closure scenarios, as Tables V and VI depict. This is attributed to the challenging nature of loop closures with varied viewpoints and a training set on consecutive frames with limited scale and pose diversity. RANSAC-based methods perform well with low outlier ratios when using high-quality descriptors like FPFH within a 10m translation. Yet, their performance declines with larger translations, or when using GEM with high outlier ratios, as the computational complexity of RANSAC increases exponentially with the outlier ratio [71]. Our PAGOR outshines TEASER++ and Quattro in graph-theoretical methods, benefiting from our distrust-and-verify framework, with only a marginal 20-millisecond time increase. Section VII-D explores this performance-efficiency trade-off in depth.

3DMAC shares our hypothesis generation and evaluation philosophy. Admittedly, the clique consisting of inliers must

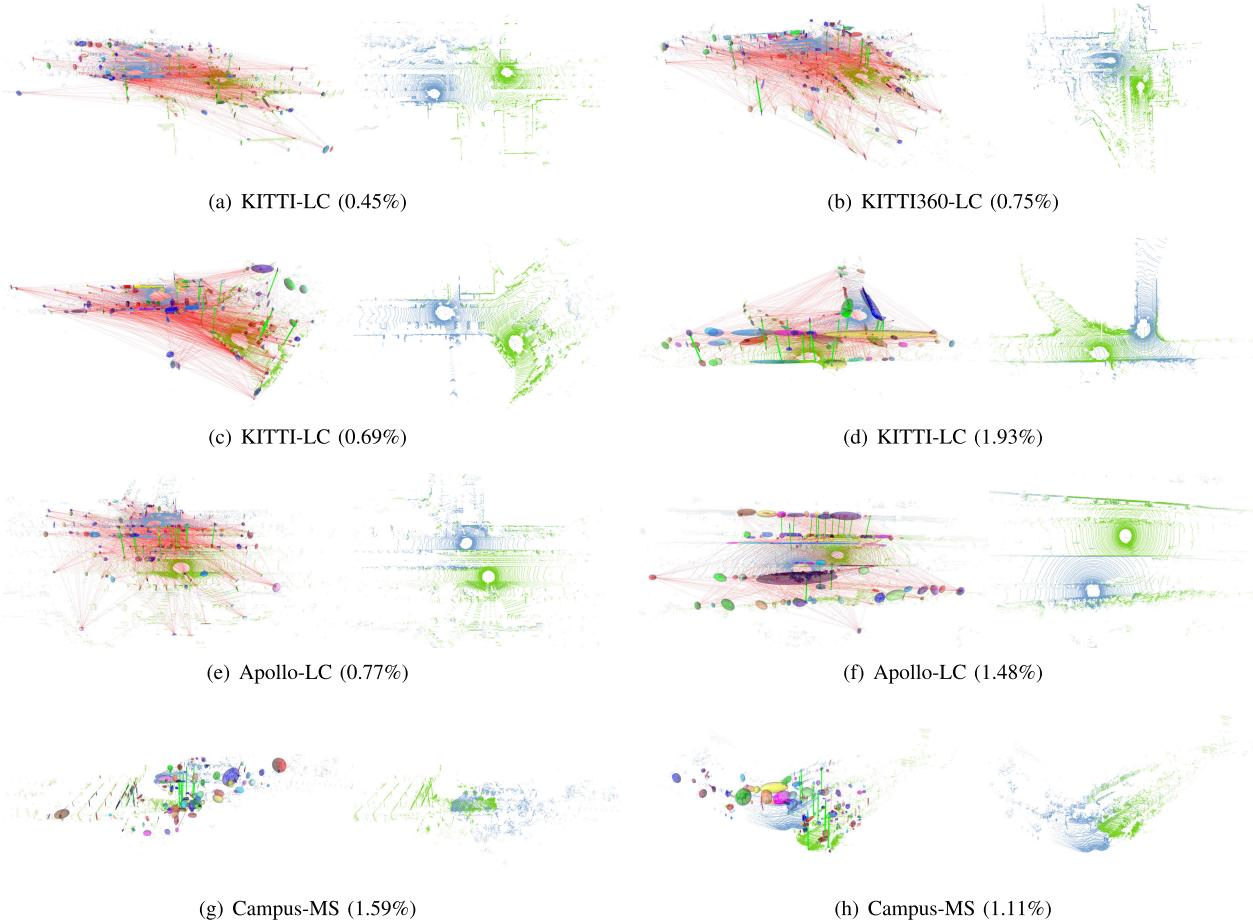


Fig. 6. Challenging cases for global registration where G3Reg could achieve successful alignment even when point clouds exhibited low inlier ratios (shown in brackets). (a)-(f) were tested with Velodyne data and (g) and (h) show registration results using Livox.

