2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

Segregator: Global Point Cloud Registration with Semantic and Geometric Cues

Pengyu Yin[1], Shenghai Yuan[1], Haozhi Cao[1], Xingyu Ji[1], Shuyang Zhang[2], and Lihua Xie[1], *Fellow, IEEE*

***Abstract*— This paper presents *Segregator*, a global point cloud registration framework that exploits both semantic information and geometric distribution to efficiently build up outlier-robust correspondences and search for inliers. Current state-of-the-art algorithms rely on point features to set up putative correspondences and refine them by employing pairwise distance consistency checks. However, such a scheme suffers from degenerate cases, where the descriptive capability of local point features downgrades, and unconstrained cases, where length-preserving (*l*-TRIMs)-based checks cannot sufficiently constrain whether the current observation is consistent with others, resulting in a complexified NP-complete problem to solve. To tackle these problems, on the one hand, we propose a novel degeneracy-robust and efficient corresponding procedure consisting of both instance-level semantic clusters and geometric-level point features. On the other hand, Gaussian distribution-based translation and rotation invariant measurements (*G*-TRIMs) are proposed to conduct the consistency check and further constrain the problem size. We validated our proposed algorithm on extensive real-world data-based experiments. The code is available: `https://github.com/Pamphlett/Segregator`.**

I. Introduction

Point cloud registration finds the pose transformation between point clouds and is widely employed in many computer vision and robotics applications ranging from object recognition to robotic grasping to satisfy the demands for accurate pose estimation. In particular, it is an essential technique in the ego-motion estimation of mobile robots [1] to solve simultaneous localization and mapping (SLAM) problems, i.e., to estimate the pose transformation between either two consecutive LiDAR frames or more distant ones in a loop closing or relocalization scenario.

Point cloud registration algorithms can be mainly divided into two categories, namely local and global registration, by whether an initial guess is needed. Local registration methods [2] set up correspondences by a heuristic nearest neighbor (NN) search and generate satisfactory pose estimation results given a good initial guess. On top of traditional local registration algorithms [2], [3], a variety of robust estimation and probabilistic modeling techniques, such as M-estimators [4] and noise modeling [5], are introduced to boost local algorithms' performance and have shown promising results in many LiDAR odometry solutions [6], [7]. However, the performance of local algorithms still depends on a good initial value, which makes them prone to partial overlapping and the emergence of dynamic obstacles and outliers.

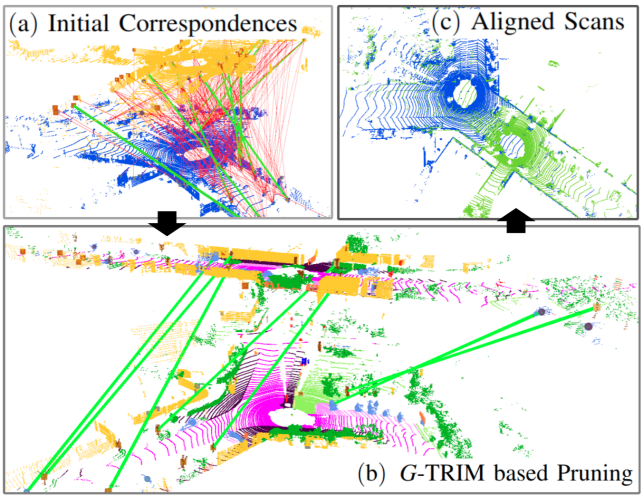


Fig. 1. The registration pipeline of Segregator with two distant point clouds as inputs. (a) The initial source and target clouds are colored yellow and blue. We managed to set up outlier-robust correspondences leveraging both point features and semantic cues (section III-B). (b) Inlier correspondences (green lines) are extracted by *G*-TRIM based graph pruning (section III-C) from massive outliers (thinner red lines in (a)). (c) Final aligned point clouds with the transformed cloud being green.

Unlike their local counterparts, global registration algorithms do not rely on any assumption of an initial guess, which is also the reason that there has been a growing interest in using them to solve loop closure and global (re)localization problems. However, these problems can not be tractably solved by traditional correspondence-free [8] or correspondence-based [9] global registration methods due to their high volumes of data points. More recently, correspondence-based methods leverage geometric primitives to set up putative correspondences to reduce problem size, and further prune built-up correspondences by employing robust estimation techniques [10], [11], [12], [13]. Still, point feature only corresponding schemes suffer from degenerate cases [11]; and using the single length preserving property cannot well constrain the consistency check problem. To sum up, there is a need to develop a degeneracy-robust correspondence establishing pipeline, as well as a more compact consistency checker for global registration problems in the autonomous driving scenario.

[1]Authors are with the Centre for Advanced Robotics Technology Innovation (CARTIN), School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. {pengyu001, shyuan, haozhi002, xingyu001, elhxie}@ntu.edu.sg

[2]Authors are with the Department of Electronic and Computer Engineering, the Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, China. {szhangcy}@connect.ust.hk

This research is supported by the National Research Foundation, Singapore under its Medium Sized Center for Advanced Robotics Technology Innovation.

