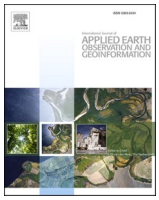
International Journal of Applied Earth Observations and Geoinformation 102 (2021) 102430

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

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



FastLCD: A fast and compact loop closure detection approach using 3D point cloud for indoor mobile mapping

Haodong Xiang [a], Wenzhong Shi [a,*], Wenzheng Fan [a], Pengxin Chen [a], Sheng Bao [a], Mingyan Nie [b]

[a] Smart Cities Research Institute, Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, PR China
[b] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, PR China

ARTICLE INFO

Keywords:
Loop closure detection
Comprehensive descriptors
Machine learning
LiDAR-based mobile mapping

ABSTRACT

In simultaneous localization and mapping (SLAM), loop closure detection is a significant yet still open problem. It contributes to construct a globally consistent and accurate map. This paper proposes a fast and compact loop closure detection method (FastLCD) based on comprehensive descriptors and machine learning to achieve reliable and precise results using 3D point cloud for indoor LiDAR mobile mapping. Comprehensive descriptors proposed in this paper encode discriminative multimodality features to describe each scan of point clouds. The specific values of descriptors of point cloud scan pairs are fed into a machine learning model. We leverage the pre-trained learning model as a classifier to distinguish whether a pair of laser scans is a loop candidate. Then, to ensure the results' precision, a novel double-deck loop candidate verification strategy is used to reject false positives. The algorithm is evaluated on datasets of some typical indoor environments. Compared with some state-of-the-art loop closure detection algorithms, the proposed FastLCD algorithm demonstrates superior performance in precision and recall rate. Moreover, the method proposed also exhibits high time efficiency, excellent generalization performance and insensitivity to threshold changes.

1. Introduction

Surveying robots are being increasingly used for mobile mapping and model reconstruction, especially in environments that lack the reliable signals of global navigation satellite system (GNSS) or other localization methods, such as indoor environments, city canyons, and underground scenes. In such environments, sensor-based localization methods are adopted, such as scan matching and visual odometry. However, these methods cause a drift in position because of the generation of cumulative errors. With increasing measurement distance, the drift also grows sharply (Thrun, 2002). Loop closure detection is a key step in SLAM to restrict these cumulative errors. It can be defined as a data association problem that aims to determine whether a place has been previously visited. It can reduce the pose estimation uncertainty and map inconsistency (Williams et al., 2011), with an aim to construct an accurate and consistent map for model reconstruction (Shi et al., 2019).

So far, several loop closure detection algorithms using 2D point cloud data (Hess et al., 2016; Himstedt et al., 2014; Zlot and Bosse, 2009) or visual data (Angeli et al., 2008; Labbe and Michaud, 2013; Labbé and Michaud, 2014; Memon et al., 2020; Ge et al., 2019) have been proposed. However, a 2D laser scanner only captures environmental information on a plane in each frame, which does not contain sufficient information (Li et al., 2017). When robots are running on a rough floor, the 2D scans might appear markedly different despite only slight changes in the position. Thus, in such cases, loop closure detection based on 2D scans is not reliable. Visual data are also widely used as they contain adequate information of environments with much lower costs. However, visual sensors are sensitive to illumination conditions (Chen et al., 2019; Lee et al., 2019), especially in indoor environments. Recently, with declining costs, 3D laser scanners have been widely applied in many fields. They capture more sufficient data than 2D laser scanners and work stably under illumination changes, even in dark environments.

In this paper, the loop closure detection problem is treated as a classification problem to identify whether two scans are captured from the same environment. We propose a loop closure detection method (FastLCD) in indoor environments based on comprehensive descriptors

* Corresponding author.
E-mail address: john.wz.shi@polyu.edu.hk (W. Shi).

https://doi.org/10.1016/j.jag.2021.102430
Received 30 April 2021; Received in revised form 12 June 2021; Accepted 30 June 2021
Available online 29 July 2021

and machine learning. Considering multimodality information in 3D point clouds, the comprehensive descriptor is proposed to describe each point cloud scan. The descriptor ratios are fed into a supervised learning model to detect loop closures. Loop results provide control information for back-end optimization in SLAM (Sünderhauf, 2012), enhancing the reliability of the adjustment network and improving the map's overall accuracy. False loop closures will have disastrous effects on SLAM results. Thus, a double-deck verification strategy is used to reject false detections to ensure the algorithm's precision.

The main contributions of this paper are:

1. The FastLCD algorithm leverages multimodality features to map a point cloud scan into a global comprehensive descriptor. The multimodality features are extracted from each single 3D LiDAR scan without any transformation or projection. Theory analysis and experiment results demonstrate the comprehensive descriptor is discriminative to location and external environments.
2. A highly-efficient supervised learning model is used as a classifier to identify loop closure candidates without priori poses. The model can also provide the estimate of the reliability of detection results. The detection results comply with Gaussian Mixture Model (GMM), which indicates the method proposed has great separability and insensitivity to threshold changes.
3. A double-deck loop verification strategy comprising cross-validation and post-verification is implemented to reject false positives to ensure the precision.

The rest of this paper is organized as follows. In Section 2, a brief review of loop closure detection methods is summarized and classified. Their problems and limitations are also stated in this section. In Section 3, FastLCD method based on comprehensive descriptors and machine learning is introduced in detail. The discrimination of multimodality features is also demonstrated. Section 4 shows the experiment results of our method and comparison algorithms in indoor environments. Section 5 discusses the performance and limitations of FastLCD algorithm. In Section 6, conclusions are stated, and directions of future work are presented.

2. Related work

Loop closure detection is a critical component towards addressing the problem of SLAM. This section briefly summarizes the previous work related to loop closure detection using point clouds. Loop closure detection algorithms can be classified into four categories according to the features adopted: (1) based on local features, (2) based on handcrafted global descriptors, (3) based on planes, objects, or semantic information, (4) based on deep learning.

Many of the traditional approaches are based on local features, such as key points. Steder et al. (2010) proposed a place recognition method based on point features extracted from 3D range data. The obtained points of interest were applied to extract features and score candidate transformations. Then, a threshold was applied to validate the candidates. In Steder et al. (2011), normal aligned radial features were applied using bag-of-words models. The fast point feature histograms (FPFHs) proposed in Rusu et al. (2009) optimized the traditional point feature histograms (PFHs) (Rusu et al., 2008). The computation of FPFH was based on the combination of geometry relations between the key points and neighbors. The FPFHs not only retained most of the descriptive power of a PFH but also could be computed online for real-time application.

Regarding methods based on handcrafted global descriptors, substantial acievements have been gained. Magnusson et al. (2009) exploited the surface representation of a normal distribution transform to create feature histograms. Here, a point cloud scan was split into several overlapping grids, and their linear, planar, and spherical properties were computed and compressed into a shape histogram. In addition, expectation maximization was used to fit a gamma mixture model to output similarity measures for automatically determining the threshold for loop closure detection. Rohling et al. (2015) proposed a fast histogram computed from the distances between the point and robot. A discrete Wasserstein metric was used to compare the two histograms, and loop closures were detected using an appropriate distance threshold. Considering structural information, a non-histogram-based global descriptor from 3D LiDAR scans, called scan context (Kim and Kim, 2018) was proposed. This approach directly recorded the 3D structure of a space that was invariant to LiDAR viewpoint changes. Scan context is also expanded to intensity scan context (H. Wang et al., 2020), considering intensity information. Granström and Schön (2010) used two types of global features, geometry features and range histograms. AdaBoost is used to learn a classifier from these features. The approach proposed in Zhuang et al. (2013) comprised local speeded-up robust features (SURFs) and global spatial features for place recognition task. M2DP (He et al., 2016) was a global descriptor produced by projecting a 3D point cloud to multiple 2D planes and computation of the signatures of the cloud on these planes. LiDAR Iris (Y. Wang et al., 2020) is a binary signature image representation. Place recognition is implemented by calculating the Hamming distance as similarities of two corresponding binary signature images.

Besides artificially designed local and global features, some algorithms based on advanced features, like planes, objects, or semantic information are proposed. Dube et al. (2017) proposed 3D segmentation methods, and realized place recognition through segment matching and geometry verification. Cupec et al. (2019) proposed an indoor place recognition approach based on matching planar surface segments and straight edges in depth images obtained from RGB-D images. Luo et al. (2016) proposed a scene recognition algorithm based on object descriptors, including the oriented, unique, and repeatable-clustered viewpoint feature histogram descriptor (Aldoma et al., 2012) and ensemble of shape functions descriptor (Wohlkinger and Vincze, 2011), which were extracted from the submaps segmented from the RGB-D range data. Furthermore, a distance metric was learned (Davis et al., 2007), with the aim of increasing the precision of place recognition under environmental changes.