TABLE V
GLOBAL REGISTRATION EVALUATION ON KITTI-LC DATASET

	Front-end	Back-end	Recall (%) on KITTI-LC ↑												Average Time (ms)			
			0-10m					10-20m					20-30m					
			00	02	05	06	08	00	02	05	06	08	00	02	05	06	08	
Deep Learning-based Method	FCGF	DGR	88.12	65.12	80.11	98.45	2.24	64.81	54.04	66.83	48.21	0.62	41.47	35.43	45.21	27.38	0.00	546.85
	FCGF	PointDSC	82.18	54.07	75.57	97.67	2.99	61.58	46.46	62.81	47.81	0.00	44.88	39.46	47.49	35.32	0.00	394.17
RANSAC-based Method	FPPH	RANSAC	99.67	94.76	99.43	99.22	100.0	73.90	67.67	77.89	95.62	91.97	25.46	29.59	32.87	78.17	58.92	230.54
	GEM	RANSAC	87.46	77.91	86.36	96.90	88.06	35.78	43.94	40.70	60.56	56.79	4.72	14.80	10.96	24.60	29.19	60.91
SOTA Loop Closure Method	STD	RANSAC	38.61	36.63	31.25	73.64	35.82	0.59	1.51	0.50	19.12	4.93	0.00	0.00	0.00	4.76	0.54	46.70
	Cont2	GMM	91.75	80.81	96.02	96.90	92.54	53.66	40.40	62.81	68.92	61.73	13.38	8.07	19.63	38.89	29.19	5.15
Graph-theoretic Method	FPPH	TEASER++	99.01	97.67	98.86	100.0	100.0	78.59	69.70	83.42	99.60	95.68	30.71	30.04	37.44	87.70	56.76	207.34
	FPPH	Quattro	98.68	99.42	100.0	100.0	99.25	89.74	75.25	90.95	99.20	94.44	44.62	33.91	50.68	92.06	61.62	207.24
	FPPH	3DMAC	99.01	93.60	98.30	99.22	99.25	66.57	50.51	65.83	88.84	84.57	14.70	19.73	21.92	63.89	41.08	203.51
	FPPH	PAGOR	100.0	100.0	100.0	100.0	100.0	92.67	86.36	90.95	100.0	99.38	46.46	43.95	53.42	96.03	73.51	283.72
	GEM	TEASER++	100.0	86.05	100.0	100.0	100.0	92.67	60.10	92.46	100.0	96.91	44.09	44.39	61.64	97.22	67.03	43.67
	GEM	Quattro	99.01	87.21	100.0	100.0	100.0	93.84	62.12	95.48	99.20	96.91	47.24	46.19	65.30	95.63	70.81	42.25
	GEM	3DMAC	100.0	74.42	100.0	100.0	98.51	86.51	56.57	87.44	99.20	88.27	38.58	34.53	48.40	91.67	60.54	60.10
	GEM	PAGOR	100.0	91.28	100.0	100.0	100.0	96.77	70.20	98.49	100.0	97.53	62.47	52.02	74.43	99.60	78.38	64.99

TABLE VI
GLOBAL REGISTRATION EVALUATION ON KITTI360-LC DATASET

	Front-end	Back-end	Recall (%) on KITTI360-LC ↑												Average Time (ms)						
			0-10m					10-20m					20-30m								
			00	02	04	05	06	09	00	02	04	05	06	09	00	02	04	05	06	09	
Deep Learning-based Method	FCGF	DGR	40.05	26.32	10.93	16.28	43.43	55.21	24.70	19.19	9.37	14.05	32.55	42.28	16.83	14.23	8.10	10.68	19.54	22.24	437.15
	FCGF	PointDSC	40.58	24.56	9.90	14.95	40.83	54.37	23.67	16.91	9.10	13.14	29.04	39.63	16.91	13.17	7.86	10.96	19.01	25.92	335.86
RANSAC-based Method	FPPH	RANSAC	98.82	98.25	98.08	93.19	97.28	99.23	85.03	84.75	73.09	70.54	65.40	75.86	48.51	43.77	30.12	35.92	30.91	28.58	203.58
	GEM	RANSAC	83.90	84.21	82.13	73.26	75.38	83.80	41.16	40.97	32.98	22.36	27.38	31.71	12.79	12.37	9.17	5.55	6.39	7.00	66.38
SOTA Loop Closure Method	STD	RANSAC	40.58	41.35	26.74	26.74	27.22	31.17	4.58	4.46	1.05	1.06	0.39	1.41	0.74	0.18	0.24	0.00	0.00	0.05	48.56
	Cont2	GMM	86.78	81.45	78.43	84.88	85.56	91.71	55.38	37.03	42.48	40.18	36.35	50.31	24.67	8.18	19.05	13.04	9.95	16.82	12.44
Graph-theoretic Method	FPPH	TEASER++	99.34	98.12	97.78	93.68	97.04	99.29	87.18	84.23	74.80	72.66	67.64	79.82	55.12	48.49	34.64	35.37	30.11	34.15	184.7
	FPPH	Quattro	99.61	97.99	98.37	97.51	96.45	99.81	91.86	84.65	83.51	83.53	74.95	88.97	61.63	54.63	45.48	50.62	42.63	45.24	184.62
	FPPH	3DMAC	98.17	98.24	96.31	89.70	95.38	98.32	76.80	73.03	59.23	58.00	54.29	64.84	38.20	29.80	20.71	23.72	19.80	19.17	183.31
	FPPH	PAGOR	100.0	100.0	99.70	97.67	99.05	99.93	94.11	93.77	91.82	88.06	84.50	91.91	68.56	65.48	53.21	53.67	49.20	49.64	212.02
	GEM	TEASER++	99.74	98.12	97.64	96.51	96.92	99.81	89.62	80.71	75.99	69.33	67.25	86.26	60.89	46.97	41.75	28.07	29.04	43.97	53.63
	GEM	Quattro	99.61	96.11	97.64	97.18	95.26	99.74	89.99	80.50	77.57	73.87	67.84	89.03	62.71	47.86	44.05	34.40	31.79	47.95	53.52
	GEM	3DMAC	99.61	95.36	96.75	93.19	95.38	98.78	84.66	66.08	64.64	48.04	56.43	78.35	46.29	31.05	29.76	19.00	21.76	36.45	61.38
	GEM	PAGOR	100.0	99.62	99.70	99.50	99.05	100.0	92.89	91.18	86.41	85.35	79.53	94.52	72.19	61.92	58.45	49.51	46.36	62.53	76.89

