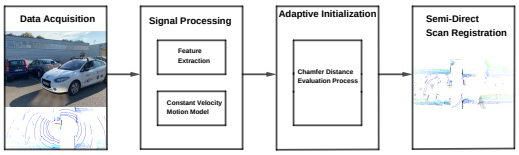
Uncertainty-Aware Adaptive Semi-Direct LiDAR Scan Matching for Ground Vehicle Positioning

Songming Chen, Vincent Frémont

Uncertainty-Aware Adaptive Semi-Direct LiDAR Scan Matching for Ground Vehicle Positioning

Songming Chen and Vincent Frémont

***Abstract*— Iterative Closest Point (ICP) algorithms are widely used in the literature for the estimation of relative transformations using 3D LiDAR point clouds. This class of algorithms proves to be efficient when the 3D data share sufficient overlapping parts and a good initial guess is provided. However, large relative motions and mutual occlusion of objects in real-road scenarios hinder traditional optimization-based ICP from achieving optimal estimation. This paper explores both direct and feature-based 3D LiDAR scan matching using the ICP framework in different contexts, such as parking, residential, urban, and highway scenarios. In order to guarantee the scan matching performances in scenarios with scarce geometric information and fast ego-vehicle motion, we propose an adaptive semi-direct scan matching method together with an alignment uncertainty quantification. The proposed semi-direct scan matching is tested on both the public KITTI and self-recorded LS2N datasets, which accomplishes the robust 6 Degrees of Freedom (DoF) pose estimation and consistent scene reconstruction. We demonstrate that the proposed approach outperforms the state-of-the-art and achieves the leading results with $68.3\,\%$ average relative fitness and $5.71\,\mathrm{cm}$ average RMSE, respectively.**



Adaptive Semi-Direct LiDAR Scan Matching for Ground Vehicles Localization and Mapping

***Index Terms*— LiDAR-based sensing, Scan matching, Ego-motion, State estimation, Uncertainty quantification**

I. Introduction

PERCEPTION of the surrounding environment and self-positioning are important functions for intelligent vehicles. Global Navigation Satellite Systems (GNSS) allow for the estimation of the vehicle global position in outdoor environments. In GNSS-denied areas, the relative position is estimated using sensors such as the Light Detection and Ranging (LiDAR) that enable the construction of maps of the surrounding environment [1]. For safety-critical functions, LIDARs are used with video cameras to have better depth information and add redundancy [2]. This is mainly attributed to LiDAR's insensitivity to illumination conditions and omnidirectional field of view, which enables scale-aware 6 DoF pose estimation and precise 3D scene mapping.

LiDAR scan matching algorithms are categorized into direct and feature-based approaches based on matching raw data or local features in the LiDAR scans. The direct LiDAR scan matching method aligns the raw point clouds without distinguishing the feature points. The correspondences are built iteratively according to the closest neighbor criteria. And the optimal transformation is estimated that aligns the source and target point clouds. However, it needs to be noticed that the direct method always requires a good initial guess to start, and it does not work efficiently for partial overlapping and noisy scans. The feature-based LiDAR scan matching methods extract local features from raw point clouds and use high-dimension descriptors for matching these features. The established corresponding feature points will then be used to estimate the relative transformation, which increases the robustness against noise and erroneous matches. Nonetheless, in repetitive and feature-less environments, the feature-based method may perform poorly due to the absence of anchor features for scan alignment.

Semi-direct approaches were proposed in [3] [4] for vision-based state estimation to overcome the problem of visual illumination changes and loss of visual feature tracking. Basically, the feature-based method provides an initial estimate for the direct method, which then improves the accuracy and reliability of the estimation. Inspired by those performances in vision-based state estimation, we come up with a so-called semi-direct LiDAR scan matching approach that combines the conceptually complementary direct and feature-based scan matching. The proposed semi-direct LiDAR scan matching algorithm ensures that the registration is robust against the high-speed or large-rotation motion, or traveling in repetitive and feature-less environments. For vehicles driving on roads, the highly isotropic ground points are inevitably scanned and matched, which may bias the state estimation. In order to reduce the registration lag effects, ground points need to be

Manuscript received July 25, 2022; revised January 5, 2023; revised February 28, 2023; revised March 7, 2023; accepted March 17, 2023. This work was carried out in the framework of the NExT Senior Talent Chair DeepCoSLAM (ANR-16-IDEX-0007), through the program Investments for the Future managed by the French National Agency for Research. The first author was also funded by China Scholarship Council.

The authors are with Nantes Université, École Centrale de Nantes, LS2N, UMR 6004, 44000 Nantes, France (e-mail: songming.chen@ec-nantes.fr; vincent.fremont@ec-nantes.fr).

removed beforehand. Thus, we first apply a simple but efficient method to identify and clear the ground points with the prior vehicle height and planar motion assumptions. Moreover, an uncertainty model of the relative transformation is also performed to evaluate the scan matching quality, which also lays the basis for pose-graph optimization [5]. The contributions of this paper are the following:

1) A robust semi-direct 3D LiDAR scan matching approach is proposed, whose adaptive initialization procedure facilitates accurate 3D point cloud registration.
2) The registration lag effect of 3D LiDAR scan matching is solved by implementing a consensus-based ground plane fitting and ground points removal. The consensus is based on the least summed distance of the potential ground points to the estimated plane.
3) An uncertainty model for the 3D LiDAR scan matching is proposed and evaluated. The main factors which degrade LiDAR scan alignment performances are also analyzed.
4) The uncertainty model is integrated into the global pose-graph optimization to reduce the pairwise registration drift and to obtain consistent scene mapping.

The rest of this paper is organized as follows: In Section II, the related work and recent advances in LiDAR-based scan matching problems is covered. Then, the proposed methodology is detailed in Section III. After that, the experimental results and corresponding analysis are shown in Section IV. Finally, the conclusion and future perspectives are given in Section V.

II. Related work

In the past few decades, several achievements in 6 DoF pose estimation using 3D LiDAR scan matching have been obtained, which reach centimeter-level precision. In order to tightly align two time-consecutive point clouds, ICP-based algorithms are well-recognized methods to estimate the relative transformation of the LiDAR scans.

A. Direct scan matching

The direct ICP scan matching algorithm is firstly introduced by [6]. The ICP algorithm iteratively searches for the nearest neighbor in the target point cloud and builds the correspondences, which refines the optimal relative transformation through iterations. In [6], the point-to-point distance is used for the closest neighbor association, which might be too greedy when the two scans are captured with strong viewpoint changes. In order to mitigate this problem, the point-to-plane distance is proposed in [7] for robust data association. Compared with the point-to-point distance metric, the point-to-plane distance metric relaxes the strict point-to-point correspondence restrictions, which is more adaptable for partially overlapping scans in practice. With the purpose of unifying the point-to-point and point-to-plane distance metrics into a probabilistic framework, Generalized-ICP (G-ICP) has been proposed in [8]. The G-ICP framework aims to align the LiDAR scan surfaces instead of distinctive individual points, which is more tolerant to incorrect correspondences and also reduces the risks of being stuck in local minima for the partially overlapped LiDAR points clouds. Moreover, since the G-ICP framework leverages the probabilistic model for LiDAR scan alignment, it maintains better robustness against the measurement noise due to manufacturing limitations or extreme working conditions.

B. Feature-based scan matching

Instead of aligning the whole raw 3D point cloud, the feature-based methods extract distinctive feature points and build the correspondences based on pre-computed descriptors. For example, a double-layer feature-based method is proposed in [9] to extract ground and vertical features that match the corresponding scan observations. Also in [10], four planar feature points are extracted in the region of interest for accurate automotive LiDAR alignment inspection. In order to efficiently compute the features, the 3D point clouds are projected into rasterized range images in [11], where the visual ORB features can be trivially extracted for state estimation and loop closure. Recently, inspired by images-based approaches, viewpoint invariant 3D descriptors such as FPFH [12] and SHOT [13] have emerged to characterize the local patches in the LiDAR scans. Sample Consensus Initial Alignment (SAC-IA) has been proposed in [12] for the rough scan alignment. This approach is robust to sensor noise but might cause overfitting in information-deprived environments such as indoor corridors, outdoor tunnels, or highways. Furthermore, the emergence of data-driven methods is reflected in the estimation of the 6 DoF pose using an end-to-end approach as in PointLoc [14].