Meanwhile, booming development in deep learning inspires many works in place recognition field. Unlike traditional methods adopting handcrafted features and metrics, deep-learning-based methods learn the features or metrics from the input data. According to the type of input data fed into deep learning models, methods can be divided into four categories: (1) handcrafted-feature-based methods, (2) image-based methods, (3) 3D-volume-based methods, (4) raw-point-cloud-based methods.

(1) Yin et al. (2017) transferred raw point cloud into a semi-handcrafted representation and considered loop closure detection as a similarity modeling problem. The semi-handcrafted representation with rotational invariance was fed into a siamese convolutional neural network (CNN) with contrastive loss for loop closure detection.
(2) OverlapNet (Chen et al., 2020) transferred point clouds to images as input. Multiple information, including depth, normal, intensity, and semantic information were used to generate images from a single LiDAR scan. The images were input into a siamese network for identifying loop closure candidates.
(3) A voxel-based representation learning method were proposed in Siva et al. (2020) to perform learning of voxels and features to represent scenes. 3D point cloud data were split into voxels. Multi-modal features extracted from those voxels were used to perform place recognition. MinkLoc3D (Komorowski, 2020) leveraged a sparse voxelized point cloud representation and sparse 3D convolutions to compute discriminative 3D point cloud descriptors.

(4) PointNetVLAD (Angelina et al., 2018) combined PointNet (Qi et al., 2017) and NetVLAD (Arandjelovic et al., 2018) to perfom end-to-end training and to extract global descriptor from 3D point cloud. The pre-classification results obtained from PointNet were input into NetVLAD with metric learning method for mapping 3D point cloud into global feature descriptors. "Lazy triplet and quadruplet" loss functions were proposed to achieve more generalization performance by augmenting differences between samples. Raw-point-cloud-based methods always need to solve the problem of permutation invariance. PointNetVLAD solved this problem by proving NetVLAD to be a symmetric function. Liu et al. (2019b) proposed LPD-Net, a large-scale place description network. This network applied adaptive feature extraction to obtain point cloud distribution information and local features. The resulting local feature vectors were processed by NetVLAD to get a global descriptor. Sun et al. (2021) proposed a place recognition algorithm taking raw point cloud data as input. This algorithm included two parts: feature extraction block using dual attention module, and feature fusion block encoding local features into a global descriptor by a NetVLAD layer. Methods proposed in Liu et al. (2019a), Xia et al. (2020), Zhang and Xiao (2019) also had similar workflow as Angelina et al. (2018). Local features were extracted from raw LiDAR scans. Then, a deep learning encoder was used to produce a compact global descriptor.

Each type of features has inherent disadvantages. Local features generally lack descriptive power and suffer from ambiguity and environment changes, while global descriptors always face problems of view-dependent and invariance. Algorithms based on planes, objects, or semantic information rely on the performance of these advanced feature extraction. Thus, multimodalities integration is an effective approach to remedy the defects of a single feature. The mining of point cloud features aims at describing environments more comprehensively and discriminatively, which founds a basis for our FastLCD method, a feasible and reliable loop closure detection algorithm. Algorithms also need to balance performance in terms of accuracy and efficiency. In addition, the majority of the existing methods rely on appropriate threshold setting, which needs to be adjusted on new datasets, while the proposed FastLCD could use a uniform threshold ignoring dataset changes.

3. The FastLCD approach

The input of this algorithm is raw point clouds without any transformation and projection. Multimodality features are extracted directly from raw 3D LiDAR scans. Then, they are concatenated into a discriminative global comprehensive descriptor, by which the computational and storage cost will reduce significantly. The descriptor ratio is calculated from a pair of comprehensive descriptors. The descriptor ratio will be checked by the pre-trained machine learning model without any priori pose information to obtain loop candidates. Then, a double-deck verification strategy comprising cross-validation and post-verification is implemented to ensure the final results' precision. The effective and efficient loop closure detection results will greatly enhance the localization and mapping tasks in application of robotics and self-driving. The algorithm architecture is shown in Fig. 1.

3.1. Multimodality feature extraction

The proposed FastLCD algorithm leverages discriminative global comprehensive descriptors encoded by multimodality features, which are extracted from each single 3D LiDAR scan, including statistics, geometry, planes, range histogram, and intensity histogram. The multimodality features are all invariant to rotation. Therefore, the FastLCD algorithm is also rotation-invariant.

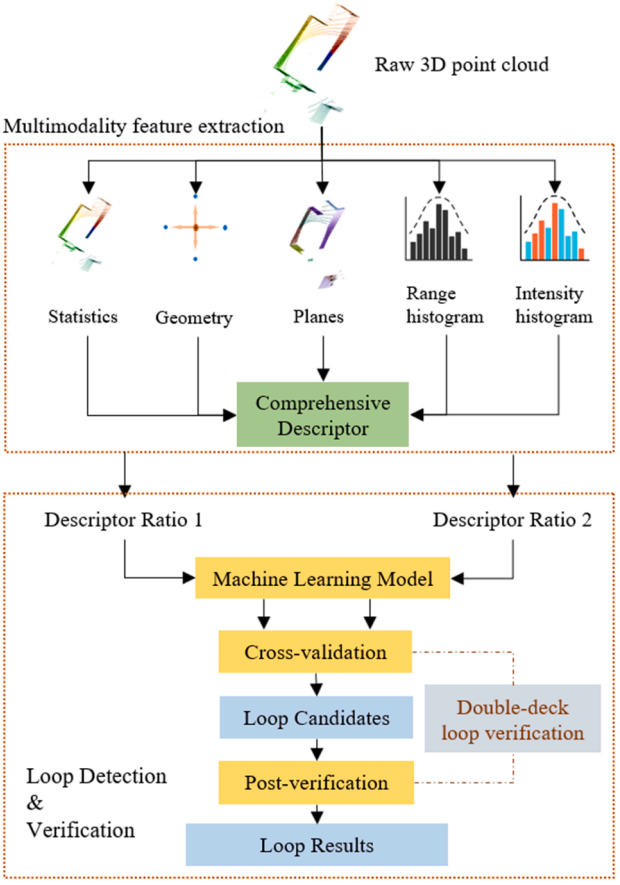


Fig. 1. Flowchart of FastLCD algorithm based on a comprehensive descriptor and machine learning. The workflow includes two main modules: multimodality feature extraction and loop detection & verfication.

Notations: Given a point cloud $P \in \mathbb{R}^{N\times 3}$, i is the point ID ($i \in [1, N]$) and (X_i, Y_i, Z_i) are the point coordinates. f_m^{id} is defined as the features, with id and m denoting the scan ID and the feature ID, respectively.

3.1.1. Statistics

The statistical features are computed using the nominal range distance, point coordinates, and point number. They can reflect the point distribution, which represents the surrounding environment intuitively.

The statistics include mean value of X, Y, Z respectively $(\overline{X}, \overline{Y}, \overline{Z})$, mean measuring distance $(\overline{r})$, maximum and minimum of distances (R_{max}, R_{min}), standard deviation (σ_R), coordinate of mass center (X, Y, Z), average distance between each point and mass center $(\overline{r})$, skewness (S_R), and kurtosis (γ_R).

3.1.2. Geometry features

Geometry features (F_g) describe the contextual information of each point on one scanline. Features of each point are computed with respect to the adjacent points. Geometry can describe local information of the point cloud. The geometry features are computed as follows:

Sum and standard deviation of distances between adjacent points $(D_{i,i+1})$ on the same scanline.

Sum and standard deviation of curvatures. A is defined as the area of a triangle formed by three points p_{i-1}, p_i, and p_{i+1}. The distances among the three points are $D_{i,i+1}$, $D_{i-1,i}$, and $D_{i-1,i+1}$. The curvature (C_i) at p_i is computed as

$$C_i = \frac{4A}{D_{i,i+1} D_{i-1,i} D_{i-1,i+1}} \tag{1}$$

where

$$A = \sqrt{s(s - D_{i,i+1})(s - D_{i-1,i})(s - D_{i-1,i+1})},$$
$$s = \frac{D_{i,i+1}D_{i-1,i}D_{i-1,i+1}}{2} \tag{2}$$

Mean value and standard deviation of range ratios on one LiDAR scan. Range ratio is defined as the specific value of ranging distances of adjacent point pairs.

Mean value and standard deviation of range differences. Range difference is defined as the difference value of ranging distances between adjacent point pairs.

3.1.3. Plane features

Planar surfaces are the most common geometric structure in man-made environments, such as ground, walls, ceilings, and furniture. These plane features can reflect the environment's structural information and complexity, as shown in Fig. 2. We define some structural features based on these plane features. The parallel planes in indoor environment always refer to ground, ceiling, or walls. Thus, the features can be designed as: (1) the number of plane features in a point cloud scan, (2) the maximum distance between any of the two parallel planes, and they are denoted as P_1 and P_2, (3) The maximum distance between any of the two parallel planes which are vertical to P_1 and P_2, (4) structure index: the ratio of two maximum distances, distance computed in (3) to distance of (2). The ratio can reflect the shape of the indoor space. Generally, the man-made environment is four-sided. The ratio ranges from 0 to 1. If it is close to 0, the environment is long and narrow, such as corridors and tunnels. Meanwhile, if it is close to 1, the environment is more like a square. In our approach, the plane feature extraction algorithm was proposed by Fan et al. (2019).