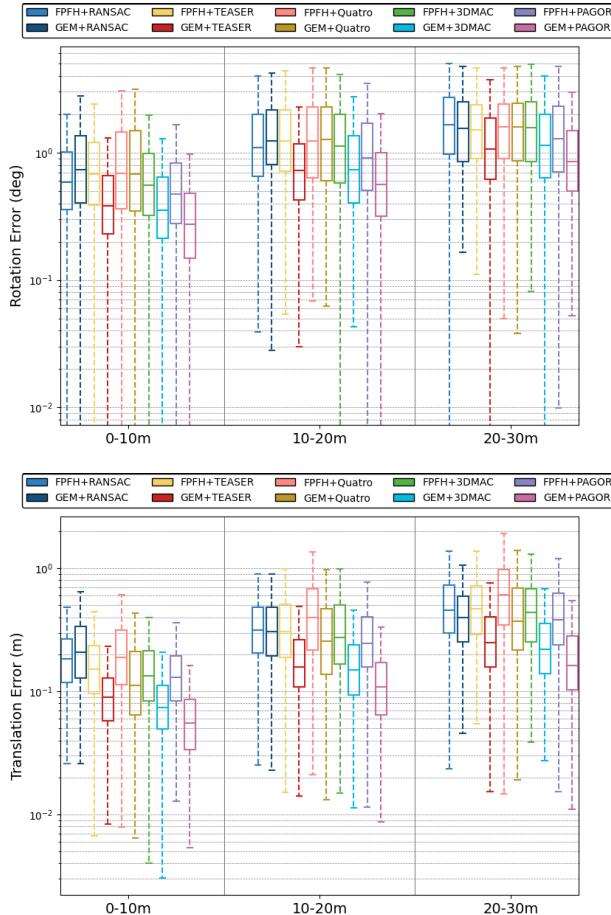


Fig. 7. Registration accuracy results. Boxplots of rotation errors and translation errors for the ten compared methods on the KITTI-LC dataset.

be the maximal clique as depicted in TEASER [13]. However, it may miss the true maximal clique due to SVD inaccuracies and an unfit evaluation function, as discussed in Section VI-B. In contrast to 3DMAC, our approach focuses on using multiple compatibility thresholds and an accurate evaluation function to select a maximum clique that is the true maximal clique.

In addition to specialized registration methods, we also test advanced loop closure systems with metric pose estimation (STD [63], Cont2 [64]) on the same datasets. We find that STD is sensitive to the physical distance of two point clouds, which we attribute to potential shortcomings in the reproducibility of their feature points (despite increasing the number to 1000). Cont2 performs fairly well with point clouds close to each other, and the speed is noticeably faster than all others, owing to its remarkably efficient scan representation. We note that both methods typically balance recall and precision of loop detection, therefore they could be too conservative for registration-only tasks.

3) Registration Accuracy: As shown in Figure 7, we compared the performance of GEM and FPFH when integrated with various back-end solutions, and we applied the same to PAGOR. Consistent with the registration recall benchmark, both GEM and PAGOR outperform the other methods in terms of accuracy. This superior performance can be attributed

TABLE VII
GLOBAL REGISTRATION EVALUATION ON APOLLO-LC DATASET

Front-end	Back-end	Recall (%) on APOLLO-LC ↑				Average Time (ms)
		0-10m	10-20m	20-30m	30-40m	
FPFH	RANSAC	99.69	92.65	67.86	35.92	272.73
	TEASER++	99.85	95.00	75.23	44.53	251.62
	Quatro	99.68	94.17	76.10	48.05	255.39
	3DMAC	99.09	84.32	51.71	23.61	257.92
	PAGOR	99.99	97.78	86.04	60.67	301.86
GEM	RANSAC	97.46	74.86	42.01	19.91	52.82
	TEASER++	99.98	98.30	91.89	77.39	38.56
	Quatro	99.99	98.17	91.71	77.91	38.63
	3DMAC	99.80	95.56	84.91	68.33	55.58
	PAGOR	100.0	99.15	95.50	84.51	64.31

to two main factors: First, unlike point-based registration methods, GEM-based registration leverages additional geometric constraints including plane-to-plane and line-to-line correspondences, followed by a distribution-to-distribution registration approach. Second, during the verification stage, the evaluation function preferentially selects the most accurate transformation, further enhancing performance.

D. Ablation Study

In this section, we validate the effectiveness of the main contributions of our method via conducting ablation studies. All results are shown in Table VIII. In Row 1, we bypass plane-aided segmentation and apply direct clustering, subsequently categorizing the results into PCL types based on eigenvalues. A comparative analysis to Row 21 reveals that plane-aided segmentation's benefits increase with translation magnitude. In translations up to 10 meters, an exception is found in sequence 02, sourced from highways with minimal planar features, suggesting that plane-aided segmentation might lead to over-segmentation and less repeatable clustering for this case. Rows 2-7 show the registration impact of various PCL type combinations, with the general trend indicating that clusters have a more pronounced effect than planes, and planes more than lines. Introducing an additional geometric type consistently improves registration accuracy.

In rows 8–13 of the dataset, we begin by considering the smallest p-values and incrementally include larger ones, evaluating them both as individual factors and in various combinations. The results show that combinations of multiple p-values consistently yield higher recall rates than either individual p-values or any smaller subsets thereof. As illustrated in Figure 8, the time required by PAGOR increases with the addition of each p-value. Notably, constructing the compatibility graph is the most time-consuming process. Moreover, our proposed graduated maximum clique (MAC) solver ensures the time complexity of finding MACs within a pyramid graph does not increase linearly.

