2021 IEEE International Conference on Robotics and Automation (ICRA 2021)
May 31 - June 4, 2021, Xi'an, China

2021 IEEE International Conference on Robotics and Automation (ICRA) | 978-1-7281-9077-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICRA48506.2021.9561176

Robust Semantic Map Matching Algorithm Based on Probabilistic Registration Model

Qingxiang Zhang, Meiling Wang, Yufeng Yue*

Abstract— **The matching and fusing 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 closet point (ICP), which typically fail in cases with large initial error. What's more, current semantic matching algorithms have high computation complexity in optimizing the transformation matrix. To address the challenge of map matching with large initial error, this paper proposes a novel semantic map matching algorithm with large convergence region. 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.**

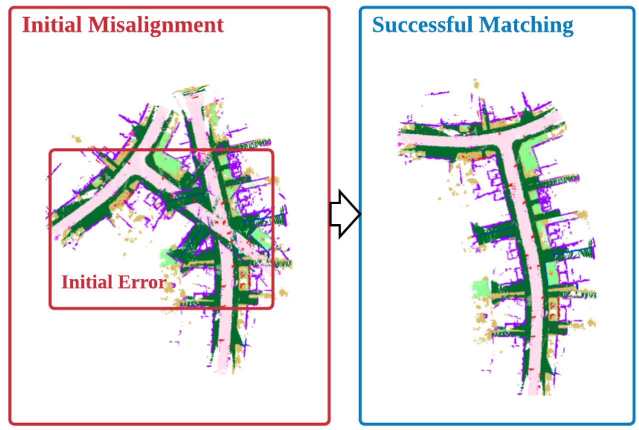


Fig. 1. The performance of the proposed algorithm on Semantic KITTI00 dataset. The initial error is 40° around z-axis, 8m along x-axis, and 6m along y-axis. The proposed algorithm is able to successfully match the two local maps in such a challenging case.

I. Introduction

Recently, the collaborative operation of a group of robots has attracted increasing attention, which promotes the development of multiple-robot mapping [1]–[3]. When there are multiple robots in the environment, the local maps need to be matched to obtain a global map, which also helps robots to perform relative localization. Currently, most existing map matching methods are based on geometric feature [4]–[7]. 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 [8]. 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 SLAM algorithms [9]–[11] 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 initial transformation is close to the ground truth, which will fail in multiple-robot map matching problems with large initial errors. Nevertheless, the latest collaborative semantic mapping research has focused on combining local maps into a global representation [12], [13], but not on how semantic information can improve the relative localization accuracy between robots. Therefore, the main objective of this paper 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 associations between corresponding points. Most existing algorithms establish data association based on closest neighbor search in either euclidean or probabilistic error metrics [14]–[16], which is unreasonable when the points to be matched are far from each other in the initial state. As a consequence of the low quality of data association, the convergence region tends to be narrow. To address this problem, the initial transformation optimization algorithm called data association credibility estimation (DACE) is proposed to ensure that the matching process is started with a set of credible data associations. Besides, 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 using Expectation Maximization(EM) method [9], [17], which increases the probability of matching points with the same semantic label. However, the error function is more complicated, thereby prolonging the running time. Therefore, a probabilistic registration model is proposed to

This work is partly supported by the National Natural Science Foundation of China under Grant 62003039, the Changjiang Scholars Programme of China under Grant No. T2014224, and the Beijing Institute of Technology Research Fund Program for Young Scholars. (*Corresponding Author: Yufeng Yue*)

Qingxiang Zhang, Meiling Wang and Yufeng Yue are with School of Automation, Beijing Institute of Technology, Beijing, 100081, China (yueyufeng@bit.edu.cn)

improve the role of long-distance data association in the optimization process.

As summarized, there still exists a substantial gap in improving the convergence region and matching robustness. In this paper, a novel large convergence region semantic map matching (LCR-SMM) algorithm is proposed. 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 is the proposing of the data association credibility estimation (DACE) method to optimize the initial transformation matrix. More specifically, a probability registration model is proposed that models initial error uncertainty. Extensive quantitative results are presented to show the robustness and accuracy of the algorithm, which extends the convergence region of map matching.

The remainder of this paper proceeds 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. Section VI concludes this paper.

II. Related works

In this section, existing algorithms related to semantic map matching are introduced, including single robot semantic mapping and map matching methods.

