

# NDT-Map-Code: A 3D global descriptor for real-time loop closure detection in lidar SLAM

Lizhou Liao<sup>1,†</sup>, Wenlei Yan<sup>1,†</sup>, Li Sun<sup>2,3\*</sup>, Xinhui Bai<sup>2</sup>, Zhenxing You<sup>2</sup>, Hongyuan Yuan<sup>2</sup>, Chunyun Fu<sup>1,4</sup>

**Abstract**—Loop-closure detection, also known as place recognition, aiming to identify previously visited locations, is an essential component of a SLAM system. Existing research on lidar-based loop closure heavily relies on dense point cloud and 360 FOV lidars. This paper proposes an out-of-the-box NDT (Normal Distribution Transform) based global descriptor, NDT-Map-Code, designed for both on-road driving and underground valet parking scenarios. NDT-Map-Code can be directly extracted from the NDT map without the need for a dense point cloud, resulting in excellent scalability and low maintenance cost. The NDT representation is leveraged to identify representative patterns, which are further encoded according to their spatial location (bearing, range, and height). Experimental results on the NIO underground parking lot dataset and the KITTI dataset demonstrate that our method achieves significantly better performance compared to the state-of-the-art.

## I. INTRODUCTION

Loop-closure detection is the key technique for eliminating the long-term drift in large-scale mapping when the GPS is not available (e.g. automated valet parking). Moreover, SLAM methods for large-scale applications require good adaptations of using lightweight maps since the trend of mapping will be based on onboard computation and crowd-sourced data. Existing lidar-based loop closure detection methods [1]–[4] usually convert a frame of point cloud into a two-dimensional global descriptor through polar-coordinate ROI partitioning, and, the rotation invariance can be achieved by column-wise shifting. The main-stream methods have two limitations: firstly, the existing methods require detailed point-level geometry information hence dense point cloud map, and 360 FOV lidar scans are necessarily used for feature representation. This will boost the requirements for onboard storage and vehicle-cloud data transmission; secondly, in contrast to on-road driving scenarios, significant adaptation is required to deal with repetitive patterns, dynamic objects, and occlusion in structures for underground parking-lots localization and mapping.

This work was supported by the Chongqing Technology Innovation and Application Development Project under Grant CSTB2022TIAD-DEX0013.

<sup>†</sup>These two authors contribute equally to this work.

<sup>1</sup>College of Mechanical and Vehicle Engineering, Chongqing University, China. E-mail: liaolizhou@cqu.edu.cn, yanwenlei@stu.cqu.edu.cn, fuchunyun@cqu.edu.cn.

<sup>2</sup>The autonomous driving division, NIO, China. E-mail: kevin.sun, xinhui.bai, zhenxing.you, hongyuan.yuan@nio.com.

<sup>3</sup>Department of Computer Science, The University of Sheffield, UK

<sup>4</sup>State Key Laboratory of Mechanical Transmission for Advanced Equipment, Chongqing University, China.

\*The corresponding author.

This paper proposes a global descriptor, NDT-Map-Code (NDT-MC), that is highly complementary with scalable NDT point clouds built through a crowd-sourced mapping way. The proposed method is devised to discover and describe the structural landmarks in consideration of NDT cells' geometrical shape, entropy, and spatial context. Our intuition is to describe the place by 'what' landmarks at 'where'. To describe 'what', we classify the geometric shapes of NDT cells. The entropy of the chosen NDT cells is considered to shortlist effective geometric patterns. To describe 'where', we propose a polar-range-height-coordinate-based ROI partitioning. Instead of focusing on structures of the largest height, we divide the entire surrounding scene structure into multiple layers according to height. Afterward, both NDT's shape types and their heights are employed to formulate a multi-layer global descriptor. For front-view lidars, we employ lidar odometry to construct and maintain sub-maps. This approach can improve the limited co-visibility of the front-view-only LiDARs.

The main contributions of our approach are:

- A novel global descriptor, NDTMC, for both underground parking scenarios and on-road driving scenes is proposed. This descriptor is complementary to scalable NDT representation, which utilizes geometric and entropy patterns in NDT and represents of scene features through multi-layer shape and context encoding;
- Extensive experiments conducted on NIO underground parking-lot dataset (i.e. a collection of eight sequences in four parking lots), coupled with six sequences in the KITTI dataset, underpin the superiority of our method over state-of-the-art approaches;
- We made the proposed method, together with an integrated real-time full-SLAM system, which is named NDTMC-LIO-SAM, publicly accessible to the community, contributing to its potential benefits, and is subject to appropriate license agreements <https://github.com/SlamCabbage/NDTMC>.

## II. RELATED WORK

Existing research on loop closure includes lidar-based methods and vision-based methods. Lidar-based methods have received increasing attention due to their robustness to lightness and illumination variance. A number of lidar-based loop closure methods have been proposed in recent years. A stream of global descriptors can be constructed as a global statics of 3D local descriptors. Several keypoint detection methods have been used, such as 3D Sift [6], Link3D [7], 3D-SURF [8], BoW3D [9], SHOT [10], and Imaging-Lidar

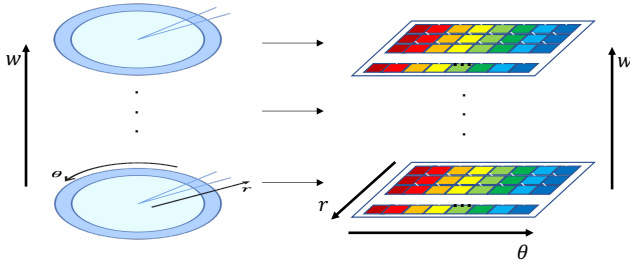


Fig. 1: Polar-range-height coordinates based ROI partitioning. We first divide the 3D space according to ring, sector, and height, and the corresponding polar-range-height coordinate axes, i.e.  $r$ ,  $\theta$ , and  $w$ , can be obtained. Afterward, the ring sector corresponding to different heights ( $w$ ) is transformed into a *Cartesian* coordinate system with  $\theta$  as the abscissa and  $r$  as the ordinate.