In Rows 14 and 15, we apply predetermined compatibility thresholds such as 0.2m, 0.4m, 0.6m, and 0.8m, or we substitute our proposed pseudo covariance matrix with a statistical covariance matrix Σ to compute thresholds. These modifications result in an increase in computational time. This can be attributed to the potential introduction of inaccurate compatibility thresholds, which introduce unnecessary graph

TABLE VIII
ABLATION STUDY ON KITTI-LC DATASET

Variables	Values	Recall Rate (%) on KITTI-LC ↑										Average Time (ms)				
		0-10m					10-20m									
		00	02	05	06	08	00	02	05	06	08	00	02	05	06	08
Plane-aided Seg	(1) w/o	100.0	95.35	100.0	100.0	100.0	89.15	70.20	92.46	99.60	94.44	40.16	46.64	63.01	95.24	70.27
	(2) Plane	92.74	50.58	87.50	99.22	75.37	48.68	26.26	47.24	85.66	37.65	15.22	8.97	17.35	59.52	8.65
	(3) Cluster	99.67	93.60	98.86	100.0	98.51	73.90	55.55	80.90	98.41	84.57	26.77	23.77	42.92	85.32	44.86
	(4) Line	66.34	41.86	50.00	85.27	74.63	30.79	26.26	26.63	47.41	62.96	9.71	15.25	10.96	23.01	43.78
	(5) Plane+Cluster	100.0	90.12	100.0	100.0	100.0	94.13	67.68	91.46	100.0	95.06	44.62	34.98	58.45	97.22	65.41
	(6) Plane+Line	99.34	65.70	94.32	100.0	99.25	78.01	57.07	73.37	97.61	82.72	31.76	39.46	38.81	79.76	57.84
	(7) Line+Cluster	100.0	90.70	100.0	100.0	100.0	90.91	63.13	91.96	99.60	95.68	45.67	44.39	61.19	90.87	71.89
GEM Type	(8) [0.99]	100.0	79.65	100.0	100.0	100.0	93.55	59.60	92.96	100.0	96.30	42.78	42.15	54.34	94.05	64.86
	(9) [0.95]	100.0	83.72	100.0	100.0	100.0	94.43	63.13	94.97	100.0	96.91	46.98	46.19	64.38	98.02	70.27
	(10) [0.9]	100.0	86.05	100.0	100.0	100.0	94.13	63.64	94.97	100.0	98.15	49.87	45.74	65.75	97.22	69.19
	(11) [0.8]	100.0	86.04	100.0	100.0	100.0	93.55	65.66	96.48	100.0	96.91	46.72	43.95	64.38	98.41	69.73
	(12) [0.99, 0.95]	100.0	86.05	100.0	100.0	100.0	96.19	64.65	96.48	100.0	97.53	53.28	48.43	68.03	98.02	71.35
	(13) [0.99, 0.95, 0.9]	100.0	88.95	100.0	100.0	100.0	96.48	66.16	97.49	100.0	98.15	59.06	50.22	73.06	98.41	74.59
	(14) [0.2m, 0.4m, 0.6m, 0.8m]	100.0	91.86	100.0	100.0	100.0	96.77	72.22	96.98	100.0	98.15	57.22	54.26	70.78	99.21	74.59
p-values	(15) Σ	100.0	91.86	100.0	100.0	100.0	96.19	66.16	97.99	100.0	98.15	57.48	52.47	73.97	99.21	76.76
	(16) $\bar{\mu}$	100.0	86.05	100.0	100.0	100.0	97.36	68.18	95.48	100.0	98.15	65.62	52.02	73.52	98.81	82.16
	(17) Tukey	100.0	90.70	100.0	100.0	100.0	97.65	69.19	98.49	100.0	97.53	61.94	52.02	74.43	99.60	77.84
	(18) Cauchy	100.0	89.53	100.0	100.0	100.0	96.77	68.18	95.48	100.0	98.15	54.33	50.22	67.58	99.21	74.05
	(19) Kernel	100.0	91.28	100.0	100.0	100.0	97.36	69.19	98.49	100.0	98.15	59.84	51.57	74.89	99.21	76.76
	(20) Function	100.0	89.53	100.0	100.0	100.0	95.01	67.17	93.47	100.0	97.53	47.24	47.53	65.30	98.02	70.81
	(21) Default	See Table II	100.0	91.28	100.0	100.0	100.0	96.77	70.20	98.49	100.0	97.53	62.47	52.02	74.43	99.60

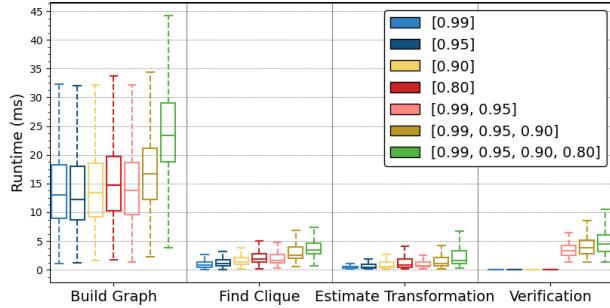


Fig. 8. The time consumption of PAGOR with different p-values combinations.

edges, thereby reducing efficiency. In Row 16, we experiment with utilizing the geometric centroid $\bar{\mu}$ of the 3D bounding box instead of the default statistical center μ when constructing TRIMs for registration. This method shows a slight performance increase of sequence 00 and 08 in the [20m, 30m] range, possibly because the symmetry of the 3D box mitigates the center uncertainty attributable to partial observations. Nevertheless, due to the ‘distrust-and-verify’ framework employed, performance fluctuations remain minimal.