3.1.4. Range histogram

Each LiDAR scan is represented as an unstructured and uneven distributed 3D point cloud and always associated with a location. With the measuring distance defining the range between each point and sensor's center, the range histogram describes the distribution of the points in a scan and reflects the environment's size and complexity.

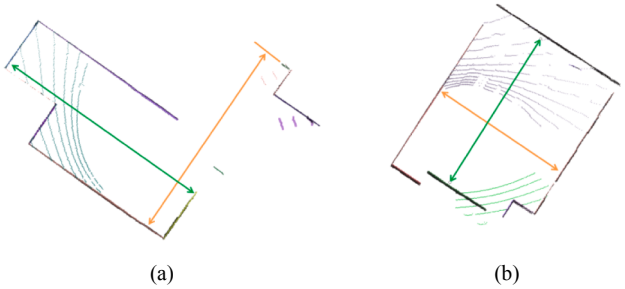
Assuming a bucket count b and a value range $R \in [R_{min}, R_{max}]$, we can divide $\{R\}$ into subintervals of size.

$$\Delta = \frac{1}{b}(R_{max} - R_{min}) \tag{3}$$

Each point falls in a corresponding bucket according to the value of R.

$$(R_{min} + k \bullet \Delta) < \mathrm{R} < [\mathrm{R_{min}} + (\mathrm{k} + 1) \bullet \Delta] \tag{4}$$

Then, the histogram for a point cloud scan P can be written as:



(a) (b)

Fig. 2. Plane features in indoor environments. The green arrow indicates the maximum distance between any of the two parallel planes, while in the vertital direction of those two planes, the maximum distance in denoted by orange arrows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$H_b = \left(h_b^0, \cdots, h_b^{b-1}\right), \quad h_b^k = count(p_b^k)/N \tag{5}$$

Theoretically, the number of valid measurement points in each LiDAR scan is relatively constant. While in practice, the number will show slight difference due to specular reflection in some surface types, or the slight vibrations of rotating devices. The normalization of point count ensures that the range histograms remain comparable under these unexpected conditions.

3.1.5. Laser intensity histogram

Similar to the range histograms, the laser intensity of the points in a scan can also be counted as a laser intensity histogram. In Fig. 3, if the two point clouds captured in the same environment, the intensity histograms are also similar. By contrast, the laser intensity histograms of different environments vary on trends and peaks.

The histogram reflects the measuring distance and object attributes. It should be indicated that intensity value of different laser scanners is defined differently. Thus, normalization of intensity value should be implemented for each point. The calculation method of intensity histogram is similar to range histogram.

3.2. Feature discriminative analysis.

Loop closure detection is to find whether the location has been revisited. To ensure the algorithm's detection performance, the features should be discriminative.

Statistics, geometry features, and range histograms are computed according to nominal distances and adjacent points on each scanline. It is obvious that the three modalities are all sensitive to locations and environments. A slight location perturbation of the laser scanner will cause changes of the ranging distances and scanlines. Even in the same environment, ranging histograms will be different when sensor's location changes, which indicates the location discriminative characteristic. The plane features describe the shape and complexity of the scenes, which will differ as scene changing. The intensity information is affected by the system and objects. The incident angles and materials of objects matter much on intensity values. The incident angles are influenced by sensor's relative locations in the environment. Besides, the materials of objects might be different in diverse environments. Thus, the intensity histogram feature is also discriminative to locations and environments.

3.3. Loop detection

After the discriminative global comprehensive descriptor ($F^{id1,id2}$) extracted, the descriptor ratios are computed by concatenating the specific values of multimodalities. Then, the descriptor ratios are organized as samples to be fed into machine learning models. Methods of calculating descriptor ratios varies for different modalities. For statistics (f_s), geometry features (f_g) and plane features (f_p), the elementwise specific values are computed as:

$$f^{id1,id2} = \begin{cases} f_m^{id1}/f_m^{id2} \\ \left(f_m^{id1}/f_m^{id2}\right)^{-1} \end{cases} \tag{6}$$

Each pair of LiDAR scans generate two samples due to the two calculation methods in formula (6). It should be indicated that a minimum value needs to be added to the denominator in case of *NaN* value. If two LiDAR scans are captured from the same environment, the values of f_m^{id1} and f_m^{id2} will be very similar, and the value of $f^{id1,id2}$ is close to 1.

As for range histogram and intensity histogram, correlation coefficients c_r and c_i are computed. Then, $F^{id1,id2}$ is computed by concatenating the specific values of each element.

$$F^{id1,id2} = f_z \oplus f_g \oplus f_p \oplus c_r \oplus c_i \tag{7}$$

For model training, a descriptor ratio and a binary label are

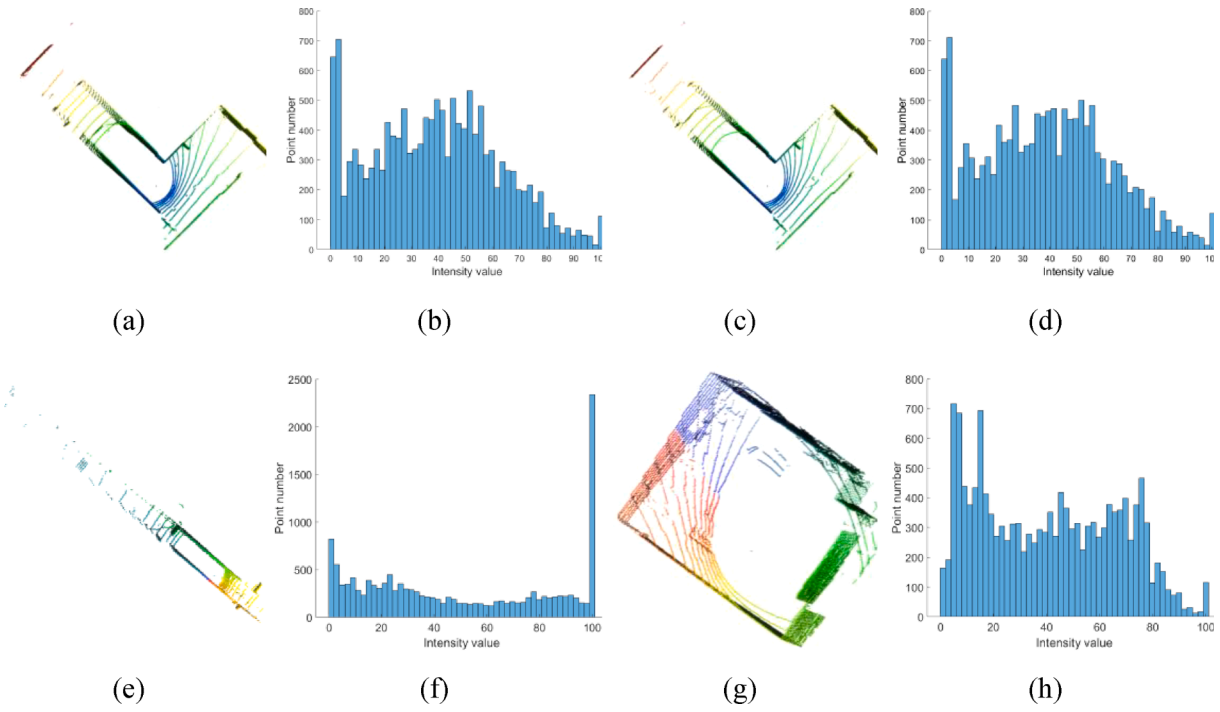


Fig. 3. Intensity histograms of different environments. (a) and (c) are at at different locations in the same corridor, (e) is a scan of a long corridor, (g) is capture in a office room. (b), (d), (f) and (h) are the intensity histogram of (a), (c), (e) and (g) respectively. (b) and (d) shows similar trends and peakness, while they are totally different from those in (f) and (h).

combined as a training sample $\{y, F_m^{id1,id2}\}$. If y is 0, the scan pair is not a loop closure, whereas the value of 1 means that the scan pair is a loop closure. Then, training samples are fed into the supervised learning model.

$$y = \begin{cases} 1 & positive \\ 0 & negative \end{cases} \tag{8}$$

Supervised learning model will learn from training samples to identify whether a scan pair is a loop candidate or not, meanwhile, the machine learning model could also provide a posterior probability of being detected as loops to estimate the reliability of the detection results.

3.4. Double-deck loop verification

Precision is the dominant indicator for evaluating loop closure detection results, due to wrong loop conditions might ruin the global map. Thus, to ensure the precision and reject false positives, a novel double-deck loop verification strategy is implemented. The loop verification contains two parts: cross-validation and post-verification.