[11]. After that, the methods of voting [12] and Bag-of-Words [13] are used to combine these descriptors for loop closure.

Compared with local descriptor methods, global-based descriptors have better reproducibility in changing scenarios [14]. More recently, M2DP [15] generates a vector descriptor by projecting a point cloud onto a plane at multiple angles. Since this method's Principal Component Analysis (PCA) of its point cloud does not guarantee that the two point clouds can be robustly aligned, the M2DP descriptor lacks rotation invariance. Methods [1], [3]–[5] perform polar-coordinate ROI partitioning on the bird-eye viewpoint cloud to generate a two-dimensional descriptor, which can achieve rotation invariance between two descriptors through matrix shift. In [1], each bin of the feature matrix retains the maximum height to represent all points in the bin; [3] and [4] include the semantic labels and maximum intensity values respectively, apart from height values; Scan Context++ [5] builds upon the work of Scan Context [1], specifically tailored for urban road scenarios. Assuming minimal roll and pitch motions, the proposed descriptor offers a robust and generalized representation that is resilient to both rotational (yaw) and translational movements. Lidar Iris [2] divides the vertical FOV into multiple layers according to the pitch angle of the point, and each layer uses the same polar-range coordinates partition as [1]. The value of a bin is assigned 0 or 1 depending on whether it is occupied or not. Finally, binary-encoded of all bins with different layers and the same polar-range partition to get the final Lidar Iris descriptor. The core idea of the Normal Distribution Descriptor (NDD) [16] method is to map point cloud data to the Range-Polar coordinate system and calculate the mean and covariance of points in the bins in each coordinate system. Through these means and covariances, NDD calculates two important descriptors, namely probability density score and entropy, which have different properties. Finally, these two descriptors are fused into a complete NDD descriptor.

Recently, deep learning methods [14], [17]–[25] have been used to learn feature descriptors, and these methods have shown significant performance improvements compared to

traditional methods. Disco [17] introduces a highly interpretable network architecture that extracts rotation-invariant place signatures in the frequency domain, resulting in an effective performance in benchmarking. However, learning-based methods show limited generalizability for novel scenes, and high computation resources are required for deployment.

From the literature, existing methods have three limitations: 1) firstly, most of the above methods are devised based on 360-degree lidars, and it is likely to fail with a front-view lidar; 2) secondly, due to limitations of public datasets, existing descriptors are primarily designed for outdoor on-road driving rather than underground parking scenes; 3) finally, existing descriptors require dense point cloud maps for feature extraction, which does not suit lightweight maps obtained by crowd-sourced mapping.

### III. METHODOLOGY

#### A. NDT represent point cloud

NDT is used to represent the point cloud due to its scalability in mapping large-scale scenes, as well as its advantages in noise and outlier removal and reducing the number of points to improve processing speed. We collected data from four parking lots, and the storage space required for NDT maps at a 2m resolution is only 0.36% of that needed for the original point clouds. These 2m resolution NDT maps are generated by dividing the point cloud into uniformly distributed 3D grids and estimating a multivariate Gaussian for each cell. This approach manages a Gaussian distribution of 3D points within each voxel, rather than individual points. A one-pass algorithm is used to calculate the mean and covariance:

$$(\bar{x}_i, \bar{y}_i, \bar{z}_i) = (\bar{x}_{i-1}, \bar{y}_{i-1}, \bar{z}_{i-1}) + \frac{1}{i}(x_i - \bar{x}_{i-1}, y_i - \bar{y}_{i-1}, z_i - \bar{z}_{i-1}) \quad (1)$$

$$C_i = C_{i-1} + \frac{i-1}{i}(x_i - \bar{x}_{i-1}, y_i - \bar{y}_{i-1}, z_i - \bar{z}_{i-1})^T \cdot (x_i - \bar{x}_{i-1}, y_i - \bar{y}_{i-1}, z_i - \bar{z}_{i-1}) \quad (2)$$

$$Cov_n = \frac{C_n}{n} \quad (3)$$

where  $(\bar{x}_i, \bar{y}_i, \bar{z}_i)$  refers to the mean value of the previous  $i$  points in the same Cell, and  $n$  is the total number of points,  $C$  is a matrix of  $3 \times 3$ , and  $Cov_n$  represents the covariance matrix of  $n$  points.

TABLE I: Comparison of storage space size of point cloud types

	Raw Point Cloud (MB)	NDT map with 2 m resolution (MB)
Rongke	2252.8	9.1
Yinwang	889.7	1.7
Yinzuo	1223.7	4.8
Lixiangguoji	1331.2	4.9