C. Hybrid scan matching

The accuracy of the standard ICP scan matching is highly dependent on the initialization process as reported in [15]. It means that a large deviation of the initial alignment may cause the ICP optimization to diverge or get stuck in local minima. Thus, hybrid methods that incorporate visual, IMU, GNSS, and range sensor information have been proposed for reliable state estimation and scene reconstruction. In [16], the multi-sensor data are tightly coupled for accurate rail vehicle localization and mapping even in degenerated environments. The visual and LiDAR sensor information are loosely coupled by covariance intersection in [17] for the robust inter-frame transformation estimation. Besides, a multi-sensor joint optimization approach has been proposed in [18] and the dual-layer optimization design ensures both local and global estimation consistency. At the same time, the RTK/IMU measurements are employed to verify the LiDAR scan matching accuracy to prevent degenerated estimation in [19]. Furthermore, the drift of the incremental LiDAR scan alignment can be eliminated by the vision-based bag-of-words place recognition technique, as demonstrated in [20].

D. Scan matching uncertainty

For the LiDAR scan matching, it is crucial to evaluate the corresponding uncertainty for the estimated relative transformation. This is beneficial for sensor fusion or weight

assignment in pose graphs to mitigate local errors. Another advantage of uncertainty modeling is to bound the estimation error within a known confidence interval, which is the prerequisite for road obstacle avoidance and interactive navigation tasks. There exist several methods for the LiDAR scan matching covariance estimation. A singular value decomposition-based method is proposed for the LiDAR-based 6 DoF pose error propagation and covariance estimation [1]. Based on the estimated covariance, the accumulated errors can then be corrected with the GNSS data. A closed-form covariance estimation method is presented in [21], which is based on the objective function linearization around the optimal estimation. In order to correctly propagate the uncertainty from the measurement space to the estimation domain, the second-order derivatives of the objective function are calculated and applied to the initial measurement noise. Nevertheless, it needs to be noted that the closed-form covariance estimation only considers the uncertainty caused by the sensor noise; thus, it could not apply to the local minima situations. Monte Carlo simulation [22] is another branch of the uncertainty modeling methods, which iteratively generates different input point clouds to test the estimated transformation sensitivity. Then several estimates could be computed with different input samples. Based on Monte Carlo simulation scan matching results, the distribution of the relative transformation estimation can be constructed. Nevertheless, brute force sampling is time-consuming and limits the application scenarios.

III. Proposed Approach

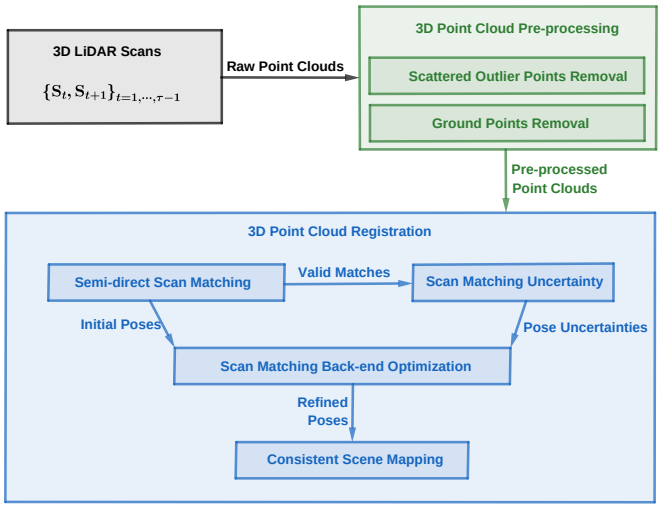


Fig. 1. Overview of the proposed semi-direct LiDAR scan matching and scene mapping framework

The pipeline of the proposed approach is shown in Fig. 1. For consecutive LiDAR scans perceived in the vehicle local frame, point cloud pre-processing is necessary before implementing the scan matching. The point cloud pre-processing includes scattered outliers and ground points removal, which will significantly ease the false correspondence matching problem. Afterwards, a robust semi-direct scan matching is applied to compute the 6 DoF motion and to estimate the relative transformation covariance. Since the scan matching-based localization is bound to drift over time, a pose-graph optimization is then used to reduce the local drift and to render a globally consistent mapping. In order to make the proposed approach easy to follow, the notations for the semi-direct scan matching are explained in Table I.

TABLE I
NOTATION TABLE FOR THE SEMI-DIRECT SCAN MATCHING

Symbol	Meaning
$(\mathbf{x}_i, \mathbf{x}_i^{'})$	The matched 3D points in the source and target point clouds
$\mathbb{V}_i$	The voxel point set for the point cloud downsampling
$\mathbf{c}_i$	The centroid of 3D points inside $\mathbb{V}_i$
$\lvert\cdot\rvert$	The cardinality of a 3D point set
$\mathbf{x}_l^G$	The potential 3D ground point
$\hat{\mathbb{G}}_t$	The estimated potential 3D ground points set
h_{prior}	The LiDAR installation height prior
$\vec{\mathbf{n}}_{\mathbf{x}_l}^G$	The unit normal vector around $\mathbf{x}_l^G$
$\mathbf{x}_{lz}^G$	The z-coordinate of $\mathbf{x}_l^G$
γ_{off}	The angle offset between $\vec{\mathbf{n}}_{\mathbf{x}_l}^G$ and vertical unit vector
$\hat{\mathbf{P}}_G$	The estimated ground plane
$\mathbf{P}_G^*$	The optimal ground plane
h_{off}	The distance offset between $\mathbf{x}_l^G$ and $\mathbf{P}_G^*$
$\{\mathbb{S}_t^{\triangle}, \mathbb{S}_{t+1}^{\triangle}\}$	The pre-processed consecutive LiDAR scans
$\mathbb{C}_{ii'}$	The inlier correspondence set
${}^{t+1}\mathbf{T}_t^*$	The optimal transformation that aligns the consecutive LiDAR scans
$\vec{\mathbf{n}}_i^{'}$	The unit normal vector around $\mathbf{x}_i^{'}$
$\mathbb{N}(\mathbf{x}_i)$	The spherical neighbor point set around $\mathbf{x}_i$
$\bar{\mathbf{x}}$	The mean position of points in $\mathbb{N}(\mathbf{x}_i)$
$\mathbf{\Sigma}(\mathbf{x}_i)$	The scattered matrix of points in $\mathbb{N}(\mathbf{x}_i)$
$\{\lambda_i^1, \lambda_i^2, \lambda_i^3\}$	The eigenvalues of the $\mathbf{\Sigma}(\mathbf{x}_i)$ eigenvalue decomposition
$\{\mathbf{e}_i^1, \mathbf{e}_i^2, \mathbf{e}_i^3\}$	The eigenvectors of the $\mathbf{\Sigma}(\mathbf{x}_i)$ eigenvalue decomposition
$\{\gamma_{21}, \gamma_{32}\}$	The salience thresholds for the ISS features detection
d_{ij}	The edge distance between the point $\mathbf{x}_i$ and $\mathbf{x}_j$
$CD(\cdot)$	The Chamfer Distance between the aligned point clouds
$\mathbf{T}_f$	The feature-based coarse transformation that aligns the consecutive LiDAR scans
$\mathbf{T}_{init}$	The initial transformation that aligns the consecutive LiDAR scans
L_{size}	The downsampling voxel size
${}^{t+1}\hat{\mathbf{T}}_t$	The estimated transformation that aligns the consecutive LiDAR scans
$\mathbf{T}_t^{-1}\mathbf{T}_{t+1}$	The ground truth transformation that aligns the consecutive LiDAR scans
$\lVert(\cdot)\rVert^2$	The Euclidean norm
$\mathbf{r} = [\alpha, \beta, \gamma]^T$	The XYZ Euler angles parameterization
$\mathbf{t} = [x, y, z]^T$	The XYZ translation parameterization
$\mathbf{R}(\cdot)$	The rotation angle matrix parameterization
$(\cdot)^\wedge$	The skew symmetric matrix of cross product multiplication
$\mathbf{D}(\cdot)$	The inconsistency indicator for uncertainty modeling
$\mathbf{\Lambda}$	The quadratic form information matrix

A. 3D Point cloud pre-processing

1) Scattered outlier points removal: Sparse and non-permanent scattered scene components, like tree leaves or bushes, may pose challenges for accurate scan matching. In our approach, a Gaussian distribution-based outlier rejection method is performed to remove the sparse and isolated points. Subsequently, the voxel grid filter with the leaf size L_{size} is applied to down-sample and to approximate the point clouds for efficient scan alignment. For the points inside each voxel

point set $\{\mathbf{x}_i : \mathbf{x}_i \in \mathbb{V}_i\}$, they are approximated with their centroid $\mathbf{c}_i = \sum_{\mathbf{x}_i \in \mathbb{V}_i} \frac{\mathbf{x}_i}{|\mathbb{V}_i|}$, where $|\cdot|$ denotes the cardinality of a set. This operation will greatly reduce the number of points without losing the raw points distribution.