Lastly, in Rows 17-20, as detailed in Section VI-B, we examine a variety of robust kernel functions. The experiments demonstrate that DCS (default), Tukey, and TLS kernels outperform other alternatives in terms of recall rate.

E. Limitation

The primary limitation of the proposed G3Reg lies in its front-end processing. As a segment-based registration method, it necessitates clear segmentation of the scene. This requirement does not imply that the scene must be inherently structured, like a town, but rather that it should be free of extensive, interconnected objects such as continuous rows of bushes along a road or walls flanking a corridor. Such repetitive and expansive features tend to introduce ambiguity in GEM matching and are illustrated in Figure 10, where our method struggles in environments that are geometrically uninformative. In these cases, GEMs are prone to significant uncertainties and exhibit low repeatability.

To mitigate these limitations, one potential strategy could be to extract FPFH within the point cloud segment of the large GEMs, which could potentially increase the inlier rate and registration recall. Alternatively, leveraging semantic information could enable better scene segmentation, an approach that we have successfully implemented in the conference version of our work [18]. Moreover, developing a classifier to detect the localizability [72] is also promising.

F. Real-World Experiments

As introduced in Section VII-A1, we also make a multi-session dataset at a university campus, named Campus-MS, which is collected using a mobile platform equipped with a solid-state LiDAR sensor. This dataset consists of five distinct robot travels (sessions), each collected in different days. To ensure accurate pose ground truth, Real-Time Kinematic (RTK) technology was employed during data collection. For every frame of one travel, we identify frames in other trips if their overlaps exceed 40%. The quantitative results in Table IX show that our method can still keep its registration efficiency and performance without changing any hyperparameters.

In order to verify the proposed G3Reg in practical applications, we embed it into a map merging task, as shown in Figure 9. Specifically, we first use LiDAR odometry to build subgraphs, then use GPS to find potential loop closures between multiple trajectories. Considering that GPS could not provide stable transformation, we use G3Reg to obtain relative poses and establish loop edges without pose priors. Finally, a global map is achieved by pairwise consistency maximization (PCM) [73] for false loop removal and global pose graph optimization. The precise map merge result demonstrates the effectiveness of G3Reg in practical applications.

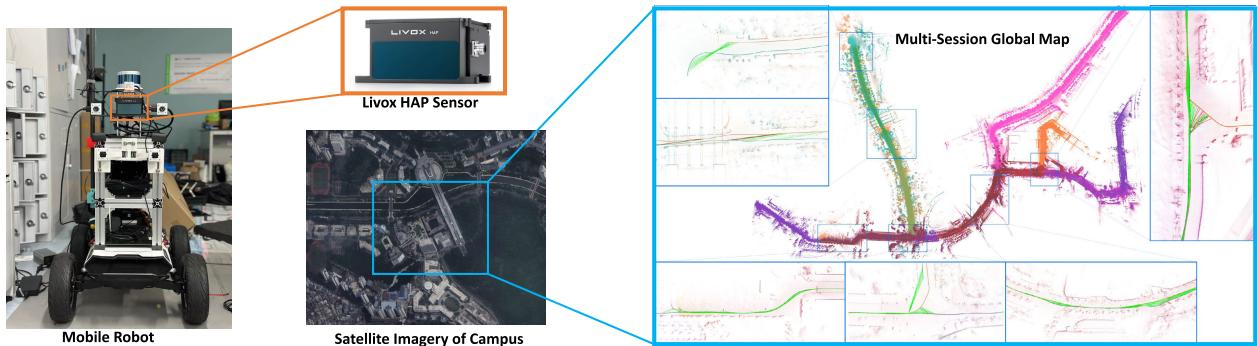


Fig. 9. The visualized results of multi-session mapping using our proposed G3Reg. The Campus-MS dataset is collected on a mobile robot platform equipped with a solid-state Livox HAP Sensor. Though the sensor type is different from Velodyne, we still use the same parameter settings for G3Reg. As a result, G3Reg provides stable transformation constraints for the multi-session mapping system (shown in green lines, best viewed in an enlarged format).

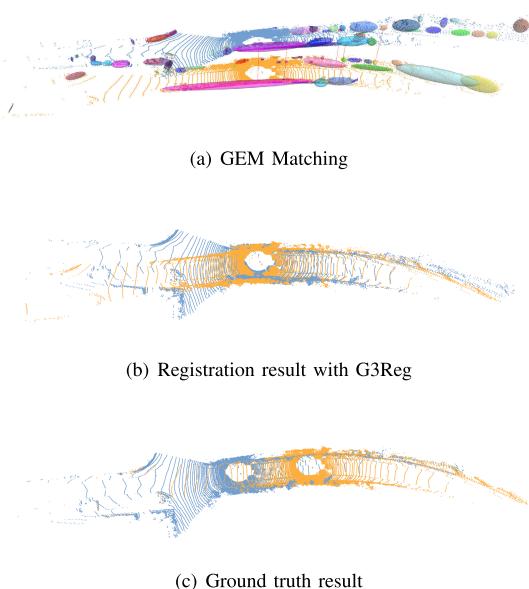


Fig. 10. Global registration failure using G3Reg. This is a very challenging case with a vehicle moving on a narrow, geometrically uninformative road. The ground truth translation is 26.18 m.

