# Place Recognition of Large-Scale Unstructured Orchards With Attention Score Maps

Fang Ou, Yunhui Li, *Member, IEEE*, and Zhonghua Miao

***Abstract*—The availability of autonomous orchard robots could alleviate the conflict caused by rising labor costs and labor shortages. The essential technical requirements are autonomous localization and mapping which rely on place recognition to explore data associations. This letter presents a novel LiDAR-based place recognition algorithm for unstructured and large-scale orchards. Concretely, we propose a discriminative global representation, spatial binary pattern (SBP), that encodes three-dimensional (3D) spatial distributed scan into an eight-bit binary pattern. In addition, an efficient two-stage hierarchical re-identification process is proposed. The attention score map is introduced for task-relevant features and preliminary candidates retrieval. The overlap re-identification is used to align a pair of descriptors to confirm the final loop closure index. Experiments on orchard and public datasets have been conducted to evaluate the performance of the proposed method, our method achieves a higher recall rate and localization accuracy. Moreover, experiments on the long-term outdoor dataset KITTI further demonstrate the generality.**

***Index Terms*—Attention score map, orchard robots, place recognition, spatial binary pattern (SBP).**

## I. Introduction

ORCHARD robots equipped with advanced sensors could play an important role in different agricultural tasks. For example, the tree trunks are detected as natural landmarks for robot localization and mapping in autonomous robots [1], [2]. In these orchard tasks, place recognition is an essential component for retrieving previously visited places. Commonly, loop closure detection is well-known place recognition in simultaneous localization and mapping (SLAM), which refers to the capability of eliminating the inevitable drift error caused by long-term state estimation [3], global localization for a kidnapped robot [4] and multi-robot mapping [5]. Place recognition in harsh outdoor environments (such as orchard and forestry) is still a growing research topic. In large-scale mountain orchards, fruit trees are usually planted in unorganized with unequal distances straight and parallel rows, and as the orchard grows, branches fall, etc., the regularity of tree and non-tree object placement is neither precise nor reliable, which determines the unstructured character. At the same time, the orchard layout makes it suitable for specific autonomous mobile robot navigation, mapping and localization.

Robots commonly used for multitasking are based on standalone Global Navigation Satellite System (GNSS). However, in complex orchards, satellite signals tend to be unreliable since the agricultural robots frequently move under the canopy blocking the signals from the receiver [6], [7]. In addition, [8] highlighted that varying outdoor light conditions, for example, direct sunlight or shadow, has a negative impact on the robots performance and point cloud is less affected by illumination changes. For these reasons, it is extremely important to research and develop intelligent solutions for orchard tasks with 3D LiDAR as a predominant sensor. 3D LiDAR can capture more stable point clouds information (e.g. height, depth, position, intensity) suitable for outdoor place recognition [9]. However, LiDAR-based place cognition in orchard still needs to overcome the major bottlenecks resulting from unstructured, unordered, large-scale, long-term and sparse range sensor data. Correspondingly, there are critical technical issues that need to be addressed urgently. Firstly, the algorithm needs to be invariant to detailed dynamics such as leaf movement, long-term weather changes and orchard growth. Secondly, the global descriptor is required to achieve the ability to noise processed and regularized to describe the spatially disordered independent point clouds. Moreover, there are multiple scenes with only slight differences in the large-scale orchard, which requires the method able to evaluate the correlation and contribution, emphasizing task-relevant content to distinguish whether there is vegetation cover. In addition, the huge demand of computing resources for place recognition in large-scale environments also severely restricts its application in the agricultural field.

Existing place descriptors focus on exploiting contextual information to compactly summarize an outdoor place, such as a fast point feature histogram (FPFH) [10] and signature of histograms of orientations (SHOT) [11]. The histogram-related descriptors with a high dimensional computational cost especially in large-scale outdoor environments, and only provides a stochastic index of the scene, which are not straightforward.

Manuscript received 7 July 2022; accepted 20 December 2022. Date of publication 6 January 2023; date of current version 17 January 2023. This letter was recommended for publication by Associate Editor A. Degani and Editor H. Moon upon evaluation of the reviewers' comments. This work was supported in part by Shanghai Agriculture Applied Technology Development Program, China under Grant 2020-02-08-00-09-F01466 and in part by the National Natural Science Foundation (NNSF) of China under Grants 61473100 and 51875331. *(Corresponding author: Zhonghua Miao.)*

The authors are with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China (e-mail: oufang@shu.edu.cn; liyunhui@shu.edu.cn; zhonghuamiao@163.com).