2) Ground points removal: When traveling on real roads, the point cloud acquired by the LiDAR sensor naturally contains many ground points. Ground points on flat roads tend to be isotropic and they encode little geometric information for the data association step. A slight deviation during iterative nearest neighbor tracking can cause ground point mismatches and distort the entire scan alignment.

To address these problems, a consensus-based method is used to segment ground points, assuming the vehicle follows a planar motion and the LiDAR installation height is known. On this basis, a set of potential ground points $\{\mathbf{x}_l^G : \mathbf{x}_l^G \in \hat{\mathbb{G}}_t\}$ are selected, incorporating all the points that locate h_{prior} along the z-axis under the vehicle roof. Hereby, h_{prior} is highly linked to the LiDAR installation height. Meanwhile, another condition for potential points selection is that the unit normal vector $\vec{\mathbf{n}}_{\mathbf{x}_l}^G$ around the potential ground point $\mathbf{x}_l^G$ needs to stay within an offset of γ_{off} from the z-axis of the LiDAR local frame. The offset γ_{off} allows the slightly sloped terrain points to participate in the ground plane estimation, considering the fact that the driving road is not always fully flat.

$$\mathbf{x}_{lz}^G < -h_{\text{prior}}, \quad [0,0,1] \cdot \vec{\mathbf{n}}_{\mathbf{x}_l}^G > cos(\gamma_{\text{off}}) \tag{1}$$

After that, the consensus-based method iteratively picks three non-collinear points $\{\mathbf{x}_i^G, \mathbf{x}_j^G, \mathbf{x}_k^G\}$ within the potential ground point set $\hat{\mathbb{G}}_t$. The selected points are used to fit the ground plane equation $\hat{\mathbf{P}}_G : Ax+By+Cz+D = 0$. The distances of the potential ground points $\{\mathbf{x}_l^G : \mathbf{x}_l^G \in \hat{\mathbb{G}}_t\}$ to the estimated plane $\hat{\mathbf{P}}_G$ are then computed and summed up. The summed distances are used to vote for the plane candidates until the convergence criteria are met. At the end of the iterations, the plane with the least summed point-to-plane distances is chosen as the optimal ground plane $\mathbf{P}_G^*$. Once the ground plane is identified, points in $\hat{\mathbb{G}}_t$ within distance threshold h_{off} to the optimal ground plane $\mathbf{P}_G^*$ are also considered as ground points. This strategy helps to increase robustness against measurement noise for ground point segmentation. Finally, after the scattered outliers and the ground points removal, the pre-processed LiDAR scans $\{\mathbb{S}_t^{\triangle}, \mathbb{S}_{t+1}^{\triangle}\}_{t=1,\cdots,\tau-1}$ can be obtained for the following 3D point cloud registration.

B. 3D Point cloud registration

1) Semi-direct scan matching: According to [15], direct scan matching with raw point clouds is less efficient for large baseline motions, while feature-based methods lose their advantages to bootstrap from local minima in feature-less environments. To complement the weakness of individual direct or feature-based methods, a hybrid semi-direct approach is proposed in this paper. The semi-direct LiDAR scan matching pipeline is detailed in Algorithm 1 at the end of Section III-B.1. Hereby, a feature-based method is leveraged to estimate a coarse but globally consistent inter-frame transformation that serves as a prior for the following multi-scale direct dense point cloud alignment. In the proposed semi-direct approach, the point-to-plane distance metric is adopted for the optimal relative transformation ${}^{t+1}\mathbf{T}_t^*$ estimation,

$${}^{t+1}\mathbf{T}_t^* = \underset{{}^{t+1}\mathbf{T}_t}{\operatorname{argmin}} \sum_{(\mathbf{x}_i, \mathbf{x}_i^{'}) \in \mathbb{C}_{ii'}} \|(\mathbf{x}_i^{'} - {}^{t+1}\mathbf{T}_t.\mathbf{x}_i)^T \vec{\mathbf{n}}_i^{'}\|^2 \tag{2}$$

where $\mathbb{C}_{ii'}$ is the inlier correspondence set incorporating the matched point pairs $\{(\mathbf{x}_i, \mathbf{x}_i^{'})\}$ in the source and target LiDAR scans. The plane unit normal vectors $\vec{\mathbf{n}}_i^{'}$ around $\mathbf{x}_i^{'}$ in the target LiDAR scan could effectively guide the scan matching process to distinguish points lying on different surfaces and discard unreliable correspondences $\{(\mathbf{x}_i, \mathbf{x}_i^{'})\}$. For the sake of reducing spurious correspondences, the selected features for the sparse point cloud alignment need to be distinctive and invariant to viewpoint changes. To this end, the salience-based intrinsic Shape Signatures (ISS) [23] keypoints extraction method is adopted, in which the salience measure is derived from the scatter matrix $\mathbf{\Sigma}(\mathbf{x}_i)$ eigenvalue decomposition.

$$\bar{\mathbf{x}} = \frac{1}{|\mathbb{N}(\mathbf{x}_i)|} \sum_{\mathbf{x}_j \in \mathbb{N}(\mathbf{x}_i)} \mathbf{x}_j \tag{3}$$

$$\mathbf{\Sigma}(\mathbf{x}_i) = \frac{1}{|\mathbb{N}(\mathbf{x}_i)|} \sum_{\mathbf{x}_j \in \mathbb{N}(\mathbf{x}_i)} (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T \tag{4}$$

where $\mathbb{N}(\mathbf{x}_i)$ are the spherical neighbors of $\mathbf{x}_i$ within a predefined radius and $|\cdot|$ denotes the number of points within the set. With the scatter matrix $\mathbf{\Sigma}(\mathbf{x}_i)$ eigenvalue decomposition, the descending magnitude eigenvalues $\{\lambda_i^1, \lambda_i^2, \lambda_i^3\}$ can be calculated, and their corresponding eigenvectors $\{\mathbf{e}_i^1, \mathbf{e}_i^2, \mathbf{e}_i^3\}$ are used to build the intrinsic reference frame. To ensure the distinctiveness of the local features, the ratios of the sequential eigenvalues of the scatter matrix $\mathbf{\Sigma}(\mathbf{x}_i)$ are required to not exceed the salience thresholds γ_{21} and γ_{32} for the pruning purpose [23].

$$\lambda_i^2/\lambda_i^1 < \gamma_{21}, \quad \lambda_i^3/\lambda_i^2 < \gamma_{32} \tag{5}$$

where γ_{21} is the upper bound on the ratio between the second and the first eigenvalue and γ_{32} is the upper bound on the ratio between the third and the second eigenvalue returned by the scatter matrix eigenvalue decomposition. The salience thresholds γ_{21} and γ_{32} guarantee that the established intrinsic reference frame exhibits maximum salience along the principal directions, which makes features more informative and recognizable from various viewpoints. As long as the ISS features are extracted from the 3D point cloud, the efficient Fast Point Feature Histogram (FPFH) descriptor [12] is used for robust data association. FPFH is basically a 33-dimensional vector that characterizes the local geometry around a point. It efficiently enriches the keypoints description and helps to establish the keypoints correspondences. Through iteratively taking three pairs of matched feature points $\{(\mathbf{x}_i, \mathbf{x}_i^{'}), (\mathbf{x}_j, \mathbf{x}_j^{'}), (\mathbf{x}_k, \mathbf{x}_k^{'})\}$ and implementing correspondence consistency (edge similarity) check, the transformation matrix $\mathbf{T}_f$ that optimally aligns the sparse feature points could be obtained by the Random Sample Consensus (RANSAC) method. The correspondence consistency check is to verify the edges distance $\{(d_{ij}, d_{ik}, d_{jk}), (d_{ij}^{'}, d_{ik}^{'}, d_{jk}^{'})\}$

formed by the features in each frame. This prevents mismatches in the environment with repeatable features. The correspondences are then considered valid if the features are not collinear

$$\begin{aligned}(\mathbf{x}_i - \mathbf{x}_j)/d_{ij} \neq (\mathbf{x}_i - \mathbf{x}_k)/d_{ik} \\ (\mathbf{x}_i^{'} - \mathbf{x}_j^{'})/d_{ij}^{'} \neq (\mathbf{x}_i^{'} - \mathbf{x}_k^{'})/d_{ik}^{'}\end{aligned} \tag{6}$$

and their formed edges have similar length as

$$0.9 < d_{ij}/d_{ij}^{'},\ d_{ik}/d_{ik}^{'},\ d_{jk}/d_{jk}^{'} < 1.1 \tag{7}$$