VIII. CONCLUSION

In this work, we present G3Reg, an effective and efficient framework for global registration of LiDAR point clouds, which addresses the limitations of conventional approaches. G3Reg introduces two primary contributions: GEMs for constructing correspondences and a distrust-and-verify scheme called PAGOR. By leveraging these innovations, G3Reg significantly improves the robustness and efficiency of global registration. Our extensive comparison experiments demonstrate that G3Reg outperforms SOTA methods. We present results across diverse hyperparameters and point cloud distributions, showcasing its applicability in various real-world scenarios. Furthermore, the promising directions for future improvement are discussed, including the integration of higher-level descriptors into GEM-based matching or the detection of localizability.

REFERENCES

- [1] P. Yin, S. Yuan, H. Cao, X. Ji, S. Zhang, and L. Xie, “Segregator: Global point cloud registration with semantic and geometric cues,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2023, pp. 2848–2854.
- [2] Z. Liu et al., “LPD-Net: 3D point cloud learning for large-scale place recognition and environment analysis,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2831–2840.
- [3] G. Pramatarov, D. De Martini, M. Gadd, and P. Newman, “BoxGraph: Semantic place recognition and pose estimation from 3D LiDAR,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Kyoto, Japan, Oct. 2022, pp. 7004–7011.
- [4] R. Dube, D. Dugas, E. Stumm, J. Nieto, R. Siegwart, and C. Cadena, “SegMatch: Segment based place recognition in 3D point clouds,” in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2017, pp. 5266–5272.
- [5] P. Yin et al., “AutoMerge: A framework for map assembling and smoothing in city-scale environments,” *IEEE Trans. Robot.*, vol. 39, no. 5, pp. 3686–3704, Oct. 2023.
- [6] Z. Yu, Z. Qiao, L. Qiu, H. Yin, and S. Shen, “Multi-session, localization-oriented and lightweight LiDAR mapping using semantic lines and planes,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2023, pp. 7210–7217.
- [7] H. Yin et al., “A survey on global LiDAR localization: Challenges, advances and open problems,” 2023, *arXiv:2302.07433*.
- [8] X. Xu et al., “RING++: Roto-translation invariant gram for global localization on a sparse scan map,” *IEEE Trans. Robot.*, vol. 39, no. 6, pp. 4616–4635, Dec. 2023.
- [9] R. B. Rusu, N. Blodow, and M. Beetz, “Fast point feature histograms (FPFH) for 3D registration,” in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2009, pp. 3212–3217.
- [10] Z. J. Yew and G. H. Lee, “RPM-Net: Robust point matching using learned features,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11824–11833.
- [11] S. Ao, Q. Hu, B. Yang, A. Markham, and Y. Guo, “SpinNet: Learning a general surface descriptor for 3D point cloud registration,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 11753–11762.
- [12] F. Poiesi and D. Boscaini, “Learning general and distinctive 3D local deep descriptors for point cloud registration,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3979–3985, Mar. 2023.
- [13] H. Yang, J. Shi, and L. Carlone, “TEASER: Fast and certifiable point cloud registration,” *IEEE Trans. Robot.*, vol. 37, no. 2, pp. 314–333, Apr. 2020.
- [14] P. C. Lusk, K. Fathian, and J. P. How, “CLIPPER: A graph-theoretic framework for robust data association,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2021, pp. 13828–13834.
- [15] J. Shi, H. Yang, and L. Carlone, “ROBIN: A graph-theoretic approach to reject outliers in robust estimation using invariants,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2021, pp. 13820–13827.
- [16] H. Lim et al., “A single correspondence is enough: Robust global registration to avoid degeneracy in urban environments,” in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2022, pp. 8010–8017.
- [17] X. Zhang, J. Yang, S. Zhang, and Y. Zhang, “3D registration with maximal cliques,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 17745–17754.
- [18] Z. Qiao, Z. Yu, H. Yin, and S. Shen, “Pyramid semantic graph-based global point cloud registration with low overlap,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2023, pp. 11202–11209.
- [19] L. Bernreiter, L. Ott, J. Nieto, R. Siegwart, and C. Cadena, “PHASER: A robust and correspondence-free global pointcloud registration,” *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 855–862, Apr. 2021.