This letter has supplementary downloadable material available at https://doi.org/10.1109/LRA.2023.3234744, provided by the authors.

Digital Object Identifier 10.1109/LRA.2023.3234744

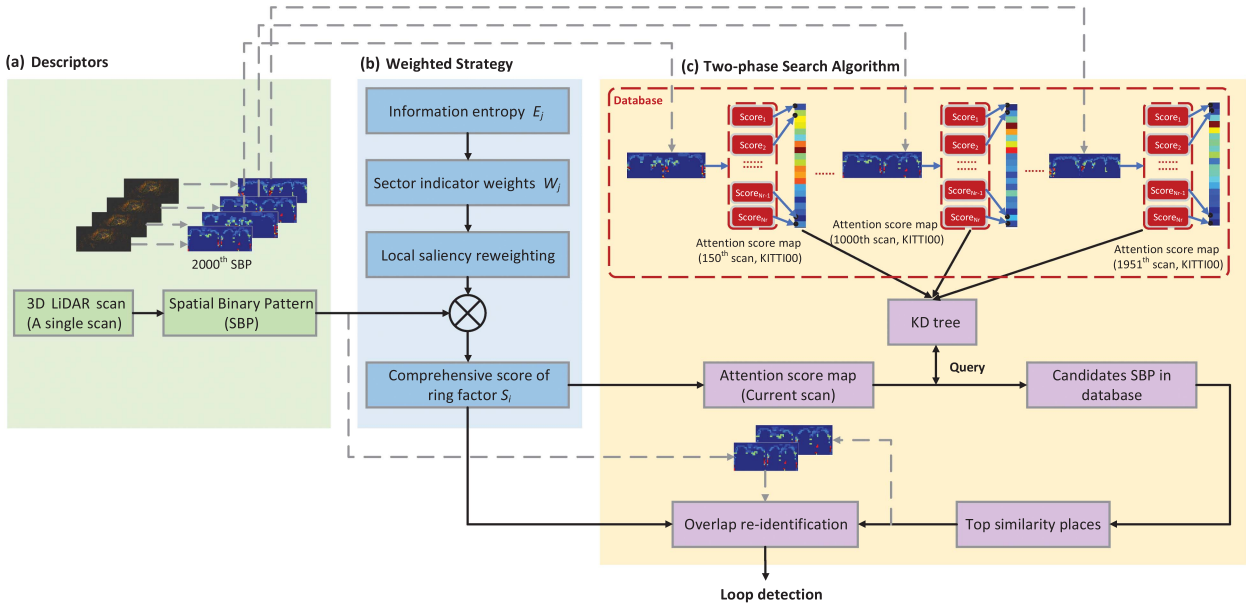



Fig. 1. Algorithm overview. (a) A place descriptor called spatial binary pattern (SBP) is extracted for LiDAR scan. (b) A information entropy weighting scheme is employed to suppress the irrelevant-features and construct an attention score map. (c) Based on the historical attention score map, the KD-Tree is constructed for loop closure candidates about the current query point cloud. Finally, compare the similarity of SBP pairs. The top similarity candidate that satisfies the overlap threshold is considered as revisiting place.

More importantly, in the process of extracting descriptors, it is unreasonable with all the point clouds being treated equally. Evaluating the contribution of local point features in global descriptor and reweighting to focus more attention on task-relevant regions is an effective scene representation strategy. For example, in place recognition tasks, focusing on salient time-varying objects, such as pedestrians and vehicles, can achieve positive retrieval results. There are already some works in the field of 2D images to focus on selecting the most important or interesting pixel and pay more attention to the discriminative local features [12], [13]. In the field of 3D point clouds, the point cloud contextual information is also essential in predicting the significance of local point features [14].

