

LCR-SMM: Large Convergence Region Semantic Map Matching Through Expectation Maximization

Qingxiang Zhang[✉], *Graduate Student Member, IEEE*, Meiling Wang[✉], *Member, IEEE*, Yufeng Yue[✉], *Member, IEEE*, and Tong Liu

Abstract—The matching and fusion of local maps generated by multiple robots can greatly enhance the performance of relative localization and collaborative mapping. Currently, existing semantic matching methods are partly based on classical iterative closest point, which typically fails in cases with large initial errors. What's more, current semantic matching algorithms have high computation complexity in optimizing the transformation matrix. To address the challenge of large initial errors and low matching efficiency, this article proposes a novel large convergence region semantic map matching algorithm. The key novelty of this work is the designing of the initial transformation optimization algorithm and the probabilistic registration model to increase the convergence region. To reduce the initial error before the iteration process, the initial transformation matrix is optimized by estimating the credibility of the data association. At the same time, a factor reflecting the uncertainty of the initial error is calculated and introduced to the formulation of the probabilistic registration model, thereby accelerating the convergence process. The proposed algorithm is performed on public datasets and compared with existing methods, demonstrating the significant improvement in terms of matching accuracy and robustness.

Index Terms—Collaborative localization, expectation maximization, map matching.

I. INTRODUCTION

MULTIROBOT systems are qualified for more complex tasks than single robots through collaborative operations [1], such as formation, exploration, and rescue [2]. Besides, simultaneous cooperation can improve the efficiency and reliability of distributed tasks, such as collaborative mapping

Manuscript received March 20, 2021; revised August 4, 2021; accepted October 25, 2021. Recommended by Technical Editor M. Xin and Senior Editor Nikos G Tsagarakis. This work was supported in part by the National Natural Science Foundation of China under Grant 62003039, in part by the Changjiang Scholars Programme of China under Grant T2014224, and in part by the CAST program under Grant YESS20200126. (*Corresponding author: Yufeng Yue*)

The authors are with the School of Automation, Beijing Institute of Technology, Beijing 100081, China (e-mail: 3120205462@bit.edu.cn; wangml@bit.edu.cn; yueyufeng@bit.edu.cn; liutong2002@bit.edu.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TMECH.2021.3124994>.

Digital Object Identifier 10.1109/TMECH.2021.3124994

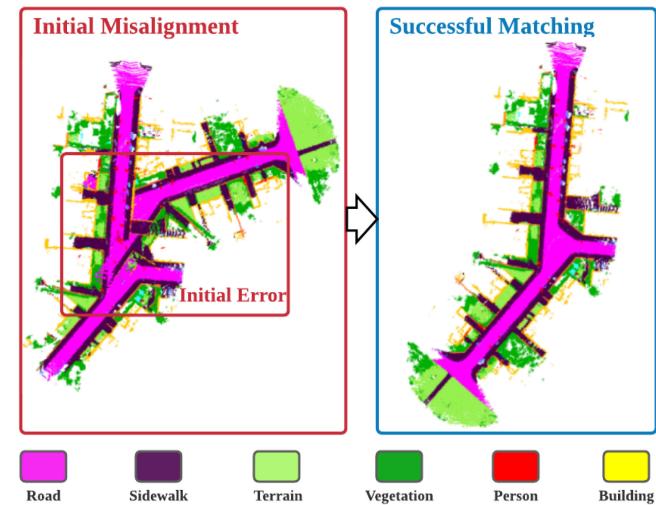


Fig. 1. Demonstration of large convergence region semantic map matching (LCR-SMM) on Semantic KITTI07 dataset. The initial error is 150° around z-axis and 8 m along the y-axis. The proposed algorithm is able to successfully match the two local maps in such a challenging case.

in large-scale environments [3]. In order to work reliably in challenging environments, the ability to perform collaborative localization based on environmental information in unstructured or global position system (GPS)-denied scenarios is necessary. However, under limited communication bandwidth, it is infeasible to share information from sensors directly. In contrast, it is more cost-effective to transmit compact maps generated by robots. Therefore, estimating the transformation matrix based on the map matching algorithm is a promising way for multi-robot localization.

Currently, most existing map-matching methods are based on geometric features [3]–[5]. Due to the lack of uniqueness of geometric features, the robustness of feature matching is low, which will cause large estimation errors when features are not rich. When semantic features are combined with geometric features, the matching accuracy and robustness will be greatly improved. In recent years, the rapid development of semantic simultaneous localization and mapping (SLAM) algorithms [6]–[9] has made it possible to design map-matching algorithms based on semantic features. However, single-robot semantic SLAM algorithms

and point cloud registration algorithms usually assume that the maps to be matched are highly overlapped and the initial transformation is close to the ground-truth, which directly applies iterative approximation to optimize the transformation [7], [10]. However, due to the nonconvexity of the map-matching problem, the iterative approximation will fail in the cases with large initial errors or low overlapping percentages, which is common for multirobot map-matching problems. Nevertheless, the latest collaborative semantic mapping research has focused on combining local maps into a global representation [11]–[13]. The main contribution of [11] is to combine local maps into a global representation, where the transformation matrix between local maps is given as a prior. In real applications, the given transformation matrix may not be accurate enough to be adopted directly. This article releases this assumption and studies the problem of how semantic information can improve the relative localization accuracy between robots. Therefore, the main objective of this article is to design a novel semantic map-matching algorithm that has a larger convergence region and is more efficient.

The first challenge of widening the convergence region is to establish correct data association between local maps with different percentages of overlapping. Most existing algorithms establish data association based on Euclidean distance or the probabilistic error metrics. For example, the iterative closest point (ICP) algorithm [14] and its variants [10], [15] establish data association between the points with closest positions. This is unreasonable when the points to be matched are far away from each other in the initial state, which means the established data association might be wrong. With low-quality data association, the iterative approximation algorithm will have a large chance to converge to a local minimum. To address this problem, the initial transformation optimization algorithm called data association credibility estimation (DACE) is proposed to be operated before the iterative optimization process, which will establish a sampling-traverse-approximation structure. The initial value within a large convergence region will be optimized to ensure that the matching process is started with a set of credible data association.

The second challenge is that the introduction of additional semantic information will increase computation complexity. Current semantic matching algorithms establish one-to-multiple data association through the expectation maximization (EM) method [7], [10], which increases the probability of matching points with the same semantic label. However, the error function is more complicated, thereby prolonging the running time. To reduce the time cost, a probabilistic registration model containing initial error uncertainty is proposed to increase the calculated probability of long-distance data association, which will improve their roles in the transformation optimization process and speed up the convergence process.

As summarized, there still exists a substantial gap in improving the convergence region and matching efficiency. This article extends [16] and proposes a novel large convergence region semantic map-matching (LCR-SMM) algorithm. The proposed algorithm is capable of performing semantic map matching and generating a consistent global map under large initial errors

(see Fig. 1). The main contributions of this work are listed as follows:

- The DACE method is proposed to optimize the initial transformation matrix, so as to extend the convergence region of the algorithm and improve the efficiency.
- A probability registration model is proposed in which the initial error uncertainty is modeled as a factor applied to the calculation of the probability of data association.
- An efficient adaptive data association strategy is introduced. The algorithm will construct one-to-multiple data association at the beginning of optimization and switch to one-to-one when it is close to convergence.

The rest of this article is organized as follows. Section II reviews the related works. Section III gives an overview of the systematic framework. In Section IV, the LCR-SMM method is introduced. Section V shows the experimental procedures and results. Finally, Section VI concludes this article.

II. RELATED WORKS

In this section, existing algorithms related to semantic map matching are introduced, including single-robot semantic mapping and map-matching methods. Since the key to map matching lies in finding the transformation between the maps, many existing point cloud registration methods are closely related and also reviewed.

A. Single-Robot Semantic Mapping

With the fast development of neural networks, many methods for semantic segmentation [17] based on deep learning are developed, which accelerates the researches on semantic SLAM [6]–[9]. In Ref. [7], a semantic SLAM method is proposed based on the EM model, which is utilized to solve the data association problem for odometry estimation. After that, Doherty *et al.* [8] modify the probabilistic data association model in [7] into a max-mixture model, which uses max-marginalization instead of the sum-marginalization to calculate the probability of one-to-multiple data association. Atanasov *et al.* [18] also give a unifying view of geometry, semantics, and data association in SLAM. These works have made great contributions to semantic mapping, but all of them focused on single robots. More specifically, they usually assume that the initial transformation is close to the ground-truth and the maps to be matched are highly overlapped, which may not hold true in the problem of aligning local maps from different robots.