A. Single Robot Semantic Mapping

With the fast development of the neural network, many methods for semantic segmentation [18] based on deep learning are developed, which accelerates the researches on semantic SLAM [9]–[11]. In [9], a semantic SLAM method is proposed based on EM model, which uses EM to solve the data association problem for odometry estimation. After that, the authors in [10] modify the probabilistic data association model in [9] into a max-mixture model, which uses max-marginalization instead of the sum-marginalization to calculate the probability of one-to-multiple data association. The authors in [19] 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 robot semantic mapping. They usually assume that the initial transformation is close to the ground truth, which may fail in the case of large initial error and data size.

B. Map Matching Algorithms

Collaborative robots can greatly improve the efficiency of environmental mapping and reconstruction [4]–[6]. The authors in [5] applies the ICP algorithm to merge octree-based occupancy grid maps. Then, a general multi-level probabilistic framework is proposed to match local maps generated by multiple robots with heterogeneous sensors in [6]. However, their work focused on geometric map matching, not semantic map matching.

Since the key to map matching lies in finding the transformation between the maps, many existing point cloud registration methods are closely related. Feature matching based algorithms extract 3D key points with descriptors [20], [21], and then match the key points using the descriptors and minimize the distances between matched points with Least-squares [22]. In recent years, many scholars have established deep learning networks to perform descriptor extraction and point cloud registration [23]–[26], and have also achieved good results. But the accuracy of these methods depends on the robustness of the methods to get 3D key points.

Another domain is based on iterative dense registration such as ICP [14], GICP [15], NDT [27], and the recent novel approach of Semantic ICP [17]. These algorithms have made extensive progress on the point cloud registration problem, but the ICP-based approaches typically fail when the initial transform deviates from the ground true value [17]. Global optimization method like BnB [28] is applied to solve large initial error problem, but it suffers from computational complexity.

III. Systematic Framework

A. The Overview of Semantic Map Matching Framework

An overview of the semantic map matching algorithm is presented in Fig. 2. The framework includes three parts: initial transformation optimization based on DACE, probability registration model, and Expectation Maximization. Firstly, initial transformation optimization provides the initial value of the iterative in Expectation Maximization. Secondly, the probability registration model combines semantic and geometric information and formulates in the form of a probabilistic formula. Thirdly, Expectation Maximization estimates data association, optimizes transformation iteratively and calculates the transformation matrix.

B. Overall Problem Formulation

Considering a semantic map matching problem, the input are two independent and overlapping semantic grid maps in the form of two sets of 3D point clouds with semantic labels. Define $\mathcal{X}^s = \{\mathbf{x}_k^s\}_{k=1}^n$ and $\mathcal{X}^t = \{\mathbf{x}_k^t\}_{k=1}^n$ as the source and target point clouds, where $\mathbf{x}^s$ and $\mathbf{x}^t$ represent the coordinate vectors of the points in the point cloud. $\mathcal{L}^s = \{l_k^s\}_{k=1}^n$ and $\mathcal{L}^t = \{l_k^t\}_{k=1}^n$ represent the semantic labels of the points in the source and target maps. 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. The objective is estimating the transformation matrix $\mathbf{T}$ that minimizes the errors between the transformed source point cloud and the target point cloud. In order to describe matching error, define $\mathcal{R} = \{\mathbf{r}_k\}_{k=1}^n$ to represent the residual, where $\mathbf{r}_k = \mathbf{x}_{t(k)}^t - \mathbf{T}(\mathbf{x}_{s(k)}^s)$. The problem can be formulated as the following maximum likelihood estimation (MLE) problem:

$$\mathbf{T} = \operatorname{argmax} p(\mathcal{X}^s, \mathcal{X}^t | \mathcal{D}, \mathcal{L}^s, \mathcal{L}^t, \mathbf{T}) \quad (1)$$

