ISPRS Journal of Photogrammetry and Remote Sensing 189 (2022) 272–288

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

A LiDAR-based single-shot global localization solution using a cross-section shape context descriptor

Dong Xu [a], Jingbin Liu [a,b,*], Yifan Liang [a], Xuanfan Lv [a], Juha Hyyppä [b]

[a] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[b] Department of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute, Masala 02430, Finland

ARTICLE INFO

Keywords:
Global localization
HD map
LiDAR
Global feature descriptor
Place recognition

ABSTRACT

Fusing LiDAR and high definition (HD) maps is a feasible way to achieve global localization in GNSS-denied environments, which is necessary in driverless vehicle and robotic technologies. This paper proposes a single-shot global localization solution that uses only a single current scan of a rotating multiple-beam LiDAR sensor to locate its own location and pose with respect to an HD map in the form of georeferenced point clouds. This single-shot global localization solution estimates the state of the current moment without the previous moment state and thus avoids the nonconvergence problems that plague filter-based methods. The proposed solution allows HD maps from diverse LiDAR sensors to be used for global localization and is more robust than existing methods. The proposed solution consists of two procedures: offline preprocessing and online global localization. In the offline procedure, diverse HD maps are preprocessed to construct a global prior map for the localization process. The online global localization procedure includes two elements: place recognition, location and pose estimation. A novel Cross-Section Shape Context (CSSC) descriptor that is highly descriptive and rotation-invariant is proposed for subsequent processes. Two strategies, two-stage similarity estimation and Nearest Cluster Distance Ratio (NCDR), based on the CSSC descriptor are proposed to improve place recognition precision. A Selective Generalized Iterative Closest Point (SGICP) algorithm is proposed to calculate location and pose accurately using the CSSC descriptor. Comprehensive experiments were performed to evaluate this solution. A comparison of the precision-recall curve of multiple scenes, particularly under changed viewpoint scenes, shows that the CSSC descriptor is more robust than existing descriptors. Experimental analysis also confirms that the proposed strategies, two-stage similarity estimation and NCDR, improve place recognition precision. Also, compared to the generalized iterative closest point algorithm, the SGICP algorithm achieved better accuracy by 31% and efficiency by 60%. The proposed solution achieves a mean relative translation error (RTE) improvement of 27% over the OneShot algorithm on the KITTI dataset. The proposed solution had an average 77% improvement over 1σ RTE relative to the benchmark in tests with the long-term localization NCLT dataset. The mean RTE of the proposed solution was 0.13 m using HD maps from different LiDAR sensors. Our code is available at: https://github.com/Dongxu05/CSSC.

1. Introduction

For driverless vehicles and robotics, global localization that fuses LiDAR and high definition (HD) maps is a critical technology that provides accurate global location and pose, particularly when the global navigation satellite system (GNSS) does not function adequately. Low accuracy and limited generalizability, however, restrict the applicability of existing global localization methods based on rotating multiple-beam LiDAR sensor data and HD maps rendered as georeferenced point clouds. In this paper, we propose a global localization solution that uses a single LiDAR scan and a novel global feature descriptor to tackle these problems.

For global localization based on LiDAR and HD maps, Monte Carlo localization (MCL) techniques are used in most localization systems. However, these MCL-based solutions (Yin et al., 2020; Sun et al., 2020; Chen et al., 2020) need multiple dynamic scans for the filter to converge, resulting in wake-up and kidnapped robot problems. The wake-up robot problem refers to instances when an autonomous mobile system boots up and does not provide instantly accurate global locations. The kidnapped robot problem occurs when an autonomous mobile system

---

* Corresponding author.
E-mail address: jingbin.liu@whu.edu.cn (J. Liu).

https://doi.org/10.1016/j.isprsjprs.2022.05.005
Received 16 January 2022; Received in revised form 8 May 2022; Accepted 13 May 2022
Available online 27 May 2022

cannot recover from a wrong previous location (Seow et al., 2018; Prestes et al., 2009). We argue that single-shot global localization solutions can resolve these challenges.

The single-shot global localization solution uses only a single LiDAR scan at a time. Unlike MCL-based methods, single-shot global localization estimates the location and pose of the current moment without the previous moment location and pose, and thus does not experience nonconvergence problems such as the wake-up robot or kidnapped robot problem. This solution includes the following two steps: 1) place recognition (Steder et al., 2011; Muhammad and Lacroix, 2011) provides an initial location from a LiDAR query scan; and 2) fine registration (Biber et al., 2003; Segal et al., 2010) estimates the six degrees of freedom (6DOF) parameters relative to the HD map based on place recognition. These steps make the single-shot global localization solution more robust than other MCL-based methods (Ratz et al., 2020).

Although the accuracy of single-shot global localization solutions is higher than that of MCL-based methods, accuracy and narrow applicability problems are still factors limiting their widespread use. Place recognition when obtained from shifting viewpoints and a low-resolution LiDAR scan is uncertain due to the low descriptiveness of global feature descriptors and the simple place-matching strategy in existing solutions. Original iterative closest point (ICP) algorithms used in the existing solution (Yin et al., 2020; Guo et al., 2019) provide terrible 6DOF parameters when a preset initial guess is used. Uncertain place recognition and terrible 6DOF parameters both yield accuracy losses to global localization. Existing solutions (Kim and Kim, 2018; Wang et al., 2020; He et al., 2016) request the HD map to be from the same type of sensor as the localization sensor to provide place recognition. Therefore, the HD maps from different LiDAR sensors, such as terrestrial laser scanners (TLSs) and 2D laser range finders, cannot be used for existing single-shot global localization solutions directly, limiting solution generalizability.

To resolve the issues of low accuracy and limited generalizability, this paper proposes a single-shot LiDAR scan global localization solution using a novel Cross-Section Shape Context (CSSC) descriptor. The contributions of this study beyond existing methods are as follows:

1. We propose an accurate global localization solution that uses only one single LiDAR scan, which consists of a high-quality CSSC descriptor and novel methods that are used in the pipeline of the global localization. The proposed offline HD map preprocessing method can process HD map data generated from diverse LiDAR sensors, effectively using all information. An integrity index is also used to monitor the integrity of each single-shot global localization result.
2. The proposed CSSC descriptor features an enhanced high descriptiveness using the elevation weight and point density weight, which describe the spatial distribution characteristics of the point cloud from two dimensions.
3. Based on the CSSC descriptor, a set of improved methods, including two-stage similarity estimation, Nearest Cluster Distance Ratio (NCDR), and the Selective Generalized Iterative Closest Point (SGICP) algorithm, are developed to improve the precision of place recognition and 6DOF estimation.

The remainder of this paper is organized as follows. Section 2 describes related works; Section 3 introduces the proposed single-shot global localization solution; and Section 4 presents experimental results. Section 5 summarizes and concludes this study.

2. Related works

In this section, we review existing studies on place recognition and 6DOF estimation that are related to global localization by fusing LiDAR scans with HD maps.

2.1. Related works on 3D LiDAR-based place recognition

Because global feature descriptors improve the robustness of place recognition by extracting the features of the entire 3D LiDAR scan, they have attracted the attention of researchers. Global feature descriptors can be divided into handcrafted and deep learning-based descriptors.

Regarding handcrafted feature descriptors, Muhammad et al. developed Z-projection (Muhammad and Lacroix, 2011), which saves the angles between the normal corresponding to each point and the Z axis to construct the feature descriptor. He et al. proposed Multiview 2D Projection (M2DP) (He et al., 2016) that projects the 3D point cloud to multiple 2D planes and obtains the feature descriptor through Singular Value Decomposition (SVD) of the distribution matrix. Scan Context (SC) (Kim and Kim, 2018), Seed (Fan et al., 2020) and LiDAR iris (iris) (Wang et al., 2020) are egocentric feature descriptors. Arc-shaped grids that are suitable for rotating LiDAR are constructed, and feature descriptors are generated by extracting the feature from the grid. Behley et al. (Behley and Stachniss, 2018) performed place recognition by range images and normal images that were extracted from point clouds. The DEscriptor of LiDAR Intensities as a Group of HisTograms (DELIGHT) (Cop et al., 2018), Intensity Signature of Histograms of OrienTations (ISHOT) (Guo et al., 2019) and Intensity Scan Context (ISC) (Wang et al., 2020) use echo intensity information to construct feature descriptor. In the same scene with different viewpoints, histogram- (Muhammad and Lacroix, 2011) and projected-based (He et al., 2016) descriptors cannot robustly perform place recognition because both descriptors do not have rotation invariance. Egocentric feature descriptors (Kim and Kim, 2018; Fan et al., 2020; Wang et al., 2019) overcame the problem of different viewpoints; however, these descriptors yield poor descriptiveness, making their place recognition performance insufficiently robust in challenging environments. LiDAR intensity information makes up for the lack of texture information. However, there is no general LiDAR intensity calibration standard, causing the place recognition solution using descriptors with LiDAR intensity to have low generalizability between different LiDAR sensors (Guo et al., 2019). Also, most of these methods use a minimum distance searching strategy for the corresponding place, leading to poor place recognition.

At present, place recognition based on deep learning is one of the research hotspots. PointNETVLAD (Angelina et al., 2018) and SeqLPD (Liu et al., 2019) extract a global descriptor for the place recognition task with an end-to-end way. These work input directly the points into the network and output the result of place recognition. Unlike the aforementioned works, other papers input the preprocessing result of the original point cloud into a network. SpoxelNet (Chang et al., 2020) encoded input spherical voxels into global descriptor vectors by extracting the structural features in both fine and coarse scales. OREOS (Schaupp et al., 2019) exploited convolutional neural network to extract compact descriptors from single 2D range image. SegMatch (Dube et al., 2017), Oneshot (Ratz et al., 2020), Locus (Vidanapathirana et al., 2020) and semantic graph (Kong et al., 2020) realized place recognition through using semantic information. The deep learning-based methods perform well in the trained environments, while they usually cannot generalize well in different environments or different LiDAR sensors.

2.2. Related works on six degrees of freedom estimation

Registration algorithms provide 6DOF parameters for LiDAR scans. Registration algorithms based on local shape descriptors (LSD) (Guo et al., 2015; Tao et al., 2020; Zhao et al., 2019; Dong et al., 2017) have been intensively studied in recent decades. LSD-based methods have three significant steps: detecting the key points, extracting the LSDs and identifying correspondences. These methods achieve good performance with registration of model points. However, these methods are not suitable for quick global localization using rotating multiple-beam LiDAR sensor for two reasons: 1) it remains a challenge to achieve 3D key point detection with a high repetition rate in sparse points (Boroson and Ayanian, 2019); and 2) it is time-consuming to calculate local feature descriptors of all key points (Yin et al., 2021).

To improve efficiency, some researchers only extracted the key points for registration in the algorithms. LOAM (Zhang and Singh, 2017) extracts edge points and plane points for 6DOF estimation. Based on the LOAM, Lego-LOAM (Shan and Englot, 2018) removes points of small objects, such as leaves moving in the wind, to enhance robustness. In addition, Lego-LOAM further optimizes ground points for high accuracy. The common characteristic of these methods is that key points are extracted from the original point cloud data before matching. The correct transformation relation between key points is obtained through continuous iterative calculation.

The key point-based methods have made marked improvements in efficiency. However, these methods lack robustness in environments where features are missing. Registration algorithms that directly operate on the original data rarely suffer from this problem. The iterative closest point (ICP) is a classical registration algorithm that directly operates on the original data. Generalized iterative closest point (GICP) (Segal et al., 2010) extends the original ICP algorithm using a distribution-to-distribution matching method and obtains higher accuracies than ICP. However, ICP and other ICP variants are sensitive to initial guesses; a poor preset initial guess leads to poor precision and a long time of convergence. Also, ICP and related variants highly depend on a nearest neighbour search to associate the closest points. The closest point, however, is not carefully checked, leading to poor registration performance in some cases.

3. Proposed global localization solution

In this section, the proposed global localization solution that includes the offline preprocessing of HD maps and online single-shot global localization is introduced, as shown in Fig. 1. The preprocessing component describes how to generate a global prior map using HD maps from different LiDAR sensors. The online global localization presents improvements in feature extraction, place recognition and 6DOF estimation. Also, the integrity of global localization and parameter analysis are introduced.

As shown in Fig. 1, an HD map is an input, and a global prior map is an output in the HD map preprocessing. The global prior map will be generated by fusing the scans and features after preprocessing. Based on the global prior map, a query scan is used as an input in online global localization. After feature extraction, place recognition and 6DOF estimation, the global location of the query scan and its integrity will be obtained.

3.1. Preprocessing of HD map

In this section, we describe how to generate a global prior map using HD maps from different LiDAR sensors by scan and feature extraction.

For extracting scans from HD maps, type judgements on the HD map must be made. In this paper, the HD map refers to the point cloud that has a global geo-referenced frame. Such HD maps can be obtained in several ways for different applications (e.g., vehicles that are equipped with high-precision integrated localization systems and rotating multi-beam LiDAR (Geiger et al., 2012); TLSs (Liu et al., 2017) and mobile laser systems (MLSs) (Liu et al., 2021). According to the type of sensor used in the HD map building process, HD maps can be divided into homogeneous and heterogeneous maps. During localization, rotating multi-beam LiDAR is typically used as a localization sensor (Prestes et al., 2009; Steder et al., 2011) due to its 360° horizontal field of view (FOV) and large vertical FOV. Therefore, the homogeneous HD map refers to the map derived from the same type of sensor as the rotating multi-beam LiDAR. The heterogeneous HD map refers to the map derived from other LiDAR systems, such as TLS and MLS.