In this paper, we investigate the combination of semantic cues, obtained using neural networks [14], [15], with Maximum Inlier Clique (MIC), a robust estimation technique, to tackle the aforementioned problems. We make the following contributions:

1) a semantic global point cloud registration framework, producing accurate pose estimations for low overlapping LiDAR scans in autonomous driving scenarios;
2) a degeneracy-robust correspondence and consistency graph construction method consisting of both semantic clusters and geometric distributions (*G*-TRIM), leading to an inlier correspondence searching algorithm with better efficiency and accuracy;
3) extensive evaluations (over 80,000 pairs of scans) are conducted on publicly available data sets, proving superior performance of the proposed framework than current state-of-the-arts.

II. Related Work

Early solutions for the point cloud registration problem adopt a framework iterating between the corresponding step and the error minimization step until certain termination criterion is satisfied [3], [5], [16]. A highly efficient but naive corresponding strategy, nearest neighbor (NN) search, is widely used to make the time complexity manageable. As stated by [1], these algorithms can produce reasonable estimation results when the following assumption is satisfied: the portion of inlier correspondences acquired by NN search improves in each iteration. If that is not the case, algorithms could be trapped in local and often unsatisfying minima. Several commonly observed phenomena could contribute to the aforementioned convergence problem, including noisy/distant measurements, degenerate environments, and dynamic obstacles.

Branch-and-bound-based (BnB) global point cloud registration methods [8], [17] avoid the correspondence establishing process by enumerating the solution space and obtaining globally optimal solutions, whereas they run in exponential time, which prohibits them from being directly used in LiDAR scans with a large collection of points. Another category of global registration methods stays in the same framework as local ones, i.e., corresponding and error minimization, but being more computationally attractive by setting up correspondence only once [18], [19], [12]. These deterministic correspondences are usually built via either learned [18] or traditional [20] feature descriptors and then fed to an outlier pruning step to find inlier correspondences.

In the correspondence setting up stage, however, degenerate cases could happen. As observed in [11], [12], point feature quality could downgrade as the size of the point cloud increases and thus those point feature-based correspondences could end up being very unreliable and contain less than two correct correspondences, making the rotation estimation impossible. Quatro [11] alleviates the degenerate problem by restricting the degree of freedom of rotation to one and assuming point clouds to be co-planar in an urban scenario. Nevertheless, the quasi-SE(3) assumption could be challenged in non-flat areas, and the method does not fundamentally resolve the degenerate problem (e.g., in the correspondence establishing step).

In terms of the outlier correspondence pruning stage, some local methods [6], [21] incorporate semantic masks to build better correspondences. A more conservative and robust way is to distinguish the inliers from the pre-built correspondences [10], [12], [17]. GORE [17] exploits geometry features to avoid using branch and bound on the whole-size initial correspondence sets. Another solution is Teaser [12], which represents correspondences by a compatibility graph where vertexes are defined as the distance between points and edges are the results of the compatibility check. Then inliers are selected by finding the Maximum Inlier Clique (MIC) of the graph, which is in general NP-complete and requires exponential time to solve. Clipper [10] further extends the formulation to a weighted scheme. However, all these methods are faced with scalability problems and do not fully exploit local geometric statistics, especially the distribution of points, resulting in searching for MIC in a bigger graph that could be computationally intensive.

III. Methodology

A. Problem Formulation

Given a pair of partially overlapping point clouds collected by the LiDAR sensor $\mathcal{P} = \{\mathbf{p}_i\}_{i=1}^{n}$ and $\mathcal{Q} = \{\mathbf{q}_j\}_{j=1}^{m}$ in two different coordinates, we seek to associate them into one common coordinate system. With unknown ground truth transformation, measurements $\mathbf{p}_i$ and $\mathbf{q}_j$ satisfy the following generative model

$$\mathbf{p}_i = \mathbf{R}\mathbf{q}_j + \mathbf{t} + \mathbf{o}_{ij}, \tag{1}$$

where index i and j denote one candidate in the raw correspondence sets $\mathcal{I}$. $\mathbf{T}$ denotes the ground truth transformation between the scans with

$$SO(3) \stackrel{\text{def}}{=} \left\{\mathbf{R} \in \mathbb{R}^{3\times 3} : \mathbf{R}^\top\mathbf{R} = \mathbf{I}, \det(\mathbf{R}) = 1\right\},$$

$$SE(3) \stackrel{\text{def}}{=} \left\{\mathbf{T} = \begin{pmatrix}\mathbf{R} & \mathbf{t}\\ \mathbf{0} & 1\end{pmatrix} : \mathbf{R} \in SO(3), \mathbf{t} \in \mathbb{R}^3\right\}.$$

In Eq. (1), $\mathbf{o}_{ij}$ can be modeled as a Gaussian distribution if $i, j \in \mathcal{I}_{GT}$ and is a random vector if not [12]. Accordingly, the objective function of point cloud registration can be formulated as minimizing the measurement-wise error in given metric space,

$$\hat{\mathbf{R}}, \hat{\mathbf{t}} = \underset{\mathbf{R}\in SO(3), \mathbf{t}\in\mathbb{R}^3}{\arg\min} \sum_{ij\in\mathcal{I}} \rho\left(e\left(\mathbf{p}_i - \mathbf{R}\mathbf{q}_j - \mathbf{t}\right)\right), \tag{2}$$

where $\rho(\cdot)$ and $e(\cdot)$ are the robust loss function and error metric respectively.

Accordingly, the point cloud registration problem can then be split into three steps: firstly, establish a set of measurement correspondences $\mathcal{I} \in [n] \times [m] := \{1, \ldots, n\} \times \{1, \ldots, m\}$; secondly, obtain inlier correspondence sets; thirdly, minimize the residual error. The proposed algorithm will also come after this framework and be presented in detail in the following subsections.