B. Collaborative Map Matching

Motivated by the strong need for efficient and reliable mapping methods in large-scale scenes, researchers have designed multirobot mapping systems to achieve the matching and merging of local maps. Saeedi *et al.* [4] give a thorough review of existing multirobot mapping methods. For example, Jessup *et al.* [5] provide an approach to merging octree-based occupancy grid maps, where the ICP algorithm is utilized to refine the transformation between 3-D maps. In Ref. [19], the 3-D map is represented as a pose graph; the transformation between different

local maps is then estimated with normal iterative closest point (NICP) [20] algorithm. In Ref. [21], the 2-D occupancy grid map is constructed as generalized Voronoi diagram (GVD) and the transformation is calculated based on the features extracted from GVD.

Considering the complexity of application scenarios, the sharing and fusing of information from heterogeneous sensors are urgently needed. Therefore, a general multilevel probabilistic framework is designed to match local maps generated by multiple robots with heterogeneous sensors in Ref. [3]. Furthermore, a collaborative dynamic mapping method based on the fusion of hybrid sensors is proposed in Ref. [22].

The aforementioned methods have made great progress in the field of multirobot mapping. However, their work mainly focuses on fusing geometric features and maps but does not explore how to make use of rich semantic information. To narrow down the gap in the fields of multirobot semantic map matching, COSEM [23] recently proposed a collaborative semantic map-matching framework, in which semantic information is utilized with the flexible semantic data association strategy. Although COSEM outperforms most geometric matching methods, the range of its convergence region is still limited by the nonconvexity of the map-matching problem. Therefore, it is necessary to design a strategy to enlarge the convergence region.

C. Point Cloud Registration Algorithms

The objective of point cloud registration algorithms is to estimate the transformation between point clouds, which can be divided into two categories: feature matching algorithms and iterative dense matching methods. Feature matching algorithms extract 3-D key points with descriptors [24], [25], and then match the key points according to the descriptors and minimize the distances between matched points with close-formed methods like least-squares [26]. But the accuracy of these methods depends on the robustness of the methods to extract 3-D key points. Iterative dense matching methods make use of all points in the point cloud and alternate from building data association to optimizing transformation. The most widely used classical method is ICP [14], which may fail to establish correct data association when the overlapping area is low or the initial transformation is not close to the ground-truth. Therefore, many scholars have modified ICP to improve its performance.

The variants of ICP can be classified into two categories. Many algorithms improve the module of establishing data association or modifying objective function, such as GICP [15], normal distributions transform (NDT) [27], and VGICP [28]. GICP minimizes the probability error metrics instead of the euclidean distance defined in ICP. Furthermore, VGICP constructs voxels and establishes data association between the points in the same voxel, which improves computational efficiency. Go-ICP [29] attempts to apply the branch and bound (BnB) algorithm to estimate the global minimum by operating ICP multiple times, which significantly increases the running time. Other researchers introduce additional information to increase matching accuracy, including NICP [20], Robust-GICP [30], and semantic ICP [10].

Cluster iterative closest point (CICP) introduces normal vectors and performs point cloud voxelization, which addresses the problem of point cloud registration with different densities. Min *et al.* [30] utilize orientational information to construct a hybrid mixture model, which increases the robustness to noise and outliers. Recently, semantic ICP [10] introduces semantic information to decide whether the data association is correct, and the matching accuracy is greatly improved.

The algorithms discussed above have greatly improved the matching accuracy and robustness, but most of them still rely on a good initial transformation matrix for a warming start. Considering large initial error and low overlap between point clouds, it is of great need to develop an algorithm with a large convergence region and higher efficiency.

III. SYSTEMATIC FRAMEWORK

In this section, the system structure of large convergence region semantic map-matching algorithm (LCR-SMM) is presented. Besides, the concepts in the semantic map-matching problem are defined and the overall problem is formulated.

A. Overview of Semantic Map-Matching Framework

An overview of the system framework is presented in Fig. 2. The algorithm for estimating transformation matrix consists of three modules: the probability registration model, the initial transformation optimization based on DACE, and the iterative optimization through EM.

The probability registration model serves as the basis of the optimization process. In this module, the semantic map-matching problem is modeled in the form of a maximum likelihood estimation (MLE) problem, then the probability of data association is expressed according to the local maps and the given transformation matrix, which combines the geometric and semantic information in the local maps. In the initial value optimization process, the probability is calculated with the transformation sampled from the search scope. In the iterative optimization process, the parameters of the transformation matrix are calculated and optimized iteratively.

The initial value optimization is the key to enlarge the convergence region. The output of probability calculation is applied to estimate the credibility of data association, which is the basis for the initial value selection. Besides, the credibility can help to modify the probability registration model in order to speed up the convergence of the iterative optimization process. Finally, the selected transformation serves as the initial value of the iterative optimization process.

The iterative optimization aims to solve the MLE problem formulated in the probability registration model in an EM framework. In the E step, the expectation of matching error is computed with the probability, which is the objective function to be minimized. In the M step, the transformation matrix is optimized to minimize the objective function. The notation symbols used in this article are defined in Table I.

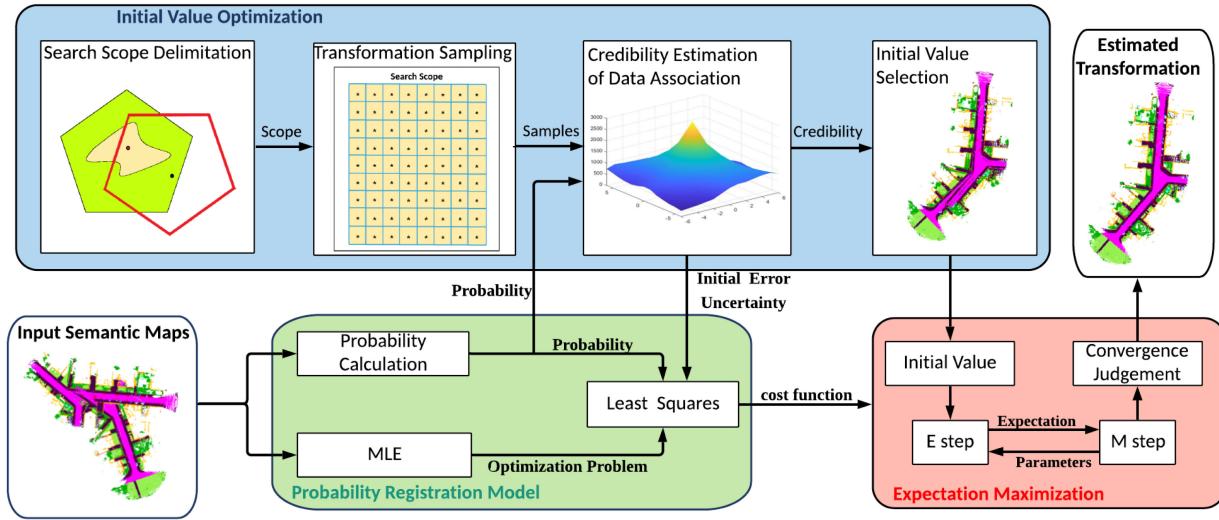


Fig. 2. Framework of large convergence region semantic map-matching algorithm, which consists of three parts: Probability registration model, initial value optimization, and expectation maximization.

TABLE I
MAIN NOTATIONS USED THROUGHOUT THE ARTICLE

Symbol	Description
\mathcal{M}	Semantic Map Description: A semantic map in the form of a 3D point cloud with semantic labels
\mathbf{x}_k	The coordinate vector of the k -th point in the point cloud
\mathcal{X}	The set of the coordinate vectors of all points $\mathcal{X} = \{\mathbf{x}_k\}_{k=1}^n$ in the point cloud
l_k	The semantic label of the point at \mathbf{x}_k in the semantic map
\mathcal{L}	The set of the semantic labels of all points $\mathcal{L} = \{l_k\}_{k=1}^n$ in the map
Transformation Description:	
\mathbf{R}	A rotation matrix
\mathbf{t}	A translation vector
$\mathbf{T}(\mathbf{R}, \mathbf{t})$	The transformation matrix consisting of
	$\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix}$
\mathbf{T}_{GT}	The ground truth value of the transformation matrix between local maps
$\mathbf{T}(\mathbf{x})$	The coordinate of \mathbf{x} transformed by the transformation matrix \mathbf{T}
$e(\mathbf{T}_1, \mathbf{T}_2)$	The error between \mathbf{T}_1 and \mathbf{T}_2 , $e(\mathbf{T}_1, \mathbf{T}_2) = \ \log(\mathbf{T}_1 \mathbf{T}_2^{-1})\ $

B. Overall Problem Formulation

Considering a semantic map-matching problem, the input is two independent and overlapping semantic grid maps in the form of two sets of 3-D point clouds with semantic labels. Then, the problem of finding the transformation matrix to align the two maps can be formulated as follows.