Cross-validation: due to the two calculation methods of descriptor ratios in formula (6), each scan pair generates two samples. Then, if one of the two samples are identified as negative, this pair of laser scans will be rejected.

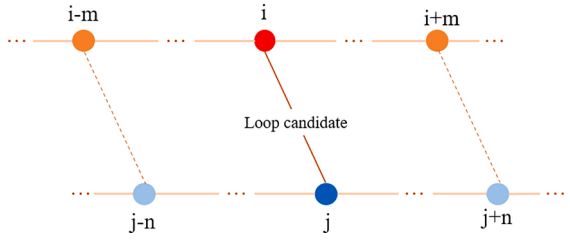


Fig. 4. Schematic of post-verification.

Post-verification: if a laser scan pair is identified as a loop candidate, it will be verified according to time-consistency and geometry-consistency. As shown in Fig. 4, in an appropriate time buffer, scan pairs are combined. Then, all these scan pairs are detected by the machine learning model to check whether they are positive or not. If they are positive, the scan i and j are verified as a loop closure.

4. Experiments

4.1. Data

We train and evaluate FastLCD algorithm on the in-house datasets and Mimap in SLAM 00 dataset (Wen et al., 2020)(Wang et al., 2018), as shown in Fig. 5.

In-house datasets: (a) The corridor A dataset is captured in a long and narrow corridor with some corners; (b) The small lecture room is an irregular lecture theatre with approximately 200 seats; (c) The large lecture room is a large irregular lecture theatre with approximately 400 seats; (d) The corridor B dataset has long and narrow corridors, corners, and a open small podium; (e) The office room dataset is captured in a square office rooms with some desks, chairs, computers, and laboratory equipment, which is much smaller than the two lecture rooms. To validate the learning model's generalization performance, the supervised machine learning model is only trained by the corridor A dataset, then tested on the other four datasets. These datasets are the most typical scenes in indoor environments. Most indoor environments are a combination of these scenes.

The in-house datasets are captured on the Hong Kong Polytechnic University campus using a backpack mobile mapping system (Fan et al., 2019). The laser scanners mounted on the backpack mobile platform are Velodyne's Puck LiDAR sensors.

Mimap in slam 00 dataset: the dataset is collected in a two-floor building scene, including data of individual rooms, non-enclosed loop corridors and stairs. The point cloud scans are captured by a Velodyne Ultra puck scanner.

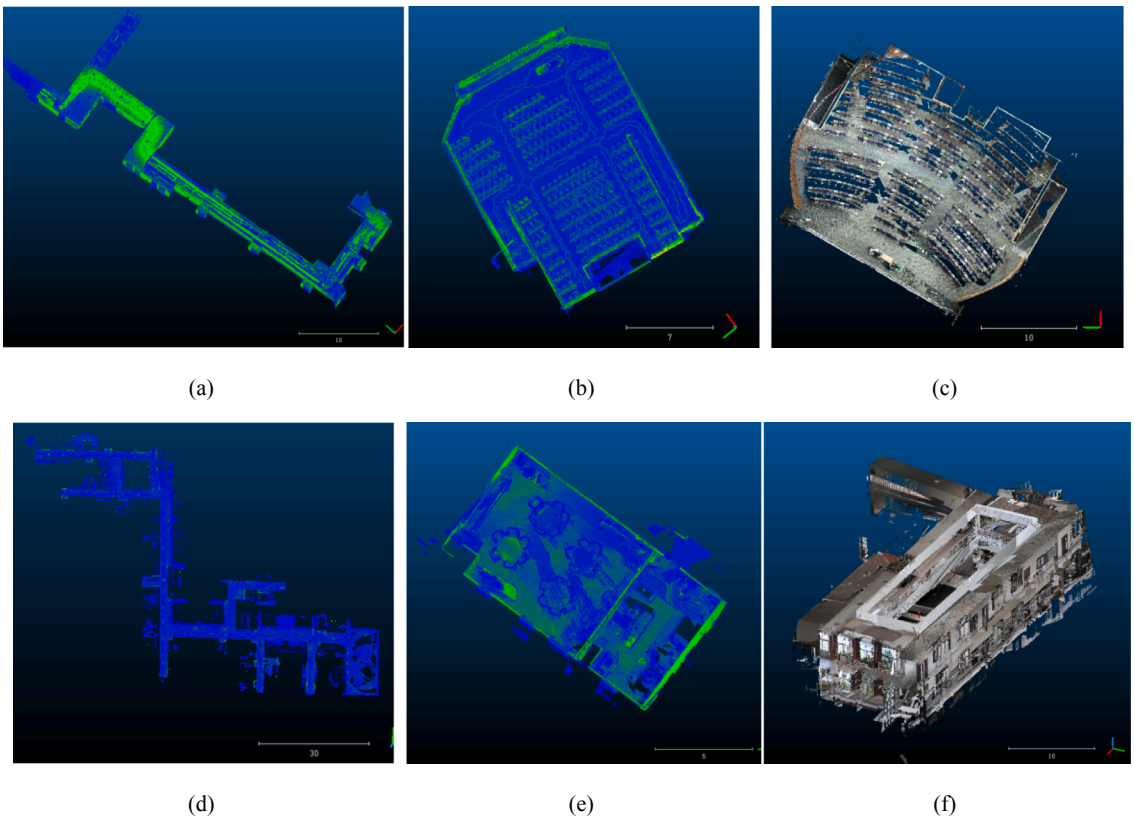


Fig. 5. (a) corridor A dataset, (b) small lecture room dataset, (c) large lecture room dataset, (d) corridor B dataset, (e) office room dataset, (f) mimap in slam 00 dataset.

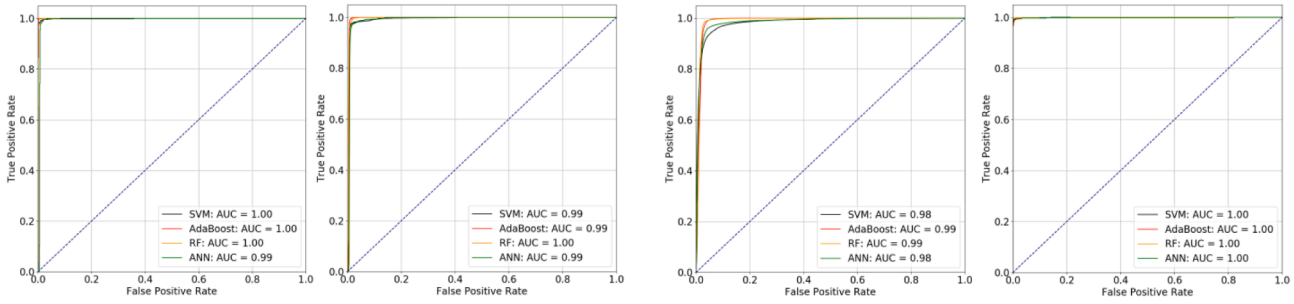
4.2. Supervised model selection

In this section, we compare the impact of different machine learning models on the algorithm's performance. Four popular machine learning models are adopted to conduct the comparative experiments, including AdaBoost, random forest (RF), support vector machine (SVM) and artificial neural network (ANN). It should be specified that back propagation neural network (BPNN), one of the ANN models, will be used. The results are shown in Fig. 6 and Table 1.

Totally the same training and test samples are used in the comparison experiment. To ensure the high precision, the four machine learning models are given the same and relatively rigorous threshold 0.8 for posterior probability. A unified threshold can better reflect the generalization performace of the models. According to Fig. 6, all the four models show excellent performances as AUCs are all close to 1. The F1-scores and AUCs evaluation results are shown in Table 1. On corridor B dataset, ANN model achieves the best performance, whereas in large lecture room dataset, small lecture room dataset and office room dataset, RF model stably outperforms SVM, AdaBoost and ANN models. Overall, RF model achieves realtively high F1-scores and AUCs on the four datasets with threshold equaling to 0.8, with high time efficiency. Thus, RF will be adopted in our FastLCD algorithm.

Table 1
F1-scores and AUCs of machine learning models on in-house datasets. The optimal results in each group of comparison experiments are shown in bold.

	Small lecture room		Large lecture room		Corridor B		Office room	
	F1	AUC	F1	AUC	F1	AUC	F1	AUC
SVM	0.98	**1.00**	0.94	0.99	0.83	0.98	0.99	**1.00**
AdaBoost	0.98	**1.00**	0.98	0.99	0.79	0.99	0.99	**1.00**
RF	**0.99**	**1.00**	**0.99**	**1.00**	0.92	0.99	**1.00**	**1.00**
ANN	0.77	0.99	0.74	0.99	**0.94**	0.98	0.96	**1.00**



(a) (b) (c) (d)

Fig. 6. ROC curves of different learning models on the in-house datasets. (a) small lecture room dataset, (b) large lecture room dataset, (c) corridor B dataset, (d) office room dataset.