B. Degeneracy-Robust Correspondences Establishment

Although not explicitly stated in the literature, setting up correspondences suffers from a *radical versus conservative* dilemma. Feature-based corresponding methods [11], [18] tend to be radical as they build one-to-one correspondence according to feature similarity. In large scenes (e.g., an urban driving scenario), the descriptive capability of local features downgrades due to either the sparse nature of LiDAR point clouds [11] or the emergence of locally similar areas. Yet all-to-all correspondence $\mathcal{I}_C = [n] \times [m]$ ensures the full inclusion of the ground truth correspondence sets while being computationally expensive. Therefore, it is neither suitable to construct putative point-level correspondences radically nor to be too conservative to set up the all-to-all correspondence. To this end, we leverage semantic and point features to seek a balance between these two.

Given the source point cloud $\mathcal{P} = \{\mathbf{p}_i\}_{i=1}^{n}$ and a semantic assignment function $\lambda : \mathbb{R}^3 \to \mathbb{N}$, where semantic labels are represented by natural numbers in $\mathbb{N}$. Each point $\mathbf{p}_i$ in the cloud could be enhanced by their per-point label $l \in \mathcal{L}$. The semantically enhanced point cloud could be represented as

$$\mathcal{S} = \{s_i | s_i = \{\mathbf{p}_i, \lambda(\mathbf{p}_i)\}, \forall \mathbf{p}_i \in \mathcal{P}\}, \tag{3}$$

where s_i carries both its coordinate and semantic label.

We use the dynamic curved-voxel clustering (DCVC) [22] algorithm to segment point clouds for each label into disjoint sub-clouds. This will produce sets of geometrically approximate points with the same semantic label

$$\begin{aligned}\mathbb{C}^l = \{C_1, \ldots, C_N | C_k \subset \mathcal{P}, \\ l = \lambda(\mathbf{p}_i) = \lambda(\mathbf{p}_j)\ \forall \mathbf{p}_i, \mathbf{p}_j \in C_k\}.\end{aligned} \tag{4}$$

For each cluster in $\mathbb{C}^l$, we compute its centroid c and its corresponding covariance matrix Σ as its geometric characteristics. The procedure is repeated for points in the target point cloud $\mathcal{Q} = \{\mathbf{q}_j\}_{j=1}^{m}$ with the same semantic label l and resulting in two semantic cluster sets $\mathbb{C}_{\mathcal{P}}^l$ and $\mathbb{C}_{\mathcal{Q}}^l$. Thereafter, all-to-all correspondence is established between these semantic clusters (SCs) with the label l

$$\mathcal{I}_{\text{SC}}^l := \left\{(C_N, C_M) \in \mathbb{C}_{\mathcal{P}}^l \times \mathbb{C}_{\mathcal{Q}}^l\right\}. \tag{5}$$

Semantic correspondences are established for every semantic class with a meaningful definition of 'instance' $\mathcal{L}_{\text{SC}} \subset \mathcal{L}$ (e.g. car, trunk, traffic sign). And the final semantic cluster-level correspondence $\mathcal{I}_{\text{SC}}$ is generated by concatenating all correspondences for each semantic category

$$\mathcal{I}_{\text{SC}} = \bigcup \mathcal{I}_{\text{SC}}^l \quad \forall l \in \mathcal{L}_{\text{SC}}. \tag{6}$$

For other more environmental semantic categories $\mathcal{L}_{\text{F}}$ (e.g., building, vegetation), we still comply with traditional feature-based setup (e.g., FPFH in Quatro [11]) to generate the other set of correspondences $\mathcal{I}_{\text{F}}$. This is because the euclidean clusters for these semantic classes are heavily viewpoint-dependent and, therefore, their geometric characteristics could be unstable. Accordingly, semantic labels are also used to eliminate cross-semantic class correspondences in $\mathcal{I}_{\text{F}}$. Finally, the final correspondence sets, $\mathcal{I}_{\text{raw}}$ is the combination of $\mathcal{I}_{\text{SC}}$ and $\mathcal{I}_{\text{F}}$.

The philosophy under the above-presented correspondence establishment procedure is threefold: Firstly, the semantic cluster is a combination of both cognitive and geometric information, which makes them a suitable choice for either high-level tasks (e.g., rule out coarse outliers by semantic label discrepancy) or low-level ones (e.g., geometric perception). Secondly, semantic cluster-based correspondences also alleviate degenerate cases as the all-to-all correspondence ensures full inclusion of the inlier correspondence sets. Thirdly, the presented formulation considers both semantic clusters and geometric features, which acts as belt and braces to ensure registration success.

One may argue that cluster centroids could also vary with viewpoint changes, and it would consequently affect the pose estimation precision as well as inlier correspondence selection. However, firstly, the proposed method can always act as a coarse alignment followed by a finer one (e.g., ICP); secondly, in the following section III-C, we demonstrate how our novel distribution-based (*G*-TRIM) consistency check alleviates the unstable centroid problem by considering both positional and distribution information.