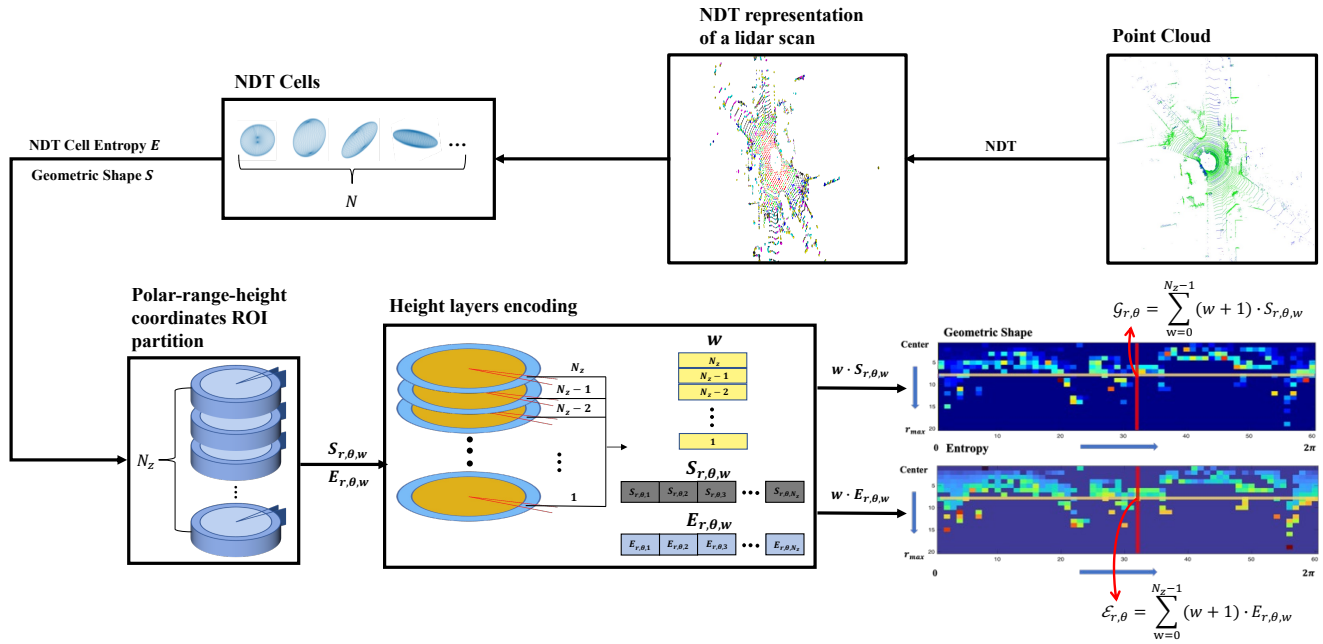


Fig. 2: The overview of NDT-MC. The construction of NDT-MC involves the following steps: 1. Converting point clouds into NDT representation. 2. Calculating geometric values and entropy for each NDT cell. 3. Computing the coordinates of each NDT cell in the Polar-Range-Height coordinate system based on its mean value. 4. Constructing the proposed descriptor using a strengthened height-layer encoding with geometric and entropy components.

### B. Polar-range-height coordinates ROI partition

Similar to Scan-Context-like methods, polar-range coordinates ROI partition is used to divide 3D space into rings and sectors:

$$r_k = \max \left\{ r_k \in \mathbb{Z} \mid r_k \leq \frac{\sqrt{x_k^2 + y_k^2}}{L_r}, 0 \leq \sqrt{x_k^2 + y_k^2} \leq R \right\} \quad (4)$$

$$\Theta \left( \begin{matrix} y_k \\ x_k \end{matrix} \right) = \begin{cases} \arctan \frac{y_k}{x_k}, & 0 < x_k, 0 \leq y_k \\ \arctan \frac{y_k}{x_k} + \pi, & x_k < 0 \\ \arctan \frac{y_k}{x_k} + 2\pi, & 0 < x_k, y_k < 0 \end{cases} \quad (5)$$

$$\theta_k = \max \left\{ \theta_k \in \mathbb{Z} \mid \theta_k \leq \frac{\Theta \left( \begin{matrix} y_k \\ x_k \end{matrix} \right)}{L_\theta}, 0 \leq \sqrt{x_k^2 + y_k^2} \leq R \right\} \quad (6)$$

where  $L_r$  and  $L_\theta$  represent the radial resolution and angular resolution of polar coordinates respectively and  $R$  represents the maximum radial distance allowed.  $\theta_k$  and  $r_k$  represent the polar and range coordinates indices of  $p_k$ , whose maximum values are  $N_\theta$  and  $N_r$  respectively.

Existing methods focus on describing the buildings for outdoor localization. Specifically, each bin of [1] retains the maximal height value to represent the point cloud in the entire ROI bin; [3] and [4] include the most semantic labels and the maximum intensity value, respectively.

In underground parking lots, the presence of dynamic objects, such as parked cars, can significantly impact scan-context-like approaches. Moreover, due to the equivalent height of ceiling, maximal-height-based descriptors show

limited effectiveness in discriminating between different locations.

To address these challenges, as illustrated in Fig. 1, we propose an alternative approach to dividing the scene  $p_k(x_k, y_k, z_k)$  into multiple layers based on height values with respect to the vehicle's base link. Unlike the method proposed in [2], we do not use pitch angles for vertical partition. Instead, we segment the scene into layers according to the height values.

$$w_k = \max \left\{ w_k \in \mathbb{Z} \mid w_k \leq \frac{z_k}{L_z}, 0 \leq z_k \leq Z \right\} \quad (7)$$

Here,  $\mathbb{Z}$  means natural number,  $L_z$  represents the height of each layer,  $Z$  refers to the maximal  $z$  value of truncated points  $[0, Z]$ , and  $w_k$  represents the layer index to which  $p_k$  belongs, whose maximum value is  $N_w$ .

### C. Classification of NDT shapes and calculation of entropy

Once the NDT point cloud is obtained, our approach analyzes each NDT cell by categorization of explicit geometric shape types, coupled with calculating the entropy of points within each NDT cell.

Our approach utilizes explicit categories to interpret the shape of each cell. For a given NDT cell, we calculate the sorted eigenvalues  $e_1 > e_2 > e_3$  from the covariance matrix, together with their corresponding eigenvectors  $v_1, v_2, v_3$ . If  $e_1 \gg e_2 \approx e_3$ , the shape of the cell will be classified as a straight line. Conversely, if  $e_1 \approx e_2 \gg e_3$ , the cell will be considered a plane. However, this shape indicator