After using type judgement, the HD map will be transformed into a series of scans for feature extraction. Because a homogeneous HD map has the same scans as scans from the localization sensor, it can be used directly for subsequent processes. The heterogeneous HD map must be converted into virtual scans that are similar to real scans from a localization sensor through virtual LiDAR. For more details about virtual LiDAR, refer to (Xu et al., 2022).

After obtaining scans from the homogeneous HD map or virtual scans from the heterogeneous HD map, these scans are used to extract features that include fingerprints and CSSC descriptors; additional details can be found in Section 3.2. A fingerprint is a feature vector and is used for forming a kd-tree, which can be used to search for place candidates quickly. The CSSC descriptor is used to estimate the similarity between candidates and query scans to obtain corresponding places.

The global prior map used for online global localization is generated by fusing the scans, fingerprints and CSSC descriptors in the HD map preprocessing component. The preprocessing time varies depending on the type and size of the HD map. In the experimental section, the preprocessing time will be discussed for the specific HD map.

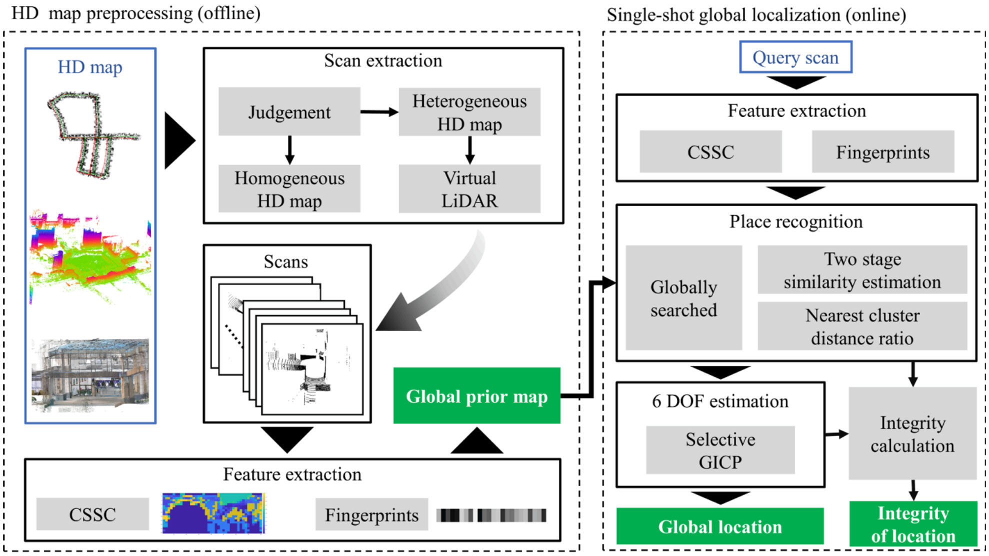


Fig. 1. Pipeline of the proposed global localization solution. In offline preprocessing, an HD map will be transformed into a global prior map for subsequent global localization. In online global localization, the global location of the query scan and its integrity will be obtained.

3.2. Feature extraction

In this section, we present methods to extract features, including CSSC descriptors and fingerprints, which are important in offline and online procedures. The CSSC descriptor describes the place in a matrix way and is used for obtaining place recognition accurately. The fingerprint can be used to search for place candidates quickly.

3.2.1. Generation of the CSSC descriptor

The proposed CSSC descriptor is inspired by the 3D Shape Context (3DSC) descriptor (Frome et al., 2004) and the measurement principle of rotating multi-beam LiDAR. The 3DSC descriptor is an LSD that was proposed for object recognition. As with other LSDs, the 3DSC descriptor selects a spherical region as the support region. The generation of 3DSC is as follows. The entire support region is divided into bins by equally spaced boundaries in the azimuth and elevation dimensions along the radial dimension. According to the number of points in each bin and the number of points in the neighbour area, a weight value is assigned to each point in the bin. The 3DSC descriptor is generated by encoding all the weighted count numbers of each bin.

Unlike the 3DSC descriptor, the CSSC descriptor is a global feature descriptor. The CSSC descriptor sets the LiDAR origin as the centre and uses the FOV of a rotating multiple-beam LiDAR system as a support region. Given the scanning principle of rotating multiple-beam LiDAR, the scan is divided into bins by horizontal and vertical angles of LiDAR along the horizontal distance, as shown in Fig. 2. The CSSC descriptor is generated by describing the shape context of all bins at the same horizontal angle and distance bins. The shape context of all bins at the same cross-section is described to generate the proposed descriptor. Therefore, the proposed descriptor is named the cross-section shape context descriptor.

After executing the division shown in Fig. 2, $N_d \times N_h \times N_v$ bins can be obtained, where N_d is the number of bins along the horizontal distance;N_h is the number of bins in the horizontal FOV; and N_v is the number of bins in the vertical FOV. In this paper, we set $Nd = 20$, $Nh = 40$, and $Nv = 8$; parameter details can be found in Section 3.6.1. The division used in this paper makes the far bin larger, compensating for the insufficient information caused by the sparsity of far points, which is caused by the principle of rotating multiple-beam LiDAR. The resolution of farther points is smaller than that of nearer points. After division, each element in the CSSC descriptor is obtained by encoding spatial distribution of the point cloud in the cross-section, as shown in Fig. 3.

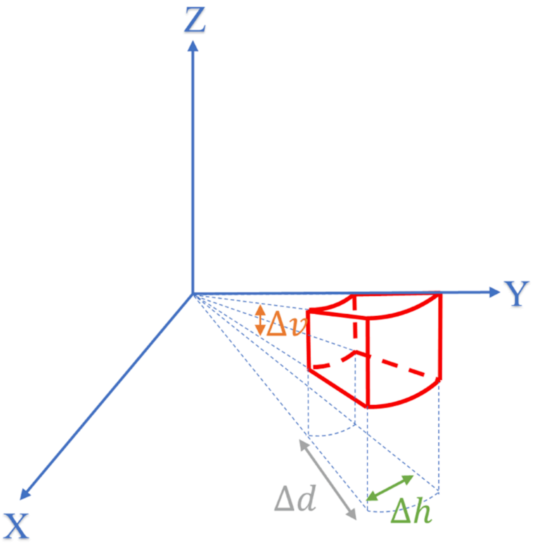


Fig. 2. Interpretation of division with the proposed method. Δd is the gap of the horizontal distance; Δh is the gap of the horizontal angle; Δv is the gap of the vertical angle.

As shown in Fig. 3, elevation and point density weights of $B_{ijk}(i \in [N_d], j \in [N_h], k \in [N_v])$ are calculated; the details of these calculations are presented in the following paragraphs. Each element in the CSSC descriptor is calculated by accumulating the products of two weights from bins at the same cross-section. The CSSC descriptor is generated after the values of all elements are obtained.

The elevation weight is related to the vertical angle of each beam, that hits the object. In the same horizontal angle and distance bins, the higher bin that has a larger vertical angle is assigned to a higher elevation weight. The power function is selected as a weighting strategy. The elevation weight of each bin can be written as:

$$E_{ijk} = \frac{2^{f\left(B_{ijk}\right)}}{255}$$

$$f\left(B_{ijk}\right) = \begin{cases} k-1\left(B_{ijk} \neq \varnothing\right) \\ 0\left(B_{ijk} = \varnothing\right) \end{cases} \tag{1}$$

where E_{ijk} is the evaluation weight of bin B_{ijk}. Because we set $Nv = 8$, 255 is calculated by $\sum_{k=1}^{8} 2^{k-1}$ and it is used to normal E_{ijk} in the calculation.

The point density weight is related to the point density of each bin. The high point density weight is assigned to the bin that includes many points.

$$D_{ijk} = \begin{cases} 1(median = 0) \\ 1\ (number_{ijk} > 2{*}median) \\ \dfrac{number_{ijk}}{2{*}median}(else) \end{cases} \tag{2}$$

where D_{ijk} is the density weight of B_{ijk}; $number_{ijk}$ is the number of points in the B_{ijk}; and *median* is the median of point numbers at the bins in the same vertical angle and horizontal distance.

Based on the elevation weight and point density weight of B_{ijk}, the value of the element M_{ij} $(i \in [N_d], j \in [N_h])$ is calculated by accumulating the products of two weights from bins in the same cross-section. After all elements have been calculated, we obtain a CSSC descriptor M:

$$M_{ij} = \sum_{k=1}^{8} E_{ijk} {*} D_{ijk}$$

$$M = \bigcup_{i=1,j=1}^{N_d,N_h,} M_{ij} \tag{3}$$

The reason for selecting the power function as the elevation weighting function is that the higher points of immobile buildings play an important role in place recognition under outdoor conditions. The redundant points of ground make little difference to the result of place recognition and the redundant points of moving vehicles have an adverse effect on the result. In the point density weight calculation procedure, the median is selected to represent the distribution of point numbers. Compared to the mean, the median can describe the central tendency accurately when outliers exist.

3.2.2. Generation of fingerprint

To find place candidates quickly, fingerprints are extracted after division of the support region within the generation of CSSC descriptors. A demonstration is shown in Fig. 4.

As shown in Fig. 4, the numbers of occupied bins (Cnt) in each ring at the same layer, where the blue bin refers to the bin at the same layer, are counted first. Then, the average (Ave), corresponding index of max Cnt (MI) and standard deviation (SD) of Cnts are calculated as sub-features. A fingerprint is generated by combining sub-features of all layers. To find appropriate candidates accurately and quickly, three single and four combined features are evaluated, as shown in Fig. 5. The execution details are as follows. Hundreds of query scans are transformed to

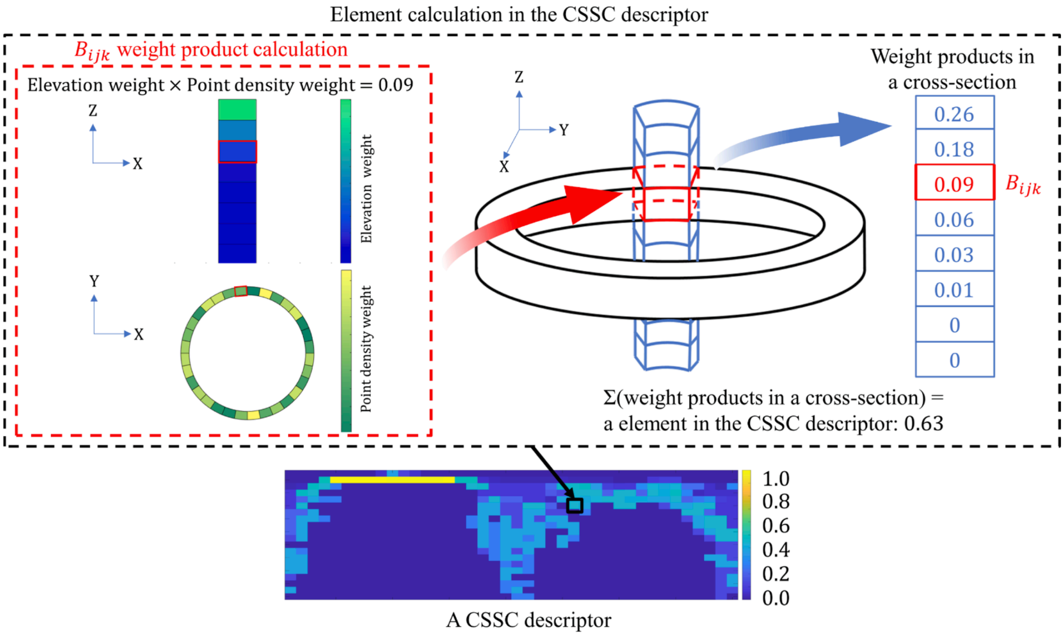


Fig. 3. Element calculation in the CSSC descriptor. Each element in the CSSC descriptor is calculated by accumulating the weight products of elevation and point density weights from bins at the same cross-section. The elevation and point density weight of the red bin (B_{ijk}) are calculated using blue bins and black bins respectively. The blue bins represent the bins at the same horizontal angle and distance as red bin. The black bins represent the bins at the same elevation and horizontal distance as red bin.