C. G-TRIM based Outlier Pruning

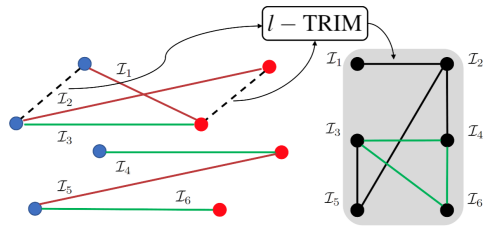
With a set of noisy correspondences, $\mathcal{I}_{\text{raw}}$, the problem of outlier pruning is to find the largest inlier correspondence sets $\hat{\mathcal{I}}$, where all entries inside satisfy the following distance constraint under the optimal transformation $\mathbf{T}$

$$\begin{aligned}\max_{\mathcal{I} \subset \mathcal{I}_{\text{raw}},\ \mathbf{T} \in \text{SE}(3)} \quad & |\mathcal{I}| \\ \text{s.t. } \|\mathbf{y}_i - \mathbf{R}\mathbf{x}_i - \mathbf{t}\|_2 \leq \epsilon, \ & \forall i \in \mathcal{I},\end{aligned} \tag{7}$$

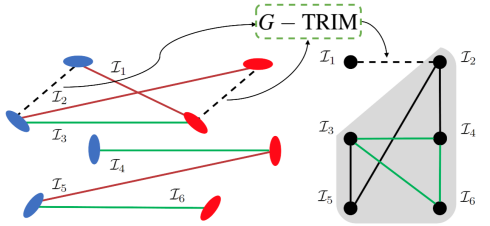
where $\mathbf{x}_i$ and $\mathbf{y}_i$ are the corresponding points, ϵ is the threshold of measurement error.

Since the ground truth transformation $\mathbf{T}$ is unknown, one cannot leverage Eq. (7) to find inlier correspondences. Rather, a graph-theoretic solution that takes advantage of the length-preserving property is proposed in Teaser [12]. To be more specific, given two pairs of points $\mathbf{a}_i$ and $\mathbf{a}_j$, $\mathbf{b}_i$ and $\mathbf{b}_j$, associated by the raw correspondence $\mathcal{I}_{\text{raw}}$. Length-based translation and rotation invariant measurements (*l*-TRIMs) are calculated as $d_{ij} := l\text{-TRIM}_{\mathbf{a}_{ij}}/l\text{-TRIM}_{\mathbf{b}_{ij}} = \|\mathbf{a}_i - \mathbf{a}_j\|_2 / \|\mathbf{b}_i - \mathbf{b}_j\|_2$. Since we work in a rigid transformation scenario, $d_{ij} = 1$ holds for no-noise cases. In real applications, a noisy measurement function is observed. Thus, two correspondences are mutually consistent with each other when the corresponding d_{ij} is very close to 1. By computing *l*-TRIMs for all possible correspondence pairs, an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where vertexes are correspondences and edges are built between mutually consistent correspondences (vertexes). Since our goal is about finding the largest correspondence inlier set, the maximum clique[1] of $\mathcal{G}$ should be found, which is NP-complete and could typically be solved in exponential time.

[1]Cliques of a graph refers to the complete subgraphs, and the maximum clique is a clique with the most vertexes [23]



(a) l-TRIM based consistency graph construction.



(b) G-TRIM based consistency graph construction.

Fig. 2. Illustration of our proposed G-TRIM-based consistency graph building method against the length-based (l-TRIM) counterpart. In (b), G-TRIM successfully resists the crossings outlier case ($\mathcal{I}_1$ and $\mathcal{I}_2$) by employing distributions (ellipsoids). Consistency graphs are shown in grey, and the final MIC is connected by green lines. G-TRIM constructs a distribution-consistent graph with fewer vertexes.

l-TRIM-based consistency check retains correspondence pairs that satisfy the length-preserving property. However, it is not a reasonable constraint for all cases. As presented in Fig. 2(a), a pair of 4-point sets are shown in red and blue with a set of noisy correspondences ($\mathcal{I}_1$-$\mathcal{I}_6$ with outliers in red and inliers in green). We focus on the first two correspondences ($\mathcal{I}_1$ and $\mathcal{I}_2$) which are in a crossings situation. In this scenario, the l-TRIM-based consistency check will treat them as inliers as swapping the corresponding sequence will not change the length. Accordingly, an edge between node $\mathcal{I}_1$ and $\mathcal{I}_2$ is constructed indicating their mutual consistency although they are indeed outlier correspondences.

We alleviate this problem by introducing a novel Gaussian distribution-based translation and rotation invariant measurement (G-TRIM). In our formulation, the 3D position of every single point is modeled as a Gaussian distribution $\mathbf{p}_i \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\mu}_i$, the centroid, and $\boldsymbol{\Sigma}_i$, the covariance matrix, are calculated by $\boldsymbol{\mu}_i = \frac{1}{|\mathcal{V}_i|}\sum \mathbf{p}_i$ and $\boldsymbol{\Sigma}_i = \frac{1}{|\mathcal{V}_i|}\sum(\mathbf{p}_i - \boldsymbol{\mu}_i)^\top(\mathbf{p}_i - \boldsymbol{\mu}_i)$. $\mathcal{V}_i$ is a small patch of neighboring points around $\mathbf{p}_i$. The point-wise measurement $\mathbf{a}_{ij}$ also follows a Gaussian distribution $\mathbf{a}_{ij} := \mathbf{a}_i - \mathbf{a}_j \sim \mathcal{N}\left(\mathbf{c}_i - \mathbf{c}_j, \boldsymbol{\Sigma}_{\mathbf{a}_i} + \boldsymbol{\Sigma}_{\mathbf{a}_j}\right)$, which we denote as Gaussian distribution-based translation and rotation invariant measurement (G-TRIM). Consistency checks are conducted by computing the Wasserstein distance between these Gaussian distributions.