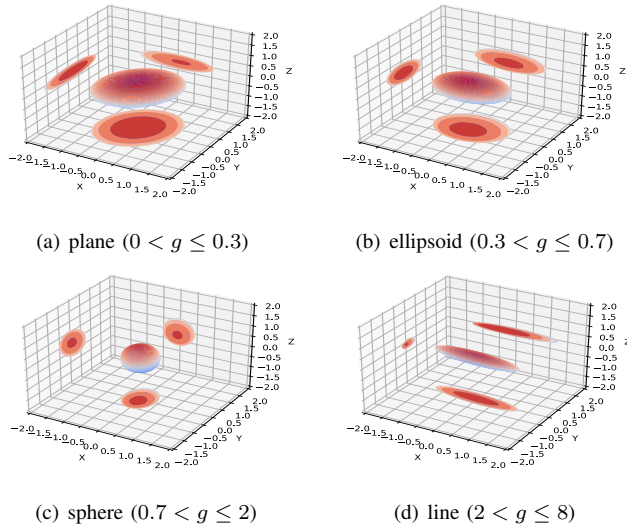


Fig. 3: The geometrical shape of NDT cells corresponding to different  $g$  values. Four subfigures illustrate representative shapes for corresponding  $g$  values, namely plane, ellipsoid, sphere, and line. In each subfigure, a 3D NDT cell is shown along with a 2D projection in the X, Y, and Z directions.

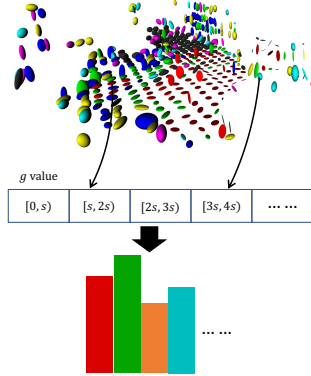


Fig. 4: Construction of geometric key.  $s$  is a segmented parameter for the  $g$  value.

requires two thresholds: 1) the ratio between  $e_1$  and  $e_2$ , and 2) the ratio between  $e_2$  and  $e_3$ . Instead of employing multiple thresholds, we propose an integrated one-dimensional NDT shape classification index:

$$g = \frac{e_1 \cdot e_3}{(e_2)^2} \quad (8)$$

Therefore, by employing a straightforward thresholding strategy on  $g$ , we can effectively classify the geometric shapes. As illustrated in Fig. 3,  $g$  enables the identification of four distinct shape categories: plane, ellipsoid, sphere, and line. To describe these explicit shape categories, we map the value of  $g$  to different shapes using the following formula:

$$S = \left\{ S \in \mathbb{Z} \mid S = \left\lceil \frac{g}{s} \right\rceil, 0 \leq g \leq g_{max} \right\} \quad (9)$$

Here,  $g_{max}$ ,  $s$ , and  $S$  denote the maximum  $g$  value selected, the segmented value we choose, and the geometric value, respectively. Taking inspiration from NDD [16], we consider

each NDT cell as a normal distribution characterized by a mean value  $\mu$  and a covariance matrix  $\Sigma$ . Consequently, we utilize entropy, a statistical measure associated with the normal distribution, to describe the NDT cell:

$$E = \frac{N}{2} (\log 2\pi + 1) + \frac{1}{2} \log |\Sigma| \quad (10)$$

where  $|\cdot|$  represents matrix determinant.  $N$  represents the dimension of the vector space, and in this particular case, the value of  $N$  is 3 because the points involved are in three-dimensional space.

Then, the Polar-Range-Height ROI partition  $r$ ,  $\theta$ ,  $w$ , geometric value  $S$ , and entropy  $E$  of each NDT cell can be obtained. Due to the variance of spatial partition between the voxel grid of NDT and ROI bins, one ROI bin likely contains multiple NDT cells. The majority value of geometric values is used for  $S$ , and sum-pooling is applied for the entropy  $E$ . The proposed descriptor NDT-MC integrates both geometric value and entropy features.

#### D. Encoding

To adapt to both indoor and outdoor scenarios, NDT-MC is devised by combining geometric values  $S_{r,\theta,w}$  and entropy  $E_{r,\theta,w}$ . Specifically, NDT-MC can be obtained by calculating  $\mathcal{G}$  and  $\mathcal{E}$  for each Region of Interest (ROI):

$$\mathcal{G}_{r,\theta} = \sum_{w=0}^{N_z-1} (w+1) \cdot S_{r,\theta,w} \quad (11)$$

$$\mathcal{E}_{r,\theta} = \sum_{w=0}^{N_z-1} (w+1) \cdot E_{r,\theta,w} \quad (12)$$

where  $N_z$  represents the number of selected height layers.

While the NDT-Map code is partly inspired by NDD [16], key differences exist. NDD calculates entropy by dividing 3D space using a polar-range partition and computing the covariance matrix within each partition. In contrast, our method uses uniformly partitioned NDT maps, offering advantages in dynamic environments like underground parking garages. Specifically, our non-destructive voxel grid manages point cloud features at different heights, and we weight the descriptor encoding by height, reducing the negative impact of dynamic vehicles. Additionally, our retrieval strategy uses a ‘coarse-to-fine’ approach with a ‘geometry key’ descriptor, compressing the original 2D descriptor into a 1D form, unlike NDD’s flattened method. These differences make our method more robust and efficient in loop detection by focusing on geometric features and employing an efficient retrieval strategy.

## IV. DESCRIPTORS MATCHING

### A. Fast Matching

**Construction of geometric key (GK):** Instead of using complete descriptors, we propose a histogram-based key called *geometric key*, to reduce the computation in similarity matching of descriptors. As shown in Fig. 4, we discrete the continuous  $g$  values of NDT representations into columns with an increment of 1. This process results in a  $1 \times N_s$