Given a set of associated feature points, the RANSAC-based scan matching does not require an initial guess and is more robust to outliers. This alleviates the problem of getting stuck at the local minima. However, in a feature-less environment with few distinctive features to be extracted, feature tracking tends to be hard to manage. In this way, we need to evaluate the estimated transformation $\mathbf{T}_f$ from the feature-based method and the transformation heuristics ${}^{t}\hat{\mathbf{T}}_{t-1}$. Hereby, the transformation heuristics come from the constant velocity motion model and the evaluation metric used is the Chamfer Distance. The Chamfer Distance $CD(\cdot)$ in (8) is a metric to measure the tightness of two aligned point clouds $\{\mathbb{S}_t, \mathbb{S}_{t+1}\}$, with the expression as

$$\begin{aligned}CD\left(\mathbb{S}_t, \mathbb{S}_{t+1}\right) = \frac{1}{|\mathbb{S}_t|}\sum_{\mathbf{x}_i \in \mathbb{S}_t} \min_{\mathbf{x}_i^{'} \in \mathbb{S}_{t+1}} \|\mathbf{x}_i - \mathbf{x}_i^{'}\|^2 \\ + \frac{1}{|\mathbb{S}_{t+1}|}\sum_{\mathbf{x}_j^{'} \in \mathbb{S}_{t+1}} \min_{\mathbf{x}_j \in \mathbb{S}_t} \|\mathbf{x}_j - \mathbf{x}_j^{'}\|^2\end{aligned} \tag{8}$$

and $|\cdot|$ stands for the cardinality of the point set. This step helps to choose a more reliable initialization point $\mathbf{T}_{init}$ for the dense alignment.

$$\mathbf{T}_{init} = \underset{{}^{t}\hat{\mathbf{T}}_{t-1}, \mathbf{T}_f}{\operatorname{argmin}} \{CD({}^{t}\hat{\mathbf{T}}_{t-1}\mathbb{S}_t^{\triangle}, \mathbb{S}_{t+1}^{\triangle}),\ CD(\mathbf{T}_f\mathbb{S}_t^{\triangle}, \mathbb{S}_{t+1}^{\triangle})\} \tag{9}$$

Then the multi-scale pyramidal dense point cloud alignment is implemented to refine the coarse initial transformation $\mathbf{T}_{init}$. The multi-scale pyramid defines a two-layer maximum correspondence distance as $3\times L_{\text{size}}$ and L_{size}, that are proportional to the downsampling voxel leaf size. The first layer point cloud alignment has the maximum correspondence distance of $3\times L_{\text{size}}$, whose convergence criteria is simple to reach. As a result, the first layer alignment further eliminates the effect of false correspondence and reduces the risk of being stuck in the local minima. Then, based on the result from the first layer, the second layer pyramid searches in a finer scale for the final tight point-cloud alignment. This design adaptively determines the nearest neighbor searching radius without fine-tuning, which greatly facilitates optimization convergence for partially overlapping point clouds.

Algorithm 1 Semi-Direct Scan Matching Algorithm

Input: Consecutive LiDAR scans $\{\mathbb{S}_t, \mathbb{S}_{t+1}\}_{t=1,\cdots,\tau-1}$, Relative transformation heuristics ${}^{t}\hat{\mathbf{T}}_{t-1}$

Output: Estimated transformation ${}^{t+1}\hat{\mathbf{T}}_t$ that tightly aligns the consecutive scans

Initialize ${}^{1}\hat{\mathbf{T}}_0 \leftarrow \mathbf{I}_4$
while New LiDAR scan arrives **do**
  Remove the scattered outliers and ground points to get the processed LiDAR scans $\{\mathbb{S}_t^{\triangle}, \mathbb{S}_{t+1}^{\triangle}\}$
  Apply the ISS keypoints detection, FPFH description to $\{\mathbb{S}_t^{\triangle}, \mathbb{S}_{t+1}^{\triangle}\}$
  Implement the keypoints correspondence consistency check and only keep the valid ones using (6), (7)
  Obtain the rough transformation $\mathbf{T}_f$ with the consistent keypoints correspondences using RANSAC
  Evaluate the Chamfer Distance $CD(\cdot)$ of ${}^{t}\hat{\mathbf{T}}_{t-1}$ and $\mathbf{T}_f$, then choose the initial guess $\mathbf{T}_{init}$ using (9)
  Conduct the multi-scale dense alignment with $\mathbf{T}_{init}$ and get the relative transformation ${}^{t+1}\hat{\mathbf{T}}_t$
  Update the LiDAR scan timestamp $t \leftarrow t+1$
end while

2) Scan matching uncertainty: In the proposed pipeline, the uncertainty quantification is integrated into the pose-graph optimization to limit the frame-to-frame estimation drift. In order to maintain computation efficiency and estimation accuracy, the derivative-free covariance estimation method is adopted to assess the scan matching quality. Owing to the point cloud pre-processing and semi-direct coarse-to-fine point cloud alignment, a considerable number of outliers are filtered out. And we manage to predict the LiDAR scan matching uncertainty in large-scale outdoor scenarios with measurement noise and large viewpoint changes. Given the two consecutive LiDAR scans in the local sensor frame $\mathbb{S}_t$ and $\mathbb{S}_{t+1}$, the estimated relative transformation ${}^{t+1}\hat{\mathbf{T}}_t$ tightly aligns the corresponding points[1] $\{\mathbf{x}_i : \mathbf{x}_i \in \mathbb{S}_t\}$ and $\{\mathbf{x}_i^{'} : \mathbf{x}_i^{'} \in \mathbb{S}_{t+1}\}$ in the point clouds. Essentially, the uncertainty estimation is based on the inconsistency indicator $\mathbf{D}({}^{t+1}\hat{\mathbf{T}}_t, \mathbf{T}_t, \mathbf{T}_{t+1})$ in (14), which reflects the overall discrepancies between the points in the inlier correspondence set $\mathbb{C}_{ii'}$. The inlier correspondence set $\mathbb{C}_{ii'}$ incorporates all valid pairwise correspondences $(\mathbf{x}_i, \mathbf{x}_i^{'})$ within a predefined euclidean distance threshold $\left\|{}^{t+1}\hat{\mathbf{T}}_t\mathbf{x}_i - \mathbf{x}_i^{'}\right\|^2 < \epsilon$. The error tolerance ϵ is linked to the downsampling voxel leaf size L_{size}, which is set as $0.1\,\text{m}$ in our case. And the information matrix can then be extracted via the local parameterization $\boldsymbol{\xi} = [\mathbf{r}, \mathbf{t}]^T = [\alpha, \beta, \gamma, x, y, z]^T$ that expresses the estimated transformation ${}^{t+1}\hat{\mathbf{T}}_t$ and true value transformation $\mathbf{T}_t^{-1}\mathbf{T}_{t+1}$ discrepancies in such manner:

$$\mathbf{T}_t^{-1}\mathbf{T}_{t+1}{}^{t+1}\hat{\mathbf{T}}_t \approx \begin{bmatrix}\mathbf{I}_3 + \mathbf{r}^{\wedge} & \mathbf{t} \\ 0 & 1\end{bmatrix} \tag{10}$$

where $\mathbf{r} = [\alpha, \beta, \gamma]^T$ represents the rotation angles[2] along the $x-y-z$ axes sequentially, and $\mathbf{t}$ stands for the corresponding

[1] $\{\mathbf{x}_i\}$ are points with the homogeneous coordinates and $\{\mathbf{x}_{i_{\text{vec}}}^{\wedge}\}$ are skew symmetric matrices of points with the vectorized cartesian coordinates

[2] For the local parameterization of transformation discrepancy, the rotation angles are far away from their singular positions of $\frac{\pi}{2}$, thus the gimbal lock issue of Euler angle parametrization is avoided

translation parts. And the approximation is made for the infinitesimal rotations angles:

$$\mathbf{R} = \mathbf{R}_z(\gamma)\mathbf{R}_y(\beta)\mathbf{R}_x(\alpha) \approx \mathbf{I}_3 + \begin{bmatrix} 0 & -\gamma & \beta \\ \gamma & 0 & -\alpha \\ -\beta & \alpha & 0 \end{bmatrix} \quad (11)$$