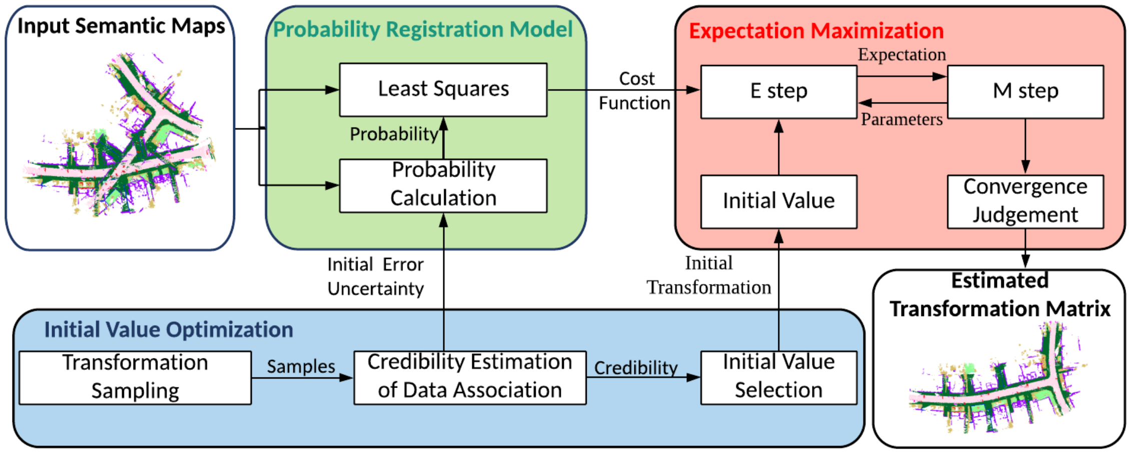


Fig. 2. The framework of Large Convergence Region Semantic Map Matching algorithm.

To solve the problem formulated above, data association must be established. Instead of constructing one-to-one data association in traditional ICP, the EM method is introduced to estimate multiple data associations. 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 associations 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 $\{\{\mathbf{x}_{k,i}^t, l_{k,i}^t\}\}_{i=1}^N$. To solve the problem with EM, the probability of data association, which is the latent variable in the MLE problem, needs to be calculated. Then the expectation of matching error can be computed and minimized by optimizing the transformation. The probability of data association can be divided into two parts: geometric and semantic data association, which is expressed as:

$$p(d_k^i|\mathbf{x}_k^s, \mathbf{x}_k^{t,i}, l_k^s, l_k^{t,i}) = p(d_k^{g,i}|\mathbf{x}_k^s, \mathbf{x}_k^{t,i}, \mathbf{T})p(d_k^{s,i}|l_k^s, l_k^{t,i}) \tag{2}$$

The probability formulated in Eq. 2 is collaborative data association established based on both semantic and geometry information. The expectation is minimized by updating the transformation matrix and the new transformation will be used in the next probability calculation.

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 Eq. 2, 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:

$$p(d_k^{s,i}|l_k^s, l_k^{t,i}) = \sum_{s_k \in \mathcal{S}} p(s_k|l_k^s)p(s_k|l_k^{t,i}) \tag{3}$$

The $p(s|l)$ in Eq. 3 represents the probability of that the label l belongs to the category s, which depends on the given semantic maps.

The geometric association probability can be described with the residual because the larger the residual error, the smaller the probability. This probability is calculated using three-dimensional Gaussian distribution model. 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. The probability of data association between two points can be calculated as:

$$p(d_{k,i}^g|\mathbf{x}_k^s, \mathbf{x}_{k,i}^t, \mathbf{T}) = \frac{I(k,i)}{2\pi^{\frac{3}{2}}|\mathbf{C}_{k,i}|^{\frac{1}{2}}} exp(-\frac{1}{2}\mathbf{r}_{k,i}^{\mathbf{T}}\mathbf{C}_{k,i}^{-1}\mathbf{r}_{k,i}) \tag{4}$$

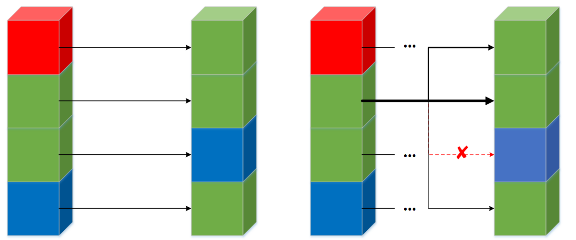
where $I(k,i)$ is the characteristic function defined in Eq. 5 and the covariance matrix $\mathbf{C}_{k,i} \triangleq \Sigma_{k,i}^t + \mathbf{R}^{\boldsymbol{T}}\Sigma_k^s\mathbf{R}$. $\mathbf{R}$ is the rotation matrix, $\Sigma_{k,i}^t$ and Σ_k^s are the covariance matrix of the two points in their respective point clouds, calculated over the distribution of the coordinates of a certain number of points near the point to be associated..

$$I(k,i) = \begin{cases} 1 & \mathbf{x}_{k,i}^t \text{ is } N \\ & \text{closest of } \mathbf{x}_k^s \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

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 associations.