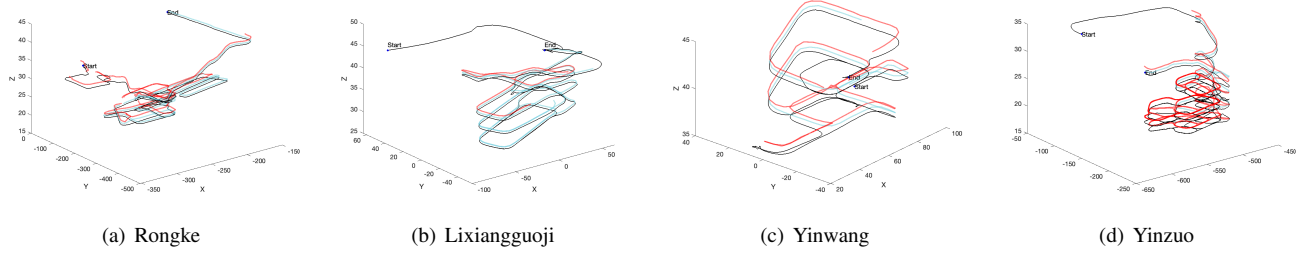


Fig. 5: A visualization of trajectories of the experimental dataset. Each sub-figure has three trajectories of different colors, in which blue, red, and black represent the trajectories of test data 1, test data 2, and the database respectively. Trajectories of the same underground parking-lot were collected at different times over different days.

vector, where  $N_s$  represents the maximum value of the selected geometric value  $S$ .

**Fast Retrieval based on kd-tree:** By using the geometry key, the time complexity of kd-tree construction can be reduced from  $O(nN_rN_\theta \log(n))$  to  $O(nN_s \log(n))$ , given the size of geometry  $N_s$  is much smaller than both  $N_r$  and  $N_\theta$ . Once the kd-tree is built, we undergo the k-nearest neighbor search for each query.

**Construction of Sector Key:** If a candidate keyframe satisfies the adaptive distance threshold, we compare its descriptor with that of the current frame using a similarity score. As mentioned earlier, rotation invariance is achieved by performing column shifts. However, iterating all columns will inevitably increase the computational complexity. Following *ScanContext*, we compute the column-wise average of the descriptor to obtain a  $1 \times N_\theta$  vector called the *sector key*. Sector keys simplify column shift approximation calculation, and descriptor matching is performed in the vicinity of the estimated column shift.

### B. Descriptors matching

Our approach uses correlation coefficient as the similarity metric.

Given two descriptors  $D_q = \{c_q^0, c_q^1, \dots, c_q^{N_\theta}\}$ ,  $D_c = \{c_c^0, c_c^1, \dots, c_c^{N_\theta}\}$ ,  $c_q^i$  and  $c_c^i$  represent the  $i^{th}$  column of  $D_q$  and  $D_c$  respectively.

Given a query descriptor  $D_q$  to  $\{c_q^0, c_q^1, \dots, c_q^{N_\theta}, c_q^0, c_q^1, \dots, c_q^{N_\theta-1}\}$ , The calculation method of correlation coefficient between the two descriptors is:

$$g_k(D_q^k, D_c) = 1 - \frac{1}{N_\theta + 1} \sum_{i=0}^{N_\theta} \left( \frac{(c_q^{i+k} - \bar{Q}) \cdot (c_c^i - \bar{C})}{\|c_q^{i+k} - \bar{Q}\| \cdot \|c_c^i - \bar{C}\|} \right) \quad (13)$$

Where  $\bar{C}$  and  $\bar{Q}$  represent the mean of all elements of  $D_c$  and  $D_q^k$  respectively.

Then, column-shifting is employed to find the appropriate yaw angle to match:

$$g_s(D_q, D_c) = \min (g_k(D_q^k, D_c)) \quad (14)$$

$$k \in \{0, 1, \dots, N_\theta - 1\} \quad (15)$$

where  $D_q^k$  means  $D_q$  shifted by  $k^{th}$ . We determine the similarity between two frame descriptors by calculating the minimum similarity distance obtained through column displacement.

## V. EXPERIMENT EVALUATION

### A. Dataset for experiments

Our experiments have two distinct scenarios. We collect a dataset for underground parking lots localization (NIO underground parking-lot dataset), while the widely-used KITTI dataset [26] is also used for the evaluation of loop closure detection for on-road driving.

**NIO underground parking-lot dataset.** This dataset is collected in real-world underground parking lots using the ET7 model of NIO's mass-produced cars, equipped with the Innovusion Falcon LiDAR(that of a FoV of  $120^\circ \times 25^\circ$ ). As shown in Fig. 5, the dataset consists of four different parking lots, located in the basements of four commercial malls, i.e. Rongke, Lixiangguoji, Yinwang, and Yinzuo. A NovAtel IMU-ISA-100C system is used to generate ground truth trajectories for loop closures. A place is defined as a submap built by a trajectory segment of 4m. For each underground parking lot, the session with the largest map area (with the largest number of places), is selected as the database. Other sessions are used for testing. Because the pipeline is designed for crowd-sourced mapping, NDT representation is used instead of raw point cloud map. If the distance between the query pose and database pose is within the range of 4m on the x-y plane and 2m on the z-axis to a database pose, it will be considered as a true positive pair. The numbers of loop closures in test data 1 and test data 2 at Rongke are 439 and 415 respectively. Similarly, the numbers of loop closures at Lixiangguoji, Yinwang, and Yinzuo are 310 and 91, 94 and 142, 144 and 454. In Fig. 5, the database and testing trajectories of the four experiments are shown.

**KITTI dataset.** The sequence data in the KITTI dataset is used for this evaluation, where point cloud data is acquired using the Velodyne HDL-64E lidar sensor. Our experiment follows the same setting with NDD [16].

TABLE II: The comparison of the F1 score and extended precision results on the NIO dataset.