After that, with the cross product $(\cdot)\times$ and skew symmetric matrix $(\cdot)^\wedge$ substitution,

$$\mathbf{r} \times \mathbf{x}_{i_{\text{vec}}} = -\mathbf{x}_{i_{\text{vec}}} \times \mathbf{r} = \mathbf{r}^\wedge \mathbf{x}_{i_{\text{vec}}} = -\mathbf{x}_{i_{\text{vec}}}^\wedge \mathbf{r} \quad (12)$$

the inconsistency indicator can be approximated as:

$$\begin{aligned} \mathbf{D}({}^{t+1}\hat{\mathbf{T}}_t, \mathbf{T}_t, \mathbf{T}_{t+1}) &= \sum_{(\mathbf{x}_i, \mathbf{x}_i^{'}) \in \mathbb{C}_{ii'}} \left\| {}^{t+1}\hat{\mathbf{T}}_t \mathbf{x}_i - \mathbf{x}_i^{'} \right\|^2 \\ &= \sum_{(\mathbf{x}_i, \mathbf{x}_i^{'}) \in \mathbb{C}_{ii'}} \left\| {}^{t+1}\hat{\mathbf{T}}_t \mathbf{x}_i - \mathbf{T}_{t+1}^{-1} \mathbf{T}_t \mathbf{x}_i \right\|^2 \\ &= \sum_{(\mathbf{x}_i, \mathbf{x}_i^{'}) \in \mathbb{C}_{ii'}} \left\| \mathbf{T}_t^{-1} \mathbf{T}_{t+1} {}^{t+1}\hat{\mathbf{T}}_t \mathbf{x}_i - \mathbf{x}_i \right\|^2 \\ &\approx \sum_{(\mathbf{x}_i, \mathbf{x}_i^{'}) \in \mathbb{C}_{ii'}} \left\| \mathbf{r} \times \mathbf{x}_{i_{\text{vec}}} + \mathbf{t} \right\|^2 \\ &= \sum_{(\mathbf{x}_i, \mathbf{x}_i^{'}) \in \mathbb{C}_{ii'}} \left\| \begin{bmatrix} -\mathbf{x}_{i_{\text{vec}}}^\wedge & \mathbf{I_3} \end{bmatrix} \boldsymbol{\xi} \right\|^2 \\ &= \boldsymbol{\xi}^T \boldsymbol{\Lambda} \boldsymbol{\xi} \end{aligned} \quad (13)$$

where $\boldsymbol{\Lambda}$ is the information matrix in the quadratic form with the expression as follows:

$$\boldsymbol{\Lambda} = \sum_{(\mathbf{x}_i, \mathbf{x}_i^{'}) \in \mathbb{C}_{ii'}} \begin{bmatrix} -\mathbf{x}_{i_{\text{vec}}}^\wedge & \mathbf{I_3} \end{bmatrix}^T \begin{bmatrix} -\mathbf{x}_{i_{\text{vec}}}^\wedge & \mathbf{I_3} \end{bmatrix} \quad (14)$$

The information matrix gives a direct reflection of the tightness of two LiDAR scans being aligned, and its inverse $\boldsymbol{\Lambda}^{-1}$ can be considered as the covariance matrix to model the scan matching uncertainty. Compared with the scan alignment itself, the covariance matrix is calculated only once at the final iteration, and its estimation time is negligible since it is derivative-free.

3) Scan matching back-end optimization: In order to reduce the LiDAR scan matching pairwise registration drift, the uncertainty quantification can be integrated into the global pose-graph optimization [5]. A pose graph is linked with edges which represent pairwise registration, and the deviated poses are adjusted and smoothed based on the edge constraints to optimize the overall discrepancies among the co-visible LiDAR scans. With such back-end optimization, the refined poses for point clouds registration can be obtained, which lays the basis for consistent scene mapping.

IV. EXPERIMENTAL RESULTS

To evaluate the performances of the proposed Semi-Direct Scan Matching (SD-SM) approach, extensive experiments have been carried out using the public KITTI [24] and self-recorded LS2N datasets at the Centrale Nantes Campus. Compared with more recent public datasets such as Nuscenes [25] and PandaSet [26], the KITTI dataset is more dedicated to the ego-motion estimation with available precise 6 DoF ground truth poses. The KITTI dataset point cloud is acquired with a Velodyne HDL-64E laser scanner, which has 64 channels with a maximum range of 120 m. Our self-recorded dataset is collected with a light-weight Velodyne VLP-16 LiDAR, which only has 16 layers and renders sparser point clouds than the HDL-64E. The scene characteristics for the experimental datasets are described in Table III, which include the scenario type, moving objects occurrence, and the number of road intersections. The parameters chosen for the proposed approach are summarized in the Table II. And all the LiDAR scans have been undistorted by compensating for the vehicle ego-motion. The two evaluation metrics are the Relative Fitness (RF) and Relative Root Mean Square Error (RMSE) of the inlier correspondences $\{\mathbb{C}_{ii'}\}$, that can be expressed as

$$RF = \frac{|\mathbb{C}_{ii'}|}{|\mathbb{S}_{t+1}|}, \quad RMSE = \frac{1}{|\mathbb{C}_{ii'}|} \sum_{(\mathbf{x}_i, \mathbf{x}_i^{'}) \in \mathbb{C}_{ii'}} \sqrt{\|\mathbf{T}\mathbf{x}_i - \mathbf{x}_i^{'}\|^2} \quad (15)$$

and $|\cdot|$ denotes the cardinality of the inlier correspondences set. The Relative Fitness (RF) measures the proportion of associated inliers $\{(\mathbf{x}_i, \mathbf{x}_i^{'}) : (\mathbf{x}_i, \mathbf{x}_i^{'}) \in \mathbb{C}_{ii'}\}$ among the full target cloud $\mathbb{S}_{t+1}$, and ideal relative fitness value is close to 1. RMSE measures the root mean square errors of all inlier correspondences, and lower RMSE value represents tighter alignment. In this section, we put the analysis emphasis on three principal parts that include the ground points removal, the semi-direct scan matching, and the uncertainty modeling.

TABLE II
THE PARAMETERS TABLE FOR THE PROPOSED SD-SM

L_{size}	h_{prior}	γ_{off}	h_{off}	γ_{21}	γ_{32}
0.1m	1.4m	$\frac{\pi}{5}$rad	0.2m	0.975	0.975

TABLE III
EXPERIMENTAL DATASET SEQUENCE SCENE CHARACTERISTICS

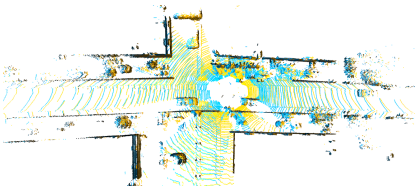
Dataset	Scenario	Moving Object Occurrence	No.Intersection
KITTI_00	residential	occasional	26
KITTI_01	highway	constant	2
KITTI_02	residential	occasional	19
KITTI_03	residential	occasional	1
KITTI_04	urban	constant	0
KITTI_05	residential	occasional	10
KITTI_06	residential	occasional	2
KITTI_07	residential	occasional	6
KITTI_08	residential	occasional	17
KITTI_09	residential	occasional	6
KITTI_10	residential	occasional	3
LS2N_00	parking	occasional	2
LS2N_01	parking	occasional	10

A. Ground points segmentation

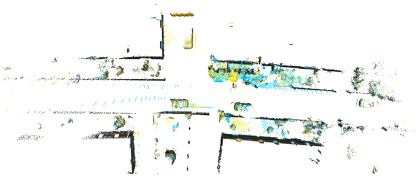
According to the KITTI sensor setup specifications[3], the LiDAR installation height is 1.73 m from the ground. This

[3]http://www.cvlibs.net/datasets/kitti/setup.php

verifies that our ground plane estimation $\mathbf{P}_G^* : -0.00x + 0.03y + 1.00z + 1.75 = 0$ is precise and reliable. This can be mainly attributed to the fact that the potential ground points set has been strictly selected and only contains few outliers. As a result, the consensus-based ground plane estimation avoids local minima, which facilitates accurate ground points segmentation. The registration lag effect, which means that erroneous matches on the ground points will disrupt the seamless alignment of the point clouds, is shown in Fig. 2. And the qualitative registration lag effect result in Fig. 2 reveals the necessity to remove the ground points before conducting the LiDAR scan matching.