Let $\mathcal{M}^s = \{\mathcal{X}^s, \mathcal{L}^s\}$ and $\mathcal{M}^t = \{\mathcal{X}^t, \mathcal{L}^t\}$ be the input semantic maps, where \mathcal{M}^s is the source map to be transformed

and \mathcal{M}^t is the target map. The initial transformation matrix, without being optimized, is symbolized with \mathbf{T}_0 . The objective is to estimate the transformation matrix \mathbf{T} that minimizes the errors between the transformed source point cloud and the target point cloud. The problem defined above is notated as $Q(\mathcal{M}^s, \mathcal{M}^t, \mathbf{T}_0)$. This problem can be formulated as the following MLE problem:

$$\hat{\mathbf{T}} = \operatorname{argmax} p(\mathcal{M}^s, \mathcal{M}^t | \mathbf{T}). \quad (1)$$

To solve the problem formulated above, hidden variable data association $\mathcal{D} = \{d_k\}_{k=1}^n$ need to be established, where $d_k = \{s(k), t(k)\}$ means that the $s(k)$ point in the source map and the $t(k)$ point in the target map refer to the same points in the global map. Data association is a variable independent of transformation. Then, the problem defined above is modified to an MLE problem with latent variable \mathcal{D} in (2)

$$\hat{\mathbf{T}} = \operatorname{argmax} p(\mathcal{M}^s, \mathcal{M}^t | \mathcal{D}, \mathbf{T}) \quad (2)$$

Instead of constructing the fixed one-to-one data association in traditional ICP, it is proposed to build one-to-multiple data association in this article. Then, the formulation of data association should be modified into $\mathcal{D} = \{D_k\}_{k=1}^n$, where $D_k = \{d_{k,i}\}_{i=1}^N$ represents multiple possible data association for one point in the source cloud. To simplify notation, it is assumed that the coordinate vectors and semantic labels are indexed according to the data association. That is, $d_{k,i}$ means $\{\mathbf{x}_k^s, l_k^s\}$ corresponds to $\{\{\mathbf{x}_{k,i}^t, l_{k,i}^t\}\}_{i=1}^N$.

C. Sub-module Definition

Probability Modeling: In order to solve the MLE problem in (2), there is a need to compute the probability of data association. The probability of data association can be divided into two parts: geometric and semantic data association. The former refers to that the corresponded points have the same position in the global map, symbolized with $d_{k,i}^g$. The latter means the corresponded

points have the same semantic label, symbolized with $d_{k,i}^{se}$. It is clear that $d_{k,i}$ is valid only when both $d_{k,i}^g$ and $d_{k,i}^{se}$ hold. Since geometric and semantic data association are independent of each other without prior information, the probability can be expressed as

$$p(d_{k,i} | \mathbf{x}_k^s, \mathbf{x}_k^t, l_k^s, l_k^t, \mathbf{T}) = \underbrace{p(d_{k,i}^g | \mathbf{x}_k^s, \mathbf{x}_k^t, \mathbf{T})}_{\text{geometric matching}} \underbrace{p(d_{k,i}^{se} | l_k^s, l_k^t)}_{\text{semantic matching}}. \quad (3)$$

The probability formulated in (3) is a collaborative data association established based on both semantic and geometry information. The detail of probability modeling will be presented in Section IV-A.

Initial Value Optimization Problem: The objective of the initial value optimization problem is to start the iterative process at the state where a high-quality data association can be established so as to enlarge the convergence region. The input is the semantic local maps and the required convergence region. The output is the optimized initial transformation within the convergence region of the iterative algorithm. Then, the problem can be defined as follows.

Considering the map-matching problem $Q(\mathcal{M}^s, \mathcal{M}^t, \mathbf{T}_0)$, the required convergence region is $\mathcal{C}_r(\mathbf{T}_{GT})$. Assuming that the convergence region of iterative optimization algorithm is $\mathcal{C}_I(\mathcal{M}^s, \mathcal{M}^t)$, both $\mathcal{C}_r(\mathbf{T}_{GT})$ and $\mathcal{C}_I(\mathcal{M}^s, \mathcal{M}^t)$ are the neighborhoods of \mathbf{T}_{GT} . The objective is to find a optimization function $f(\mathcal{M}_1, \mathcal{M}_2, \mathbf{T})$, subject to that for an arbitrary transformation matrix $\mathbf{T}_0 \in \mathcal{C}_r(\mathbf{T}_{GT})$; it is always true that $\mathbf{T}_{\text{init}} = f(\mathcal{M}^s, \mathcal{M}^t, \mathbf{T}_0) \in \mathcal{C}_I(\mathcal{M}^s, \mathcal{M}^t)$.

To describe the map conveniently, the 3-D rectangular coordinate system is established with the xoy plane parallel to the ground. In common map-matching problems, the initial errors mainly exist in rotation errors around the z -axis and translation errors on the xoy plane. The errors in other directions will not be considered in order to improve efficiency. The approach to constructing the optimization function, which is the DACE algorithm, will be detailed in Section IV-B.

Iterative Optimization Problem: After obtaining the optimized transformation matrix \mathbf{T}_{init} , the map-matching problem $Q(\mathcal{M}^s, \mathcal{M}^t, \mathbf{T}_0)$ is transformed to $Q(\mathcal{M}^s, \mathcal{M}^t, \mathbf{T}_{\text{init}})$, which can be solved with an iterative approximation optimization algorithm. Since both data association and transformation are unknown, the iteration is operated in the coordinate descent form as follows:

$$\text{E step: } \hat{\mathcal{D}}^{j+1} = \underset{\mathcal{D}}{\operatorname{argmax}} p(\mathcal{M}^s, \mathcal{M}^t | \mathbf{T}^j, \mathcal{D}) \quad (4)$$

$$\text{M step: } \hat{\mathbf{T}}^{j+1} = \underset{\mathbf{T}}{\operatorname{argmax}} p(\mathcal{M}^s, \mathcal{M}^t | \mathcal{D}^{j+1}, \mathbf{T}). \quad (5)$$

Equations (4) and (5) are the E step and M step of EM, respectively. The transformation matrix is updated in the M step and the new transformation will be used in the next E step. The iteration will stop when converged. The iterative optimization process will be presented in detail in Section IV-C.

IV. LARGE CONVERGENCE REGION SEMANTIC MAP MATCHING

In this section, the LCR-SMM algorithm is described and formulated. The section is divided into three subsections: probabilistic registration model, DACE for initial transformation optimization, and semantic map matching using EM.

A. Probabilistic Registration Model

In (3), the overall probability is divided into semantic and geometric data association probability. If the set of all the categories of semantic labels is expressed as $\mathcal{S} = \{s_k\}_{k=1}^n$, the semantic association probability can be expressed as

$$p(d_{k,i}^{se} | l_k^s, l_k^t) = \sum_{s_k \in \mathcal{S}} p(s_k | l_k^s) p(s_k | l_k^t) \quad (6)$$

where $p(s|l)$ represents the probability that the label l belongs to the category s .

The geometric association probability can be described with the distance between corresponded points, because the larger the distance, the smaller the probability will be. In order to describe the problem conveniently, define $\mathcal{R} = \{\{\mathbf{r}_{k,i}\}_{i=1}^N\}_{k=1}^n$ to represent the residual, where $\mathbf{r}_{k,i} = \mathbf{x}_{k,i}^t - \mathbf{T}(\mathbf{x}_k^s)$. The residual is computed according to the data association. This probability is calculated using a three-dimensional Gaussian distribution model as follows:

$$p(d_{k,i}^g | \mathbf{x}_k^s, \mathbf{x}_k^t, \mathbf{T}) = \frac{1}{2\pi^{\frac{3}{2}}|\mathbf{C}_{k,i}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{r}_{k,i}^T \mathbf{C}_{k,i}^{-1} \mathbf{r}_{k,i}\right). \quad (7)$$

The covariance matrix of $\mathbf{r}_{k,i}$ is denoted as $\mathbf{C}_{k,i} \triangleq \Sigma_{k,i}^t + \mathbf{R} \Sigma_k^s \mathbf{R}^T$. \mathbf{R} is the rotation matrix, $\Sigma_{k,i}^t$ and Σ_k^s are the covariance matrix of the two points in their respective point clouds, calculated over the distribution of the coordinates of a certain number of points near the point to be associated.