B. DACE Algorithm for Initial Transformation Optimization

The credibility of the data association between two points can be expressed as the probability that the two points refer



(a) geometric data association (b) collaborative data association

Fig. 3. The difference between geometric data association and 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.

to the same point in the global map. Then the credibility of the data association between two point clouds can be defined as:

$$c(\mathcal{X}^s, \mathcal{X}^t, \mathbf{T}) = \sum_{d_k \in \mathcal{D}} p(d_k | \mathbf{x}_k^s, \mathbf{x}_k^t, l_k^s, l_k^t, \mathbf{T}) \tag{6}$$

Because multiple data association for a certain point will be considered, the Eq. 6 is modified to:

$$c(\mathcal{X}^s, \mathcal{X}^t, \mathbf{T}) = \sum_{D_k \in \mathcal{D}} \sum_{d_k^i \in D_k} p(d_k^i | \mathbf{x}_k^s, \mathbf{x}_k^{t,i}, l_k^s, l_k^{t,i}, \mathbf{T}) \tag{7}$$

With the formulation of data association credibility, the initial value of the transformation can be optimized by sampling a certain number of transformation matrices and selecting the transform with the best data association.

Considering the semantic map matching problem formulated above, $\mathcal{X}^s$ and $\mathcal{X}^t$ represent the source map and target map. Assuming that the possible range of the ground truth transformation is $\mathcal{T}_{GT}$. The problem required to be solved is to find a sampling method $S : \mathcal{T}_{GT} \longrightarrow \mathcal{T}$, 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:

$$\begin{aligned} \mathcal{T} &= argmin\, E_{\mathbf{T}_{GT}}[d_{min}(\mathbf{T}_{GT}, \mathbf{T}_k | \mathcal{T})] \\ &= argmin \sum_{\mathbf{T}_{GT} \in \mathcal{T}_{GT}} d_{min}(\mathbf{T}_{GT}, \mathbf{T}_k | \mathcal{T}) p(\mathbf{T}_{GT}) \end{aligned} \tag{8}$$

Without the prior estimation, the probability distribution of ground truth transformation in $\mathcal{T}_{GT}$ is uniform. In this case, it is obvious that the best sampling method is to divide $\mathcal{T}_{GT}$ into multiple sub-regions and take the matrices at the center of each sub-region as the samples. In common map matching problems for ground robots, the initial error mainly exists in the rotation error around the z-axis and the translation error on the xOy plane. Assuming that the number of branches is n_r, n_x, n_y, the bounds of the rotation and translation error are $\overline{r}$, $\underline{r}$ and $\overline{x}$, $\underline{x}$, $\overline{y}$, $\underline{y}$. If the information is insufficient, the range of ground truth rotation angle can be set to $[-\pi, \pi]$. Since it is assumed that the local maps are overlapping, the bounds of the components of translation vector along x-axis and y-axis can be defined as:

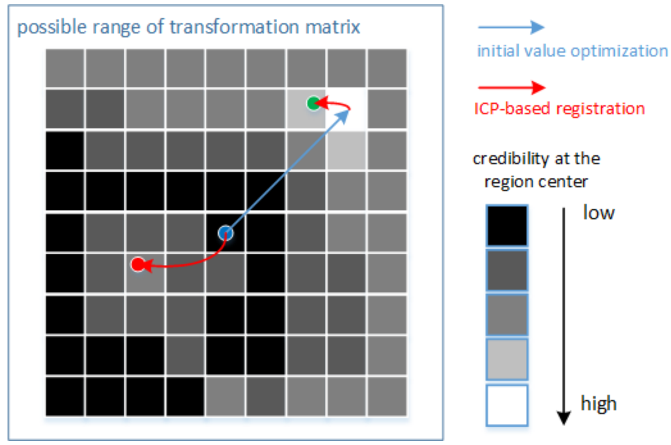


Fig. 4. The difference between the map matching process of traditional ICP-based algorithm and the algorithm with initial value optimization. The red point represents the local minimum and the green point represents the global minimum.