(a) The ground points erroneous matches distract the point clouds from seamless alignment



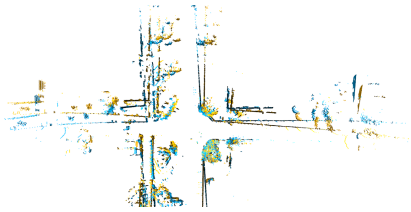
(b) The registration lag effect is eased by ground points removal

Fig. 2. Qualitative results for the registration lag effect caused by the ground points erroneous matches for the KITTI sequence 05

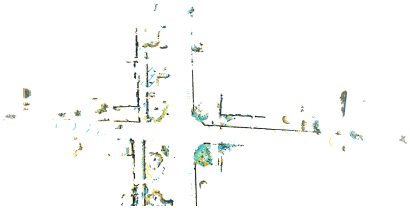
B. Registration results benchmarking

In order to benchmark the performances of the proposed semi-direct scan matching method, a thorough quantitative evaluation is conducted in various scenarios. The scenario types are listed as residential (KITTI 00, 02-03, 05-10), highway (KITTI 01), urban (KITTI 04), and outdoor parking (LS2N 00-01) areas. In these different driving scenarios, the vehicle ego-motion varies a lot. And it provides high-speed motion on the highway and mild motion in the parking area for us to investigate the performance of different LiDAR scan matching approaches. Table. IV lists the RF and RMSE metric values of the state-of-the-art scan matching methods as well as our approach, including Direct ICP [7] (D-ICP), Feature-based ICP [12] (F-ICP), Generalized ICP [8] (G-ICP) and our Semi-Direct Scan Matching (SD-SM). The scan registration with the relative fitness below $30\,\%$ is considered as invalid and the RMSE value is not computed in that case.

From Table. IV, it is shown that the dense point cloud alignment using the direct ICP is generally superior to the sparse feature-based ICP. Particularly, for the sequence of LS2N 00, direct ICP outperforms other approaches in terms of the fitness metric. It is due to the fact that the ego-vehicle moves slowly in the parking area. Since more raw point cloud information is leveraged for scan matching, the direct ICP has better performance when the subsequent scans that share adequate

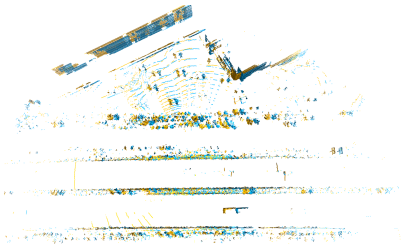


(a) Point cloud registration stuck in local minima for the D-ICP

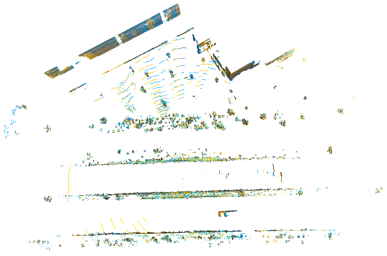


(b) Point cloud registration boosted from local minima for the SD-SM

Fig. 3. Point cloud registration results visualization at the road intersection for frame 133-134 in the KITTI sequence 05



(a) Point cloud registration lag in the featureless environment for the F-ICP



(b) Robust point cloud registration with the adaptive initialization strategy for the SD-SM

Fig. 4. Point cloud registration results visualization on the highway for frame 634-635 in the KITTI sequence 01

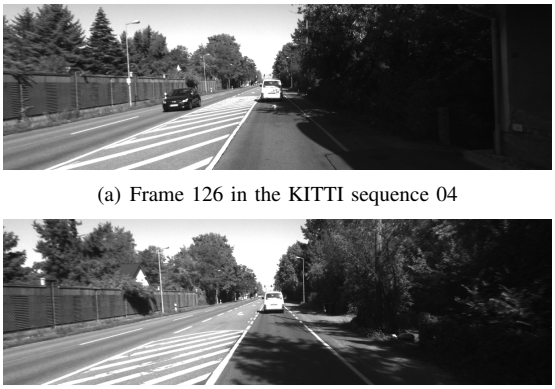
overlapping areas. The sparsity of the point clouds obtained by 16-layer LiDAR is another reason for the poor performance of the feature-based method. However, when the ego-vehicle experiences dramatic viewpoint changes, the direct ICP loses its advantages, especially with inadequate initialization of the identity matrix transformation. Due to the unanticipated decrease of overlaps, the inherent iterative nearest neighbor association strategy of D-ICP is vulnerable and tends to be stuck in the local minima, see Fig. 3(a). On the contrary, high-dimensional feature descriptors are invariant to viewpoint changes. And the feature-based scan matching is more robust for the large baseline motion across consecutive frames. The putative descriptor-based keypoints correspondences are constructed beyond iterative nearest neighbor searching and

TABLE IV
REGISTRATION RESULTS BENCHMARKING WITH METRICS OF RF(%) AND RMSE (cm)

Method / Dataset	D-ICP		F-ICP		G-ICP		SD-SM	
	RF	RMSE	RF	RMSE	RF	RMSE	RF	RMSE
KITTI_00	47.7%	5.83	27.0%	–	59.7%	5.68	**77.2%**	**5.53**
KITTI_01	26.8%	–	24.8%	–	27.2%	–	**49.4%**	**6.23**
KITTI_02	38.1%	6.18	29.9%	–	57.7%	5.94	**73.2%**	**5.75**
KITTI_03	47.2%	6.08	27.7%	–	62.8%	5.83	**67.2%**	**5.75**
KITTI_04	31.7%	**5.91**	30.9%	6.03	31.6%	5.92	**65.4%**	6.05
KITTI_05	43.8%	5.80	28.9%	–	61.4%	5.64	**76.4%**	**5.51**
KITTI_06	27.6%	–	19.8%	–	40.9%	5.84	**62.3%**	**5.75**
KITTI_07	56.8%	5.49	33.4%	5.98	62.4%	5.36	**78.7%**	**5.26**
KITTI_08	39.1%	6.05	24.5%	–	50.4%	5.91	**70.7%**	**5.68**
KITTI_09	28.3%	–	21.4%	–	52.6%	6.07	**68.8%**	**5.85**
KITTI_10	49.5%	5.95	30.7%	6.38	69.8%	5.62	**76.0%**	**5.53**
LS2N_00	**74.7%**	5.37	69.9%	5.56	74.3%	5.41	72.7%	**5.34**
LS2N_01	44.8%	6.11	32.9%	6.40	44.9%	6.07	**49.5%**	**6.02**
Average	42.8%	5.88	30.9%	6.07	53.5%	5.77	**68.3%**	**5.71**

association loop, which reduces the risks of being stuck in the local minima. With pre-established correspondences, the initialization-free sampling consensus-based method is applied to reject potential outliers and to obtain the optimal relative transformation. Nevertheless, it also needs to be mentioned that the feature-based scan matching is highly dependent on the feature detection procedure. This may lead to inaccurate or biased registration in the feature-less or feature-repetitive environments such as corridors or highways, see Fig. 4(a). The drawbacks of direct and feature-based scan matching reveal the necessity for their combination to provide more robust estimation in information-deprived environments. From the benchmarking results, it can be noticed that our semi-direct approach outperforms the state-of-the-art methods by a considerable margin in various scenarios. It is shown that the proposed method achieves the leading results with $68.3\,\%$ average relative fitness and $5.71\,\mathrm{cm}$ average RMSE distance, respectively. It can be seen that our semi-direct method implements the scan matching in a coarse-to-fine manner. Thus, it is less sensitive to unmodelled artifacts such as moving objects, undergoing the view occlusion, viewpoint changes, and information-deprived environments, see Fig. 3(b). For instance, it can be seen from Fig. 5 that a van constantly appears in front of the ego-vehicle for the KITTI sequence 04. The existence of moving objects in the scene greatly degrades the performance of both direct and feature-based ICP methods, see the 7th row of the Table. IV. In this case, the D-ICP and F-ICP tend to be stuck in the local minima, which may partially align the point clouds and provide low fitness registration results.

Indeed, the undergoing environments and realized ego-trajectories have a deep impact on the performance of LiDAR scan matching. To be more specific, our approach obtains promising results in the highway scenarios (KITTI 01), which is very challenging for direct and feature-based methods because of the relatively low frame rate compared with the high-speed ego-motion (see the 4th row of Table. IV). The adaptive initialization strategy in (9) ensures that the starting point for the pose optimization is not biased. With a reliable initialization point considering the constant velocity motion model, it takes fewer iterations for multi-scale dense point-cloud alignment to converge. And it is also more likely to obtain the global minima even for the fast motion, see Fig. 4(b).



(a) Frame 126 in the KITTI sequence 04

(b) Frame 143 in the KITTI sequence 04

Fig. 5. Moving objects in the scene degrade the state estimation performance for the KITTI sequence 04

C. Uncertainty modeling

TABLE V
THE OUTLIER FRAME NUMBER FOR POSE UNCERTAINTY ESTIMATION

Method / Sequence	Monte-Carlo simulation	Inconsistency indicator
KITTI_04	17, 126, 188 233, 239, 254	18, 90, 110 126, 132, 143
KITTI_05	233, 500, 592 1389, 1500	27, 233, 500 1393, 1500

The pose uncertainty estimation helps to bound the pose error within a known confidence interval. In this part, the proposed uncertainty modeling approach is benchmarked with the Monte-Carlo simulation-based method in [22]. The detailed implementation of the Monte-Carlo simulation-based scan matching uncertainty modeling can be seen from Algorithm 2.