Method	Rongke1	Rongke2	Lixiangguoji1	Lixiangguoji2	Yinwang1	Yinwang2	Yinzuo1	Yinzuo2
NDT-MC	<b>0.831/0.521</b>	<b>0.778/0.553</b>	<b>0.894/0.726</b>	<b>0.882/0.509</b>	<b>0.912/0.744</b>	<b>0.927/0.845</b>	<b>0.923/0.628</b>	<b>0.933/0.702</b>
Scan Context	0.551/—	0.360/—	0.896/—	0.833/—	0.747/—	0.855/—	0.418/—	0.913/—
Lidar Iris	0.641/0.362	0.743/0.518	0.867/0.566	0.878/0.518	0.754/0.599	0.861/0.525	0.743/0.415	0.862/0.672
STD	0.171/—	0.061/—	0.176/—	0.142/0.511	0.139/—	0.156/—	0.054/0.503	0.136/0.501
NDT-Hist	0.116/—	0.114/—	0.377/—	0.330/—	0.321/—	0.258/—	0.296/—	0.378/—
NDD	0.611/—	0.276/—	0.852/0.635	0.866/0.507	0.881/0.725	0.689/0.399	0.843/0.615	0.879/0.434

TABLE III: The comparison of the F1 score and extended precision results on the KITTI dataset.

Method	00	02	05	06	07	08
SC	0.924/0.891	0.690/0.516	0.859/0.902	0.932/0.982	0.482/0.630	0.608/0.667
ISC	0.856/0.737	0.675/0.510	0.847/0.813	0.937/0.921	0.506/0.634	0.719/0.710
IRIS	0.873/0.909	0.813/ <b>0.860</b>	0.922/0.925	0.936/0.971	0.585/0.710	0.534/0.665
M2DP	0.885/0.911	0.616/0.500	0.802/0.799	0.945/0.920	0.515/0.589	0.022/0.500
OverlapTransformer	0.915/0.842	0.801/0.646	0.853/0.839	0.948/0.915	0.438/0.520	0.375/—
STD	0.544/0.549	0.394/0.505	0.734/0.563	0.897/0.578	<b>0.756/0.639</b>	0.603/0.508
NDD <sup>5</sup>	0.943/ <b>0.963</b>	0.851/0.592	0.947/0.941	0.989/0.976	0.659/ <b>0.713</b>	<b>0.851/0.661</b>
NDD [16]	0.943/0.963	0.846/0.710	0.945/0.934	0.996/0.998	0.644/0.733	0.896/0.904
NDT-MC	<b>0.954/0.942</b>	<b>0.871/0.854</b>	<b>0.952/0.949</b>	<b>0.993/0.993</b>	0.615/ <b>0.713</b>	0.736/ <b>0.752</b>

TABLE IV: The runtime performance on KITTI.

Method	Descriptor Extraction(ms)	Query(ms)
STD	10.214	13.399
SC	1.199	1.798
IRIS	5.922	1023.213
NDT-MC	<b>0.116</b>	<b>0.161</b>

### B. Experiment on NIO underground parking-lot dataset.

In the experiment, the database frames are extracted by cropping the NDT submap across the mapping trajectories. The testing observation is extracted from the local NDT map built on the fly by a Lidar mapper. There are three sessions of data were collected, one for the database and the other for testing. For each query observation, Eq. 13 and Eq. 14 are used to retrieve the most similar location from the database. Inter-session poses within 4 meters (with z differences not exceeding 2m) will be considered closed loops. In this experiment, the dense point cloud is not applicable. We use the mean values of NDT cells as the input of SOTA methods.

The proposed method is evaluated on eight testing datasets collected in four underground parking lots. For comparison, state-of-the-art place recognition methods are implemented, i.e., Scan Context [1], Lidar Iris [2], NDD [16], STD [27] and NDT-Histograms [28]. Since Scan Context-like methods are not designed for underground parking lots, we make the following adaptations: points below 2m w.r.t. lidar coordinate system are used for descriptor representation. For Scan Context and Lidar Iris, the mean values of NDT Submap as the input. For a fair comparison, the proposed method, Scan Context, and Lidar Iris are set with the same parameters, i.e.,  $N_r=40$ ,  $N_\theta=60$ ,  $Z=3m$ ,  $L_z=1m$  and  $R=80m$ . Since NDD and STD require the raw point cloud as input, both the observation and the database use the raw point cloud submap stitched by LIO. We use two widely-used evaluation metrics as NDD, namely F1 score, and Extended Precision (EP).

In the NIO parking lot dataset, performance evaluations of

methods such as M2DP, OverlapTransformer, and Intensity Scan Context have not been conducted. M2DP exhibits a low Top-1 recall in scenes like Rongke1, Rongke2, and Liyangguoji1 due to its reliance on dense point clouds, while the sparse NDT data in our dataset adversely affects its performance. The OverlapTransformer method, relying on distance images and designed to run on CPU due to the sparsity of NDT point cloud data, is consequently expected to underperform on the NIO dataset. Lastly, the Intensity Scan Context method relies on specific intensity values, and variations in LiDAR sensor intensity values across different hardware platforms may impact its performance.

### C. Experiment on KITTI.

In this experiment, we define a true positive detection if the distance between the query and matched database frame node is less than 5 meters. Note, that consecutive frames within a certain range will not be considered as positive pair.

During the testing phase on the KITTI dataset, we evaluate the proposed NDT-MC as loop-closure detection component for a lidar SLAM system. As a comparison, other existing global descriptors are also evaluated, including Scan Context [1], Intensity Scan Context [4], Lidar Iris [2], M2DP [15], NDD [16] and STD [27]. Default parameters are used in SC<sup>1</sup>, ISC<sup>2</sup>, IRIS<sup>3</sup>, M2DP<sup>4</sup>, NDD<sup>5</sup>, and STD<sup>6</sup>. The same evaluation metrics, i.e. F1 score and Extended Precision (EP), are used in this experiment.

In this paper, the parameters of NDT-MC are set as follows:  $N_r = 20$ ,  $N_\theta = 60$ ,  $N_w = 6$ ,  $g_{max} = 2.4$ , and the maximum point cloud range is 80m. The proposed descriptor is a  $(2 \times 20) \times 60$  matrix.