$$\overline{x} = \overline{x}^t + \overline{r}^s, \underline{x} = \underline{x}^t - \overline{r}^s \tag{9}$$

$$\overline{y} = \overline{y}^t + \overline{r}^s, \underline{y} = \underline{y}^t - \overline{r}^s \tag{10}$$

where $\overline{x}^t$, $\underline{x}^t$, $\overline{y}^t$, $\underline{y}^t$ represent the bounds of the x,y coordinate of the points in target map and $\overline{r}^s$ represents the upper bound of the distance from the points in the source cloud to the origin of coordinates. However, it is suggested to narrow down the range based on the prior information to improve the efficiency. The sets of rotation vector and translation vector of the center of each divided sub-region are:

$$\mathcal{R} = \{(0, 0, \underline{r} + i\Delta r) | i = 1, 3, 5...2n_r - 1\} \tag{11}$$

$$\mathcal{A} = \left\{(\underline{x} + i\Delta x, \underline{y} + i\Delta y, 0) | i = 1, 3, 5...2n_t - 1\right\} \tag{12}$$

Where $\Delta x = \frac{\overline{x} - \underline{x}}{n_x}, \Delta y = \frac{\overline{y} - \underline{y}}{n_y}, \Delta r = \frac{\overline{r} - \underline{r}}{n_r}$. Then the sampled transformation matrices are:

$$\mathcal{T} = \{\mathbf{T}(\mathbf{r}, \mathbf{t}) | \mathbf{r} \in \mathcal{R}, \mathbf{t} \in \mathcal{A}\} \tag{13}$$

So the initial transform can be determined by:

$$\mathbf{T}_{init}^* = \underset{\mathbf{T}_{init} \in \mathcal{T}}{argmax}\, c(\mathcal{X}^s, \mathcal{X}^t, \mathbf{T}_{init}) \tag{14}$$

The advantage of initial value optimization is shown in Fig. 4, algorithm can avoid being trapped into local minimums in many cases and the convergence process will be shortened.

The reliability of the optimization above can be estimated through the calculated credibility of data association. When the calculated maximum value of credibility is much larger than other calculation results, the calculation is reliable. Therefore, the initial error uncertainty can be defined as:

$$k_{init} = \frac{c_{avg}(\mathcal{X}^s, \mathcal{X}^t, \mathcal{T})}{c(\mathcal{X}^s, \mathcal{X}^t, \mathbf{T}_{init}^*) - c_{avg}(\mathcal{X}^s, \mathcal{X}^t, \mathcal{T})} \theta \tag{15}$$

where $c_{avg}(\mathcal{X}^s, \mathcal{X}^t, \mathcal{T})$ represents the average value of the five largest results of the data association credibility calculation, and θ is a parameter related to the calculation accuracy.

C. Semantic Map Matching Using EM

E step: In the E step, the probability of the latent variable is computed, which has been formulated in section III. However, the fact that the uncertainty of the initial value of the iteration has an effect on the residual at the beginning of the iteration is not considered. Hence, the initial error uncertainty factor can be defined as:

$$k_{cov} = \begin{cases} -\dfrac{k_{init} - 1}{ir_{th}} ir + k_{init} & ir < ir_{th} \\ 1 & ir > ir_{th} \end{cases} \tag{16}$$

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 calculation method of the covariance matrix after considering the uncertainty of the initial error is $\Sigma_{cov} = k_{cov}\Sigma$. 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. That is:

$$w_k^i = p(d_k^i | \mathbf{x}_k^s, \mathbf{x}_k^{t,i}, l_k^s, l_k^{t,i}, \mathbf{T}) \tag{17}$$

Then, the cost function can be expressed in least square error form.

$$f_{cost} = \sum_{k=1}^{n} \sum_{i=1}^{N} (w_k^i \left\| \mathbf{x}_k^{t,i} - \mathbf{T}(\mathbf{x}_k^s) \right\|_{\boldsymbol{C}_k}) \tag{18}$$

M step: The cost function can be minimized with the ceres solver [29]. When the algorithm is nearly converged, building one-to-one data association can improve the efficiency. Thus, the parameter N in Eq. 5 is required to change over the iteration process(see Eq. 19).

$$N = \begin{cases} 1 & \text{once } d((\mathbf{T}^i, \mathbf{T}^{i-1}) < \epsilon' \\ N_{init} & \text{before } d((\mathbf{T}^i, \mathbf{T}^{i-1}) < \epsilon' \end{cases} \tag{19}$$

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

V. Experimental Results

In this section, the performance of the large-convergence-region semantic map matching method is validated through experiments performed on semantic KITTI datasets [30]. The experiment program is a cmake project running on a personal laptop with an Intel Core i7-6500U CPU @2.50GHz, 8GB RAM, and the OS is Ubuntu 18.04.