$$
\begin{aligned}
d^2\left(\mathbf{a}_{ij}, \mathbf{b}_{ij}\right) &= W\left(\mathbf{a}_{ij}, \mathbf{b}_{ij}\right) \\
&= \left\|\mu_{\mathbf{a}_{ij}} - \mu_{\mathbf{b}_{ij}}\right\|_2^2 \\
&+ \operatorname{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{a}_{ij}} + \boldsymbol{\Sigma}_{\mathbf{b}_{ij}} - 2\left(\boldsymbol{\Sigma}_{\mathbf{a}_{ij}}^{1/2}\boldsymbol{\Sigma}_{\mathbf{b}_{ij}}\boldsymbol{\Sigma}_{\mathbf{a}_{ij}}^{1/2}\right)^{1/2}\right).
\end{aligned} \tag{8}
$$

We remark here that the first and second term in (8) captures positional and shape variations respectively. Reasons for choosing the Wasserstein distance are twofold: First, the Wasserstein distance is a perfect distance metric for distributions, while some other distribution discrepancy measurements, like KL-divergence, are not in a symmetric form and can not be used as metric level measurements; Second, Wasserstein distance has an elegant closed-form solution, while measured entries are both Gaussian [24].

It is worth noting that the above G-TRIM based consistency check could successfully resist the crossings case (Fig. 2) as $W\left(\mathbf{a}_{ij}, \mathbf{b}_{ij}\right) \neq W\left(\mathbf{a}_{ij}, \mathbf{b}_{ji}\right)$ as we presented in Fig. 2(b), the formulation prevents the insertion of the edge between correspondences $\mathcal{I}_1$ and $\mathcal{I}_2$. Such crossings cases, however, could be commonly observed in $\mathcal{I}_{raw}$, especially for the all-to-all correspondence case. Moreover, G-TRIM-based consistency check captures more geometric information compared to its length-based counterpart as it measures positional as well as the structural similarity between corresponding measurements. Considering the mixer of information could 1. partially solve the unstable centroid problem mentioned in III-B as two corresponding centroids with tiny positional variation could have similar neighboring distribution and thus pass the consistency check; 2. serves in a belt and braces way with both length and distribution preserving property to solve the unconstrained problem.

Finding the inlier correspondence sets $\hat{\mathcal{I}}$ in Eq. (7) can be solved by incorporating the parallel maximum clique algorithm [25] to the consistency graph built by G-TRIM.

D. Pose Estimation

With inlier correspondence sets $\hat{\mathcal{I}}$, we formulate the objective function in Eq. (2) into the following truncated least squares (TLS) form to resist potential outliers further [26]

$$
\hat{\mathbf{R}}, \hat{\mathbf{t}} = \underset{\mathbf{R}\in SO(3), \mathbf{t}\in\mathbb{R}^3}{\arg\min} \sum_{ij\in\hat{\mathcal{I}}} \min\left(\left\|\mathbf{p}_i - \mathbf{R}\mathbf{q}_j - \mathbf{t}\right\|_2, c_{ij}\right), \tag{9}
$$

with c_{ij} the truncation parameter. Eq. (9) are then solved by leveraging Black-Rangarajan duality [27] and graduated non-convexity (GNC) [26].

IV. Experimental Results

In this section, we conduct comparative experiments to test the proposed algorithm against current state-of-the-arts. The proposed algorithm is implemented in C++. All experiments are conducted on a PC with an Intel Core 2.60GHz i5 CPU, 32Gb RAM, and Geforce RTX3060 GPU. We follow the parameter settings in [22] for clustering and empirically set the noise bound c_{ij} in 9 to 0.2.

Benchmark Data Set We choose the publicly available KITTI data set [28] to conduct all experiments. Experiments are divided into two parts, namely the robustness test, and the loop closing test. Detailed experiment settings can be found in the following sections (IV-A and IV-B) respectively.

Baseline Methods Three state-of-the-arts are chosen to be baseline, consisting both local and global registration methods, namely V-GICP [29], TEASER++ [12] and Quatro [11]. We leverage the open-sourced version of each algorithm. As TEASER++ [12] was not intend to be a full registration

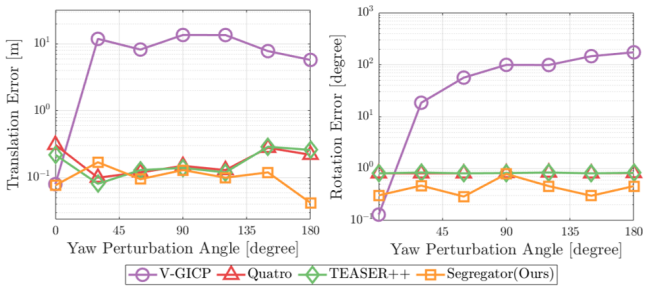


Fig. 3. Pose estimation with yaw angle perturbation.

method (more of a robust solver), we apply the FPFH-based correspondence built by Quatro [11] to it. Moreover, all other methods employed OpenMP [30] in the MIC estimation, so we set the thread number to 4 accordingly when evaluating V-GICP [29] for a fair comparison.