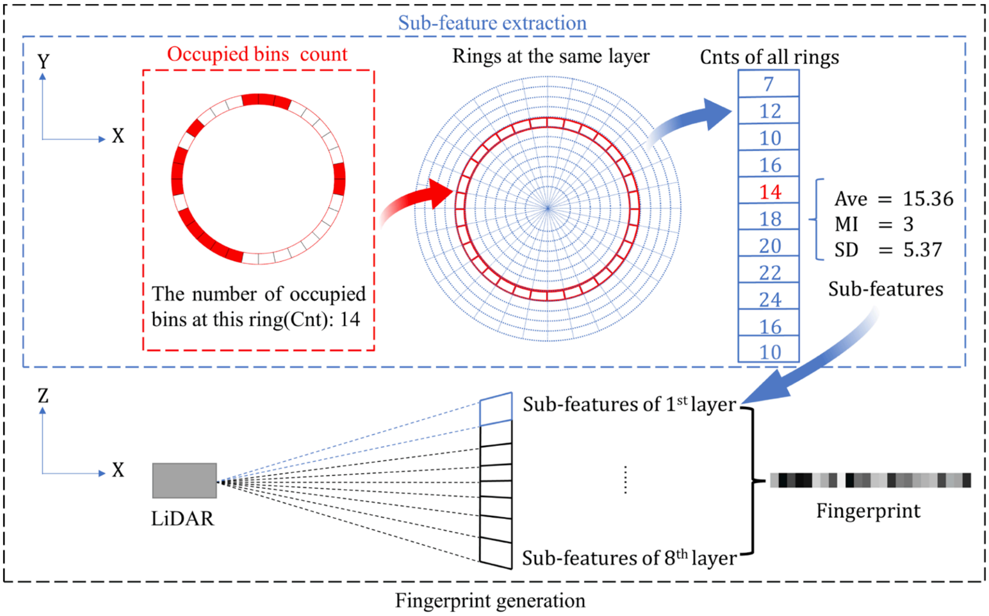


Fig. 4. Demonstration of fingerprints generation. The sub-feature of a layer is extracted by analyzing the numbers of occupied bins of each ring in this layer. Here, the rings at the same layer refer to the rings at the same vertical angle. A fingerprint is generated by combining sub-features of all layers.

different fingerprints and used for candidate searching. The recall curve can be obtained for a different number of candidates.

Fig. 5 shows that the recall curve of Ave + SD covers the largest area, which indicates that Ave + SD is appropriate for candidate searching. Therefore, Ave + SD is selected as fingerprints in this study.

3.3. Place recognition

In this section, we introduce improved place recognition, including two-stage similarity estimation and NCDR strategies, which provides an initial location of the query scan from LiDAR. After extracting the fingerprint from a query scan, place candidates can be obtained in a global and efficient way through kd-tree. Two-stage similarity estimation is conducted to calculate similarities between CSSC descriptors of the query scan and candidates. The NCDR strategy was developed to deny inaccurately matched candidates. Corresponding places can be found through the improved place recognition.

3.3.1. Two-stage similarity estimation

The two-stage method is used to improve the similarity estimation performance. The proposed method is shown in Fig. 6. The inputs are the query scan and candidate scan. The two stages include coarse angle

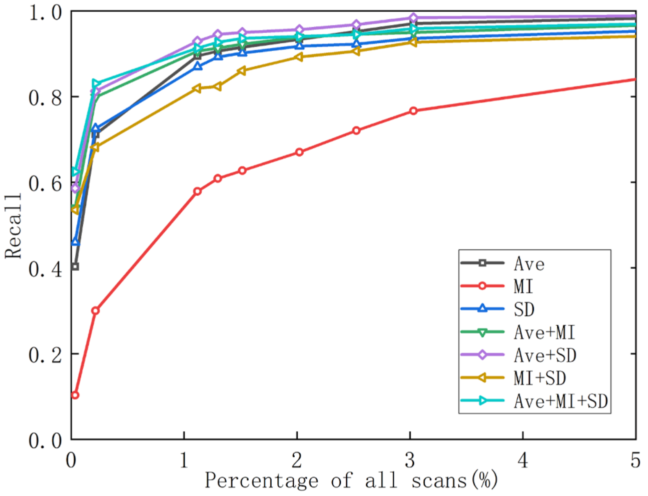


Fig. 5. Recall curve of different features with different percentage of all scans; The percentage of all scans refers to the number of candidates as a percentage of the all scans.

transformation and semi-semantic ICP alignment.

As shown in Fig. 6, the coarse angle between two viewpoints is provided in a rapid way in the first stage. The rapid method used in this study can be interpreted as follows. The coarse angle is obtained by multiplying the horizontal angle gap by the index difference of the corresponding columns. The horizontal angle gap is calculated according to $360/N_h$. We consider each column within the CSSC descriptor as a discrete distribution. The index difference of the corresponding columns is determined using Jensen–Shannon (JS) divergence, which measures the similarity of two discrete distributions. Before the second stage, the query scan is transformed with a coarse angle.

As shown in Fig. 6, the semi-semantic ICP is used to provide an initial guess for 6DOF estimation and improve the robustness of place recognition in the second stage. The semi-semantic ICP can be interpreted as follows. The points that have high elevation and point density weights typically belong to immobile objects, such as buildings and trees, and thus, these points are retained to calculate the initial guess for 6DOF estimation. Because the classes of retained points are not clear sufficient, the proposed method is equivalent to semi-semantic ICP. The specific calculation steps are as follows. While generating CSSC descriptors, the two weights of each point can be determined. The points that have high elevation and point density weights are retained to be projected onto the x-y plane, and the ICP is used to calculate 3DOF (Δx, Δy, yaw angle), which can be used as an initial guess for 6DOF estimation. Also, we minimize the influence of rotation and translation between two scans in the same place on the similarity of the descriptors by transforming the query scan according to the 3DOF estimation, thus improving the robustness of place recognition.

After semi-semantic ICP, M^q and M^c are generated. M^q refers to the CSSC descriptor of the transformed query scan, and M^c refers to that of the candidate. The similarity is estimated as follows:

$$dis = \frac{1}{N_h}\sum\nolimits_{j=1}^{N_h}\left(1-\frac{c_j^q \bullet c_j^c}{\|c_j^q\|\|c_j^c\|}\right) \tag{4}$$

where *dis* is the similarity between two descriptors, c_j^q is the *jth* clowns of M^q, and c_j^c is the *jth* clowns of M^c.

The range of *dis* is [0,1], and a smaller distance indicates that the two CSSC descriptors are more similar. The two-stage similarity estimation provides an initial guess for accurate 6DOF estimation. Also, it is beneficial to improve the robustness of place recognition.

3.3.2. Nearest cluster distance ratio

After obtaining candidates, the conventional method seeks the corresponding place by calculating similarity and checking whether the maximum similarity is larger than the threshold value, thus bringing low robustness. In this section, NCDR is proposed to improve place recognition precision by denying wrong matches. The steps are as follows.

Based on the Euclidean distance, candidates are divided into several clusters. The *dis* between the CSSC descriptors of all candidates and the query scan are calculated. In a cluster, if the different value between maximum and minimum *dis* is larger than the threshold value, the candidates in this cluster are inaccurate matches and will be deleted. These steps are repeated until all clusters are checked. We obtain the minimum *dis* of each retained cluster, and *ratio* is calculated as follows:

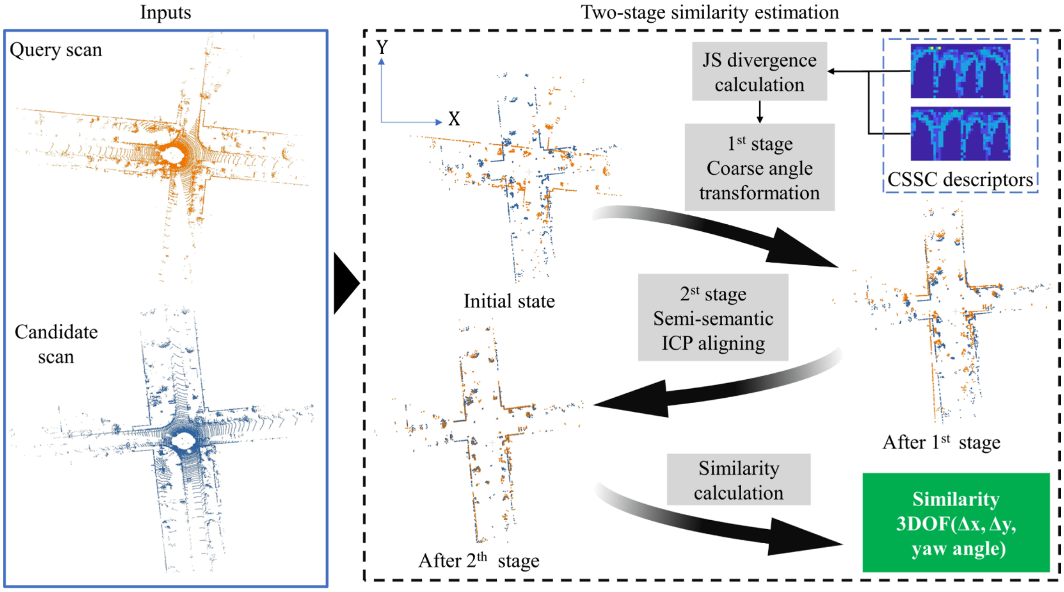


Fig. 6. Demonstration of two-stage similarity estimation. The query scan and candidate scan are inputs. The first stage provides a coarse angle between two viewpoints of scan by determining the index difference of corresponding columns in CSSC descriptors. The second stage provides Δx, Δy and yaw angle using semi-semantic ICP aligning. Then we estimate the similarity between the transformed query scan and candidate scan.

$$ratio = \frac{dis_{1st}}{dis_{2st}} \tag{5}$$

where dis_{1st} is the first smallest *dis*, and dis_{2st} is the second smallest *dis*. The range of *ratio* is [0,1], and a smaller *ratio* indicates that place recognition is more robust. In Fig. 7, positive and negative examples are shown.

As shown in Fig. 7, ten candidates and their corresponding *dis* are obtained in both examples through global searching and two-stage similarity estimation. In the positive example, dis_{1st}, dis_{2st} and *ratio* are 0.124, 0.321 and 0.386, respectively. In the negative example, the minimum distance is 0.183, and the maximum distance is 0.418 in the cluster. The different value in this cluster is 0.235. Therefore, this cluster is deleted. In the retained clusters, dis_{1st}, dis_{2st} and *ratio* are 0.342, 0.372 and 0.919, respectively. The ratio of the positive example is far lower than the ratio of the negative example.

3.4. 6DOF estimation

In this section, we describe how to estimate the 6DOF of a query scan based on the place recognition. We briefly introduce the GICP algorithm and then propose a selective GICP that derives from the closest point checking method. The initial guess is provided by place recognition for later iteration optimization.

We consider the estimation of the transformation $\mathbf{T}$, which aligns a set of points $\mathrm{Q} = \{q_0, \cdots, q_N\}$ (the point cloud of a query scan) with respect to another set of points $\mathrm{C} = \{c_0, \cdots, c_M\}$ (the point cloud of the corresponding place). The correspondence between Q and C is given by the nearest neighbour search: $c_i = \mathbf{T}q_i$.

The GICP algorithm is an accurate registration algorithm that adopts a distribution-to-distribution method. This algorithm describes points in a normal distribution model, $q_i \sim N\left(\widehat{q}_i, C_i^Q\right)$ and $c_i \sim N\left(\widehat{c}_i, C_i^C\right)$. C_i^Q and C_i^C are covariance matrices associated with the measured points q_i and c_i. The transformation error can be written as:

$$d_i = c_i - \mathbf{T}q_i \tag{6}$$

d_i is a linear combination of q_i and c_i. Therefore, it can be drawn from a normal distribution:

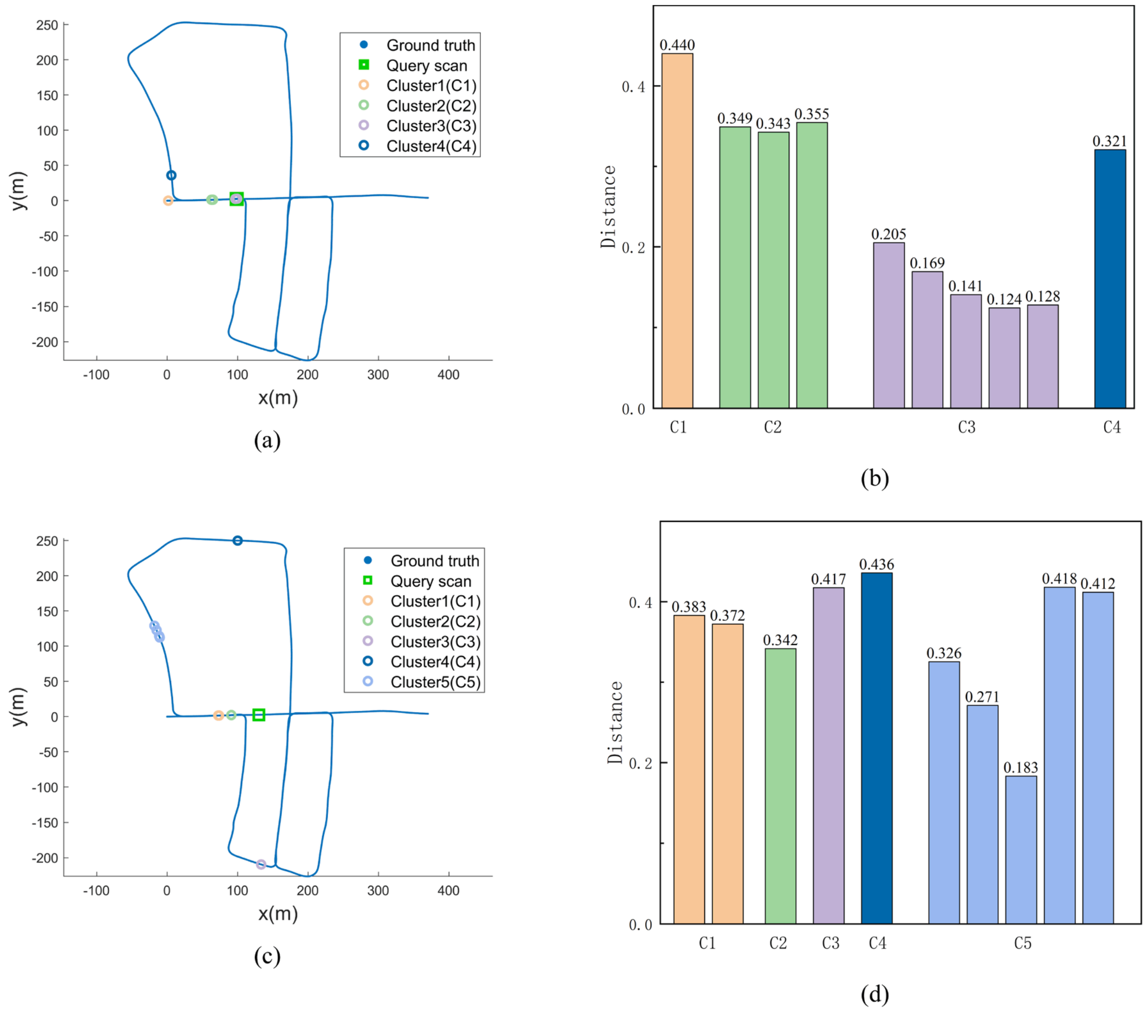
$$d_i \sim N\left(\widehat{c}_i - \mathbf{T}\widehat{q}_i,\ C_i^C + \mathbf{T}C_i^Q\mathbf{T}^T\right)$$