Semantic KITTI00, KITTI07 and KITTI09 datasets are selected to test the algorithm. The map generated from a part of the dataset is splitted into two local maps to simulate the maps generated by two robots. Since there is no available algorithms directly address semantic map matching problem, several geometric matching algorithms ICP [14], GICP [15], NDT [27], VGICP [31], and semantic matching algorithm SICP [17] are introduced as 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 error. 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.

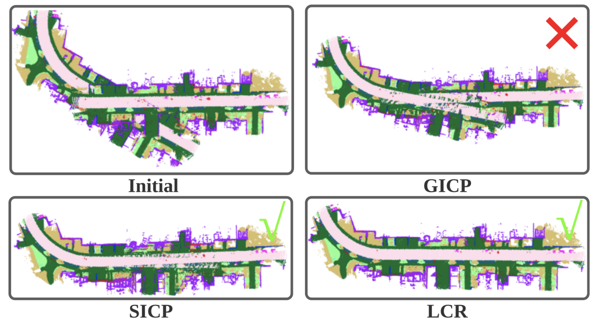
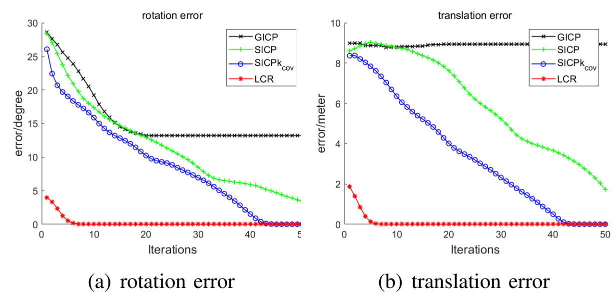


Fig. 5. The matching results on Semantic KITTI 09 dataset.



(a) rotation error (b) translation error

Fig. 6. The rotation error(left) and translation error(right) changing over the iterations on Semantic KITTI 09 dataset.

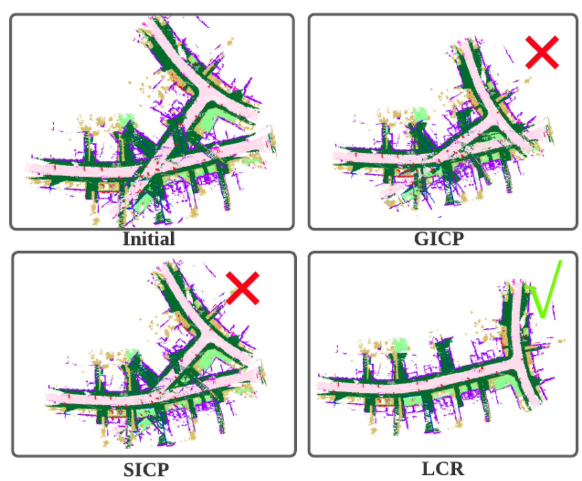


Fig. 7. The matching results on Semantic KITTI 00 dataset.

The first experiment is performed on Semantic KITTI09 dataset. A rotation error of $-30°$ around z-axis and a translation error of 9m along the y-axis are introduced. GICP provided the transformation far from the ground truth (see Fig. 5). This is because the map is 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 process of LCR-SMM, the modified SICP algorithm with k_{cov} (SICPk_{cov}, $k_{cov} = 20$), SICP, and GICP is shown in Fig. 6. It can be seen that our algorithm greatly improves the convergence speed in both rotation and translation. The reason is that the improvement of data association modeling can speed up optimization and improve the convergence of the algorithm. The initial value optimization can also greatly reduce the error before the iteration. In a more challenging experiment on Semantic KITTI00 dataset, a rotation error of 40° around the z-axis and a translation error of 8m along the x-axis and 6m along y-axis are introduced. In this case, only LCR-SMM algorithm successfully converges to a global minimum (see Fig. 7). Even SICP fails to estimate the accurate transformation matrix. 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.