4.3. Ablation studies

4.3.1. Feature selection

RF model has the ability to weight the importance of each feature element. Feature selection is significant to reduce data dimensions and save computation cost. Fig. 7 depicts the feature importance analysis result. The range histogram plays the dominant role among all feature elements. Then, the intensity histogram, skewness, and sum value of distances between adjacent points also show outstanding influences, while plane features hardly affect the results.

The feature ablation study results are demonstrated in Table 2. Ablation of plane features does not make noticeable impacts on results' precision and recall rate, which is consistent with the feature importance analysis results shown in Fig. 7. It should be emphasized that range histogram ablation totally ruined the algorithm's performance. The other four types of features all influenced more on recall rate, while precision remains stable without obvious loss.

Besides feature importance analysis, feature selection also relies on correlation analysis results. In this experiment, a chisquare test will be used to perform correlation analysis. After feature selection, the updated descriptor shows the best performance with the precision, recall rate and F1-score achieving 0.98, 0.92 and 0.95, respectively. Thus, the FastLCD adopts 15 features ultimately, including mean measuring distance, maximum of ranging distances, mean value of X coordinates, standard deviation of all ranging distances, mean distance between each point and mass center, kurtosis, skewness, sum and standard deviation of adjacent point distances, sum, and standard deviation of curvatures, mean value and standard deviation of all range differences, range histogram and intensity histogram.

4.3.2. Double-deck verification ablation study

In this section, we will study the impact of the double-deck verification step. The ablation experiment results are demonstrated in the Table 3, in which the precision and F1-score are compared. We can see the precision of the results increases to varying degrees after adding verification steps. However, due to the discriminative descriptor and excellent supervised learning model, even the double-deck verification step is ablated, the precision and F1-score still remain relatively stable without much loss. After all, the verification step indeed increases the precision of the results more or less, which aims to make the algorithm more robust.

Table 2
Result of feature ablation experiments.

	Precision	Recall	F1
Complete	0.98	0.87	0.92
Statistics ablation	0.98	0.55	0.70
Geometry features ablation	0.98	0.77	0.86
Plane features ablation	0.98	0.86	0.92
Range histogram ablation	0.82	0.06	0.12
Intensity histogram ablation	0.97	0.85	0.91
After feature selection	0.98	0.92	0.95

Table 3
Double-deck verification ablation experiment results.

	Without double-deck verification		Double-deck verification	
	F1	Precision	F1	Precision
Small lecture room	0.99	0.98	1.00	1.00
Large lecture room	0.96	0.94	0.99	0.99
Corridor B	0.95	0.98	0.95	0.98
Office room	0.99	1.00	1.00	1.00

4.4. Loop closure detection results

4.4.1. FastLCD results

In-house datasets: we compare our FastLCD algorithm with some state-of-the-art methods on the in-house indoor datasets. Comparison results are shown in Fig. 8 and Table 4, where M2DP (He et al., 2016), FastHistogram (Rohling et al., 2015), ScanContext (Kim and Kim, 2018), LiDAR Iris (Y. Wang et al., 2020), Yin (Yin et al., 2017) are adopted. We can find that the comparison algorithms are difficult to achieve stable performance on all the four datasets. It is obvious that FastLCD algorithm over-performs the state-of-the-art methods on the four datasets as the AUCs almost equal to 1. The specific F1-scores and AUCs are demonstrated in Table 4. Though on corridor B dataset, the F1-score and AUC both rank second with 0.95 and 0.99 respectively, on the other three datasets, FastLCD all obtains superior performance. Because the experiments are trained only on corridor A dataset, the FastLCD's superior results on the four datasets indicate the great generalization ability.

Mimap in slam 00 dataset: compared with the five algorithms in mimap in slam 00 dataset, FastLCD still outperforms stably, with 0.94 F1-score and 1.00 AUC. Yin method also shows the same great performance as our method. However, ScanContext and LiDAR Iris are not

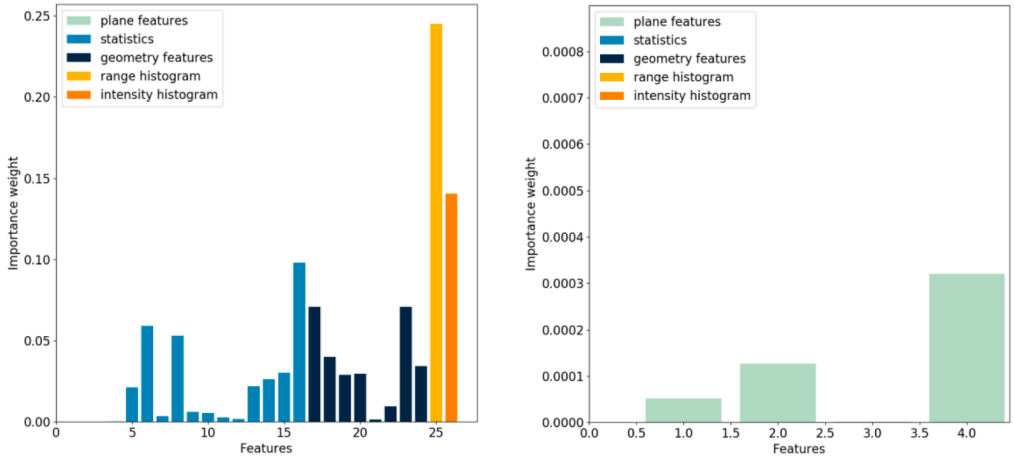


Fig. 7. Importance weight of features evaluated by RF model. (a) feature importance bar graph of all feature elements, (b) feature importance bar graph of plane features.

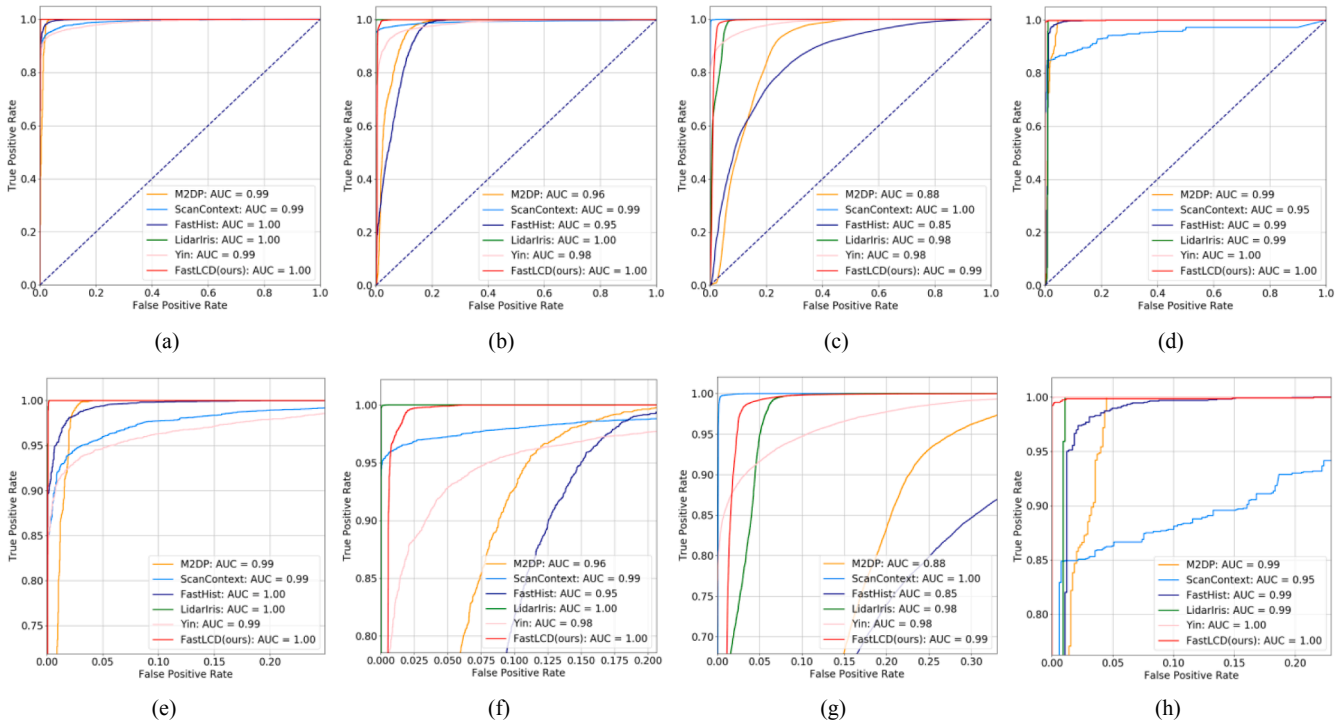


Fig. 8. (a), (b), (c), and (d) are ROC curves of FastLCD and state-of-the-art algorithms in small lecture room, large lecture room, corridor B, and office room, respectively. The pictures (e), (f), (g), (h) are the zoom in parts of (a), (b), (c), and (d).