Error Metrics For all qualitative experiments, we adopt the relative pose error (RPE) [4] to evaluate the accuracy of the estimated pose $\hat{\mathbf{T}}$ against the ground truth $\mathbf{T}$. Note $\Delta\mathbf{T} = \hat{\mathbf{T}} \cdot \mathbf{T}^{-1}$ and Δx, Δy, Δz are the positional entries in $\Delta\mathbf{T}$, then the translation error and rotation error are calculated by

$$e_{trans} = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2},$$

$$e_{rot} = \arccos\left(trace(\Delta\mathbf{T})/2 - 1\right).$$

A. Robustness Test

The robustness factor is defined as the ability to perform normally under perturbations. For this purpose, we test whether algorithms are robust to translation, rotation, and also imperfect semantic masks. Accordingly, we manipulate the first and the 10th frame (approximately 5 meters apart with no rotational misalignment) of KITTI sequence 05 by adding different yaw angles perturbations as initial values and observe the performance of each algorithm. Results are shown in Fig. 3. V-GICP [29], as the only local registration method, gets the most accurate pose estimation result in the case of no perturbation. However, its performance drastically collapsed when perturbation started to increase due to the naive NN-based correspondence establishing. All global methods are capable of generating reasonable pose estimations even under the most serious yaw perturbation, as all these methods contain a well-designed outlier-rejecting module. TEASER++ and Quatro behave very similarly due to that they are provided with the same FPFH-generated correspondences and have similar length-based inlier correspondence selection procedures. But it is also worth noting that the proposed algorithm, Segregator, is the only global registration method that can resist all yaw angle perturbations while also generating rather accurate pose estimation results as local state-of-the-art V-GICP. This is because our pose estimation measurements consist of both point-level features and instance-level semantic clusters, constituting a hierarchical measurement pool. The proposed mix-level correspondence setup (section III-B) thus complements to each other and anchors each inlier measurement more precisely.

Furthermore, we investigate the performance of Segregator against semantic label deterioration. Namely, how would the

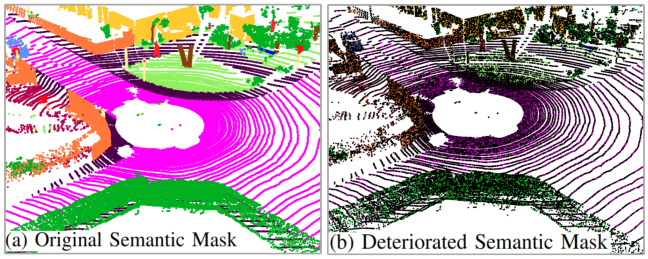


Fig. 4. Illustration of the original semantic mask inferred by SPVNAS [15] and the deteriorated one. We randomly set 90% of the labels to unclassified (shown as black in (b)). Even with such a severely deteriorated semantic mask, Segregator could produce reasonable pose estimation results with a translation error of 0.27 m and a rotation error of 1.69 degrees due to the employment of both geometric and semantic features.

quality of the semantic mask affect the quality of pose estimation. We gradually increase the noise ratio in the semantic label predicted by SPVNAS [15] by randomly setting part of original labels as unclassified (marked black in Fig. 4(b)). For each contamination level, we run our algorithm ten times and get the average translation and rotation error. As the results in Table I have shown, the proposed method can provide satisfactory pose estimation result even under very challenging semantic mask deterioration with up to 90% of noise. This is because semantic labels guide the formulation of object level representations, rather to be involved in a metric level computation. Thus Segregator won't fail as long as the semantic-guided clustering process don't fail.

TABLE I
REGISTRATION ACCURACY WITH LABEL DETERIORATION

Deterioration Rate (%)	10	30	50	70	90
e_{trans} [m]	0.09	0.04	0.31	0.06	0.27
e_{rot} [deg]	0.20	0.22	0.21	1.70	1.70

B. Protocol-based Loop Closing Test

In this section, algorithms are tested in a loop closure condition, which refers to registering two loop closure candidates. Loops add metric level constraints in the pose graph and reduce the cumulative error (a.k.a. odometry drift). However, building wrong loops could result in catastrophic failure. To have an all-round testing basis, we extract all possible loop candidates in several sequences in KITTI. This data is set up by running the following protocol through the ground truth pose $\mathbf{T}_k$ of each scan,

$$\mathcal{X}_k = \{\mathbf{T}_i : r_1 \leq \|\mathbf{t}_k - \mathbf{t}_i\|_2 \leq r_2, |i-k| \geq m, \forall \mathbf{T}_i \in \mathcal{T}\}, \tag{10}$$

where r_1 and r_2 are predefined translation thresholds; m is a loop closing parameter to rule out neighboring frames; $\mathcal{T}$ is the whole pose set. Thus all potential loop closure candidates are found by searching for all geometrically proximate scans but with a relatively large acquiring time offset. Statistics of resulted data set are presented in Table III. We include both unidirectional and reverse loops. In terms of quantity, the protocol-based loop closing data set contains 83,989 pairs of loop closure candidates. According to the translation discrepancy ($\|\mathbf{t}_k - \mathbf{t}_i\|_2$ in Eq. (10)), we separate