The probability of data association between two points can be calculated using (3)–(7). Since it is inefficient to consider every possible data association, only the association constructed on N closest points will be used to calculate the probability. Therefore, (7) is modified into

$$p(d_{k,i}^g | \mathbf{x}_k^s, \mathbf{x}_k^t, \mathbf{T}) = \frac{I(k, i)}{2\pi^{\frac{3}{2}}|\mathbf{C}_{k,i}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{r}_{k,i}^T \mathbf{C}_{k,i}^{-1} \mathbf{r}_{k,i}\right) \quad (8)$$

where $I(k, i)$ is the characteristic function defined in (9).

$$I(k, i) = \begin{cases} 1 & \mathbf{x}_{k,i}^t \text{ is } N \text{ closest of } \mathbf{x}_k^s \\ 0 & \text{otherwise} \end{cases}. \quad (9)$$

The difference between pure geometric data association and collaborative data association is shown in Fig. 3. It can be seen that the method formulated above is more likely to establish correct data association.

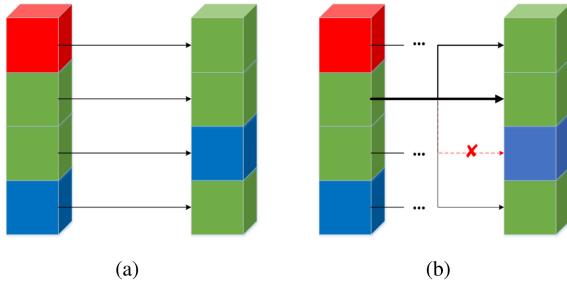


Fig. 3. Difference between (a) geometric data association and (b) collaborative data association. The color of the cube means the category of semantic label. The thickness of the line reflects the probability of data association and the red line means the association denied with the semantic message.

B. DACE Algorithm for Initial Transformation Optimization

Since adopting iterative approximation may lead the algorithm to a local minimum, a sampling-traverse optimization algorithm is proposed. In the “sampling” part, a certain number of transformation matrices are selected to represent the matrices in their neighborhoods. In the “traverse” part, the algorithm searches for the transformation with the highest data association credibility from the samples.

1) Search Scope Delimitation: Before operating the transformation matrix sampling, the range of search scope, symbolized as \mathcal{T}_{ss} , should be determined. As discussed in Section III, the result of DACE optimization needs to fall into the convergence region of the iterative optimization algorithm. This is possible only when (10) holds.

$$\mathcal{T}_{ss} \cap \mathcal{C}_I(\mathcal{M}^s, \mathcal{M}^t) \neq \emptyset \quad (10)$$

where $\mathcal{C}_I(\mathcal{M}^s, \mathcal{M}^t)$ refers to the convergence region of the iterative optimization. That is to say, $Q(\mathcal{M}^s, \mathcal{M}^t, \mathbf{T}_0)$ can be solved with the iterative optimization if and only if \mathbf{T}_0 satisfies $\mathbf{T}_0 \in \mathcal{C}_I(\mathcal{M}^s, \mathcal{M}^t)$. Since we have no prior knowledge about $\mathcal{C}_I(\mathcal{M}^s, \mathcal{M}^t)$ except that it is the neighborhood of \mathbf{T}_{GT} , (10) is true only if $\mathbf{T}_{GT} \in \mathcal{T}_{ss}$ holds.

To determine search scope, we assume that $\mathbf{T}_i = \mathbf{T}(\mathbf{R}_z(\theta_i), \mathbf{t}_i)$, where $\mathbf{R}_z(\theta)$ is the rotation matrix representing a rotation of θ around the z -axis

$$\mathbf{R}_z(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (11)$$

Considering an arbitrary transformation \mathbf{T}_0 in the required convergence region $\mathcal{C}_r(\mathbf{T}_{GT})$, the new operations \oplus/\ominus are defined as

$$\mathbf{T}_1 \oplus (\ominus) \mathbf{T}_2 = \mathbf{T}(\mathbf{R}_z(\theta_1 \pm \theta_2), \mathbf{t}_1 \pm \mathbf{t}_2). \quad (12)$$

Then, the symmetry region of $\mathcal{C}_r(\mathbf{T}_{GT})$ with respect to \mathbf{T}_{GT} can be described with

$$\mathcal{C}_r^{\text{sym}}(\mathbf{T}_{GT}) = \{\mathbf{T}_{GT} \oplus \mathbf{T}_{GT} \ominus \mathbf{T}, \mathbf{T} \in \mathcal{C}_r(\mathbf{T}_{GT})\}. \quad (13)$$

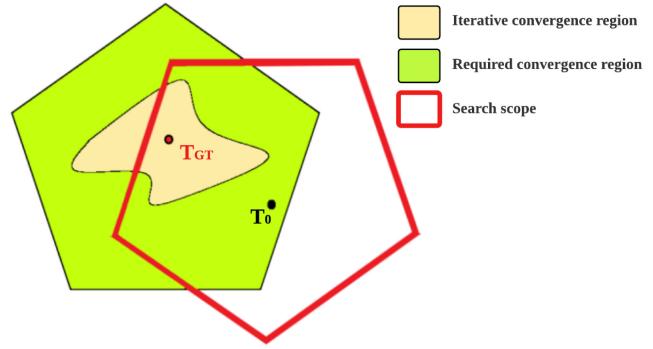


Fig. 4. Search scope with the same size as required convergence region always contains the ground truth value when the initial transformation is within the required convergence region.

Then, the search scope can be represented as

$$\mathcal{T}_{ss} = \mathcal{C}_r^{\text{sym}}(\mathbf{T}_0) = \{\mathbf{T} \oplus \mathbf{T}_0 \ominus \mathbf{T}_{GT}, \mathbf{T} \in \mathcal{C}_r^{\text{sym}}(\mathbf{T}_0)\}. \quad (14)$$

The search scope defined above always satisfies (10) and the proof is as follows:

$$\text{Proof: } \forall \mathbf{T}_0 \in \mathcal{C}_r(\mathbf{T}_{GT}), \quad \mathbf{T}_0^{\text{sym}} \triangleq \mathbf{T}_{GT} \oplus \mathbf{T}_{GT} \ominus \mathbf{T}_0 \in \mathcal{C}_r^{\text{sym}}(\mathbf{T}_{GT}), \quad \text{so} \quad \mathbf{T}_{GT} = \mathbf{T}_0^{\text{sym}} \oplus \mathbf{T}_0 \ominus \mathbf{T}_{GT} \in \mathcal{C}_r^{\text{sym}}(\mathbf{T}_0) = \mathcal{T}_{ss}. \quad \text{Since} \quad \mathbf{T}_{GT} \in \mathcal{C}_I(\mathcal{M}^s, \mathcal{M}^t), \quad \mathcal{T}_{ss} \cap \mathcal{C}_I(\mathcal{M}^s, \mathcal{M}^t) \neq \emptyset.$$

The relationship between the required convergence region and search scope is shown in Fig. 4; the search scope has the same size as the required convergence region.

According to (14), the range of \mathcal{T}_{ss} is decided by the shape and size of the required convergence region, which depends on the robustness requirements of the user and the range of initial errors. The latter can be estimated by evaluating the semantic mapping algorithm with quantities of experimental results. If the information is insufficient for estimating the range, $\mathcal{C}_r(\mathbf{T}_{GT}) = \mathcal{T}_{ss} = SE(3)$. In that case, the range of search scope on ground truth rotation angle can be set to $[-\pi, \pi]$, but the bounds of translation errors are nonexistent. However, maps without any overlapping area can not be matched. Hence, it can be assumed that the maps overlap. The upper and lower bounds of the components of the translation vector along the x -axis and y -axis can be defined as

$$\bar{x} = \underline{x}^t + \bar{r}^s, \underline{x} = \underline{x}^t - \bar{r}^s \quad (15)$$

$$\bar{y} = \underline{y}^t + \bar{r}^s, \underline{y} = \underline{y}^t - \bar{r}^s \quad (16)$$

where $\bar{x}^t, \underline{x}^t, \bar{y}^t, \underline{y}^t$ represent the bounds of the x, y coordinate of the points in the target map and \bar{r}^s represents the upper bound of the distance from the points in the source cloud to the origin of coordinates. However, the size of the search scope will prolong the operating time of the initial transformation optimization. This is inefficient especially when the initial errors are small. So it is suggested to narrow down the range of the required convergence region based on prior information.

2) Transformation Matrix Sampling: Assuming that the search scope of the ground-truth transformation is \mathcal{T}_{ss} , the problem required to be solved is to find a sampling method $S : \mathcal{T}_{ss} \rightarrow \mathcal{T}$, where $\mathcal{T} = \{\mathbf{T}_k\}_{k=1}^n$ is the set of a certain

number of sampled transformation matrices, to minimize the minimum distance between one of the samples and the ground truth transformation. Hence, the problem can be formulated as

$$\begin{aligned}\hat{\mathbf{T}} &= \operatorname{argmin} E_{\mathbf{T}_{GT}}[e(\mathbf{T}_{GT}, \mathbf{T}_k) | \mathcal{T}] \\ &= \operatorname{argmin} \sum_{\mathbf{T}_{GT} \in \mathcal{T}_{GT}} e(\mathbf{T}_{GT}, \mathbf{T}_k | \mathcal{T}) p(\mathbf{T}_{GT}).\end{aligned}\quad (17)$$

Without the prior estimation, the probability distribution of ground-truth transformation in \mathcal{T}_{ss} is uniform. In this case, it is obvious that the best sampling method is to divide \mathcal{T}_{ss} into multiple subregions and take the matrices at the center of each subregion as samples.