$$= N\left(0,\ C_i^C + \mathbf{T}C_i^Q\mathbf{T}^T\right) \tag{7}$$

The optimal transformation matrix $\mathbf{T}$ can be determined using maximum likelihood estimation (MLE):

$$\mathbf{T} = \underset{\mathbf{T}}{\mathrm{argmax}} \prod_i p(d_i) = \underset{\mathbf{T}}{\mathrm{argmax}} \sum_i log(p(d_i)) \tag{8}$$

which can be simplified to



(a)

(b)

(c)

(d)

Fig. 7. Examples for interpreting NCDR. (a) shows a positive example and (b) shows the *dis* between query scan and all candidates. (c) shows a negative example, (d) shows the *dis* between query scan and all candidates.

$$\mathbf{T} = \underset{\mathbf{T}}{\operatorname{argmin}} \sum_i d_i^T \left(C_i^C + \mathbf{T} C_i^Q \mathbf{T}^T \right)^{-1} d_i \tag{9}$$

The covariance matrix of each point is typically estimated by its k neighbours. Following the suggestion in (Segal et al., 2010), each covariance matrix is regularized by replacing its eigenvalue with (1, 1, ϵ). This regularization makes the GICP work as a plane-to-plane ICP. There is a case where the two points are false matches, but the distributions of adjacent points are the same. The closest point is not carefully checked, leading to terrible registration performance in this case.

We improved GICP by adding the closest point checking process. In the generation of the CSSC descriptor, each point in the scan is assigned elevation and point density weights. According to the weights, the closest point will be checked. The different weights result in the nearest neighbour's search again. Fig. 8 shows the correspondence models used in GICP and SGICP.

As shown in Fig. 8, the neighbour point distributions of the three points are the same, and the covariance matrices of these points are the same. After a transformation, GICP finds an inaccurate point via closest point searching. An inaccurate transformation error is generated by Equation (6). The incorrect transformation error and the same covariance matrices are used to optimize the transformation matrix by Equation (9), leading to poor results. Compared to GICP, SGICP executes the closest point checking process first based on the weights. The checking process is equal to a selective process of the closest point. After the selective process, SGICP identifies the corresponding closest point to optimize the transformation matrix. The transformation matrix will obtain better optimization than optimization in GICP. In addition, SGICP has better efficiency than GICP because SGICP uses a selective process of the closest point to avoid unnecessary calculation of covariance matrices. The SGICP is listed as Algorithm. 1.

Algorithm. 1. SGICP

```
Input: Two point clouds: $Q = \{q_0, \cdots, q_N\}$, $C = \{c_0, \cdots, c_M\}$
    Two covariances: $C^Q = \left\{C_0^Q, \cdots, C_N^Q\right\}$, $C^C = \{C_0^C, \cdots, C_M^C\}$
    Initial guess: $\mathbf{T}$
Output: The optimal transformation $\widehat{\mathbf{T}}$
while $\mathbf{T}$ $not\ converged$ do
  $e = [], J = []$
  for $i \in \{0, \cdots N\}$ do
    while $WeightJudge(q_i, c_i) = false$ do
      $c_i = FindCloestPointInC(\mathbf{T} q_i)$
    end
  $e_i, J_i \leftarrow Cost(d_i^T \left(C_i^C + \mathbf{T} C_i^Q \mathbf{T}^T\right)^{-1} d_i)$
  $e \leftarrow e \cup e_i,\ J \leftarrow J \cup J_i$
  end
    $\delta T \leftarrow -\left(J^T J\right)^{-1} J^T e\ \nabla\ Gause\ Newton\ update$
    $\mathbf{T} \leftarrow \mathbf{T} \boxplus \delta \mathbf{T}$
end
return $\mathbf{T}$
$\widehat{\mathbf{T}} = \mathbf{T}$
```

After obtaining the optimal transformation, the *score* of SGICP can be written as:

$$score = \frac{1}{n} \sum_{i=1}^{n} \left(c_i - \widehat{\mathbf{T}} q_i\right) \tag{10}$$

which describes the transformation accuracy to some extent but does not linearly describe the transformation error. Based on the transformation and known location and pose of the candidate, the global location and pose of the query scan are obtained.

3.5. *Integrity of global localization*

In this section, an index that reflects the integrity of global localization is presented, avoiding inaccurate global locations under some unforeseen conditions.

The proposed global localization solution includes place recognition and 6DOF estimation. The similarity between the CSSC descriptors of the two scans (*dis*) and the NCDR (*ratio*) both describe the performance of place recognition. The *score* of SGICP describes the performance of 6DOF estimation to some extent. Therefore, *dis*, *ratio* and *score* are used to compute the integrity of global localization. The integrity index can be written as:

$$\begin{aligned} WCS &= W_1(1 - dis)(1 - ratio) + W_2(1 - score) \\ Li &= \begin{cases} 1(WCS \geq Thr) \\ 0(WCS < Thr) \end{cases} \end{aligned} \tag{11}$$

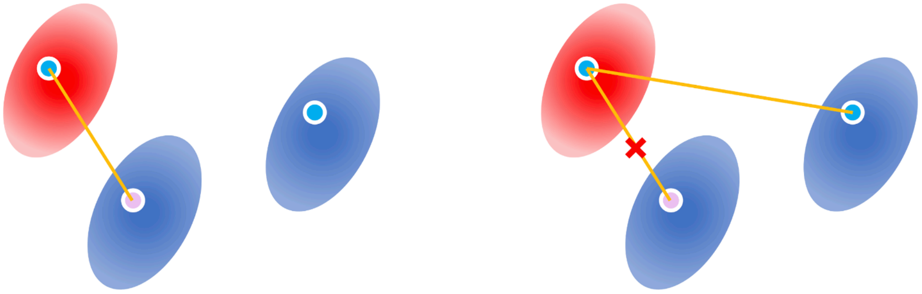
where *Li* is the integrity of global localization; *WCS* is the weighted combined score; *Thr* is the threshold value of *WCS*; W_1 is the weight of place recognition; and W_2 is the weight of 6DOF estimation. *dis*, *ratio* and *score* are indices of place recognition and 6DOF estimation.

Equal weights are used to calculate the integrity. Because two indices are obtained after place recognition and one index is obtained after 6DOF estimation, W_1 and W_2 are set equal to 0.67 and 0.33, respectively.

After obtaining a global location, *dis*, *ratio* and *score* are also obtained. According to Equation (11), *Li* can be calculated to describe the integrity of the global localization. When *Li* is 1, the global location is reliable. In contrast, when *Li* is 0, the global location is not reliable. *Thr* is discussed in Section 3.6.2.

3.6. *Parameters analyses*

In this section, the parameter analyses of the CSSC descriptor and global localization integrity are presented.



(a) (b)

Fig. 8. Demonstration of GICP and SGICP. The red circles indicate a source point and the blue circles indicate target points. (a) GICP finds a wrong point via closest point searching; (b) SGICP identifies the corresponding closest point after the weight checking.

3.6.1. Parameter analysis of CSSC descriptors

N_d, N_h and N_v are important parameters of the CSSC descriptor and have strong impacts on the descriptiveness of descriptors. N_d, N_h and N_v affect the size of bins within the CSSC descriptor. Larger N_d, N_h and N_v make the bin smaller, indicating that the CSSC descriptor is more descriptive but also more sensitive to noise. In addition, it takes more time to calculate the similarity between two CSSC descriptors. In contrast, smaller N_d, N_h and N_v with larger bins are more robust to noise and spend less time estimating similarity. In this study, noise refers to the points of small objects, such as leaves moving in the wind and the grass, which have a misleading effect on place recognition.

In the most egocentric feature descriptor (Kim and Kim, 2018; Fan et al., 2020); N_d is typically set equal to 20. The length of each bin is set at 4 m. In previous work (Xu et al., 2022), a setting for the length of the bin was tested. The larger bins (a length of 8 m) hide many details, affecting the descriptiveness of the descriptor. The descriptor has comparable performance when the length of each bin is set at 4 m, 2 m and 1 m. However, when the length of each bin is set at 4 m, execution time during similarity estimation decreases. Therefore, the 4 m-length bin is used in this study.

N_h is unfixed in these global descriptors. SC (Kim and Kim, 2018) set N_h to 60 for noise resistance and efficiency. However, iris (Wang et al., 2020) sets N_h to 360 and conducts LoG-Gabor filtering for high descriptiveness. To determine the appropriate N_h, parameter experiments are performed. We examine the performance of the CSSC descriptor in the partial KITTI sequence 00 for different N_h; details of this dataset are described in Section 4.1.1. The precision-recall (PR) curve is used to assess the performance; evaluation criteria are described in Section 4.1.2. Fig. 9 shows the PR curves for various N_h.

Fig. 9 shows that the best performance is obtained when N_h equals 40. Too small bins are sensitive to noise, leading to poor place recognition. The CSSC descriptor uses elevation and point density weights to describe the spatial distribution characteristics of the point cloud from two dimensions, enhancing the descriptiveness of the descriptor. For the proposed CSSC descriptor, the noise resistance with a larger bin is preferable, instead of high descriptiveness with smaller bins. Therefore, we set $N_d = 20$ and $N_h = 40$ in this study. If place recognition is required in a new dataset with higher resolution, the CSSC descriptor will have more descriptiveness with marginally larger N_d and N_h. Because the resolution is sufficiently high, a bin that is too large will hide many details. In contrast, if place recognition is required in a new dataset with lower resolution, marginally smaller N_d and N_h will be good choices.

Nv is set according to the analysis of frequently used rotating multi-beam LiDAR sensors in localization. The vertical FOV of rotating multi-beam LiDAR is limited. For example, the vertical angle field values of Velodyne HDL-64E (64 beam LiDAR), Velodyne HDL-32E (32 beam LiDAR) and Velodyne VLP-16C (16 beam LiDAR) are 26.8°, 40° and 30°, respectively. Considering that the proposed descriptor should be general to the frequently used LiDAR sensors, we set $Nv = 8$ to ensure that there are at least two laser beams in each partition within the bin, thus resisting the effects of noise. If Nv is larger than 8, there will not be a sufficient number of laser beam pairs in each partition within the bin when using Velodyne VLP-16C. If Nv is smaller than 8, a partition that is too large within the bin will hide many details.

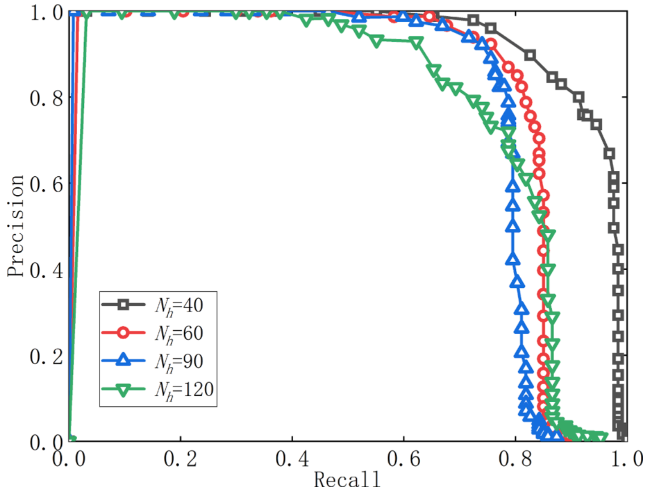


Fig. 9. PR curves with different parameters in the partial KITTI 00 sequences.