TABLE II

SUCCESS RATE RESULTS ON KITTI

Sequence	KITTI														
	00			05			06			07			08		
Hardness	*easy*	*medium*	*hard*	*easy*	*medium*	*hard*	*easy*	*medium*	*hard*	*easy*	*medium*	*hard*	*easy*	*medium*	*hard*
Num. of Pairs	8503	7527	21267	5155	5243	11600	1128	1113	2825	1311	1169	2877	2064	2750	9457
V-GICP [29]	95.6	34.1	14.9	82.6	34.6	25.3	94.6	66.7	35.1	80.0	36.5	0.8	0.24	0.0	4.9
TEASER++ [12]	97.0	65.4	41.1	96.4	73.4	48.0	99.8	94.3	83.2	91.1	49.2	30.7	96.6	74.8	49.5
Quatro [11]	96.9	66.2	41.2	96.3	73.7	48.1	99.8	93.9	83.2	91.3	51.0	31.9	96.8	75.2	49.8
Segator†(Ours)	98.6	88.3	75.4	77.8	72.2	59.0	91.0	74.6	63.8	**100.0**	95.5	81.4	90.3	84.4	78.2
Segregator (Ours)	**99.7**	**89.3**	**80.7**	**98.7**	**93.2**	**81.8**	**100.0**	**99.6**	**98.3**	**100.0**	**98.1**	**88.0**	**96.9**	**94.1**	**88.8**

†for ablation study.

TABLE III

STATISTICS OF EVALUATION DATA SET

Sequence	KITTI				
	00	05	06	07	08
Num. of Scans	4541	2761	1101	1101	4071
Num. of Pairs	37297	21998	5066	5357	14271
Loop Direction	*Both*	*Both*	*Same*	*Same*	*Reverse*
% Reverse loop	3%	5%	0%	0%	100%

TABLE IV

ATE, ARE AND TIMING ON KITTI SEQUENCE 06*

Method	Success Rate	APE [m/°]	Total Time [s]	AIC
V-GICP	55.4%	**0.07/0.10**	**0.09**	-
TEASER++	89.0%	0.36/0.67	0.15	3025
Quatro	89.0%	0.36/0.66	0.15	3025
Segregator (Ours)	**99.0%**	0.09/0.27	0.10	2687
Segregator-c2f (Ours)	**99.0%**	**0.07**/0.12	0.15	-

* Evaluations are conducted on successfully registered pairs only.

the whole data set into three categories, namely easy ($3-5$m), medium ($8-10$m) and hard ($10-15$m). The categories and measures are inspired by the recent Hilti SLAM challenge scoring scheme. Apart from all other baseline methods, we also include a trimmed version of our proposed method, which is Segator in Table II. We replaced the presented *G*-TRIM (III-C) based pruning with an ordinary length-based one (*l*-TRIM) to have a clear view of its effectiveness. All registration results are considered to be successful when the translation error is smaller than 2 meters and the rotation error is under $5°$. We compute the ratio between successful registered frames and the number of whole pairs to be registered. Furthermore, the ground truth data for KITTI 08 [28] is found to be relatively noisy so we use the pose data in Semantickitti [31] estimated by Suma++ [32].

Matching success rate results for the aforementioned algorithms is shown in Table II. It is observed that our proposed method, Segregator, outperforms all other baselines by a margin especially on harder cases. While all algorithms are working comparably well on *easy* cases, it is worth noting that even the state-of-the-art local registration method V-GICP [29], as we mentioned earlier, can hardly deal with *medium* and *hard* situations for its naive correspondence setup. As for global baselines, TEASER++ [12] and Quatro [11] have very similar overall performance with Quatro [11] being slightly better. That is because these two algorithms are fed with exactly the same correspondence sets. Quatro [11] employed a Quasi-SE(3) assumption and relaxed the number of correspondence that is needed to estimate the rotation to only one, which we thought is the reason for the performance boost. As for our ablation, Segator, we observe the overall performance falls with the trimmed *G*-TRIM part thus shown validity of our proposition.

Furthermore, we evaluate the running time and average pose error (APE), of each algorithm on KITTI [28] sequence 06 with results presented in Table IV. Only successfully registered pairs are included in the pose error calculation for a meaningful comparison. V-GICP [29] achieved the least ATE and ARE as well as the total run-time. It is typical for local registration methods to outperform global ones in terms of accuracy as they employ more points in one scan and thus can anchor every point more precisely. Our original method, Segregator, achieves the best performance among all global registration methods by using semantic clusters and geometric distribution information to search for more geometrically corresponding measurements. We further install Segregator into a simple coarse to fine scheme (Segregator-c2f) where Iterative Closest Points (ICP) [2] takes the estimation result from Segregator as an initial value. The simple integration obtains comparable precision compared to V-GICP [29] while being far more stable in terms of success rate. Moreover, Segregator is more computationally attractive than all other global registration methods by being faster in correspondence generation while also having a smaller size MIC problem to solve. Average inlier count (AIC) refers to the average amount of edges in the consistency graph. *G*-TRIM, serves as a more constrained consistency checker, successfully ruling out distribution inconsistent and crossings cases III-C.

V. CONCLUSIONS

In this paper, we present a global point cloud registration algorithm leveraging both semantic and geometric information dubbed as Segregator. It is proved to be more robust in registering distant scans, which makes it a suitable choice in loop closure and relocalization scenario. In the future, we plan to extend Segregator with probabilistic modeling technique in the pose estimation part for more accurate estimations.