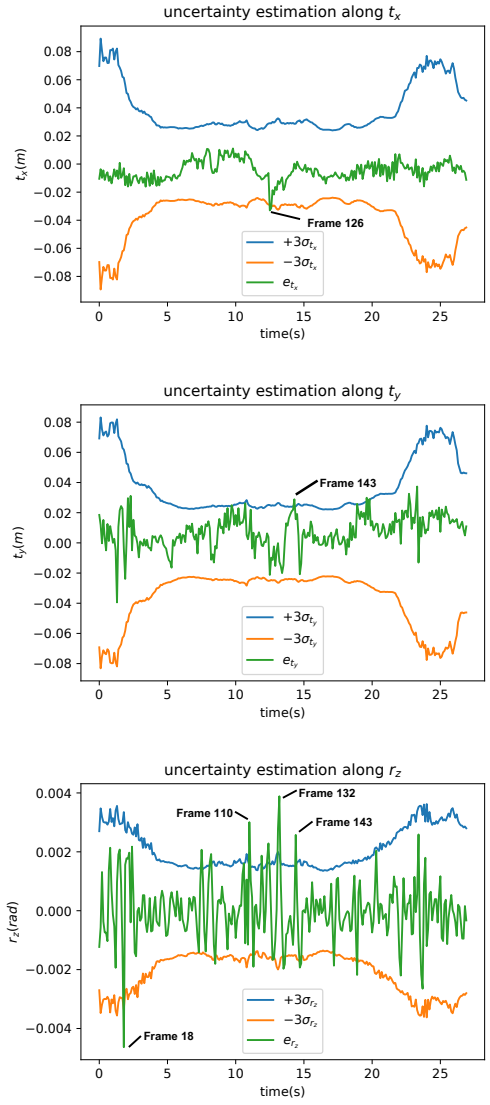


Fig. 6. Inconsistency indicator-based scan matching uncertainty along $\boldsymbol{t_x}$, $\boldsymbol{t_y}$ and $\boldsymbol{r_z}$ for the KITTI sequence 04

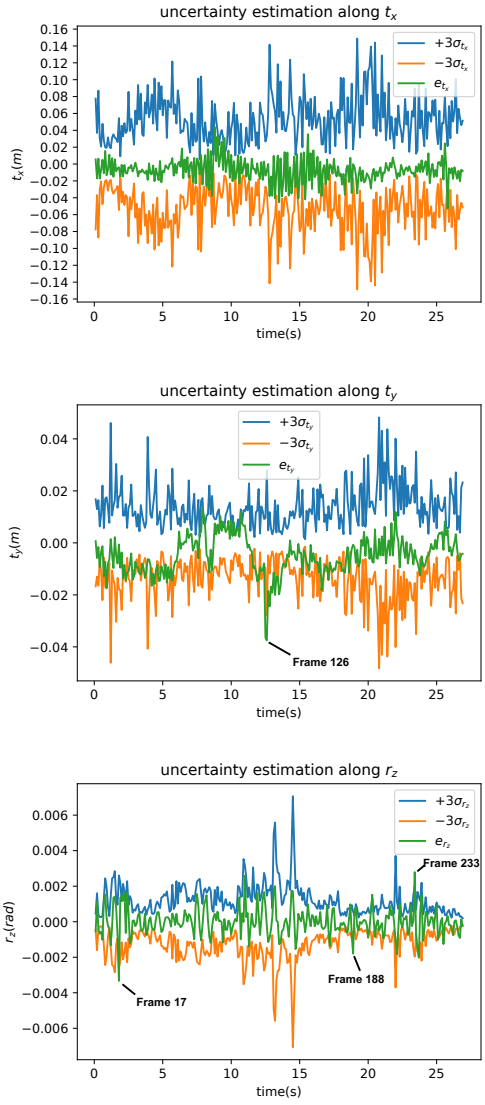


Fig. 7. Monte-Carlo simulation-based scan matching uncertainty estimation along $\boldsymbol{t_x}$, $\boldsymbol{t_y}$ and $\boldsymbol{r_z}$ for the KITTI sequence 04

It can be seen from Fig. 6 that our predicted confidence interval could accurately bound the estimation errors during most of the time. We attribute this primarily to our pre-conducted outdoor ground points removal, scattered outliers removal and multi-level semi-direct scan matching that incorporate more reliable correspondences for the pose estimation. As it can be inferred from (14), more valid correspondences will result in a more confident state estimation, which is in line with the principle of maximum likelihood estimation. It can also be seen from (14) that the farther the inliers $\{\mathbf{x}_i\}$ locate from the local sensor frame, the more confidence we gain from the pairwise correspondences. While it is observed that few overshoots occur at the road intersection or due to some moving objects in the scene, see Fig. 5 and Fig. 8. In the KITTI dataset, the LiDAR point clouds were captured at the frequency of 10 Hz, on which basis the relation between the frame number and timestamp could be established. The ego-motion at the road intersection in Fig. 8 will cause inadequate overlapping of subsequent LiDAR scans for state estimation, and the scan matching convergence can not be guaranteed in this situation. Besides, the existence of the moving object in Fig. 5 further complicates the point cloud registration process and may lead to inconsistency in uncertainty estimation. The LiDAR scans' inadequate overlapping and moving objects' contamination also pose challenges for the Monte-Carlo sampling methods, which can be seen from Fig. 7. The outlier frame numbers of two typical sequences KITTI 04 (moving objects contamination) and KITTI 05 (road intersection partial overlapping) are summarized in Table. V for better visualization. Apart from these few overshoots, the Monte-Carlo sampling performs very well in estimating the pose estimation uncertainty. The different voxel sizes' downsampling will approximate the raw point clouds at different scales, which helps the Monte-Carlo simulation to boost from the local minima while maintaining the estimation accuracy. Compared with the sampling-based Monte-Carlo simulation scan matching uncertainty estimation,

(a) Frame 233 in the KITTI sequence 05

(b) Frame 500 in the KITTI sequence 05

Fig. 8. Partially overlapping point clouds at the road intersection degrade the scan matching performance for the KITTI sequence 05

Algorithm 2 Monte-Carlo simulation-based scan matching uncertainty estimation

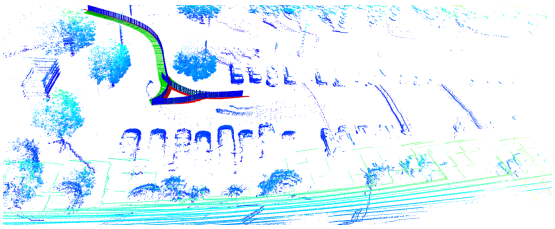
Input: Consecutive LiDAR scans $\{\mathbb{S}_t, \mathbb{S}_{t+1}\}_{t=1,\cdots,\tau-1}$
Output: The distribution of estimated motion parameters

1: **while** New LiDAR scan arrives **do**
2: Remove the scattered outliers and ground points to get the pre-processed LiDAR scans $\{\mathbb{S}_t^{\triangle}, \mathbb{S}_{t+1}^{\triangle}\}$
3: Downsample the pre-processed LiDAR scans with the voxel sizes of $0.05\,\mathrm{m}$, $0.1\,\mathrm{m}$, $0.2\,\mathrm{m}$, $0.4\,\mathrm{m}$, and $0.8\,\mathrm{m}$
4: Conduct semi-direct scan matching iteratively five times with the different pre-processed input LiDAR scans
5: Obtain different estimated transformations with different downsampled input LiDAR scans
6: Calculate the distribution of the estimated motion parameters that align $\{\mathbb{S}_t^{\triangle}, \mathbb{S}_{t+1}^{\triangle}\}$
7: **end while**

our proposed method achieves comparable performances. It can be seen from Algorithm 2 that the pre-processed LiDAR scans are downsampled with five different voxel sizes for Monte-Carlo simulation. And then the semi-direct LiDAR scan matching is conducted iteratively five times with different downsampled LiDAR point clouds to obtain the final distribution of the estimated motion parameters. However, for the proposed method, the semi-direct scan matching is conducted only once. And the covariance matrix is calculated at the final iteration based on the inconsistency indicator, which is derivative-free. Thus, the proposed method is much more computationally efficient than the Monte-Carlo simulation uncertainty estimation.