Assuming that the sampling intervals in three dimensions are Δr , Δx , and Δy , the upper and lower bounds of the search scope are \bar{r} , \underline{r} , \bar{x} , \underline{x} , \bar{y} , and \underline{y} . Then, the number of branches are

$$n_x = \frac{\bar{x} - \underline{x}}{\Delta x}, n_y = \frac{\bar{y} - \underline{y}}{\Delta y}, n_r = \frac{\bar{r} - \underline{r}}{\Delta r}.\quad (18)$$

The sets of rotation matrices and translation vectors of the center of each divided subregion are

$$\mathcal{R} = \{\mathbf{R}_z(\theta) | \theta = \underline{r} + i\Delta r, i = 1, 3 \dots 2n_r - 1\}\quad (19)$$

$$\mathcal{A} = \{\mathbf{t} | \mathbf{t} = (\underline{x} + i_x \Delta x, \underline{y} + i_y \Delta y, 0)^T\}$$

$$i_x = 1, 3 \dots 2n_x - 1, i_y = 1, 3 \dots 2n_y - 1\}.\quad (20)$$

Then, the sampled transformation matrices are

$$\mathcal{T} = \{\mathbf{T}(\mathbf{R}, \mathbf{t}) | \mathbf{R} \in \mathcal{R}, \mathbf{t} \in \mathcal{A}\}.\quad (21)$$

3) Formulation of the Data Association Credibility: In order to describe the credibility of data association between two points, the variable c expressed as the probability that the two points refer to the same point in the global map is defined in (22).

$$c(\mathbf{x}_k^s, \mathbf{x}_{k,i}^t, l_k^s, l_{k,i}^t, \mathbf{T}) = p(d_{k,i} | \mathbf{x}_k^s, \mathbf{x}_{k,i}^t, l_k^s, l_{k,i}^t, \mathbf{T}).\quad (22)$$

Then, the credibility of the data association between two semantic maps can be defined as

$$c(\mathcal{M}^s, \mathcal{M}^t, \mathbf{T}) = \sum_{d_k \in \mathcal{D}} p(d_k | \mathbf{x}_k^s, \mathbf{x}_k^t, l_k^s, l_k^t, \mathbf{T}).\quad (23)$$

Since multiple data association for a given point will be considered, (23) is modified to

$$c(\mathcal{M}^s, \mathcal{M}^t, \mathbf{T}) = \sum_{D_k \in \mathcal{D}} \sum_{d_{k,i} \in D_k} p(d_{k,i} | \mathbf{x}_k^s, \mathbf{x}_{k,i}^t, l_k^s, l_{k,i}^t, \mathbf{T}).\quad (24)$$

4) Initial Value Selection: With the method proposed above, a traverse search on the set of sampled transformation matrices is performed and the credibility of data association will be computed with every sample. The initial value can be determined by selecting the transformation matrix with the largest data association credibility [see (25)].

$$\hat{\mathbf{T}}_{\text{init}} = \operatorname{argmax}_{\mathbf{T}_{\text{init}} \in \mathcal{T}} c(\mathcal{X}_s, \mathcal{X}_t, \mathbf{T}_{\text{init}}).\quad (25)$$

While the maximum of credibility is used to select the initial transformation, the remaining calculation results are utilized to estimate the reliability of the optimization. When the calculated

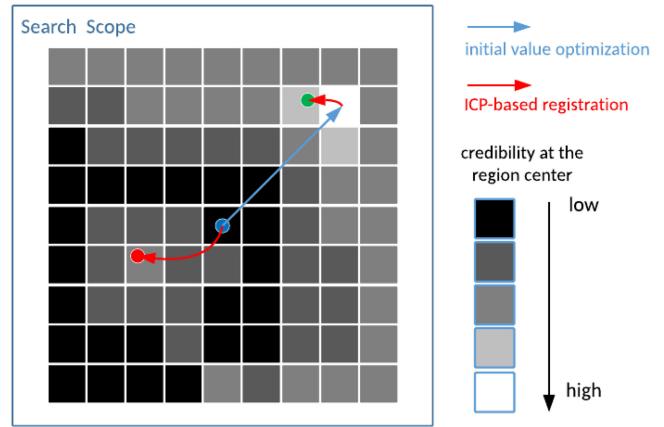


Fig. 5. Difference between map-matching process of traditional ICP-based algorithm and the algorithm with initial value optimization. The red point represents the local minimum and the green point represents the global minimum.

maximum value of credibility is much larger than other calculation results, it is more likely that the selected initial value is close to the global minimum, so the calculation is reliable. If not, more than one subregions may contain the global minimum. Besides, it is obvious that the reliability is also related to the sampling density. Then, the reliability factor of DACE optimization is defined as

$$k_{\text{DACE}} = \frac{c(\mathcal{M}^s, \mathcal{M}^t, \mathbf{T}_{\text{init}}) - c_m(\mathcal{M}^s, \mathcal{M}^t, \mathbf{T})}{c_m(\mathcal{M}^s, \mathcal{M}^t, \mathbf{T})} \theta\quad (26)$$

$c_m(\mathcal{X}_s, \mathcal{X}_t, \mathbf{T})$ represents the average value of the m largest results of the data association credibility calculation, where m can be set to 5–10% of the number of samples. θ is a parameter describing sampling density defined in (27).

$$\theta = \frac{n_r n_x n_y}{(\bar{x} - \underline{x})(\bar{y} - \underline{y})(\bar{z} - \underline{z})}.\quad (27)$$

The advantage of initial value optimization is shown in Fig. 5, which can avoid being trapped into local minimums in many cases and the convergence process will be shortened. Since the DACE algorithm reduces the initial error, the reliability of the optimization also reflects the uncertainty of initial error. The initial error uncertainty can be used to modify the objective function to be minimized in Section IV-C.

C. Semantic Map Matching Using EM

The optimization problem in the form of EM is presented in Section III. However, in practice, the maximum likelihood in (5) is usually expressed as the minimization of the expectation of matching error in (28).

$$\begin{aligned}\hat{\mathbf{T}}^{j+1} &= \operatorname{argmin} E_{\mathcal{D}^j}(\mathcal{R} | \mathcal{M}^s, \mathcal{M}^t, \mathbf{T}^{j+1}) \\ &= \sum_{d_{i,k}^j \in \mathcal{D}^j} p(d_{i,k}^j | \mathcal{M}^s, \mathcal{M}^t, \mathbf{T}^j) \|\mathbf{r}_{k,i}\|\end{aligned}\quad (28)$$

E step: In the E step, the probability of the latent variable is computed, which has been formulated in Sections III and IV-A.

The calculation method of the covariance matrix in Section III only considers the uncertainty estimated from the spatial distribution of points in the map. However, the uncertainty of the initial value of the iteration plays a significant role in the residual at the beginning of the iteration. Once the iteration begins with a large initial error, the points in the overlapping area are far from each other, which causes a large residual. In order to increase the probability of establishing correct data association, factor k_{DACE} in (26) is introduced to modify the probability registration model. Since the reliability of DACE is inversely correlated with the effect of the initial error, k_{DACE}^{-1} can be utilized to describe the uncertainty. Considering the effect of the initial error decreases gradually in the iteration process, the covariance factor k_{cov} is defined as

$$k_{\text{cov}} = \begin{cases} -\frac{k_s k_{\text{DACE}}^{-1} - 1}{ir_{\text{th}}} ir + k_s k_{\text{DACE}}^{-1} & ir \leq ir_{\text{th}} \\ 1 & ir > ir_{\text{th}} \end{cases} \quad (29)$$

ir represents the number of iterations, and ir_{th} is a parameter representing the minimum number of iterations required to eliminate the initial error of the iteration. The speed factor k_s can be determined according to the calculation efficiency requirements. If the algorithm is required to converge faster, k_s can be set larger. It is worth noting that k_{cov} equals $k_s k_{\text{DACE}}^{-1}$ at the beginning of the iteration and equals 1 after ir_{th} iterations. The calculation method of the covariance matrix after considering the uncertainty of the initial error is $\mathbf{C}_{k,i}^{\text{cov}} = k_{\text{cov}} \mathbf{C}_{k,i}$. With the probability, the cost function can be constructed by computing the expectation, where the probability computed above works as the weight of each error.

$$w_k^i = p(d_k^i | \mathbf{x}_k^s, \mathbf{x}_k^{t,i}, l_k^s, l_k^{t,i}, \mathbf{T}) \quad (30)$$

where the probability is calculated with $\mathbf{C}_{k,i}^{\text{cov}}$ instead of $\mathbf{C}_{k,i}$.