REFERENCES

[1] F. Pomerleau, F. Colas, R. Siegwart, *et al.*, "A review of point cloud registration algorithms for mobile robotics," *Foundations and Trends® in Robotics*, vol. 4, no. 1, pp. 1–104, 2015.

[2] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in *Sensor fusion IV: control paradigms and data structures*, vol. 1611. Spie, 1992, pp. 586–606.

[3] D. Chetverikov, D. Svirko, D. Stepanov, and P. Krsek, "The trimmed iterative closest point algorithm," in *2002 International Conference on Pattern Recognition*, vol. 3. IEEE, 2002, pp. 545–548.

[4] F. Pomerleau, F. Colas, R. Siegwart, and S. Magnenat, "Comparing icp variants on real-world data sets," *Autonomous Robots*, vol. 34, no. 3, pp. 133–148, 2013.

[5] A. Segal, D. Haehnel, and S. Thrun, "Generalized-icp." in *Robotics: science and systems*, vol. 2, no. 4. Seattle, WA, 2009, p. 435.

[6] Y. Pan, P. Xiao, Y. He, Z. Shao, and Z. Li, "Mulls: Versatile lidar slam via multi-metric linear least square," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 11 633–11 640.

[7] K. Chen, B. T. Lopez, A.-a. Agha-mohammadi, and A. Mehta, "Direct lidar odometry: Fast localization with dense point clouds," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2000–2007, 2022.

[8] J. Yang, H. Li, D. Campbell, and Y. Jia, "Go-icp: A globally optimal solution to 3d icp point-set registration," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 11, pp. 2241–2254, 2015.

[9] Q.-Y. Zhou, J. Park, and V. Koltun, "Fast global registration," in *European conference on computer vision*. Springer, 2016, pp. 766–782.

[10] P. C. Lusk, K. Fathian, and J. P. How, "Clipper: A graph-theoretic framework for robust data association," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 13 828–13 834.

[11] H. Lim, S. Yeon, S. Ryu, Y. Lee, Y. Kim, J. Yun, E. Jung, D. Lee, and H. Myung, "A single correspondence is enough: Robust global registration to avoid degeneracy in urban environments," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 8010–8017.

[12] H. Yang, J. Shi, and L. Carlone, "Teaser: Fast and certifiable point cloud registration," *IEEE Transactions on Robotics*, vol. 37, no. 2, pp. 314–333, 2020.

[13] J. Shi, H. Yang, and L. Carlone, "Robin: a graph-theoretic approach to reject outliers in robust estimation using invariants," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 13 820–13 827.

[14] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss, "Rangenet++: Fast and accurate lidar semantic segmentation," in *2019 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2019, pp. 4213–4220.

[15] H. Tang, Z. Liu, S. Zhao, Y. Lin, J. Lin, H. Wang, and S. Han, "Searching efficient 3d architectures with sparse point-voxel convolution," in *European conference on computer vision*. Springer, 2020, pp. 685–702.

[16] J. Serafin and G. Grisetti, "Nicp: Dense normal based point cloud registration," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 742–749.

[17] A. P. Bustos and T.-J. Chin, "Guaranteed outlier removal for point cloud registration with correspondences," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 12, pp. 2868–2882, 2017.

[18] C. Choy, J. Park, and V. Koltun, "Fully convolutional geometric features," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 8958–8966.

[19] J. Briales and J. Gonzalez-Jimenez, "Convex global 3d registration with lagrangian duality," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4960–4969.

[20] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (fpfh) for 3d registration," in *2009 IEEE international conference on robotics and automation*. IEEE, 2009, pp. 3212–3217.

[21] A. Zaganidis, L. Sun, T. Duckett, and G. Cielniak, "Integrating deep semantic segmentation into 3-d point cloud registration," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 2942–2949, 2018.

[22] P. Zhou, X. Guo, X. Pei, and C. Chen, "T-loam: truncated least squares lidar-only odometry and mapping in real time," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2021.

[23] C. Bron and J. Kerbosch, "Algorithm 457: finding all cliques of an undirected graph," *Communications of the ACM*, vol. 16, no. 9, pp. 575–577, 1973.

[24] C. Villani, *Optimal transport: old and new*. Springer, 2009, vol. 338.

[25] R. A. Rossi, D. F. Gleich, and A. H. Gebremedhin, "Parallel maximum clique algorithms with applications to network analysis," *SIAM Journal on Scientific Computing*, vol. 37, no. 5, pp. C589–C616, 2015.

[26] 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 Robotics and Automation Letters*, vol. 5, no. 2, pp. 1127–1134, 2020.

[27] M. J. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision," *International journal of computer vision*, vol. 19, no. 1, pp. 57–91, 1996.

[28] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[29] K. Koide, M. Yokozuka, S. Oishi, and A. Banno, "Voxelized gicp for fast and accurate 3d point cloud registration," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 11 054–11 059.

[30] R. Chandra, L. Dagum, D. Kohr, R. Menon, D. Maydan, and J. McDonald, *Parallel programming in OpenMP*. Morgan kaufmann, 2001.

[31] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "Semantickitti: A dataset for semantic scene understanding of lidar sequences," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9297–9307.

[32] X. Chen, A. Milioto, E. Palazzolo, P. Giguere, J. Behley, and C. Stachniss, "Suma++: Efficient lidar-based semantic slam," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 4530–4537.