Table 4
F1-scores and AUCs of FastLCD and state-of-the-art algorithms on in-house datasets. The optimal results in each group of comparison experiments are shown in bold.

	Small lecture room		Large lecture room		Corridor B		Office room	
	F1	AUC	F1	AUC	F1	AUC	F1	AUC
M2DP (He et al., 2016)	0.75	0.99	0.82	0.96	0.79	0.88	0.85	0.99
ScanContext (Kim and Kim, 2018)	0.75	0.99	0.87	0.99	0.86	**1.00**	0.83	0.95
FastHistogram (Rohling et al., 2015)	0.95	**1.00**	0.86	0.95	0.76	0.85	0.95	0.99
LiDAR Iris (Y. Wang et al., 2020)	0.9	**1.00**	0.95	**1.00**	**0.96**	0.98	0.99	0.99
Yin (Yin et al., 2017)	0.9	0.99	0.91	0.98	0.66	0.98	0.92	**1.00**
FastLCD (ours)	**0.99**	**1.00**	**0.99**	**1.00**	0.95	0.99	**1.00**	**1.00**

suitable for this scene, only getting 0.56 and 0.53 F1-score respectively, which are almost useless for indoor mapping. FastLCD and Yin method both need training samples. Totally the same training samples are used, while the training time of FastLCD is significantly shorter than that of Yin method. The other four methods all define distances to measure the differentiation between a pair of point clouds. According to the results in Tables 4 and 5, FastLCD shows superior performance to the five methods on in-house datasets and mimap in slam 00 dataset.

M2DP, FastHistogram, ScanContext and LiDAR Iris are four popular approaches based on advanced handcraft features, while Yin method uses a siamese CNN-based network, in which it is also trained by corridor A dataset. M2DP processes 3D point clouds by projecting them onto 2D planes, which will lose some 3D information. FastHistogram and Yin methods only use ranging distance histograms as features and set thresholds experimentally. However, FastHistogram uses ranging distance histograms directly, while Yin learns deep features by a siamese CNN-based network using histograms as input. ScanContext considers the geometry characteristics and transforms the 3D point cloud into a 2D feature image. Similarly, LiDAR Iris also generates 2D binary feature maps based on geometry information. All these methods considers part of the characteristics of 3D point cloud with information loss. Our method integrates geometry, ranging distances and intensity information to a discriminative global comprehensive descriptor, by which our algorithm exhibits superior and stable performance on difference datasets.

Table 5
F1-scores and AUCs of FastLCD and state-of-the-art algorithms on Mimap in slam 00 dataset. The best results in each group of comparison experiments is shown in bold.

	F1	AUC
M2DP (He et al., 2016)	0.91	0.97
ScanContext (Kim and Kim, 2018)	0.56	0.65
FastHistogram (Rohling et al., 2015)	0.81	0.9
LiDAR Iris (Y. Wang et al., 2020)	0.53	0.64
Yin (Yin et al., 2017)	**0.94**	**1.00**
FastLCD (ours)	**0.94**	**1.00**

Thresholds of the experiments are all set accordingly. We can find the comparison algorithms' performances all rely on the appropriate threshold setting, which brings in extra work, and makes the algorithm not robust. Besides, overdependence on thresholds would also weaken the generalization performance of the methods. By contrast, our algorithm shows stable results using a uniform threshold, which will be analyzed in the Section 4.4.2.

4.4.2. Separability and threshold sensitivity studies

If AUC equals 1, it is confirmed that there is an appropriate threshold making the model a perfect classifier. However, the defect of the

comparison methods is that the thresholds may differs on different datasets. Due to the comparison methods all depend on appropriate thresholds, much effort is needed to adjust the parameters and thresholds for a relatively good result. If the algorithm is sensitive to the threshold, it needs to be updated accordingly when a new dataset comes. Thus, this section is to study the threshold sensitivity of our algorithm.

As shown in Fig. 9, the bar graphs depict the posterior probability distribution of being recognized as loop closures. The X axis is the posterior probability, while the Y axis is the number of laser scan pairs to be detected. The color of the bars refers to the ground truth. The orange bars are the correct loops, while the blue bars are not loop closures.

Loop closure detection could be treated as a binary classification problem. We can find that the probability distribution complies with GMM. It is obvious that the probability distributions of correct loops on the four experiments are all concentrated in [0.8, 1], while probability of negative samples is concentrated in [0, 0.4]. The probability distribution shows a clear trend of U-shape canyon. It proves the superior separability and insensitivity to threshold of the proposed algorithm. Low threshold sensitivity makes it easy to determine the threshold. According to Fig. 9, we can find a uniform threshold 0.8 for posterior probability could be adopted on the four datasets.

According to the probability distribution, we can also analyze the reliability of the results. Due to the true positive probability's concentrated distribution in [0.8, 1], it is demonstrated that the results are highly reliable. Moreover, the posterior probability also can be used as the estimation of the loop closure detection results' reliability. Our method decides the threshold based on the posterior probability, while aforementioned state-of-the-art algorithms all use the distance metrics. Posterior probability will be more stable than distance metrics on different datasets.

4.5. Time efficiency

This experiment is conducted to compute the time efficiency of the proposed FastLCD approach. The algorithm includes three key steps—the comprehensive descriptor extraction, supervised learning model training, and loop detection & verification. Model training could be conducted offline. Thus, time efficiency experiment will focus on the comprehensive descriptor extraction and loop detection & verification. We test the time efficiency on a system equipped with an Intel i7-7700 CPU with 3.6 GHz running Windows 10 x64 operating system. It should be emphasized that the code has not been optimized by time

Table 6
Descriptor extraction time cost of FastLCD on in-house datasets.

Datasets	Data amount (scan)	Time cost in total (s)	Time cost (ms/scan)
Small lecture room	9545	790	82
Large lecture room	16,538	1474	89
Corridor B	19,801	1784	90
Office room	1948	166	85

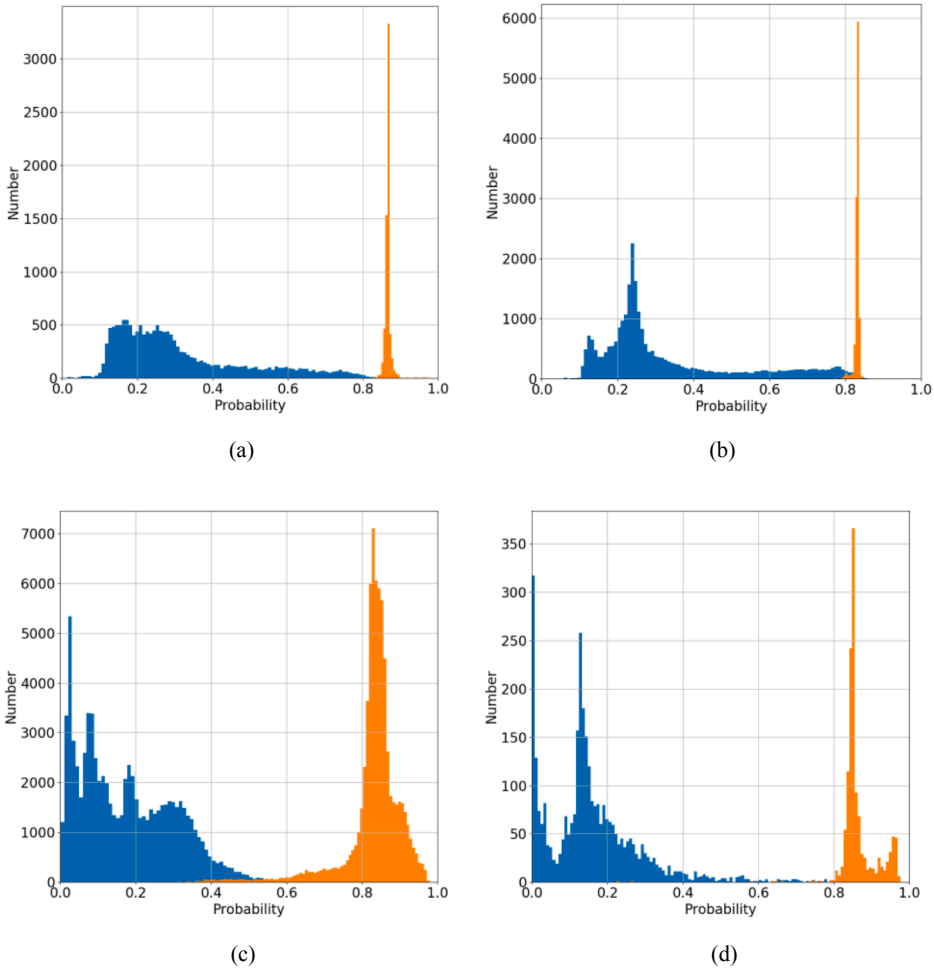


Fig. 9. The posterior probability distribution of being recognized as loops on the in-house datasets, (a), (b). (c) and (d) are the bar graph of probability distribution on small lecture room, large lecture room, corridor B and office room, respectively.