- [20] J. Yang, H. Li, D. Campbell, and Y. Jia, "Go-ICP: A globally optimal solution to 3D ICP point-set registration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 11, pp. 2241–2254, Nov. 2016.
- [21] X. Li, J. K. Pontes, and S. Lucey, "PointNetLK revisited," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 12763–12772.
- [22] Y. Aoki, H. Goforth, R. A. Srivatsan, and S. Lucey, "PointNetLK: Robust & efficient point cloud registration using PointNet," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7156–7165.
- [23] Y. Zhong, "Intrinsic shape signatures: A shape descriptor for 3D object recognition," in *Proc. IEEE 12th Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Oct. 2009, pp. 689–696.
- [24] A. Mian, M. Bennamoun, and R. Owens, "On the repeatability and quality of keypoints for local feature-based 3D object retrieval from cluttered scenes," *Int. J. Comput. Vis.*, vol. 89, nos. 2–3, pp. 348–361, Sep. 2010.
- [25] X. Bai, Z. Luo, L. Zhou, H. Fu, L. Quan, and C.-L. Tai, "D3Feat: Joint learning of dense detection and description of 3D local features," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6359–6367.
- [26] K. Liu, K. Ok, and N. Roy, "VoluMon: Weakly-supervised volumetric monocular estimation with ellipsoid representations," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2021, pp. 5686–5693.
- [27] L. Nicholson, M. Milford, and N. Sunderhauf, "QuadricSLAM: Dual quadrics from object detections as landmarks in object-oriented SLAM," *IEEE Robot. Autom. Lett.*, vol. 4, no. 1, pp. 1–8, Jan. 2019.
- [28] F. Tombari, S. Salti, and L. Di Stefano, "Unique signatures of histograms for local surface description," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2010, pp. 356–369.
- [29] H. Wang et al., "RoReg: Pairwise point cloud registration with oriented descriptors and local rotations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 8, pp. 1–18, Aug. 2023.
- [30] H. Yang, P. Antonante, V. Tzoumas, and L. Carlone, "Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1127–1134, Apr. 2020.
- [31] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [32] Á. Parra Bustos and T.-J. Chin, "Guaranteed outlier removal for point cloud registration with correspondences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2868–2882, Dec. 2018.
- [33] O. Enqvist, K. Josephson, and F. Kahl, "Optimal correspondences from pairwise constraints," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 1295–1302.
- [34] J. Yang, X. Zhang, S. Fan, C. Ren, and Y. Zhang, "Mutual voting for ranking 3D correspondences," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–18, 2023. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/10105460>
- [35] T. Bailey, E. M. Nebot, J. Rosenblatt, and H. F. Durrant-Whyte, "Data association for mobile robot navigation: A graph theoretic approach," in *Proc. IEEE Int. Conf. Robot. Automat.*, vol. 3, Apr. 2000, pp. 2512–2517.
- [36] P. C. Lusk, D. Parikh, and J. P. How, "GraffMatch: Global matching of 3D lines and planes for wide baseline LiDAR registration," *IEEE Robot. Autom. Lett.*, vol. 8, no. 2, pp. 632–639, Feb. 2023.
- [37] L. Carlone, "Estimation contracts for outlier-robust geometric perception," *Found. Trends Robot.*, vol. 11, nos. 2–3, pp. 90–224, 2023, doi: [10.1561/2300000077](https://doi.org/10.1561/2300000077).
- [38] D. Chetverikov, D. Svirkov, D. Stepanov, and P. Krsek, "The trimmed iterative closest point algorithm," in *Proc. Int. Conf. Pattern Recognit.*, 2002, pp. 545–548.
- [39] L. Peng, M. C. Tsakiris, and R. Vidal, "ARCS: Accurate rotation and correspondence search," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11143–11153.
- [40] J. Wu et al., "Quadratic pose estimation problems: Globally optimal solutions, solvability/observability analysis, and uncertainty description," *IEEE Trans. Robot.*, vol. 38, no. 5, pp. 3314–3335, Oct. 2022.
- [41] Q.-Y. Zhou, J. Park, and V. Koltun, "Fast global registration," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2016, pp. 766–782.
- [42] T.-J. Chin, Z. Cai, and F. Neumann, "Robust fitting in computer vision: Easy or hard?" in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 701–716.
- [43] M. J. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision," *Int. J. Comput. Vis.*, vol. 19, no. 1, pp. 57–91, Jul. 1996.
- [44] A. Segal, D. Haehnel, and S. Thrun, "Generalized-ICP," in *Robotics: Science and Systems*, vol. 2. Cambridge, MA, USA: MIT Press, 2009, p. 435.
- [45] M. Oh et al., "TRAVEL: Traversable ground and above-ground object segmentation using graph representation of 3D LiDAR scans," *IEEE Robot. Autom. Lett.*, vol. 7, no. 3, pp. 7255–7262, Jul. 2022, doi: [10.1109/LRA.2022.3182096](https://doi.org/10.1109/LRA.2022.3182096).
- [46] H. Freeman and R. Shapira, "Determining the minimum-area enclosing rectangle for an arbitrary closed curve," *Commun. ACM*, vol. 18, no. 7, pp. 409–413, Jul. 1975.
- [47] O. Perron, "Zur theorie der matrices," *Mathematische Annalen*, vol. 64, no. 2, pp. 248–263, Jun. 1907.
- [48] G. Frobenius, F. G. Frobenius, F. G. Frobenius, and G. Mathematician, *Über Matrizen Aus Nicht Negativen Elementen*. Berlin, Germany: Berlin-Brandenburg Academy of Sciences and Humanities, 1912.
- [49] H. Wolkowicz and G. P. H. Styan, "Bounds for eigenvalues using traces," *Linear Algebra Appl.*, vol. 29, pp. 471–506, Feb. 1980.
- [50] H. Weyl, "Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung)," *Mathematische Annalen*, vol. 71, no. 4, pp. 441–479, Dec. 1912.
- [51] R. A. Rossi, D. F. Gleich, and A. H. Gebremedhin, "Parallel maximum clique algorithms with applications to network analysis," *SIAM J. Sci. Comput.*, vol. 37, no. 5, pp. C589–C616, Jan. 2015.
- [52] F. Dellaert and G. Contributors. (2022). *Borglab/Gtsam*. [Online]. Available: <https://github.com/borglab/gtsam>
- [53] C. Holmes and T. D. Barfoot, "An efficient global optimality certificate for landmark-based SLAM," *IEEE Robot. Autom. Lett.*, vol. 8, no. 3, pp. 1539–1546, Mar. 2023.
- [54] H. Yang and L. Carlone, "Certifiably optimal outlier-robust geometric perception: Semidefinite relaxations and scalable global optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 2816–2834, Mar. 2023.
- [55] P. Agarwal, G. D. Tipaldi, L. Spinello, C. Stachniss, and W. Burgard, "Robust map optimization using dynamic covariance scaling," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2013, pp. 62–69.
- [56] Z. Zhang, "Parameter estimation techniques: A tutorial with application to conic fitting," *Image Vis. Comput.*, vol. 15, no. 1, pp. 59–76, Jan. 1997, doi: [10.1016/S0262-8856\(96\)01112-2](https://doi.org/10.1016/S0262-8856(96)01112-2).
- [57] C. Choy, W. Dong, and V. Koltun, "Deep global registration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2514–2523.
- [58] X. Bai et al., "PointDSC: Robust point cloud registration using deep spatial consistency," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15859–15869.
- [59] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.
- [60] Y. Liao, J. Xie, and A. Geiger, "KITTI-360: A novel dataset and benchmarks for urban scene understanding in 2D and 3D," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3292–3310, Mar. 2023.
- [61] W. Lu, G. Wan, Y. Zhou, X. Fu, P. Yuan, and S. Song, "DeepVCP: An end-to-end deep neural network for point cloud registration," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 12–21.
- [62] C. Choy, J. Park, and V. Koltun, "Fully convolutional geometric features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8958–8966.
- [63] C. Yuan, J. Lin, Z. Liu, H. Wei, X. Hong, and F. Zhang, "BTC: A binary and triangle combined descriptor for 3-D place recognition," *IEEE Trans. Robot.*, vol. 40, no. 1, pp. 1580–1599, Oct. 2024.
- [64] B. Jiang and S. Shen, "Contour context: Abstract structural distribution for 3D LiDAR loop detection and metric pose estimation," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, London, U.K., May 2023, pp. 8386–8392.
- [65] M. A. Uy and G. H. Lee, "PointNetVLAD: Deep point cloud based retrieval for large-scale place recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4470–4479.
- [66] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," *Sensor fusion IV, Control paradigms data Struct.*, vol. 1611, pp. 586–606, Jul. 1992.
- [67] M. Magnusson, A. Lilienthal, and T. Duckett, "Scan registration for autonomous mining vehicles using 3D-NDT," *J. Field Robot.*, vol. 24, no. 10, pp. 803–827, 2007.
- [68] P. Zhou, X. Guo, X. Pei, and C. Chen, "T-LOAM: Truncated least squares LiDAR-only odometry and mapping in real time," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9446309>