TABLE I
SUCCESS RATIOS IN 441 GROUPS OF EXPERIMENTS(%)

Experiment	ICP	GICP	NDT	VGICP	SICP	LCR
KITTI 00	16.33	37.41	4.54	4.76	22.68	**95.92**
KITTI 07	19.95	36.05	2.72	5.67	37.19	**95.46**
KITTI 09	16.78	26.08	4.31	19.05	62.81	**92.06**

TABLE II
SUCCESS RATIOS IN 125 GROUPS OF EXPERIMENTS(%)

Experiment	ICP	GICP	NDT	VGICP	SICP	LCR
KITTI 00	28.80	69.60	16.00	15.20	52.80	**100.00**
KITTI 07	28.80	58.40	9.60	20.00	70.40	**100.00**
KITTI 09	18.40	48.00	15.20	42.40	68.80	**99.20**

B. Statistic Testing Results

In order 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. The translation error range from -9m to 9m in steps of 3m is introduced along the x and y axis, 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 ratio of different algorithms is summarized in Tab. I. Successful map matching is defined as when the rotation error is less than 5 degrees and the translation error is less than 2 meters. 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 is not too large, the success ratio of 125 sets is counted when initial rotation error range from -15° to 15° and translation error is between $-6m$ and $6m$. As shown in Tab. II, the success rate for all algorithms has increased given smaller initial error, 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.

TABLE III
MEAN ROTATION ERRORS IN SUCCESSFUL CASES(DEGREE)

Experiment	ICP	GICP	NDT	VGICP	SICP	LCR
KITTI 00	1.454	0.302	1.025	0.843	0.225	**0.103**
KITTI 07	1.918	1.058	1.124	0.663	0.124	**0.019**
KITTI 09	1.667	0.175	0.994	0.442	0.278	**0.012**

TABLE IV
MEAN TRANSLATION ERRORS IN SUCCESSFUL CASES(METER)

Experiment	ICP	GICP	NDT	VGICP	SICP	LCR
KITTI 00	1.293	0.195	0.652	0.328	0.120	**0.047**
KITTI 07	1.097	0.393	0.527	0.281	0.055	**0.006**
KITTI 09	1.146	0.113	0.662	0.149	0.109	**0.008**

2) Mean Error: We also present the mean error to show the overall accuracy with 441 initial errors. Since the matching error is too large when benchmark algorithms fail, we filter out failed outliers and only count the mean error when the matching is successful. Therefore, only rotation errors and translation errors in successful cases are selected to calculate mean error, which is shown in Tab. III and Tab. IV. 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.

VI. CONCLUSION

This paper has presented a semantic map matching algorithm with large convergence region. More specifically, the proposed DACE method can optimize the initial transformation and enlarge the convergence region by avoiding being trapped into the local minimum. In addition, the probabilistic registration model with initial error uncertainty is introduced to speed up the convergence. The results have demonstrated that the proposed algorithm achieves high robustness and accuracy. In summary, the proposed algorithm has raised a novel perspective of collaborative perception and localization for multiple robots, which would greatly improve the autonomous operation ability for collaborative robots. In the future, we plan to the further improve 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 3D reconstruction.

REFERENCES

[1] Y. Yue and D. Wang, *Collaborative Perception, Localization and Mapping for Autonomous Systems*. Springer, 2021.

[2] Y. Wang, M. Shan, Y. Yue, and D. Wang, "Autonomous target docking of nonholonomic mobile robots using relative pose measurements," *IEEE Transactions on Industrial Electronics*, pp. 1–11, 2020.

[3] Y. Wang, Y. Yue, M. Shan, L. He, and D. Wang, "Formation reconstruction and trajectory replanning for multi-uav patrol," *IEEE/ASME Transactions on Mechatronics*, pp. 1–11, 2021.

[4] S. Saeedi, M. Trentini, M. Seto, and H. Li, "Multiple-robot simultaneous localization and mapping: A review," *Journal of Field Robotics*, 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 Systems Journal*, vol. 11, no. 3, pp. 1723–1732, 2017.

[6] Y. Yue, C. Yang, Y. Wang, P. C. N. Senarathne, J. Zhang, M. Wen, and D. Wang, "A multilevel fusion system for multirobot 3-d mapping using heterogeneous sensors," *IEEE Systems Journal*, vol. 14, no. 1, pp. 1341–1352, 2019.

[7] Y. Yue, C. Yang, J. Zhang, M. Wen, Z. Wu, H. Zhang, and D. Wang, "Day and night collaborative dynamic mapping in unstructured environment based on multimodal sensors," in *2020 IEEE international conference on robotics and automation (ICRA)*, pp. 2981–2987, IEEE, 2020.