M Step: In the M step, with the weights and residuals, the cost function can be expressed in a least square form.

$$f_{\text{cost}} = \sum_{k=1}^n \sum_{i=1}^N \left(w_k^i \left\| \mathbf{x}_k^{t,i} - \mathbf{T}(\mathbf{x}_k^s) \right\|_{\mathbf{C}_{k,i}^{\text{cov}}} \right). \quad (31)$$

The cost function can be minimized with the ceres solver [31] by optimizing the transformation matrix. The new transformation will be used in the next E step. The iteration will stop when (32) holds.

$$e(\mathbf{T}^i, \mathbf{T}^{i-1}) < \epsilon \quad (32)$$

where ϵ is the threshold for judging the algorithm to converge. To ensure the convergence, ϵ is set as a small value ($1e-6$). \mathbf{T}^i and \mathbf{T}^{i-1} refer to the transformation matrix estimated in the iteration of this time and last time. The value of the threshold should be determined according to the required accuracy of the map-matching task. When the algorithm is nearly converged, the matching error between maps is already small, so there is no need to build a one-to-multiple data association. Instead, building a one-to-one data association can improve efficiency. Thus, the parameter N in (9) is required to change over the iteration process

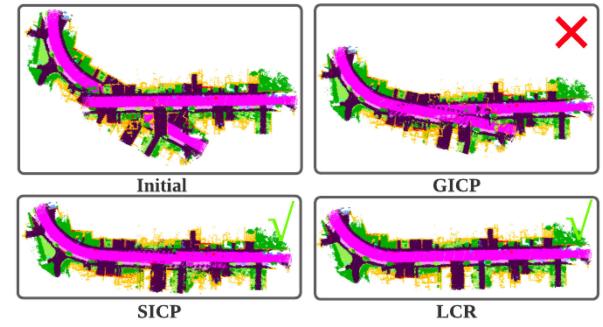


Fig. 6. Matching results on Semantic KITTI 09 dataset.

[see (33)].

$$N = \begin{cases} 1 & \text{once } e(\mathbf{T}^i, \mathbf{T}^{i-1}) < \epsilon' \\ N_{\text{init}} & \text{before } e(\mathbf{T}^i, \mathbf{T}^{i-1}) < \epsilon' \end{cases} \quad (33)$$

where ϵ' is the threshold for judging that the algorithm is close to convergence, which can be set to several times larger than ϵ . The estimated transform can be obtained at the end of the iteration.

V. EXPERIMENTAL RESULTS

In this section, the performance of the proposed LCR-SMM method is validated through experiments performed on semantic KITTI datasets [32]. The input semantic maps are established by reading the 3-D coordinates and semantic labels from the dataset and transforming them to the world coordinate system. The executed algorithm is a C++ program running on a personal laptop with an Intel Core i7-6500 U CPU @2.50 GHz, 8 GB RAM, and the OS is Ubuntu 18.04.

Semantic KITTI00, KITTI07, and KITTI09 datasets are selected to test the algorithm. The selected fragments contain the most common elements in the Semantic KITTI dataset. We also perform experiments on maps from other sequences and receive similar results. The map generated from the dataset is split into two local maps to simulate the maps generated by two robots. Since few available algorithms directly address semantic map matching, several geometric matching algorithms, including ICP [14], GICP [15], NDT [27], VGICP [28], and semantic matching algorithms SICP [10] and COSEM [23] are introduced as the comparison baseline. The maximum number of iterations is set to 50.

A. Matching Accuracy

As discussed above, the most significant improvement of our method compared to the existing algorithms is the matching accuracy with large initial errors. To verify that, this part aims to show the matching accuracy of our algorithm. More specifically, we walked through the optimization process of each algorithm to illustrate the internal principle.

The first experiment is performed on the Semantic KITTI09 dataset. A rotation error of -30° around the z -axis and a translation error of 9 m along the y -axis are introduced and the overlapping percentage is 67.1%. GICP provides the transformation far from the ground-truth (see Fig. 6). This is because the map is

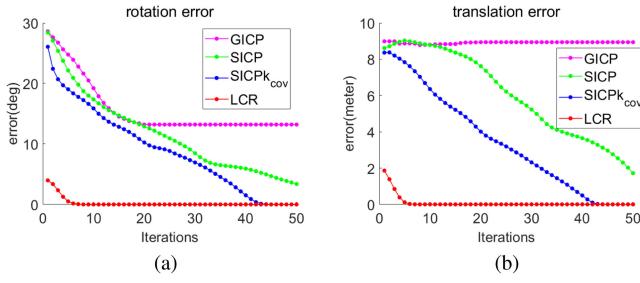


Fig. 7. (a) Rotation error and (b) translation error changing over the iterations on Semantic KITTI 09 dataset.

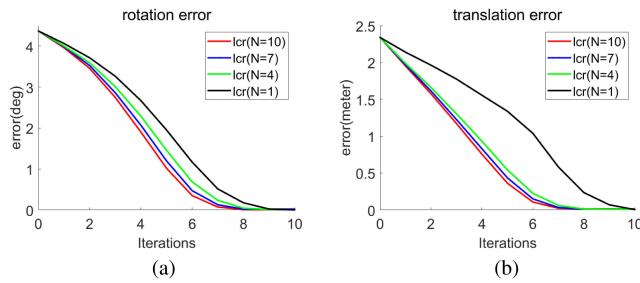


Fig. 8. Rotation error (left) and translation error (right) changing over the iterations with different data association strategy.

lack of geometry feature. SICP and LCR-SMM perform better because they make full use of semantic information to establish correct data association. The convergence processes of LCR-SMM and the modified SICP algorithm with k_{cov} (SICP_{k_{cov}}, $k_{cov} = 20$), SICP, and GICP are shown in Fig. 7. It can be seen that our algorithm greatly improves the convergence speed in both rotation and translation. The reason is that the improved data association model can speed up optimization and increase the convergence of the algorithm. To analyze the advantage of one-to-multiple data association, LCR-SMM is running with different N values, and the convergence process is shown in Fig. 8. As the value of N increases, the convergence speed is accelerated. The results above reveal that one-to-multiple data association can speed up the convergence. However, since the increasing of N will increase the complexity of objective function, it is inefficient to set N too large.

In the Semantic KITTI07 dataset, more challenging initial errors are introduced. More specifically, a rotation error of 150° around the z -axis and a translation error of 8 m along the y -axis are introduced and the overlapping percentage is 64.1%. In this case, only the LCR-SMM algorithm successfully converges to a global minimum (see Fig. 9). This is because the initial error is too large for other algorithms to establish correct data association, while our method reduces the initial error by applying DACE.

To simulate the situation with more robots, the experiment of matching four fragments of a semantic map is performed, and the result is shown in Fig. 10. The overlapping between 1 and 2, 2 and 3, and 3 and 4 are 56.4, 53.2, and 60.0%, respectively. As shown in the figure, LCR-SMM succeeds in estimating the transformation between the four fragments.

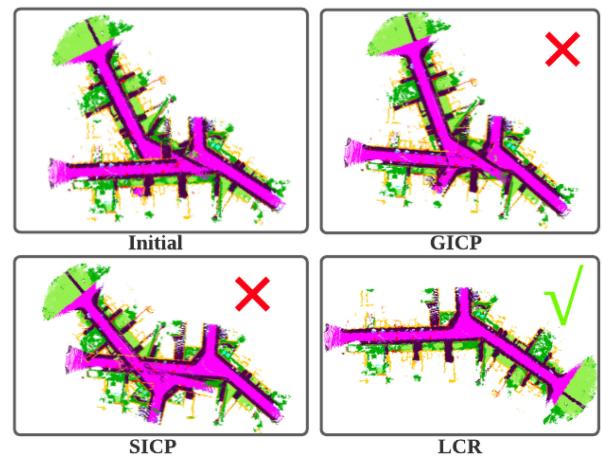


Fig. 9. Matching results on Semantic KITTI 07 dataset.

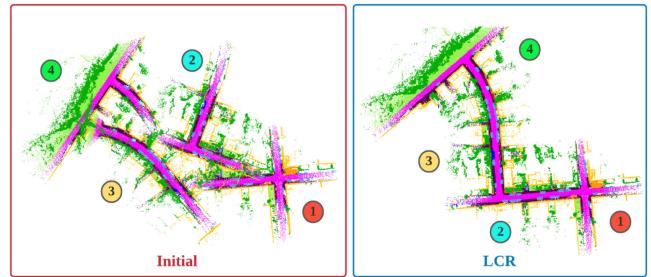


Fig. 10. Matching results of four partial overlapped maps.