In the proposed approach pipeline, the uncertainty quantification is integrated into the global pose-graph optimization [5]. This helps to reduce the pairwise registration drift and to obtain consistent scene mapping, where the point clouds are registered with high fidelity. In our case, in order to accelerate the pose-graph optimization convergence, only the pose nodes in the graph are adjusted and the environment 3D points are not involved. By adjusting the pose nodes in the pose graph, the overall cost function will be minimized and the refined 6 DoF pose can be obtained. The scene mapping is in fact the registered point clouds with the refined 6 DoF pose. It can be seen from Fig. 9 that, the scene mapping is consistent and the vehicle poses are smooth even with a sharp turning in the trajectory.

(a) The main parking scene visualization with the front camera



(b) The main parking scene reconstruction with the VLP-16 LiDAR

Fig. 9. The main parking area scene mapping and self-positioning with the VLP-16 LiDAR at the Centrale Nantes Campus (LS2N 00)

V. Conclusion and Future Work

In this paper, an adaptive semi-direct LiDAR scan matching approach is proposed to overcome the weakness of individual direct or feature-based methods. The semi-direct LiDAR scan matching guarantees convergence in challenging environments such as undergoing high-speed motion and traveling in repetitive, feature-less environments. On this basis, extensive registration results on city, residential, highway, and parking scenarios with the relative fitness and RMSE metrics are presented and discussed. The superiority of the proposed semi-direct LiDAR scan matching method is validated with both HDL-64E and VLP-16 Velodyne LiDARs. At the same time, scan matching uncertainty is modeled as well to evaluate the final convergence accuracy. Furthermore, we also analyze the possible sources that may lead to scan matching divergence in various scenarios. It is demonstrated that ground points and dynamic objects such as vehicles or pedestrians are the main causes of the estimation accuracy decrease. It is also noticed that significant errors frequently occur near road intersections and in highway scenarios, where it is more likely to come across dynamic vehicles and the geometric information is not adequate for reliable state estimation.

The main contribution of this paper lies in the front-end improvement of LiDAR-based state estimation and uncertainty modeling. For future work, we will focus on integrating the GNSS measurements [27] in the back-end to correct the odometry drifts for the large-scale scene [28] application. Moreover, the semantic information can be leveraged for the moving objects segmentation [29], which increases the automotive perception system. More efficient outliers rejection mechanisms for partially overlapping point clouds that incorporate the IMU

pre-integration will also be studied to further ameliorate the scan matching accuracy and robustness with the presence of several moving objects in the scene.

REFERENCES

[1] T. Li, L. Pei, Y. Xiang, Q. Wu, S. Xia, L. Tao, X. Guan, and W. Yu, "P3-LOAM: PPP/LiDAR loosely coupled SLAM with accurate covariance estimation and robust RAIM in urban canyon environment," *IEEE Sensors Journal*, vol. 21, no. 5, pp. 6660–6671, 2021.

[2] Y. Li and J. Ibanez-Guzman, "LiDAR for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 50–61, 2020.

[3] N. Krombach, D. Droeschel, and S. Behnke, "Combining feature-based and direct methods for semi-dense real-time stereo visual odometry," in *Proceedings of the International Conference on Intelligent Autonomous Systems*, 2016, pp. 855–868.

[4] X. Zhao, L. Liu, R. Zheng, W. Ye, and Y. Liu, "A robust stereo feature-aided semi-direct SLAM system," *Robotics and Autonomous Systems*, vol. 132, p. 103597, 2020.

[5] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G2O: A general framework for graph optimization," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2011, pp. 3607–3613.

[6] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[7] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image and Vision Computing*, vol. 10, no. 3, pp. 145–155, 1992.

[8] A. Segal, D. Haehnel, and S. Thrun, "Generalized-ICP." in *Proceedings of Robotics: science and systems*, 2009, pp. 435–443.

[9] Z. Wang, J. Fang, X. Dai, H. Zhang, and L. Vlacic, "Intelligent vehicle self-localization based on double-layer features and multi-layer LIDAR," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 4, pp. 616–625, 2020.

[10] S. Oh, J.-H. You, A. Eskandarian, and Y.-K. Kim, "Accurate alignment inspection system for low-resolution automotive LiDAR," *IEEE Sensors Journal*, vol. 21, no. 10, pp. 11 961–11 968, 2021.

[11] W. Ali, P. Liu, R. Ying, and Z. Gong, "A feature based laser SLAM using rasterized images of 3D point cloud," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24 422–24 430, 2021.

[12] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2009, pp. 3212–3217.

[13] F. Tombari, S. Salti, and L. Di Stefano, "Unique signatures of histograms for local surface description," in *Proceedings of the European Conference on Computer Vision*, 2010, pp. 356–369.

[14] W. Wang, B. Wang, P. Zhao, C. Chen, R. Clark, B. Yang, A. Markham, and N. Trigoni, "PointLoc: Deep pose regressor for LiDAR point cloud localization," *IEEE Sensors Journal*, vol. 22, no. 1, pp. 959–968, 2021.

[15] 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.

[16] Y. Wang, Y. Lou, W. Song, Z. Tu, Y. Wang, and S. Zhang, "Simultaneous localization of rail vehicles and mapping of surroundings with LiDAR-inertial-GNSS integration," *IEEE Sensors Journal*, vol. 22, no. 14, pp. 14 501–14 512, 2022.

[17] S. Chen and V. Frémont, "A loosely coupled vision-LiDAR odometry using covariance intersection filtering," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, 2021, pp. 1102–1107.

[18] K. Wang, C. Cao, S. Ma, and F. Ren, "An optimization-based multi-sensor fusion approach towards global drift-free motion estimation," *IEEE Sensors Journal*, vol. 21, no. 10, pp. 12 228–12 235, 2021.

[19] Y. Wang, Y. Lou, W. Song, and Z. Tu, "A tightly-coupled framework for large-scale map construction with multiple non-repetitive scanning LiDARs," *IEEE Sensors Journal*, vol. 22, no. 4, pp. 3626–3636, 2022.

[20] W. Ali, P. Liu, R. Ying, and Z. Gong, "A life-long SLAM approach using adaptable local maps based on rasterized LIDAR images," *IEEE Sensors Journal*, vol. 21, no. 19, pp. 21 740–21 749, 2021.

[21] A. Censi, "An accurate closed-form estimate of ICP's covariance," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2007, pp. 3167–3172.

[22] A. G. Buch, D. Kraft *et al.*, "Prediction of ICP pose uncertainties using monte carlo simulation with synthetic depth images," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2017, pp. 4640–4647.

[23] Y. Zhong, "Intrinsic shape signatures: A shape descriptor for 3D object recognition," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2009, pp. 689–696.

[24] 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.

[25] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuScenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 621–11 631.

[26] P. Xiao, Z. Shao, S. Hao, Z. Zhang, X. Chai, J. Jiao, Z. Li, J. Wu, K. Sun, K. Jiang, Y. Wang, and D. Yang, "Pandaset: Advanced sensor suite dataset for autonomous driving," in *Proceedings of the IEEE International Intelligent Transportation Systems Conference*, 2021, pp. 3095–3101.

[27] N. Zhu, D. Betaille, J. Marais, and M. Berbineau, "GNSS integrity monitoring schemes for terrestrial applications in harsh signal environments," *IEEE Intelligent Transportation Systems Magazine*, vol. 12, no. 3, pp. 81–91, 2020.

[28] Z. Yan, L. Sun, T. Krajník, and Y. Ruichek, "EU long-term dataset with multiple sensors for autonomous driving," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2020, pp. 10 697–10 704.

[29] S. Chen, H. Sun, and V. Frémont, "A semantic-guided LiDAR-vision fusion approach for moving objects segmentation and state estimation," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems*, 2022, pp. 4308–4313.

Songming Chen received the M.S. degree in Automatic Control and Robotics from the École Centrale de Nantes, France, in 2019. And then he obtained the Ph.D. degree in Automation and Robotics from the École Centrale de Nantes, within the ARMEN Team of LS2N Laboratory, UMR CNRS 6004, France, in 2022. His research interests belong to multi-sensor based perception and state estimation for autonomous vehicles.

Vincent Frémont received a Ph.D. degree in Automatic Control and Computer Science from the École Centrale de Nantes, France, in 2003. He is currently a Full Professor at École Centrale de Nantes, within the ARMEN Team of LS2N Laboratory, UMR CNRS 6004. His research interests belong to perception systems for autonomous mobile robotics with an emphasis on computer vision, machine learning and multi-sensor fusion.