[8] Z. Wu, J. Zhang, Y. Yue, M. Wen, Z. Jiang, H. Zhang, and D. Wang, "Infrastructure-free global localization in repetitive environments: An overview," in *IECON 2020 - 46th Annual Conference of the IEEE Industrial Electronics Society*, pp. 626–631, IEEE, 2020.

[9] S. L. Bowman, N. Atanasov, K. Daniilidis, and G. J. Pappas, "Probabilistic data association for semantic slam," in *2017 IEEE international conference on robotics and automation (ICRA)*, pp. 1722–1729, IEEE, 2017.

[10] K. Doherty, D. Fourie, and J. Leonard, "Multimodal semantic slam with probabilistic data association," in *2019 international conference on robotics and automation (ICRA)*, pp. 2419–2425, IEEE, 2019.

[11] K. J. Doherty, D. P. Baxter, E. Schneeweiss, and J. J. Leonard, "Probabilistic data association via mixture models for robust semantic slam," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1098–1104, IEEE, 2020.

[12] Y. Yue, C. Zhao, Z. Wu, C. Yang, Y. Wang, and D. Wang, "Collaborative semantic understanding and mapping framework for autonomous systems," *IEEE/ASME Transactions on Mechatronics*, pp. 1–11, 2020.

[13] Y. Yue, C. Zhao, R. Li, C. Yang, J. Zhang, M. Wen, Y. Wang, and D. Wang, "A hierarchical framework for collaborative probabilistic semantic mapping," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 9659–9665, IEEE, 2020.

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

[15] A. Segal, D. Haehnel, and S. Thrun, "Generalized-icp.," in *Robotics: science and systems*, vol. 2, p. 435, Seattle, WA, 2009.

[16] Z. Wu, Y. Yue, M. Wen, J. Zhang, G. Peng, and D. Wang, "MSTSL: Multi-sensor based two-step localization in geometrically symmetric environments," in *2021 International Conference on Robotics and Automation (ICRA)*, IEEE, to be published, 2021.

[17] S. A. Parkison, L. Gan, M. G. Jadidi, and R. M. Eustice, "Semantic iterative closest point through expectation-maximization.," in *BMVC*, p. 280, 2018.

[18] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660, 2017.

[19] N. Atanasov, S. L. Bowman, K. Daniilidis, and G. J. Pappas, "A unifying view of geometry, semantics, and data association in slam.," in *IJCAI*, pp. 5204–5208, 2018.

[20] F. Tombari, S. Salti, and L. Di Stefano, "Performance evaluation of 3d keypoint detectors," *International Journal of Computer Vision*, vol. 102, no. 1-3, pp. 198–220, 2013.

[21] H. Yang and L. Carlone, "A polynomial-time solution for robust registration with extreme outlier rates," *arXiv preprint arXiv:1903.08588*, 2019.

[22] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-d point sets," *IEEE Transactions on pattern analysis and machine intelligence*, no. 5, pp. 698–700, 1987.

[23] W. Lu, G. Wan, Y. Zhou, X. Fu, P. Yuan, and S. Song, "Deepvcp: An end-to-end deep neural network for point cloud registration," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 12–21, 2019.

[24] J. Zhou, X. Ma, L. Liang, Y. Yang, S. Xu, Y. Liu, and S.-H. Ong, "Robust variational bayesian point set registration," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 9905–9914, 2019.

[25] Y. Zhao and P. A. Vela, "Good feature matching: Toward accurate, robust vo/vslam with low latency," *IEEE Transactions on Robotics*, vol. 36, no. 3, pp. 657–675, 2020.

[26] A. Kurobe, Y. Sekikawa, K. Ishikawa, and H. Saito, "Corsnet: 3d point cloud registration by deep neural network," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 3960–3966, 2020.

[27] M. Magnusson, *The three-dimensional normal-distributions transform: an efficient representation for registration, surface analysis, and loop detection*. PhD thesis, Örebro universitet, 2009.

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

[29] S. Agarwal, K. Mierle, *et al.*, "Ceres solver," 2012.

[30] 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.

[31] K. Koide, M. Yokozuka, S. Oishi, and A. Banno, "Voxelized gicp for fast and accurate 3d point cloud registration," tech. rep., EasyChair, 2020.