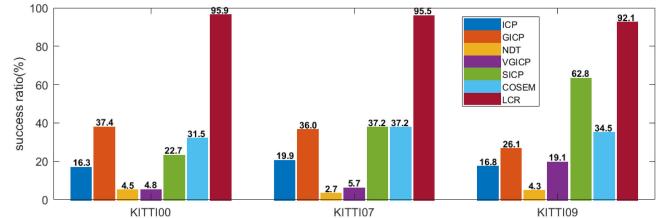


Fig. 11. Success ratios in 441 groups of experiments.

B. Statistic Testing Results

To quantitatively analyze the robustness and accuracy of the algorithm, experiments are performed on local maps generated from semantic KITTI00, KITTI07, and KITTI09 datasets with 441 different initial errors, and the overlapping percentages are 62.3, 64.1, and 67.1%, respectively. The translation errors range from -9 to 9 m in steps of 3 m is introduced along the x - and y -axes, and the map rotates from -30° to 30° in steps of 7.5°.

1) *Robustness to Initial Errors*: For all 441 initial errors, the success ratios of different algorithms are shown in Fig. 11. Successful map matching is defined as when the rotation error is less than 5° and the translation error is less than 2 m. In all cases, the performance of LCR-SMM is significantly better than other algorithms. In most cases, the success ratios of geometric matching algorithms are lower than semantic ones. Among geometric matching algorithms, GICP has the best performance, but the gap with LCR-SMM is still very large. To verify the superiority of our method in the situation where the initial error

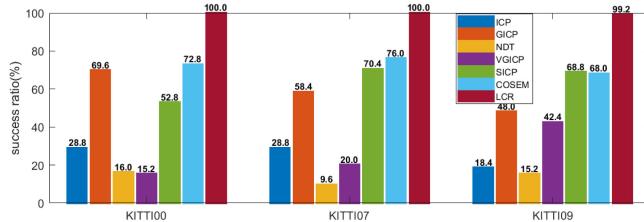


Fig. 12. Success ratios in 125 groups of experiments.

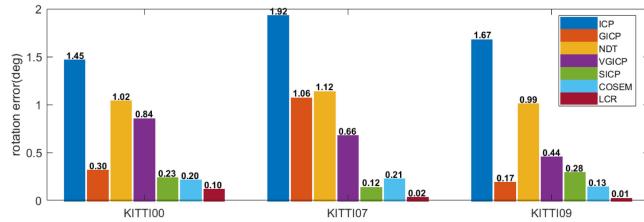


Fig. 13. Mean rotation errors in successful cases.

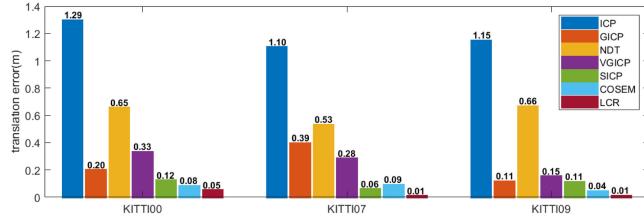


Fig. 14. Mean translation errors in successful cases.

is not too large, the success ratios of 125 sets are counted when initial rotation errors range from -15° to 15° and translation errors are between -6 and 6 m. As shown in Fig. 12, the success rate for all algorithms has increased given smaller initial errors, where LCR-SMM still has the highest success ratio. The overall results demonstrate that LCR-CMM has a larger convergence baseline and is far more robust than other algorithms.

2) Mean Errors: We also present the mean errors to show the overall accuracy with 441 initial errors. Since the matching errors are too large when benchmark algorithms fail, we filter out failed outliers and only count the mean errors when the matching is successful. Therefore, only rotation errors and translation errors in successful cases are selected to calculate mean errors, which are shown in Figs. 13 and 14. It can be learned that the errors of semantic matching algorithms are less than geometric ones, which proves introducing semantic information improves matching accuracy. Furthermore, LCR-SMM method has the least matching error. This is because our method optimizes the initial value and improves the registration model, which greatly increases the accuracy.

C. Execution Efficiency

The mean runtime of each algorithm is summarized in Table II. In general, the geometric-matching algorithms cost less time than semantic-matching algorithms with EM. This is because the semantic one-to-multiple data association strategy increases the number of terms in the cost function, therefore

TABLE II
MEAN RUNTIME OF MAP-MATCHING ALGORITHMS (S)

map	ICP	GICP	NDT	VGICP	SICP	COSEM	LCR
00	37.1	28.6	28.6	3.2	450.9	41.0	99.5
07	33.4	26.1	25.7	3.8	466.9	48.6	107.8
09	23.3	20.8	29.3	3.4	405.4	36.5	101.7
mean	31.3	25.2	27.9	3.5	441.1	42.0	103.0

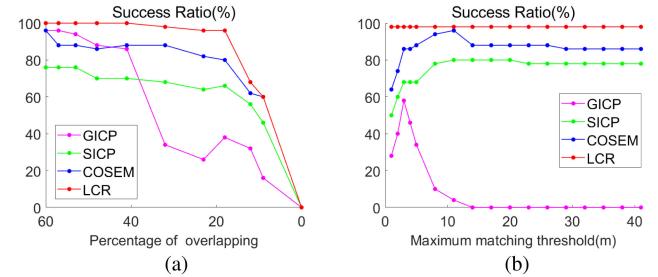


Fig. 15. Robustness to overlapping area and maximum matching threshold. (a) Success ratios changing over different overlapping areas. (b) Success ratios changing over different maximum matching thresholds.

increasing the time of optimization. Although VGICP achieves high efficiency, its success ratio is not comparable with GICP. Due to the additional initial value optimization process, LCR-SMM is more time-consuming than COSEM and geometric matching algorithms, but much more efficient than SICP. This is because our method reduces the initial errors and accelerates the optimization process. Considering the significant improvement in robustness and accuracy compared to COSEM, the time cost of LCR-SMM is acceptable. In multirobot applications, relative localization does not need to be performed in high frequency and can be executed occasionally.

D. Robustness to Overlapping Area

To analyze the robustness of the algorithm to the overlapping area, the semantic maps from the KITTI00 dataset are cropped into 10 different patterns so that the overlapping percentages vary from 0 to 60%, with the rotation errors of -7.5° or 7.5° around the z-axis and the translation errors vary from -4 to 4 m in steps of 2 m along the x-axis and y-axis applied to them. Since other algorithms perform poorly in this robustness testing, only GICP, SICP, COSEM, and LCR-SMM are reported in Fig. 15(a). As the overlapping area shrinks, the success ratio of GICP decreases rapidly while COSEM and LCR-SMM succeed in more than 80% cases even if the overlapping is less than 30%. The reason why semantic methods are more robust is that the semantic information supplements the features to be matched. More specially, the success ratio of LCR-SMM is barely changed when the overlapping ratio is larger than 20%. This is because LCR-SMM shortens the distance between the overlapping area from local maps in the initial state.

E. Robustness to Maximum Matching Threshold

The maximum matching threshold is an important parameter in iterative optimization-based algorithms, which determines the maximum distance between the associated points. We perform

850 groups of experiments on the KITTI00 dataset with GICP, SICP, COSEM, and LCR-SMM to test the robustness of the algorithms to the maximum matching threshold. The initial errors are the same as those in Section V-D. The success ratios of the four algorithms with the change of maximum matching threshold are shown in Fig. 15(b). When the threshold is larger than 10 m, the success ratios of SICP, COSEM, and LCR-SMM are stable while GICP failed. This is because the errors in the objective function of these methods are weighted with a factor negative related to the matching distance. The effect of long-distance association in the optimization is neglected. When the threshold is less than 5 m, only LCR-SMM can work stably. This is because LCR-SMM reduces the initial error, the map matching can be completed even if the algorithm becomes “short-sighted.”

VI. CONCLUSION

This article has presented an LCR-SMM. More specifically, the proposed DACE method operates a sampling-traverse process to optimize the initial transformation. To reduce the possibility of being trapped into the local minimum, we perform transformation matrix sampling and select the sample with the largest data association credibility, which enlarges the convergence region. In addition, the probabilistic registration model is introduced to improve the effect of long-distance data association, which speeds up the convergence and further improves the efficiency. The results have demonstrated that the proposed algorithm achieves high robustness and accuracy. In summary, the proposed algorithm has raised a novel perspective of matching dense semantic maps. In the future, we plan to further improve the DACE method to shorten the runtime of initial transformation optimization. In addition to collaborative mapping, we will modify the algorithm so that it can be used in more fields, such as the semantic SLAM and semantic 3-D reconstruction.