3.6.2. Parameter analysis of the threshold value of global localization integrity

The threshold value of the weighted combined score is used to calculate the integrity that reflects the robustness of global localization. In this section, the details of setting the threshold values in different conditions (i.e., different LiDAR sensors) and the requirements of localization precision are discussed.

The threshold value of the weighted combined score can be written as:

$$Thr = 0.67 \bullet Thr_{pr} + 0.33 \bullet \left(1 - precision_{req}\right) \tag{12}$$

where, Thr is the threshold value of the weighted combined score;Thr_{pr} is the threshold value of place recognition; and $precision_{req}$ contains the requirements of localization precision.

Thr_{pr} is related to the type of LiDAR sensor present, *dis* and *ratio* both express the performance of place recognition. In Xu et al. (2022), the performances of place recognition with different LiDAR sensors with different *dis* are analysed. To ensure the precision of place recognition, a smaller *dis* is required when LiDAR with fewer beams is used. For example, when 64-beam LiDAR is used, *dis* is set as 0.13 to ensure place recognition precision. When 16-beam LiDAR is used, *dis* is set as 0.07 to ensure place recognition precision. In the experiment evaluating the NCDR strategy, we find that when *ratio* is 0.5, precision improves markedly. Therefore, when 64-beam LIDAR is used, Thr_{pr} can be set equal to 0.435 (0.87 × 0.5). When a 0.5-m localization accuracy is required, $precision_{req}$ is 0.5. Thr is thus calculated to be 0.456 using Formula (12). Table 1 shows the integrity analysis of the global localization result.

As shown in the Table 1, we find that the proposed integrity accurately reflects the robustness of global localization under different conditions and precision requirements.

4. Experiments

In this section, the CSSC descriptor, two-stage similarity estimation and NCDR strategy are evaluated. The SGICP for 6DOF estimation is subsequently evaluated. Finally, the entire global localization solution is assessed using homogeneous and heterogeneous HD maps.

4.1. Experiments setup

In this section, the details of the datasets applied in this study's experiments are described, and evaluation criteria are introduced. All experiments were performed on a computer with a Ryzen 9 3900x CPU and 64 GB of RAM.

Table 1
Percentage of correct location under different conditions and precision requirements.

Dataset	$precision_{req}(m)$			
	0.2	0.3	0.4	0.5
KITTI (64-beam LiDAR)	94.8%	98.4%	99.4%	98.3%
NCLT (32-beam LiDAR)	92.2%	94.1%	94.8%	92.1%

4.1.1. Datasets

The experiments were performed with three popular open-source datasets: the KITTI dataset (Geiger et al., 2012), NCLT dataset (Carlevaris-Bianco et al., 2016); WHU-TLS campus dataset (Dong et al., 2020), and a custom dataset created for this study. These datasets have a clear degree of differentiation, such as the type of LiDAR used and changed or unchanged viewpoints at the same place. KITTI and NCLT datasets are collected by rotating multiple-beam LiDAR. The WHU-TLS campus dataset and the proposed own dataset are collected by TLS and 2D range finders, respectively, as shown in Fig. 10.

In the KITTI dataset, the KITTI sequences 00, 05 and 08, containing 4541, 2761 scans and 4071 scans, respectively, are selected. A 64-beam LiDAR (Velodyne HDL-64E) is used for data collection. In sequences 00 and 05, the viewpoints were unchanged when the vehicle revisited the place. In sequence 08, the viewpoints were changed.

The NCLT dataset contains repetitive measurements of different times along similar routes and is obtained in a challenging environment. This dataset can be used to evaluate global localization solutions based on a global prior map (Kim et al., 2019). A 32-beam LiDAR (Velodyne HDL-32E) is used for data collection. The NCLT-01-08, NCLT-05-26 and NCLT-09-28 datasets, containing 28,127 scans, 26,544 scans and 23,394 scans, respectively, are selected for evaluation. In this dataset, the robot moves in both the same and opposite directions at revisited places.

The WHU-TLS campus dataset (Dong et al., 2020) was captured at Wuhan University using the RIEGL VZ-400. RIEGL VZ-400 is a TLS that obtains a point cloud with millimetre accuracy. The custom dataset was captured on the first floor of a building using Navvis m3, which is a trolley-based MLS that consists of three 2D laser range finders and other sensors.

There is inevitably motion artefacts in the point cloud that are acquired by a rotating LiDAR sensor mounted above the moving platform. Fortunately, the frequency of LiDAR is high. In one single-shot LiDAR scan period, the moving platform can be thought of as moving at a constant speed. Therefore, the motion artefact can be corrected using the IMU or wheel odometry. All data used in the proposed experiments have been corrected.

4.1.2. Evaluation criteria

The performances of the proposed improved methods are evaluated by the PR curve, relative translation error (RTE) and relative rotation error (RRE).

The PR curve is obtained by calculating the precision and recall under different thresholds.

$$\text{Precision} = \frac{number\ \ of\ \ correct\ \ matches}{total\ \ number\ \ of\ \ matches}$$

$$\text{Recall} = \frac{number\ \ of\ \ correct\ \ matches}{total\ \ number\ \ of\ \ corresponding\ \ matches} \tag{13}$$

where the *number of correct matches* is the number of pairs whose Euclidean distance is smaller than 4 m and whose descriptor similarity is less than the threshold; the *total number of matches* is the number of pairs in which their descriptor similarity is less than the threshold; and the *total number of corresponding matches* is the number of pairs whose Euclidean distance is smaller than a threshold value. The threshold value is set at 4 m according to Kim and Kim (2018), Wang et al. (2019).

The RTE and RRE are given by:

$$\Delta \boldsymbol{H}_r = \boldsymbol{H}_r(\boldsymbol{H}_G)^{-1} = \begin{bmatrix} \Delta\boldsymbol{R} & \Delta\boldsymbol{T} \\ 0^T & 1 \end{bmatrix}$$

$$\text{RRE} = arccos\left(\frac{tr(\Delta\boldsymbol{R}) - 1}{2}\right) \tag{14}$$

$$\text{RTE} = \|\Delta\boldsymbol{T}\|$$

where $\boldsymbol{H}_r$ is the estimated transformation from the proposed solution; $\boldsymbol{H}_G$ is the corresponding ground-truth transformation from the computer vision method; and $tr(\Delta\boldsymbol{R})$ is the trace of $\Delta\boldsymbol{R}$.

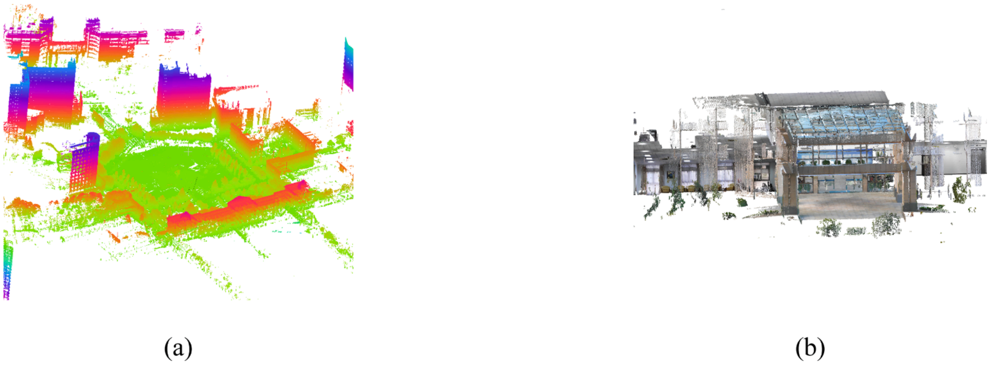
4.2. Evaluation of place recognition based on the CSSC descriptor

In this section, the strategies used in place recognition, including two-stage similarity estimation and NCDR, are evaluated. In the experiments that evaluate two-stage similarity estimation, the performance of the CSSC descriptor is also evaluated. For each experiment, the proposed methods are compared against different benchmarks.

4.2.1. Evaluation of two-stage similarity estimation

In the experiments that evaluate the two-stage similarity estimation, the KITTI and NCLT datasets are used. The scans in all sequences are sampled at approximately equidistant 2-m intervals. To conduct a more complete experiment, the similarities between the selected scan and all the remaining scans are calculated. Three other global feature descriptors (M2DP (He et al., 2016), SC (Kim and Kim, 2018) and iris (Wang et al., 2020) are used for evaluation. The corresponding similarity estimations used in these original papers are used. We implemented SC and M2DP in MATLAB, and iris in C++. The default parameters of the available codes are used. A comparison between the CSSC descriptor with the original similarity estimation and the CSSC descriptor with two-stage similarity estimation (CSSC-TS) is also conducted. The performances of all descriptors are evaluated using a 4-m threshold value of Euclidean distance, and the PR curve is shown in Fig. 11. The area under the curve (AUC) is a quantitative value that is more intuitive for evaluating the descriptiveness of a descriptor, and the AUCs of different methods are listed in Table 2.

Fig. 11 and Table 2 show that the SC descriptor obtains the best performance in KITTI sequences 00 and 05 from 64-beam LiDAR under unchanged viewpoint conditions. In KITTI sequence 08, the viewpoints are changed. In the NCLT-01-08, NCLT-05-26 and NCLT-09-28, the


(a)

(b)

Fig. 10. (a) WHU-TLS campus dataset. (b) Custom dataset created for this study.

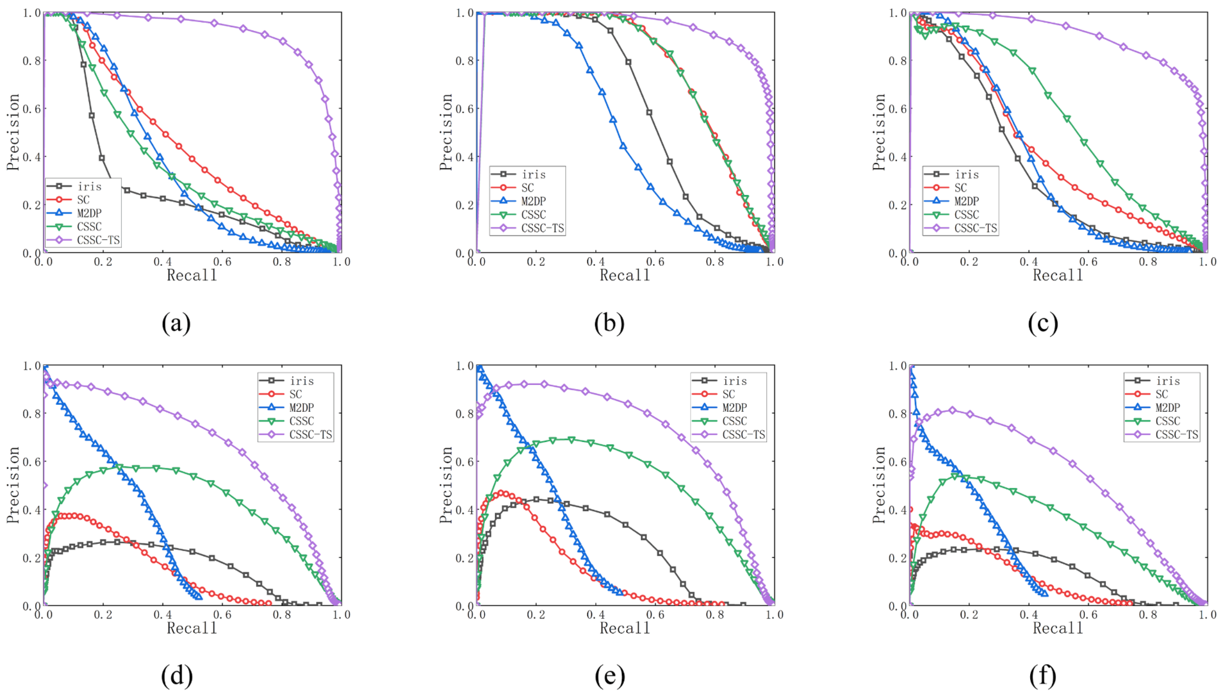


Fig. 11. The PR curve of different descriptors. (a) KITTI sequence 00; (b) KITTI sequence 05; (c) KITTI sequence 08; (d) NCLT-01-08; (e) NCLT-05-26; (f) NCLT-09-28.

Table 2
AUC of different descriptor in KITTI and NCLT datasets.

Dataset	iris	SC	M2DP	CSSC	CSSC-TS
KITTI 00	0.274	0.448	0.370	0.368	**0.908**
KITTI 05	0.618	0.776	0.499	0.763	**0.926**
KITTI 08	0.348	0.425	0.375	0.556	**0.896**
NCLT-01-08	0.164	0.142	0.269	0.420	**0.671**
NCLT-05-26	0.242	0.141	0.240	0.498	**0.718**
NCLT-09-28	0.133	0.113	0.195	0.325	**0.536**

resolution of the point cloud is lower than that of the KITTI datasets, and the robot moves in both the same and opposite directions at revisited places. The CSSC descriptor achieves the best performance with these datasets, which indicates that the elevation and point density weights used in this study enhance the descriptiveness of the CSSC descriptor, making it more robust than the other descriptors in challenging environments.

As shown in Fig. 11 and Table 2, the CSSC descriptor with two-stage similarity estimation achieves better performance than other descriptors in all datasets. Two-stage similarity estimation minimizes the influence of rotation and translation between two scans in the same place on the similarity of the descriptors, thus improving the performance of place recognition. This result indicates that the CSSC descriptor with two-stage similarity provides a robust initial location for query scans from LiDAR.