<sup>1</sup><https://github.com/irapkaist/scancontext>  
<sup>2</sup><https://github.com/wh200720041/fiscoam>  
<sup>3</sup><https://github.com/JoestarK/lidar-Iris>  
<sup>4</sup><https://github.com/LiHeUA/M2DP>  
<sup>5</sup><https://github.com/zhouruihao1001/NDD>  
<sup>6</sup><https://github.com/hku-mars/STD>

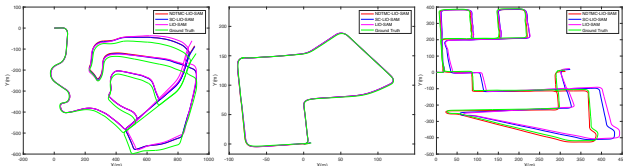


Fig. 6: The comparison of trajectories on KITTI sequence 02, 07, and 08. The trajectories of NDTMC-LIO-SAM, SC-LIO-SAM, LIO-SAM, and ground truth are represented in red, blue, rose red, and green, respectively.

#### D. Evaluation as an Integrated full-SLAM system.

Different from most of the state-of-the-art descriptors, our approach can deploy in real-time using a normal CPU. We integrate NDT-MC with LIO-SAM [29] and set  $K = 10$ .

Afterward, we compare our method with open-source algorithms SC-LIO-SAM in terms of real-time loop closure performance. we compare the real-time loop closure performance between both our integrated systems, i.e., NDT-MC with LIO-SAM, and SC-LIO-SAM. For a fair comparison, parameters, i.e. a maximum distance of 80 meters, a similarity distance threshold of 0.6, and a submap update frequency of 1Hz are used in both our approach and the baseline. A video demo of the integrated full SLAM system can be accessed through the hyperlink provided below [https://youtu.be/xCtWR1EKcfk?si=J-\\_TAYcQmW1g\\_Juv](https://youtu.be/xCtWR1EKcfk?si=J-_TAYcQmW1g_Juv).

#### E. Run-time performance.

We also conduct a comparison of runtime performance for descriptor extraction and retrieval on the KITTI. Our method is benchmarked with Scan Context, Lidar Iris, and STD. All comparisons were performed on a desktop equipped with an i9-10900K CPU @ 3.70GHz and 32GB of memory. As shown in Table IV, our approach shows a supreme performance in terms of both the descriptor extraction and query time.

#### F. Results analysis

TABLE II presents the F1 Score and EP of the proposed method and the compared methods and NDT point clouds of 2m resolution are investigated. The use of "—" in the table signifies the inability to compute the Extended Precision at various thresholds. This is because these methods did not achieve 100 % precision at different thresholds, hence preventing the derivation of the  $R_{P100}$  (recall at a precision of 100 %) value. The Extended Precision (EP) is calculated as  $0.5 \times (R_{P100} + P_{R0})$ ,  $P_{R0}$  is the precision at minimum recall. On the eight testing trajectories of four testing scenes, our approach achieves the highest F1 Score and EP. Specifically, our approach with an NDT point cloud resolution achieves the highest performance across all eight test datasets, demonstrating significant advancements compared to other baseline methods.

NDT-MC presents superior performance in underground parking scenarios due to its effective utilization of height information, its ability to capture complex geometric structures, and its capacity to handle dynamic environments. En-

coding and weighting height information in a limited vertical range enhances loop closure accuracy. NDT-MC categorizes NDT cells to capture geometric intricacies, and it emphasizes permanent structure shapes over non-static elements to ensure robust performance in dynamic environments. This approach provides a unique and stable feature identification for prolonged and accurate localization, surpassing methods that rely on less informative high z-values, especially in indoor and dynamic scenes.

TABLE III presents the comparison results of F1 Score and EP on the KITTI dataset. In the six test scenarios, our method NDT-MC achieved the highest F1 Score in sequences 00, 02, 05, and 06, and the highest EP in sequences 05, 06, 07, and 08. Additionally, our method ranked second in EP in sequences 00 and 02, and second in F1 Score in sequences 07 and 08. Specifically, our method has shown significant advances in the experiment on the KITTI dataset. Our method achieves superior performance on sequences 00, 02, 05, and 06 over the six sequences. In this table, NDD without superscript represents the test results from the NDD paper [16], while NDD with superscript indicates the results obtained by the provided code.

We provide standalone comparisons of global descriptors and matching performance on both the NIO dataset and KITTI datasets, as detailed in Sections V-B and V-C. Note that in these comparisons, no prior poses were utilized. Our technique demonstrated superior results on the KITTI dataset, evident from the relatively high F1 Scores and Extended Precision. On the NIO dataset, our method's performance was particularly noteworthy. Additionally, in Section V-D, we provided the assessment results of our proposed technique within an integrated SLAM framework. For this experiment, our proposed global descriptor and matching were combined with the selection of a prior pose for loop-closure detection. We named this system NDTMC-LIO-SAM. Its performance was benchmarked against a leading open-source Lidar SLAM system, SC-LIO-SAM, by comparing overall trajectory accuracy. This is a system-level comparison that takes into consideration integration and run-time factors. From Fig. 6, the proposed NDT-MC-LIO-SAM and the SC-LIO-SAM achieve similar performance on sequences 02 and 07. Especially on sequence 08, the trajectory of our method significantly outperforms SC-LIO-SAM.

## VI. CONCLUSION

This paper proposed a real-time loop-closure detection approach based on a geometry-only global descriptor, which shows good generalizability in a variety of real-world scenes. In terms of indoor scenes, we used a mass-produced front-view LiDAR for implementation and evaluation. To overcome the limitations of FOV and visibility, we proposed to use an instantly-built SLAM map during the localization process. By this means, spatial-temporal mapping can largely eliminate occlusions and improve co-visibility between databases and query observations. Additionally, our

approach leverages a lightweight NDT point cloud representation and encodes explicit geometric shape information by applying shape classification to NDT cells.