REFERENCES

- [1] Y. Yue and D. Wang, *Collaborative Perception, Localization and Mapping for Autonomous Systems*. Basingstoke, U.K.: Springer Nature, vol. 2, 2021.
- [2] Y. Wang, Y. Yue, M. Shan, L. He, and D. Wang, “Formation reconstruction and trajectory replanning for multi-UAV patrol,” *IEEE/ASME Trans. Mechatronics*, vol. 26, no. 2, pp. 719–729, Apr. 2021.
- [3] Y. Yue *et al.*, “A multilevel fusion system for multirobot 3-D mapping using heterogeneous sensors,” *IEEE Syst. J.*, vol. 14, no. 1, pp. 1341–1352, Mar. 2020.
- [4] S. Saeedi, M. Trentini, M. Seto, and H. Li, “Multiple-robot simultaneous localization and mapping: A review,” *J. Field Robot.*, vol. 33, no. 1, pp. 3–46, 2016.
- [5] J. Jessup, S. N. Givigi, and A. Beaulieu, “Robust and efficient multirobot 3-D mapping merging with Octree-based occupancy grids,” *IEEE Syst. J.*, vol. 11, no. 3, pp. 1723–1732, Sep. 2017.
- [6] I. Kostavelis and A. Gasteratos, “Semantic mapping for mobile robotics tasks: A survey,” *Robot. Auton. Syst.*, vol. 66, pp. 86–103, 2015.
- [7] S. L. Bowman, N. Atanasov, K. Damilidis, and G. J. Pappas, “Probabilistic data association for semantic SLAM,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2017, pp. 1722–1729.
- [8] K. Doherty, D. Fourie, and J. Leonard, “Multimodal semantic SLAM with probabilistic data association,” in *Proc. Int. Conf. Robot. Automat.*, 2019, pp. 2419–2425.
- [9] K. J. Doherty, D. P. Baxter, E. Schneeweiss, and J. J. Leonard, “Probabilistic data association via mixture models for robust semantic slam,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 1098–1104.
- [10] S. A. Parkison *et al.*, “Semantic iterative closest point through expectation–maximization,” in *Proc. Brit. Mach. Vis. Conf.*, Sep. 2018, pp. 1–17.
- [11] Y. Yue, C. Zhao, Z. Wu, C. Yang, Y. Wang, and D. Wang, “Collaborative semantic understanding and mapping framework for autonomous systems,” *IEEE/ASME Trans. Mechatronics*, vol. 26, no. 2, pp. 978–989, Apr. 2021.
- [12] Y. Yue *et al.*, “A hierarchical framework for collaborative probabilistic semantic mapping,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 9659–9665.
- [13] V. Balaska, L. Bampis, I. Kansizoglu, and A. Gasteratos, “Enhancing satellite semantic maps with ground-level imagery,” *Robot. Auton. Syst.*, vol. 139, 2021, Art. no. 103760.
- [14] P. J. Besl and N. D. McKay, “Method for registration of 3-D shapes,” in *Sensor Fusion IV: Control Paradigms and Data Structures. International Society for Optics and Photonics*, 1992, vol. 1611, pp. 586–606.
- [15] A. Segal, D. Haehnel, and S. Thrun, “Generalized-ICP,” in *Proc. Robot.: Sci. Syst.*, Seattle, WA, USA, 2009, vol. 2, p. 435.
- [16] Q. Zhang, M. Wang, and Y. Yue, “Robust semantic map matching algorithm based on probabilistic registration model,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2021, pp. 5289–5295.
- [17] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, “PointNet: Deep learning on point sets for 3D classification and segmentation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 652–660.
- [18] N. Atanasov *et al.*, “A unifying view of geometry, semantics, and data association in SLAM,” in *Proc. 27th Int. Joint Conf. Artif. Intell.*, 2018, pp. 5204–5208.
- [19] T. M. Bonanni, B. Della Corte, and G. Grisetti, “3-D map merging on pose graphs,” *IEEE Robot. Automat. Lett.*, vol. 2, no. 2, pp. 1031–1038, Apr. 2017.
- [20] J. Serafin and G. Grisetti, “NICP: Dense normal based point cloud registration,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2015, pp. 742–749.
- [21] S. Saeedi, L. Paull, M. Trentini, M. Seto, and H. Li, “Group mapping: A topological approach to map merging for multiple robots,” *IEEE Robot. Automat. Mag.*, vol. 21, no. 2, pp. 60–72, Jun. 2014.
- [22] Y. Yue *et al.*, “Day and night collaborative dynamic mapping in unstructured environment based on multimodal sensors,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 2981–2987.
- [23] Y. Yue, M. Wen, C. Zhao, Y. Wang, and D. Wang, “COSEM: Collaborative semantic map matching framework for autonomous robots,” *IEEE Trans. Ind. Electron.*, 2021, early access, doi: [10.1109/TIE.2021.3070497](https://doi.org/10.1109/TIE.2021.3070497).
- [24] C.-S. Chen, Y.-P. Hung, and J.-B. Cheng, “RANSAC-based DARCES: A new approach to fast automatic registration of partially overlapping range images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 11, pp. 1229–1234, Nov. 1999.
- [25] H. Yang and L. Carlone, “A polynomial-time solution for robust registration with extreme outlier rates,” *Robotics: Sci. Syst.*, 2019.
- [26] K. S. Arun, T. S. Huang, and S. D. Blostein, “Least-squares fitting of two 3-D point sets,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 5, pp. 698–700, Sep. 1987.
- [27] P. Biber and W. Strasser, “The normal distributions transform: A new approach to laser scan matching,” in *IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, vol. 3, pp. 2743–2748, 2003.
- [28] K. Koide, M. Yokozuka, S. Oishi, and A. Banno, “Voxelized GICP for fast and accurate 3D point cloud registration,” *IEEE Int. Conf. Robot. Automat. (ICRA)*, IEEE, pp. 11054–11059, 2021.
- [29] J. Yang, H. Li, D. Campbell, and Y. Jia, “Go-ICP: A globally optimal solution to 3D ICP point-set registration,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 11, pp. 2241–2254, Nov. 2016.
- [30] Z. Min, J. Wang, and M. Q.-H. Meng, “Robust generalized point cloud registration with orientational data based on expectation maximization,” *IEEE Trans. Automat. Sci. Eng.*, vol. 17, no. 1, pp. 207–221, Jan. 2020.
- [31] S. Agarwal *et al.*, “Ceres solver,” 2012. [Online]. Available: <http://ceres-solver.org>
- [32] J. Behley *et al.*, “Semantic KITTI: A dataset for semantic scene understanding of LiDAR sequences,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 9296–9306.



Qingxiang Zhang received the B.S. degree in electrical engineering and automation, in 2020, from the Beijing Institute of Technology, Beijing, China, where he is currently working toward the Ph.D. degree in intelligent navigation with the School of Automation.

His research interests include the semantic understanding and collaborative localization for robot systems.



Meiling Wang received the B.S. and M.S. degrees in automation and Ph.D. degree in navigation, guidance, and control from Beijing Institute of Technology, Beijing, China, in 1992, 1995, and 2007, respectively.

She was with the University of California San Diego as a Visiting Scholar in 2004. Since 1995, she has been with the Beijing Institute of Technology, where she is currently the Director of Integrated Navigation and Intelligent Navigation Laboratory. Her research interests include advanced technology of sensing and detecting and vehicle intelligent navigation.

Dr. Wang was elected as the Yangtze River Scholar Distinguished Professor in 2014.



Tong Liu received the M.S. and Ph.D. degrees in control theory and engineering from the Department of Automation, Beijing Institute of Technology, Beijing, China, in 2002 and 2013, respectively.

Since 2002, he has been with the Department of Automatic Control, Beijing Institute of Technology, first as a Lecturer, and then, since 2018, as an Associate Professor. His research interests include new angular accelerometer, robot indoor positioning technology, and unmanned

vehicles.



Yufeng Yue (Member, IEEE) received the B.Eng. degree in automation from the Beijing Institute of Technology, Beijing, China, in 2014, and the Ph.D. degree in electrical and electronic engineering from Nanyang Technological University, Singapore, in 2019.

He is currently an Associate Professor with School of Automation, Beijing Institute of Technology. He has authored or coauthored a book in Springer, and more than 40 journal/conference papers, including *IEEE/ASME Transactions on Mechatronics*, *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS*, *IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY*, *IEEE Transactions on Multimedia TMM*, *IEEE International Conference on Robotics and Automation*, and *IEEE/RSJ International Conference on Intelligent Robots and Systems*. His research interests include perception, mapping and navigation for collaborative autonomous systems.

Dr. Yue was the Associate Editor for 2020–2021 *IEEE/RSJ International Conference on Intelligent Robots and Systems*. He has received the 2020 IEEE ICARCV Best Paper Award.