To determine what will occur if different threshold values of Euclidean distance are used, related experiments are also conducted. Because the scans in all sequences are sampled at approximately equidistant 2-m intervals, the threshold values of Euclidean distance are set at 3, 4, 6, 8 and 10 m. The AUC curves of different descriptors with different threshold values are shown in Fig. 12.

Fig. 12 shows that the AUC of all descriptors decreases with the increase in the threshold value of Euclidean distance. The larger threshold value makes place recognition more difficult because the two places are further apart. In all datasets, the CSSC-TS has the best performance at different threshold values, which indicates that the proposed descriptor has more descriptiveness and is more robust under challenging conditions.

The initial guess (Δx, Δy, yaw angle) is also provided by two-stage similarity estimation. Table 3 shows the error of the initial guess under KITTI sequences 05 and 08.

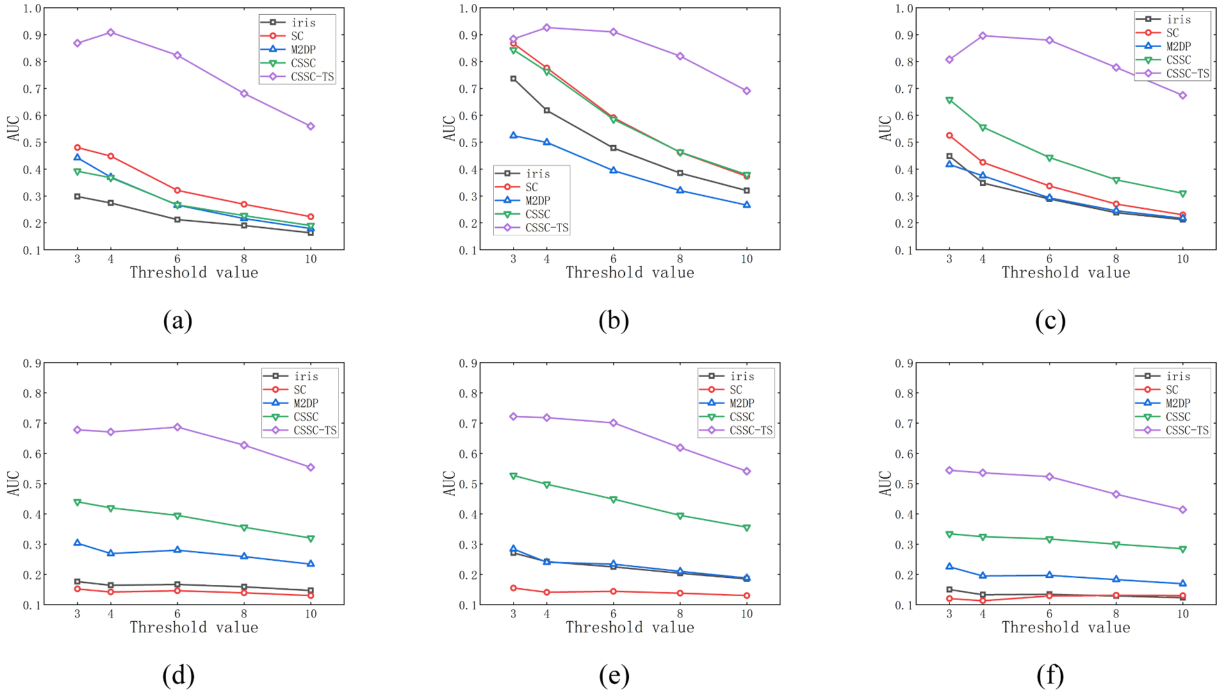
Table 3 shows that the two-stage similarity estimation provides an accurate initial guess in terms of the mean value. However, there are unacceptable errors in the results, such as the 6.08-m error of Δx and 89.9° error of the yaw angle. These inaccurate results will be filtered by the integrity of global localization. The initial guess helps reduce the error of the 6DOF estimation. Additional details can be found in Section 4.3.

4.2.2. *Evaluation of NCDR*

In the experiments that evaluate the proposed NCDR strategy, the KITTI sequences 05 and 08 are used. Most scans in both sequences are used to construct global prior maps, and the others are used as query scans to test the NCDR strategy, as shown in Fig. 13. In KITTI sequence 05, scans 1300–1600 and 2300–2650 are used as query scans, and the others are used to construct a global prior map. In KITTI sequence 08, scans 1400–1850 are used as query scans, and the others are used to construct a global prior map.

Query scans include some scans obtained in places where there is no global prior map and some scans obtained in places that are on the constructed map. This process is helpful when evaluating the strategy of finding the corresponding place. The conventional strategy is selected for evaluation. In the conventional strategy, similarity estimation is conducted between the query scan and all candidates. If the minimum *dis* between the CSSC descriptors of the query scan and the candidate is smaller than the threshold value, the corresponding place can be obtained. The precision curve of the conventional strategy and the proposed NCDR strategy with different ratio values is shown in Fig. 14.

Fig. 14 shows that the NCDR strategy can markedly improve the place recognition precision under both sequences. The precision becomes increasingly better with a decreasing ratio. When the ratio is<0.2, the place recognition precision is up to 1. When the ratio is between 0.6 and 0.3, there is small gap in precision improvement. When the ratio is



(a) (b) (c)

(d) (e) (f)

Fig. 12. The AUC curve of different descriptors. (a) KITTI sequence 00; (b) KITTI sequence 05; (c) KITTI sequence 08; (d) NCLT-01-08; (e) NCLT-05-26; (f) NCLT-09-28.

Table 3
Error of initial guess in KITTI sequences.

Sequences	Δx (m)		Δy (m)		Yaw angle (°)	
	mean	max	mean	max	mean	max
05	0.24	6.08	0.18	5.35	0.57	3.02
08	0.66	5.67	0.29	6.15	0.77	89.90

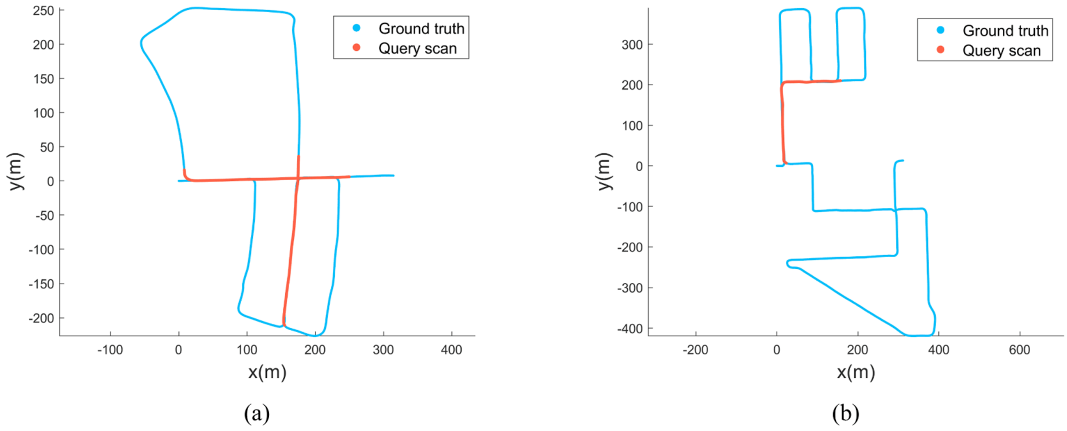
0.9, the precision improves the least. However, there is still a large improvement in precision compared to the conventional method, which indicates that NCDR effectively improves the place recognition precision.

4.3. Evaluation of SGICP

In the SGICP evaluation experiment, KITTI sequences 05 and 08 were used. Because 6DOF estimation is executed after place recognition, the experiment of evaluating the proposed SGICP is conducted over query scans and their corresponding candidates. The query scans are the same as the query scans that are selected in Section 4.2.2. The corresponding candidates can be obtained according to the ground truth of the global location. The execution times required for 6DOF estimation are 4888 and 1788 in the KITTI sequences 05 and 08, respectively. Two other 6DOF estimation methods are selected for evaluation, including the original ICP used in the existing solution and the original GICP. The initial guess is provided by two-stage similarity estimation. In addition, SGICP without an initial guess (SGICP-NI) is also compared for evaluation. Table 4 shows the related results.

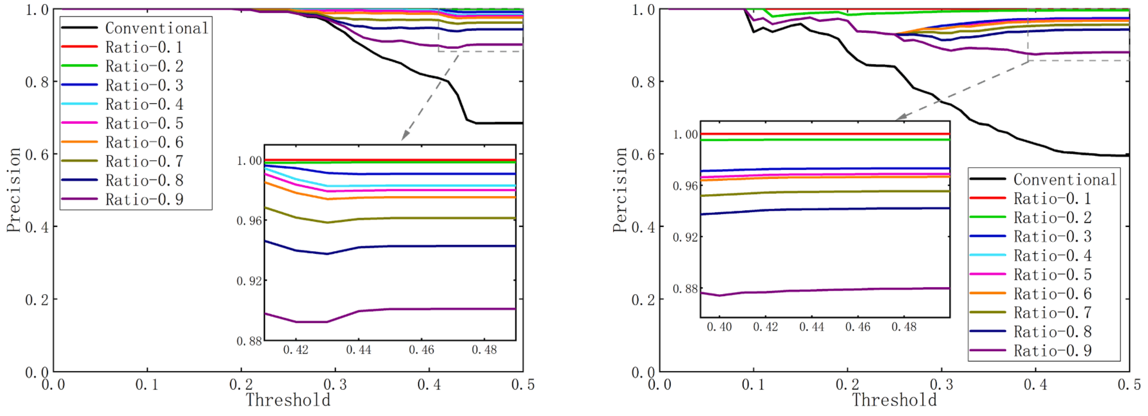
Table 4 shows that in terms of RTE, SGICP achieves the best performance. In challenging KITTI sequence 08, SGICP yields a large improvement compared to GICP. In terms of RRE, there are marginal differences between the three methods because the initial angular guess provided by two-stage similarity estimation is sufficiently accurate. By comparing the results of SGICP and SGICP-NI, we find that the initial guess leads to better precision.

In terms of time, the mean times of ICP, GICP and SGICP are 1.063 s, 0.356 s and 0.141 s, respectively. SGICP can estimate 6DOF in a shorter



(a) (b)

Fig. 13. The demonstration of query scan in KITTI dataset. (a) KITTI sequence 05; (b) KITTI sequence 08.



(a)

(b)

Fig. 14. The curve of precision with different ratio in KITTI dataset. (a) KITTI sequence 05; (b) KITTI sequence 08.

Table 4
Comparison of different 6DOF estimation methods.

Sequence	RTE				RRE			
	ICP	GICP	SGICP	SGICP-NI	ICP	GICP	SGICP	SGICP-NI
5	0.17	0.04	0.04	0.08	0.78	0.82	0.79	0.81
8	0.71	0.67	0.46	0.72	1.32	1.13	1.12	1.13

time than others. The cautious nearest point searching procedure, based on two weights, gives SGICP higher accuracy and better efficiency.

4.4. Evaluation of global localization solution

In this section, the proposed global localization solution is evaluated using homogeneous and heterogeneous HD maps. The experiments focus on the accuracy evaluation with homogeneous HD maps. Using heterogeneous HD maps, the applicability of the proposed solution is assessed. The execution time of the proposed solution is discussed. The threshold value of the integration is set according to the parameter analysis in Section 3.6.2.

4.4.1. Accuracy comparison with homogeneous HD maps

To date, few studies have focused on using only one LiDAR scan for global localization. OneShot (Ratz et al., 2020) is a similar type of solution that uses only a single 3D LiDAR scan for instant global localization. Unfortunately, OneShot is not an open-source project currently. KITTI sequence 00 was selected to evaluate the performance of OneShot in the paper. Therefore, the comparison with OneShot is executed in KITTI sequence 00. The same implemented details as those in OneShot are performed. The scans of seconds 340 to 397 are query scans and those of seconds 0 to 300 are for global prior map building. The results of OneShot are extracted from the original paper and are shown in Fig. 15 with the results of the proposed solution.

Fig. 15 shows that 75% of the proposed global localization solution RTEs are distributed in the range of 0 m to 0.1 m, and 6% of the proposed global localization solution RTEs are distributed rarely in a range larger than 0.2 m. Thirty-two percent of OneShot RTEs are distributed in the range of 0 m to 0.1 m, and 15% of OneShot RTEs are distributed in the range larger than 0.2 m. In KITTI sequence 00, the mean RTE of OneShot was computed as 0.11 m in this study. In KITTI sequence 00, the mean RTE of the proposed solution is 0.08 m. Compared to OneShot, the proposed solution achieves a mean RTE improvement of 27% on KITTI sequences 00, which indicates that the proposed solution has better performance than OneShot.