- [69] M. Weinmann, B. Jutzi, and C. Mallet, "Semantic 3D scene interpretation: A framework combining optimal neighborhood size selection with relevant features," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 2, pp. 181–188, Aug. 2014.
- [70] J. Behley et al., "Towards 3D LiDAR-based semantic scene understanding of 3D point cloud sequences: The SemanticKITTI dataset," *Int. J. Robot. Res.*, vol. 40, nos. 8–9, pp. 959–967, Aug. 2021.
- [71] S. Choi, T. Kim, and W. Yu, "Performance evaluation of ransac family," *J. Comput. Vis.*, vol. 24, no. 1, pp. 271–300, 1997.
- [72] J. Nubert, E. Walther, S. Khattak, and M. Hutter, "Learning-based localizability estimation for robust LiDAR localization," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2022, pp. 17–24.
- [73] J. G. Mangelson, D. Dominic, R. M. Eustice, and R. Vasudevan, "Pairwise consistent measurement set maximization for robust multi-robot map merging," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 2916–2923.

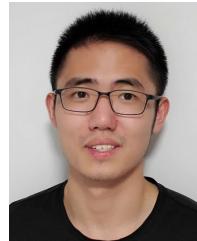


and computer vision.

Zhijian Qiao (Graduate Student Member, IEEE) received the B.Eng. degree in automation from Northeastern University, Shenyang, China, in 2019, and the M.S. degree in control engineering from Shanghai Jiao Tong University, Shanghai, China, in 2022. He is currently pursuing the Ph.D. degree with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong. His research interests include robotics and autonomous driving, with a focus on state estimation, localization and mapping,



Zehuan Yu received the B.Eng. degree in mechanical engineering from Harbin Institute of Technology, Harbin, China, in 2019, and the M.S. degree in control science and engineering from Harbin Institute of Technology, Shenzhen, China, in 2022. He is currently pursuing the Ph.D. degree with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong. His research interests include robot perception, SLAM, and computer vision.



Binqian Jiang (Graduate Student Member, IEEE) received the B.Eng. degree in information engineering from Southeast University, Nanjing, China, in 2019. He is currently pursuing the Ph.D. degree with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong. His research interests include state estimation, sensor fusion, and SLAM.



Huan Yin (Member, IEEE) received the B.Eng. degree in electronic and instrumentation and the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2016 and 2021, respectively.

He was a Visiting Scholar with the University of Technology Sydney, Australia, in 2018. After graduation, he was a Research Fellow with the Advanced Robotics Centre, National University of Singapore. He then joined the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, as a Research Assistant Professor. His research interests include robot perception, localization and mapping, and autonomous navigation.



Shaojie Shen received the B.Eng. degree in electronic engineering from The Hong Kong University of Science and Technology, Hong Kong, in 2009, and the M.S. degree in robotics and the Ph.D. degree in electrical and systems engineering from the University of Pennsylvania, Philadelphia, PA, USA, in 2011 and 2014, respectively.

In September 2014, he joined the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, as an Assistant Professor, and was promoted to an Associate Professor in July 2020. His research interests include robotics and unmanned aerial vehicles, with a focus on state estimation, sensor fusion, computer vision, localization and mapping, and autonomous navigation in complex environments.