Table 7
Loop detection and verification time cost of FastLCD on in-house datasets.

Datasets	Data amount (laser pair)	With verification or not	Time cost in total (s)	Time cost (ms/detection)
Small lecture room	22,104	With	25.26	1.14
		Without	0.68	0.03
Large lecture room	38,246	With	41.12	1.08
		Without	1.19	0.03
Corridor B	148,536	With	176.00	1.18
		Without	4.59	0.03
Office room	4576	With	5.60	1.22
		Without	0.14	0.03

Table 8
Time cost (s) comparison of FastLCD and the state-of-the-art methods on in-house datasets.

	Small lecture room	Large lecture room	Corridor B	Office room
M2DP (He et al., 2016)	1097.68	1984.56	2574.13	214.28
ScanContext (Kim and Kim, 2018)	998.23	1730.00	2071.34	202.59
FastHistogram (Rohling et al., 2015)	1587.20	2756.33	3298.16	306.66
LiDAR Iris (Y. Wang et al., 2020)	1408.33	2439.55	2970.15	351.67
Yin (Yin et al., 2017)	1345.40	2234.17	2813.60	252.00
FastLCD	**815.26**	**1515.12**	**1960.00**	**171.60**

acceleration technologies, such as Compute Unified Device Architecture (CUDA) and the multithreading. The results are shown in Table 6.

As shown in Table 6, the time cost of feature extraction is stable on different datasets. About 85 ms is needed to process each LiDAR scan, in which around 15,000 points are stored. While in Table 7, compared with feature extraction, loop detection is much more time saving. Only about 1 ms is needed for each laser scan pair detection, which includes loop candidate determined by supervised learning model and double-deck verification. To check the time cost of the double-deck verification, we remove this step. Then, we can find the time efficiency increase sharply to 0.03 ms per detection. Compared with the state-of-the-art methods, our algorithm is the most time-efficient, with less than 100 ms to detect a loop.

Time efficiency comparison experiments are also performed. The results are shown in Table 8. The experiments use open-source code of the comparison algorithms on the aforementioned equipment. High time efficiency is a decisive factor for the feasibility of loop closure detection algorithm. All the algorithms in Table 8 show fast and efficient characteristics, while, obviously, FastLCD costs the least time among all comparison algorithms.

Overall, considering the computation cost and practical feasibility, it is not necessary to conduct loop closure detection when every single LiDAR scan is captured. Therefore, our FastLCD algorithm can realize real-time loop closure detection in SLAM, if the data capture frequency of the scanner is set appropriately.

5. Discussion and limitations

According to the experiment results of proposed FastLCD algorithm, some characteristics and limitations are summarized.

(1) *Precision, generalization performance, and time efficiency*. The proposed method shows superior performance to some state-of-the-art algorithms, being more accurate, reliable, and robust. The AUCs of FastLCD algorithm on the different datasets are all close to 1, which indicates that it has excellent generalization performance. As for the time cost, feature extraction and loop detection both shows high time efficiency. If the data capture frequency of the scanner is set appropriately, FastLCD can realize real-time loop closure detection in SLAM.

(2) *Feature importance*. The five types of features have varying degrees of influences. Range histogram plays the dominant role, while plane features impact a little. Statistics, intensity histogram and geometry features matter much on recall rate. According to feature importance analysis and correlation analysis, we select some significant feature elements into the comprehensive descriptors.

(3) *Separability and threshold sensitivity*. The posterior probability distribution of being recognized as loops almost perfect complies with GMM, which indicates the separability of FastLCD is excellent. The probability distributions of being detected as loops on different datasets all concentrate in the same and narrow range. It shows FastLCD is insensitive to threshold. Thus, the loop results detected is reliable and robust.

(4) *Limitations*. The FastLCD algorithm is designed for indoor environments. If the algorithm is extended to outdoor environments and large-scale scenes, some optimization and experiments should be conducted. Besides, the computation cost will grow sharply with the measuring distance and time increasing. Thus, some computation cost optimization and acceleration technologies could be adopted to make it more feasible and robust for SLAM.

6. Conclusions

This paper proposes a fast and compact loop closure detection method FastLCD based on comprehensive descriptors and machine learning. to achieve reliable and precise results using 3D point cloud for indoor LiDAR mobile mapping. FastLCD algorithm extracts multi-modality features from each single 3D LiDAR scan without any transformation and projection, to map a LiDAR scan into a discriminative comprehensive descriptor. A machine learning model with a double-deck loop verification strategy is used not only to identify loop closures without priori poses, but also to provide estimates of the reliability of detection results. Experiments show the algorithm can detect loop results reliably and precisely. The algorithm also shows great performance on separability, threshold insensitivity, and generalization. Besides, the high time efficiency makes it possible to realize real-time loop closure detection for SLAM in indoor mapping.

In the future, as the proposed approach is only designed in indoor environments for LiDAR mobile mapping, it can be extended to outdoor and large-scale environments. Furthermore, a deep learning model could be used to uncover some hidden features.

CRediT authorship contribution statement

Haodong Xiang: Conceptualization, Methodology, Writing – original draft. **Wenzhong Shi:** Supervision, Writing - review & editing, Funding acquisition. **Wenzheng Fan:** Writing - review & editing. **Pengxin Chen:** Validation. **Sheng Bao:** Data curation. **Mingyan Nie:** Data curation.

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.

Acknowledgment

This work was supported by The Hong Kong Polytechnic University [1-ZVN6, 4-BCF7]; The State Bureau of Surveying and Mapping, P.R.

China [1-ZVE8]; and Hong Kong Research Grants Council [T22-505/19-N].

References

Aldoma, A., Tombari, F., Rusu, R.B., Vincze, M., 2012. OUR-CVFH – Oriented, unique and repeatable clustered viewpoint feature histogram for object recognition and 6DOF pose estimation. In: Joint DAGM (German Association for Pattern Recognition) and OAGM Symposium, pp. 113–122. https://doi.org/10.1007/978-3-642-32717-9_12.

Angeli, A., Filliat, D., Doncieux, S., Meyer, J.-A., 2008. Fast and incremental method for loop-closure detection using bags of visual words. IEEE Trans. Robot. 24 (5), 1027–1037. https://doi.org/10.1109/TRO.2008.2004514.

Angelina, M., Gim, U., Lee, H., 2018. PointNetVLAD: deep point cloud based retrieval for large-scale place recognition supplementary material. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4470–4479. https://doi.org/10.1109/CVPR.2018.00470.

Arandjelovic, R., Gronat, P., Torii, A., Pajdla, T., Sivic, J., 2018. NetVLAD: CNN architecture for weakly supervised place recognition. IEEE Trans. Pattern Anal. Mach. Intell. 40 (6), 1437–1451. https://doi.org/10.1109/TPAMI.2017.2711011.

Chen, S., Wu, J., Wang, Y., Zhou, L., Lu, Q., Zhang, Y., 2019. Robust loop-closure detection with a learned illumination invariant representation for robot vSLAM. In: 2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM), pp. 342–347. https://doi.org/10.1109/ICARM.2019.8833730.

Chen, X., Läbe, T., Milioto, A., Röhling, T., Vysotska, O., Haag, A., Behley, J., Stachniss, C., 2020. OverlapNet: loop closing for LiDAR-based SLAM. In: Robotics: Science and Systems. https://doi.org/10.15607/rss.2020.xvi.009.

Cupec, R., Filko, D., Nyarko, E.K., 2019. Place recognition based on planar surfaces using multiple RGB-D Images taken from the same position. In: 2019 European Conference on Mobile Robots (ECMR), pp. 1–8. https://doi.org/10.1109/ECMR.2019.8870915.

Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S., 2007. Information-theoretic metric learning. In: 24th International Conference on Machine Learning(ICML), pp. 209–216. https://doi.org/10.1145/1273496.1273523.

Dube, R., Dugas, D., Stumm, E., Nieto, J., Siegwart, R., Cadena, C., 2017. SegMatch: Segment based place recognition in 3D point clouds. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 5266–5272. https://doi.org/10.1109/ICRA.2017.7989618.

Fan, W., Shi, W., Xiang, H., Ding, K., 2019. A novel method for plane extraction from low-resolution inhomogeneous point clouds and its application to a customized low-cost mobile mapping system. Remote Sens. 11, 2789. https://doi.org/10.3390/rs11232789.

Ge, X., Hu, H., Wu, B., Ding, K., 2019. Image-Guided Registration of Unordered Terrestrial Laser Scanning Point Clouds for Urban Scenes. IEEE Transactions on Geoscience and Remote Sensing 57 (11), 9264–9276. https://doi.org/10.1109/TGRS.2019.2925805.

Granström, K., Schön, T.B., 2010. Learning to close the loop from 3D point clouds. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2089–2095. https://doi.org/10.1109/IROS.2010.5651013.