For localization in both underground parking and on-road driving scenarios, NDT-MC combines the height of cells with their shape and entropy. Our method is thoroughly tested on eight datasets collected from four indoor underground parking lots and the most widely-used KITTI dataset. Experimental results demonstrate significant advantages of our method in terms of numerous metrics of accuracy and efficiency. We also integrate the proposed global descriptor as a real-time loop closure detection component with a lidar mapping system. Codes are available publicly to the public.

## REFERENCES

- [1] G. Kim and A. Kim, "Scan context: Egocentric spatial descriptor for place recognition within 3d point cloud map," pp. 4802–4809, 2018.
- [2] Y. Wang, Z. Sun, C.-Z. Xu, S. E. Sarma, J. Yang, and H. Kong, "Lidar iris for loop-closure detection," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 5769–5775.
- [3] L. Li, X. Kong, X. Zhao, T. Huang, W. Li, F. Wen, H. Zhang, and Y. Liu, "Ssc: Semantic scan context for large-scale place recognition," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 2092–2099.
- [4] H. Wang, C. Wang, and L. Xie, "Intensity scan context: Coding intensity and geometry relations for loop closure detection," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2095–2101.
- [5] G. Kim, S. Choi, and A. Kim, "Scan Context++: Structural Place Recognition Robust to Rotation and Lateral Variations in Urban Environments," in *IEEE Transactions on Robotics*, vol. 38, no. 3, pp. 1856–1874, 2022.
- [6] P. Scovanner, S. Ali, and M. Shah, "A 3-dimensional sift descriptor and its application to action recognition," in *Proceedings of the 15th ACM international conference on Multimedia*, 2007, pp. 357–360.
- [7] Y. Cui, Y. Zhang, J. Dong, H. Sun, and F. Zhu, "Link3d: Linear keypoints representation for 3d lidar point cloud," *arXiv preprint arXiv:2206.05927*, 2022.
- [8] J. Knopp, M. Prasad, G. Willems, R. Timofte, and L. V. Gool, "Hough transform and 3d surf for robust three dimensional classification," in *European Conference on Computer Vision*. Springer, 2010, pp. 589–602.
- [9] Y. Cui, X. Chen, Y. Zhang, J. Dong, Q. Wu, and F. Zhu, "Bow3d: Bag of words for real-time loop closing in 3d lidar slam," *IEEE Robotics and Automation Letters*, 2022.
- [10] S. Salti, F. Tombari, and L. Di Stefano, "Shot: Unique signatures of histograms for surface and texture description," *Computer Vision and Image Understanding*, vol. 125, pp. 251–264, 2014.
- [11] T. Shan, B. Englot, F. Duarte, C. Ratti, and D. Rus, "Robust place recognition using an imaging lidar," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 5469–5475.
- [12] M. Bosse and R. Zlot, "Place recognition using keypoint voting in large 3d lidar datasets," in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 2677–2684.
- [13] B. Steder, M. Ruhnke, S. Grzonka, and W. Burgard, "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*. IEEE, 2011, pp. 1249–1255.
- [14] K. Vidanapathirana, P. Moghadam, B. Harwood, M. Zhao, S. Sridharan, and C. Fookes, "Locus: Lidar-based place recognition using spatiotemporal higher-order pooling," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 5075–5081.
- [15] L. He, X. Wang, and H. Zhang, "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 (IROS)*. IEEE, 2016, pp. 231–237.
- [16] R. Zhou, L. He, H. Zhang, X. Lin, and Y. Guan, "Ndd: A 3d point cloud descriptor based on normal distribution for loop closure detection," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 1328–1335.
- [17] X. Xu, H. Yin, Z. Chen, L. Li, Y. Wang and R. Xiong "Disco: Differentiable scan context with orientation," in *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2791–2798, 2021.
- [18] J. Ma, J. Zhang, J. Xu, R. Ai, W. Gu, and X. Chen, "Overlaptransformer: An efficient and yaw-angle-invariant transformer network for lidar-based place recognition," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6958–6965, 2022.
- [19] J. Ma, G. Xiong, and J. Xu, and X. Chen "CVTNet: A Cross-View Transformer Network for LiDAR-Based Place Recognition in Autonomous Driving Environments," in *IEEE Transactions on Industrial Informatics*, vol. 20, no. 3, pp. 4039–4048, 2024.
- [20] A. Dewan, T. Caselitz, and W. Burgard, "Learning a local feature descriptor for 3d lidar scans," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 4774–4780.
- [21] H. Yin, X. Ding, L. Tang, Y. Wang, and R. Xiong, "Efficient 3d lidar based loop closing using deep neural network," in *2017 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2017, pp. 481–486.
- [22] R. Dube, A. Cramariuc, D. Dugas, H. Sommer, M. Dymczyk, J. Nieto, R. Siegwart, and C. Cadena, "Segmap: Segment-based mapping and localization using data-driven descriptors," *The International Journal of Robotics Research*, vol. 39, no. 2-3, pp. 339–355, 2020.
- [23] M. A. Uy and G. H. Lee, "Pointnetvlad: Deep point cloud based retrieval for large-scale place recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4470–4479.
- [24] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "Netvlad: Cnn architecture for weakly supervised place recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5297–5307.
- [25] Z. Zhou, C. Zhao, D. Adolphsson, S. Su, Y. Gao, T. Duckett, and L. Sun, "Ndt-transformer: Large-scale 3d point cloud localisation using the normal distribution transform representation," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 5654–5660.
- [26] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.
- [27] C. Yuan, J. Lin, Z. Zou, X. Hong, and F. Zhang, "Std: Stable triangle descriptor for 3d place recognition," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 1897–1903.
- [28] T. Stoyanov, M. Magnusson, H. Andreasson, and A. J. Lilienthal, "Fast and accurate scan registration through minimization of the distance between compact 3d ndt representations," *The International Journal of Robotics Research*, vol. 31, no. 12, pp. 1377–1393, 2012.
- [29] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and R. Daniela, "Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 5135–5142.