For a comprehensive evaluation of the proposed solution, experiments are also performed with the challenging NCLT datasets. However, no study has evaluated global localization that uses one LiDAR scan over NCLT datasets, nor is there an open source single-shot global localization solution. Therefore, the combined framework of the SC descriptor (Kim and Kim, 2018) and faster GICP (Koide, et al., 2021) is selected as the benchmark based on the following reasons. The SC descriptor is a modern egocentric global feature descriptor that uses a two-phase search algorithm that efficiently finds corresponding places. Also, the initial guess of the yaw angle is provided by the SC descriptor. For 6DOF estimation, faster GICP is picked for its good accuracy and efficiency. NCLT-01-08 is selected to construct a global prior map, and the scans in NCLT-05-26 and NCLT-09-28 are selected as query scans, as shown in Fig. 16.

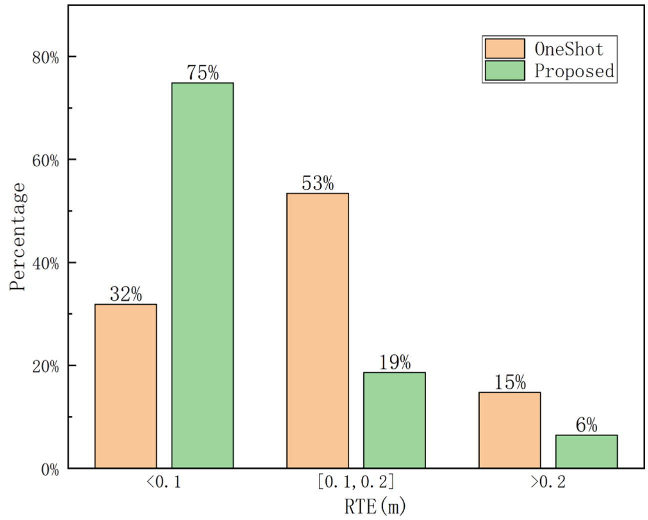
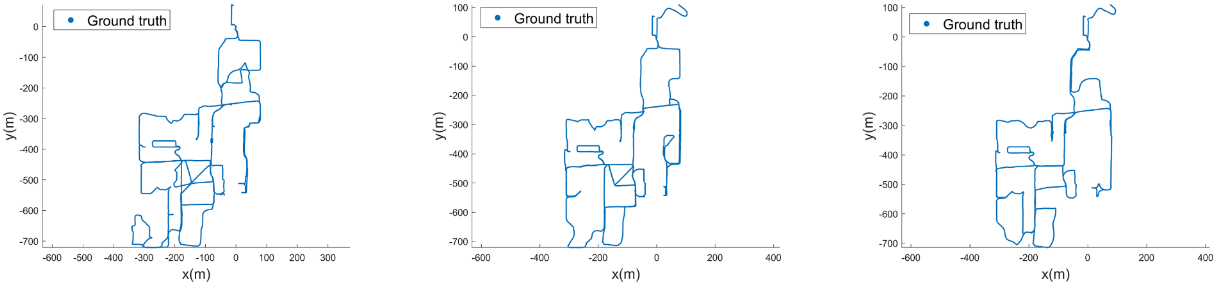


Fig. 15. The percentage of different methods RTE distribution in KITTI sequence 00.

To visualize the experimental results more intuitively, the



(a) (b) (c)

Fig. 16. Ground truth of NCLT datasets. (a) NCLT-01-08; (b) NCLT-05-26; (c) NCLT-09-28.

cumulative distribution function (CDF) figures of both solutions are shown in Fig. 17 and Fig. 18. Comparative results are also listed in Table 5.

Fig. 17, Fig. 18 and Table 5 show that in terms of the mean, one standard deviation (1σ), and the distribution of RTE, the proposed solution achieve better performance than the benchmark. Relative to the benchmark, the proposed solution achieves an average 39% improvement over the mean of RTE and an average 75% improvement over the mean of RRE in both datasets. The proposed solution has an average 77% improvement over 1σ of RTE and an average 93% improvement over 1σ of RRE in both datasets.

The proposed global localization algorithm achieves excellent performance due to improvements in place recognition and 6DOF estimation methods. Compared with the SC descriptor, the CSSC descriptor achieves better descriptiveness in challenging environments, as shown in Section 4.2.1. Two-stage similarity estimation and the NCDR strategy improve place recognition precision. These factors allow the proposed solution to produce more accurate place recognition results. In the 6DOF estimation, the SC descriptor provides an initial guess of the yaw angle, while two-stage similarity estimation provides a 3D initial guess (Δx, Δy, yaw angle). Also, SGICP has higher accuracy than GICP. These improvements make the proposed solution more accurate than the benchmark.

4.4.2. *Applicability of the proposed solution with heterogeneous HD maps*

In this section, the applicability of the proposed solution with heterogeneous HD maps is evaluated. A 16-beam LiDAR (Velodyne VLP-16C) is used to collect query scans in the same scene as the heterogeneous HD maps. Using manual registration, the location of the query scan in the same geo-referenced frame as the HD map can be obtained. Unfortunately, because the pose of LiDAR cannot be measured accurately during data collection, we cannot obtain the real pose of LiDAR in the georeferenced frame. Therefore, the RTE is only evaluated in the following experiment.

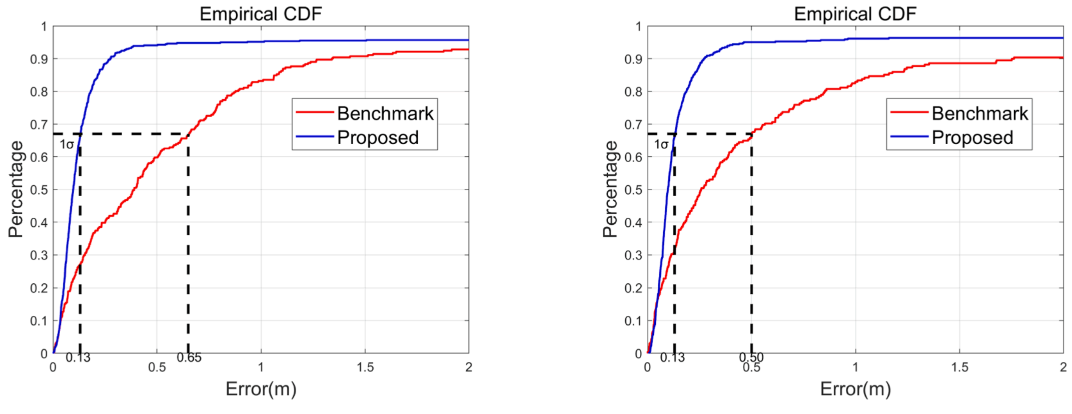
First, heterogeneous HD maps must be converted into virtual scans that are similar to real scans from VLP-16C through virtual LiDAR, as shown in Fig. 19. The point cloud and location of a virtual scan are generated in a semi-automatic way. More details can be found in (Xu et al., 2022).

After obtaining the virtual scans, the global prior map is acquired by fusing scans, CSSC descriptors and fingerprints. The estimated global location of a query scan is obtained with place recognition and 6DOF estimation. In the WHU-TLS campus dataset, the mean RTE of the proposed solution is 0.18 m. In the custom dataset, the mean RTE of the proposed solution is 0.07 m. Results show that the proposed global localization solution can obtain accurate locations using the heterogeneous HD map. This indicates that the proposed solution allows for the use of different types of LiDAR during HD map creation and localization separately. Combined with the results in Section 4.4.1, the proposed solution is found to be widely available for use with point clouds that acquired by different sensors, such as rotating multiple-beam LiDAR, TLS and MLS.

4.4.3. *Execution time of the proposed solution*

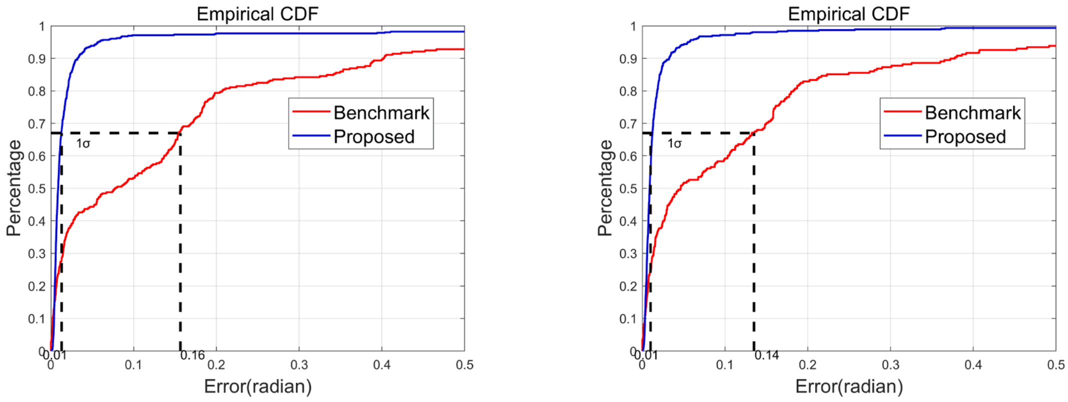
In this section, the execution time of the procedures, including offline HD map preprocessing and online global localization, are discussed.

The execution time of homogeneous HD map preprocessing is equal to the time of feature extraction. In KITTI datasets, the mean time for extracting features from one scan is 4.37 ms. In NCLT datasets, the mean time for extracting features from one scan is 1.87 ms. In a VLP-16C dataset, the mean time for extracting features from one scan is 1.41 ms. The time of the feature extraction depends on the number of points



(a) (b)

Fig. 17. The empirical CDF of RTE in NCLT datasets. (a) NCLT-05-26; (b) NCLT-09-28.



(a) (b)

Fig. 18. Empirical CDF of RRE in NCLT datasets. (a) NCLT-05-26; (b) NCLT-09-28.

Table 5
Comparison with LiDAR global localization benchmark.

NCLT-Sequence	methods	RTE				RRE			
		mean	1σ	<0.1	<0.5	mean	1σ	<0.1	<0.5
05–26	proposed	**0.70**	**0.13**	**52%**	**94%**	**0.05**	**0.01**	**97%**	**99%**
	benchmark	1.23	0.65	23%	60%	0.17	0.16	53%	93%
09–28	proposed	**1.07**	**0.13**	**51%**	**95%**	**0.03**	**0.01**	**97%**	**99%**
	benchmark	1.65	0.50	26%	66%	0.15	0.14	59%	94%

(a) (b)

Fig. 19. Different LiDAR scans in the same scene. (a) A real scan from VLP-16C; (b) a virtual scan from virtual LiDAR.

in one LiDAR scan.

The time of heterogeneous HD map preprocessing includes the time of data conversion and feature extraction. With the WHU-TLS campus dataset, 109.05 million points are contained, and 88 locations are generated to set virtual LiDAR. Virtual scans require 80.47 s to generate results. With the custom dataset, 21.83 million points are contained, and 26 locations are generated to set virtual LiDAR. Virtual scans require 2.29 s to generate results. Thus, we find that offline preprocessing of HD maps is not time-consuming. The execution time of online global localization is shown in Fig. 20.

Fig. 20 shows that place recognition requires a long time in most cases. In a few cases, 6DOF takes more time. The mean time consumed for place recognition is 0.86 s, and the mean time consumed for 6DOF estimation is 0.6 s. The mean time consumed for global localization using a single shot scan is 1.46 s. In general, the proposed solution provides global location at an approximate frequency of 1 HZ.

According to the evaluation of the proposed global localization solution, we find that the proposed solution provides accurate global locations by fusing rotating multiple-beam LiDAR and diverse HD maps that include homogeneous and heterogeneous maps. Offline preprocessing improves the applicability of the proposed solution by converting homogeneous and heterogeneous HD map maps into global prior maps. The global prior map is used to calculate the global location of the query scan from rotating multiple-beam LiDAR. The wide applicability of the proposed solution makes some existing heterogeneous HD maps available for global localization, avoiding repeated collection of HD maps. In terms of computation time, we find that the HD map from diverse LiDAR sensors can be used to execute global localization after a moment offline preprocessing. In online global localization, the proposed solution provides a global location at a frequency of 1 HZ in most

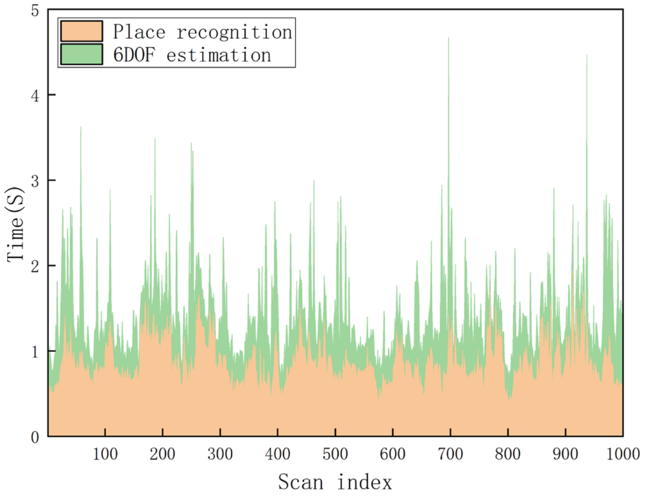


Fig. 20. Execution time of the proposed solution.

cases. The frequency of global localization can be further improved by limiting the search space according to the translation provided by LiDAR odometry. In general, the proposed solution provides global locations in an accurate and efficient way and has broad applicability with respect to using point clouds acquired by different LiDAR sensors.