He, L., Wang, X., Zhang, H., 2016. M2DP: a novel 3D point cloud descriptor and its application in loop closure detection. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 231–237. https://doi.org/10.1109/IROS.2016.7759060.

Hess, W., Kohler, D., Rapp, H., Andor, D., 2016. Real-time loop closure in 2D LIDAR SLAM. In: IEEE, pp. 1271–1278. https://doi.org/10.1109/ICRA.2016.7487258.

Himstedt, M., Frost, J., Hellbach, S., Bohme, H.J., Maehle, E., 2014. Large scale place recognition in 2D LIDAR scans using Geometrical Landmark Relations. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 5030–5035. https://doi.org/10.1109/IROS.2014.6943277.

Kim, G., Kim, A., 2018. Scan context: egocentric spatial descriptor for place recognition within 3D point cloud map. IEEE Int. Conf. Intell. Robot. Syst. 4802–4809. https://doi.org/10.1109/iros.2018.8593953.

Komorowski, J., 2021. MinkLoc3D: Point cloud based large-scale place recognition. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 1790–1799. https://doi.org/10.1109/WACV48630.2021.00183.

Labbe, M., Michaud, F., 2013. Appearance-based loop closure detection for online large-scale and long-term operation. IEEE Trans. Robot. 29 (3), 734–745. https://doi.org/10.1109/TRO.2013.2242375.

Labbé, M., Michaud, F., 2014. Online global loop closure detection for large-scale multi-session graph-based SLAM. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2661–2666. https://doi.org/10.1109/IROS.2014.6942926.

Lee, S., Jo, H.G., Cho, H.M., Kim, E., 2019. Visual loop closure detection over illumination change. In: 2019 16th International Conference on Ubiquitous Robots (UR), pp. 77–80. https://doi.org/10.1109/URAI.2019.8768627.

Li, J., Zhan, H., Chen, B.M., Reid, I., Lee, G.H., 2017. Deep learning for 2D scan matching and loop closure. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 763–768. https://doi.org/10.1109/IROS.2017.8202236.

Liu, Z., Suo, C., Zhou, S., Xu, F., Wei, H., Chen, W., Wang, H., Liang, X., Liu, Y.H., 2019a. SeqLPD: sequence matching enhanced loop-closure detection based on large-scale point cloud description for self-driving vehicles. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1218–1223. https://doi.org/10.1109/IROS40897.2019.8967875.

Liu, Z., Zhou, S., Suo, C., Yin, P., Chen, W., Wang, H., Li, H., Liu, Y., 2019b. LPD-Net: 3D point cloud learning for large-scale place recognition and environment analysis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 2831–2840. https://doi.org/10.1109/ICCV.2019.00292.

Luo, R.C., Ee, V.W.S., Hsieh, C.K., 2016. 3D point cloud based indoor mobile robot in 6-DoF pose localization using Fast Scene Recognition and Alignment approach. In: 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), pp. 470–475. https://doi.org/10.1109/MFI.2016.7849532.

Magnusson, M., Andreasson, H., Nüchter, A., Lilienthal, A.J., 2009. Automatic appearance-based loop detection from three-dimensional laser data using the normal distributions transform. J. F. Robot. 26, 892–914. https://doi.org/10.1002/rob.20314.

Memon, A.R., Wang, H., Hussain, A., 2020. Loop closure detection using supervised and unsupervised deep neural networks for monocular SLAM systems. Rob. Auton. Syst. 126, 103470. https://doi.org/10.1016/j.robot.2020.103470.

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017. PointNet: Deep learning on point sets for 3D classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 652–660. https://doi.org/10.1109/CVPR.2017.16.

Rohling, T., Mack, J., Schulz, D., 2015. A fast histogram-based similarity measure for detecting loop closures in 3-D LIDAR data. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 736–741. https://doi.org/10.1109/IROS.2015.7353454.

Rusu, R.B., Blodow, N., Beetz, M., 2009. Fast Point Feature Histograms (FPFH) for 3D registration. In: 2009 IEEE International Conference on Robotics and Automation, pp. 3212–3217. https://doi.org/10.1109/robot.2009.5152473.

Rusu, R.B., Marton, Z.C., Blodow, N., Beetz, M., 2008. Learning informative point classes for the acquisition of object model maps. In: 2008 10th International Conference on Control, Automation, Robotics and Vision, pp. 643–650. https://doi.org/10.1109/ICARCV.2008.4795593.

Shi, W., Ahmed, W., Li, N., Fan, W., Xiang, H., Wang, M., 2019. Semantic geometric modelling of unstructured indoor point cloud. ISPRS Int. J. Geo-Inform. 8, 9. https://doi.org/10.3390/ijgi8010009.

Siva, S., Nahman, Z., Zhang, H., 2020. Voxel-based representation learning for place recognition based on 3D point clouds. IEEE/RSJ Int. Conf. Intell. Robot. Syst. 8351–8357 https://doi.org/10.1109/IROS45743.2020.9340992.

Steder, B., Grisetti, G., Burgard, W., 2010. Robust place recognition for 3D range data based on point features. In: In: 2010 IEEE International Conference on Robotics and Automation, pp. 1400–1405. https://doi.org/10.1109/ROBOT.2010.5509401.

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. In: 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1249–1255. https://doi.org/10.1109/iros.2011.6094638.

Sun, Qi, Liu, Hongyan, He, Jun, Fan, Zhaoxin, Xiaoyong, Du., 2020. DAGC: Employing Dual Attention and Graph Convolution for Point Cloud based Place Recognition. In: Proceedings of the 2020 International Conference on Multimedia Retrieval (ICMR '20). Association for Computing Machinery, New York, NY, USA, pp. 224–232. https://doi.org/10.1145/3372278.3390693.

Sünderhauf, N., 2012. Robust Optimization for Simultaneous Localization and Mapping. Technischen Universitat Chemnitz.

Thrun, S., 2002. Robotic Mapping: A Survey. Carnegie Mellon University.

Wang, C., Hou, S., Wen, C., Gong, Z., Li, Q., Sun, X., Li, J., 2018. Semantic line framework-based indoor building modeling using backpacked laser scanning point cloud. ISPRS J. Photogramm. Remote Sens. 143, 150–166. https://doi.org/10.1016/j.isprsjprs.2018.03.025.

Wang, H., Wang, C., Xie, L., 2020a. Intensity scan context: coding intensity and geometry relations for loop closure detection. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 2095–2101. https://doi.org/10.1109/ICRA40945.2020.9196764.

Wang, Y., Sun, Z., Xu, C.Z., Sarma, S.E., Yang, J., Kong, H., 2020b. LiDAR iris for loop-closure detection. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5769–5775. https://doi.org/10.1109/IROS45743.2020.9341010.

Wen, C., Dai, Y., Xia, Y., Lian, Y., Tan, J., Wang, C., Li, J., 2020. Toward efficient 3-D colored mapping in GPS-/GNSS-denied environments. IEEE Geosci. Remote Sens. Lett. 17 (1), 147–151. https://doi.org/10.1109/LGRS.885910.1109/LGRS.2019.2916844.

Williams, B., Klein, G., Reid, I., 2011. Automatic relocalization and loop closing for real-time monocular SLAM. IEEE Trans. Pattern Anal. Mach. Intell. 33 (9), 1699–1712. https://doi.org/10.1109/TPAMI.2011.41.

Wohlkinger, W., Vincze, M., 2011. Ensemble of shape functions for 3D object classification. In: 2011 IEEE International Conference on Robotics and Biomimetics, pp. 2987–2992. https://doi.org/10.1109/ROBIO.2011.6181760.

Xia, Yan, Yusheng, Xu., Li, Shuang, Wang, Rui, Juan, Du., Cremers, Daniel, Stilla, Uwe, 2021. SOE-Net: A Self-Attention and Orientation Encoding Network for Point Cloud Based Place Recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11348–11357.

Yin, H., Ding, X., Tang, L., Wang, Y., Xiong, R., 2017. Efficient 3D LIDAR based loop closing using deep neural network. In: 2017 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 481–486. https://doi.org/10.1109/ROBIO.2017.8324463.

Zhang, W., Xiao, C., 2019. PCAN: 3D attention map learning using contextual information for point cloud based retrieval. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12436–12445. https://doi.org/10.1109/CVPR.2019.01272.

Zhuang, Y., Jiang, N., Hu, H., Yan, F., 2013. 3-D-laser-based scene measurement and place recognition for mobile robots in dynamic indoor environments. IEEE Trans. Instrum. Meas. 62 (2), 438–450. https://doi.org/10.1109/TIM.2012.2216475.

Zlot, R., Bosse, M., 2009. Place recognition using keypoint similasrities in 2D lidar maps. In: Experimental Robotics. Springer Tracts in Advanced Robotics, pp. 363–372. https://doi.org/10.1007/978-3-642-00196-3_42.