5. Conclusion

To date, global localization under GNSS-denied environments remains a challenge in terms of accuracy and usability. This paper proposes a single-shot LiDAR scan global localization solution for driverless vehicles and robots, which provides an instant global location by fusing LiDAR and HD maps and has no wake-up or kidnapped problems. This paper proposes a novel CSSC descriptor, two-stage similarity estimation and NCDR strategies to enhance the robustness of place recognition and the SGICP algorithm for improving the precision of location and pose. Also, the proposed preprocessing procedure makes the proposed solution accept HD maps acquired by different LiDAR sensors. The proposed CSSC descriptor is more robust than existing descriptors, as shown by the comparison of the PR curve of multiple scenes, particularly under changed viewpoints and low-resolution conditions. Experimental analysis also confirms that the proposed strategies, two-stage similarity estimation and NCDR, improve place recognition precision. Also, the SGICP algorithm improves accuracy compared to the GICP algorithm by 31% and efficiency by 60%. Using the homogeneous KITTI dataset, the proposed global localization method achieves a mean RTE improvement of 27% relative to existing methods. Using a long-term localization dataset, the proposed solution achieves an average 77% improvement over 1σ of RTE and an average 93% improvement over 1σ of RRE compared to the benchmark method. Using the heterogeneous WHU-TLS campus dataset and a custom dataset, the proposed solution achieves 0.18 m and 0.07 m RTEs, respectively. Many experiments confirm that the proposed solution achieves higher accuracies and broader generalizability than existing solutions.

The proposed solution provides accurate global locations by fusing LiDAR and HD maps. However, the proposed solution cannot meet real-time requirements. In the future, we plan to integrate the proposed solution with other localization systems to develop real-time global localization solution. We also plan to research the impact of the different LiDAR sensors on the accuracy of global localization, further improving the applicability of the proposed solution.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported in part by the Natural Science Fund of China with Project No. 41874031 and 42111530064, the National Key Research Development Program of China with project No. 2021YFB2501102, Shenzhen Science and Technology Program No. JCYJ20210324123611032 and Academy of Finland with Project No. 337656.

References

Angelina, M., Gim, U., Lee, H., 2018. Pointnetvlad: Deep point cloud based retrieval for large-scale place recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4470–4479.

Behley, J., Stachniss, C., 2018. Efficient Surfel-Based SLAM using 3D Laser Range Data in Urban Environments. IN: Robotics: Science and Systems, vol. 2018, p. 59.

Biber, P., Straßer, W., Gris, W.S.I., 2003. The normal distributions transform: A new approach to laser scan matching. In: Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453). vol. 3, IEEE, pp. 2743–2748.

Boroson, E.R., Ayanian, N., 2019. 3D keypoint repeatability for heterogeneous multi-robot SLAM. In: Proc. - IEEE Int. Conf. Robot. Autom., vol. 2019-May, pp. 6337–6343, https://doi.org/10.1109/ICRA.2019.8793609.

Carlevaris-Bianco, N., Ushani, A.K., Eustice, R.M., 2016. University of Michigan North Campus long-term vision and lidar dataset. Int. J. Rob. Res. 35 (9), 1023–1035. https://doi.org/10.1177/0278364915614638.

Chang, M.Y., Yeon, S., Ryu, S., Lee, D., 2020. SpoxelNet: Spherical voxel-based deep place recognition for 3D point clouds of crowded indoor spaces. In: IEEE Int. Conf. Intell. Robot. Syst., pp. 8564–8570. https://doi.org/10.1109/IROS45743.2020.9341549.

Chen, X., Läbe, T., Nardi, L., Behley, J., Stachniss, C., 2020. Learning an Overlap-based Sensor Model for 3D LiDAR Localization. In: Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst., pp. 4602–4608, [Online]. Available: http://www.ipb.uni-bonn.de/pdfs/chen2020iros.pdf.

Cop, K.P., Borges, P.V.K., Dube, R., 2018. Delight: An Efficient Descriptor for Global Localisation Using LiDAR Intensities. In: Proc. - IEEE Int. Conf. Robot. Autom., pp. 3653–3660. https://doi.org/10.1109/ICRA.2018.8460940.

Dong, Z., Yang, B., Liu, Y., Liang, F., Li, B., Zang, Y., 2017. A novel binary shape context for 3D local surface description. ISPRS J. Photogramm. Remote Sens. 130, 431–452. https://doi.org/10.1016/j.isprsjprs.2017.06.012.

Dong, Z., Liang, F., Yang, B., Xu, Y., Zang, Y., Li, J., Wang, Y., Dai, W., Fan, H., Hyyppä, J., Stilla, U., 2020. Registration of large-scale terrestrial laser scanner point clouds: A review and benchmark. ISPRS J. Photogramm. Remote Sens. 163, 327–342.

Dube, R., Dugas, D., Stumm, E., Nieto, J., Siegwart, R., Cadena, C., 2017. SegMatch: Segment based place recognition in 3D point clouds. In: Proc. - IEEE Int. Conf. Robot. Autom., pp. 5266–5272. https://doi.org/10.1109/ICRA.2017.7989618.

Fan, Y., He, Y., Tan, U.X., 2020. Seed: A segmentation-based egocentric 3D point cloud descriptor for loop closure detection. In: IEEE Int. Conf. Intell. Robot. Syst., pp. 5158–5163. https://doi.org/10.1109/IROS45743.2020.9341517.

Frome, A., Huber, D., Kolluri, R., Bülow, T., Malik, J., 2004. Recognizing objects in range data using regional point descriptors. In: Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 3023, pp. 224–237, https://doi.org/10.1007/978-3-540-24672-5_18.

Geiger, A., Lenz, P., Urtasun, R., 2012. Are we ready for autonomous driving? the KITTI vision benchmark suite. In: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 3354–3361. https://doi.org/10.1109/CVPR.2012.6248074.

Guo, J., Borges, P.V.K., Park, C., Gawel, A., 2019. Local Descriptor for Robust Place Recognition Using LiDAR Intensity. IEEE Robot. Autom. Lett. 4 (2), 1470–1477. https://doi.org/10.1109/LRA.2019.2893887.

Guo, Y., Sohel, F., Bennamoun, M., Wan, J., Lu, M., 2015. A novel local surface feature for 3D object recognition under clutter and occlusion. Inf. Sci. (Ny) 293, 196–213. https://doi.org/10.1016/j.ins.2014.09.015.

He, L., Wang, X., Zhang, H., 2016. M2dp: A novel 3D point cloud descriptor and its application in loop closure detection. In: IEEE Int. Conf. Intell. Robot. Syst., vol. 2016-Novem, pp. 231–237, https://doi.org/10.1109/IROS.2016.7759060.

Kim, G., Kim, A., 2018. Scan Context: Egocentric Spatial Descriptor for Place Recognition Within 3D Point Cloud Map. In: IEEE Int. Conf. Intell. Robot. Syst., pp. 4802–4809. https://doi.org/10.1109/IROS.2018.8593953.

Kim, G., Park, B., Kim, A., 2019. SCI - LiDAR Localization Using Scan Context Image. IEEE Robot. Autom. Lett. 4 (2), 1948–1955.

Koide, K., Yokozuka, M., Oishi, S., et al., 2021. Voxelized gicp for fast and accurate 3d point cloud registration. In: 2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 11054–11059.

Kong, X., et al., 2020. Semantic graph based place recognition for 3D point clouds. In: IEEE Int. Conf. Intell. Robot. Syst., pp. 8216–8223. https://doi.org/10.1109/IROS45743.2020.9341060.

Liu, Z., et al., 2019. SeqLPD: Sequence Matching Enhanced Loop-Closure Detection Based on Large-Scale Point Cloud Description for Self-Driving Vehicles. In: IEEE Int. Conf.

Intell. Robot. Syst., pp. 1218–1223. https://doi.org/10.1109/IROS40897.2019.8967875.

Liu, J., Liang, X., Hyyppä, J., Yu, X., Lehtomäki, M., Pyörälä, J., Zhu, L., Wang, Y., Chen, R., 2017. Automated matching of multiple terrestrial laser scans for stem mapping without the use of artificial references. Int. J. Appl. Earth Obs. Geoinf. 56, 13–23.

Liu, J., Xu, D., Hyyppa, J., Liang, Y., 2021. A survey of applications with combined BIM and 3D laser scanning in the life cycle of buildings. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 5627–5637.

Muhammad, N., Lacroix, S., 2011. Loop closure detection using small-sized signatures from 3D LIDAR data. In: 9th IEEE Int. Symp. Safety, Secur. Rescue Robot. SSRR 2011, pp. 333–338, 2011, https://doi.org/10.1109/SSRR.2011.6106765.

Prestes, E., Ritt, M., Führ, G., 2009. An improved particle filter for sparse environments. J. Brazilian Comput. Soc. 15 (3), 55–64. https://doi.org/10.1007/bf03194506.

Ratz, S., Dymczyk, M., Siegwart, R., Dube, R., 2020. OneShot Global Localization: Instant LiDAR-Visual Pose Estimation. In: Proc. - IEEE Int. Conf. Robot. Autom., pp. 5415–5421. https://doi.org/10.1109/ICRA40945.2020.9197458.

Schaupp, L., Burki, M., Dube, R., Siegwart, R., Cadena, C., 2019. OREOS: Oriented Recognition of 3D Point Clouds in Outdoor Scenarios. In: IEEE Int. Conf. Intell. Robot. Syst., pp. 3255–3261. https://doi.org/10.1109/IROS40897.2019.8968094.

Segal, A.V., Haehnel, D., Thrun, S., 2010. Generalized-ICP. Robot. Sci. Syst. 5, 161–168. https://doi.org/10.15607/rss.2009.v.021.

Seow, Y., Miyagusuku, R., Yamashita, A., Asama, H., 2017. Detecting and solving the kidnapped robot problem using laser range finder and wifi signal. In: 2017 IEEE Int. Conf. Real-Time Comput. Robot. RCAR 2017, vol. 2017-July, pp. 303–308, https://doi.org/10.1109/RCAR.2017.8311878.

Shan, T., Englot, B., 2018. LeGO-LOAM: Lightweight and Ground-Optimized Lidar Odometry and Mapping on Variable Terrain. In: IEEE Int. Conf. Intell. Robot. Syst., no. October, pp. 4758–4765, https://doi.org/10.1109/IROS.2018.8594299.

Steder, B., Ruhnke, M., Grzonka, S., Burgard, W., 2011. Place recognition in 3D scans using a combination of bag of words and point feature based relative pose estimation, pp. 1249–1255, https://doi.org/10.1109/iros.2011.6094638.

Sun, L., Adolfsson, D., Magnusson, M., Andreasson, H., Posner, I., Duckett, T., 2020. Localising Faster: Efficient and precise lidar-based robot localisation in large-scale environments, pp. 4386–4392, https://doi/ogr/10.1109/icra40945.2020.9196708.

Tao, W., Hua, X., Wang, R., Xu, D., 2020. Quintuple local coordinate images for local shape description. Photogramm. Eng. Remote Sensing 86 (2), 121–132. https://doi.org/10.14358/PERS.86.2.121.

Vidanapathirana, K., Moghadam, P., Harwood, B., Zhao, M., Sridharan, S., Fookes, C., 2020. Locus: LiDAR-based place recognition using spatiotemporal higher-order pooling," arXiv.

Wang, Y., Sun, Z., Yang, J., Kong, H., 2019. LiDAR Iris for loop-closure detection, arXiv, pp. 5769–5775.

Wang, H., Wang, C., Xie, L., 2020. Intensity Scan Context: Coding Intensity and Geometry Relations for Loop Closure Detection, arXiv, pp. 2095–2101.

Wang, Y., Sun, Z., Xu, C.Z., Sarma, S.E., Yang, J., Kong, H., 2020. LiDAR iris for loop-closure detection. In: IEEE Int. Conf. Intell. Robot. Syst., pp. 5769–5775. https://doi.org/10.1109/IROS45743.2020.9341010.

Xu, D., Liu, J., Hyyppä, J., Liang, Y., Tao, W., October 2021. A heterogeneous 3D map-based place recognition solution using virtual LiDAR and a polar grid height coding image descriptor. ISPRS J. Photogramm. Remote Sens. 183, 1–18. https://doi.org/10.1016/j.isprsjprs.2021.10.020.

Yin, H., Wang, Y., Ding, X., Tang, L., Huang, S., Xiong, R., 2020. 3D LiDAR-Based Global Localization Using Siamese Neural Network. IEEE Trans. Intell. Transp. Syst. 21 (4), 1380–1392. https://doi.org/10.1109/TITS.2019.2905046.

Yin, D., Zhang, Q., Liu, J., Liang, X., Wang, Y., Chen, S., Maanpää, J., Hyyppä, J., Chen, R., 2021. Interest point detection from multi-beam light detection and ranging point cloud using unsupervised convolutional neural network. IET Image Process. 15 (2), 369–377.

Zhang, J., Singh, S., 2017. Low-drift and real-time lidar odometry and mapping. Auton. Robots 41 (2), 401–416. https://doi.org/10.1007/s10514-016-9548-2.

Zhao, B., Le, X., Xi, J., 2019. A novel SDASS descriptor for fully encoding the information of a 3D local surface. Inf. Sci. (Ny) 483, 363–382. https://doi.org/10.1016/j.ins.2019.01.045.