Aiming at the unstructured and large-scale orchard place recognition, the novel LiDAR-based place recognition method is presented in Fig. 1. In Section III, we first explain how the density and geometry information of point clouds can be used to distinguish objects in orchards and propose a global descriptor spatial binary pattern (SBP) for place recognition. To further improve the efficiency of our algorithm, a two-phase search algorithm is performed. The attention score map based on information entropy is introduced for finding possible candidates for loop closure detection. The overlap re-identification is used to identify precise loop index among candidates. To summarize, the contributions of the letter include:

- We propose a novel 3D LiDAR scan representation so called SBP, which integrates geometry and density characteristics utilizing a more efficient bin encoding function.
- A novel reweighting strategy evaluates the contribution scores of radial (from center to maximum sensing range) point clouds to construct an attention score map. The attention score map is a low-dimensional representation of a LiDAR scan from the perspective based on information entropy, highlighting task-relevant features for preliminary retrieval possible revisited scenes.
- To achieve a feasible search time, we propose a two-stage search algorithm, including an attention score map for loop closure candidates and SBP pairwise overlap as a further metric, thus avoiding searching all databases for loop closure detection.
- A thorough evaluation on three different scale orchards and public datasets is conducted, the proposed approach exhibits competitive performance.

## II. Related Work

### A. Handcrafted 3D Descriptors

LiDAR presents strong robustness to perceptual changes, the early phase of LiDAR-based outdoor place recognition focused on 2D range data [15], [16]. As 3D LiDAR appeared, many works have been conducted for place recognition with 3D point clouds. Early approaches of 3D analysis mainly adopt statistics ideas of histograms such as PFH [17], SHOT [18], shape context [19], or spin image [20]. The general idea is described as, finding a keypoint, separating nearby points into bins, and encoding a pattern of surrounding bins into a histogram. The method based on the point cloud contextual information projects the point cloud to a bird's-eye view (BEV), the scan context [3] proposes a 2D descriptor based on the maximum height within the bin. ISC [9] utilizes binary-operation fast geometric retrieval and intensity structure re-identification. LiDAR-Iris [21] can obtain binary signature images after multiple LoG-Gabor filtering and threshold operations on LiDAR-Iris images. M2DP [22]

compressed a LiDAR scan into a global descriptor on multiple planes that is robust to noisy input. These handcraft methods focused on acquiring local features and aggregating them equally as global descriptors which are not efficient for complex scenes. In particular, in a large-scale environment, iterative registration to retrieve the best registration for the query point cloud in the full feature database will result in huge computational pressure.

### B. Deep Learning Descriptors

Recent advances in machine learning have opened up new possibilities to deal with the weaknesses of handcrafted presentations for place recognition with LiDAR. Some learning-based methods tried to learn descriptors directly from 3D points, PointNet [23] directly handles point cloud and uses a symmetric function to make the output invariant to the order permutation of the input points, PointNet++ [24] leverages neighborhoods at multiple scales to capture local structures. PointNetVLAD [25] is a deep network combining PointNet and NetVLAD [26] to extract the global descriptor for retrieval task. Though PointNetVLAD is more efficient to add a NetVLAD layer to the global feature than just using the vanilla PointNet architecture, it does not discriminate the local features which positively contribute to the final global feature representations. Usually, deep learning requires a large amount of training data for the model. However, obtaining well-characterized point cloud in large-scale and complex orchards is a daunting task. Unlike our algorithm, which considers registration from the perspective of spatial distribution and contribution, they are general orchard property. Moreover, the training of deep learning models similar to "black box," which does not explore the intrinsic structural properties of the data, usually tends to only be effective for a specific scene.

### C. Attention Module for Descriptors

In order to reduce the influence of task-irrelevant local features, researchers encode discriminative global representations via attention modules. The existing attention strategies are either completely data-driven or based on artificial prior. Kim et al. [27] introduces a CRN that can effectively predict the importance of feature map regions based on image context. Zhu et al. [28] proposes APANet that aggregates multi-scale regional features weighted by cascaded attention blocks. Concurrently, some approaches perform feature filtering with prior semantic-guided [29], [30], they retain the local features with specified semantics for visual localization. SRALNet [31] combines semantic priors and data-driven finetuning to enhance the differential local weighting scheme. However, it overlooks the image contextual information of local features. Aiming at the characteristics of single vegetation objects and highly similar scenes in orchards. We exploit the voxel distribution priors which are associated with point cloud contextual information to predict the importance of feature map regions based on attention entropy.

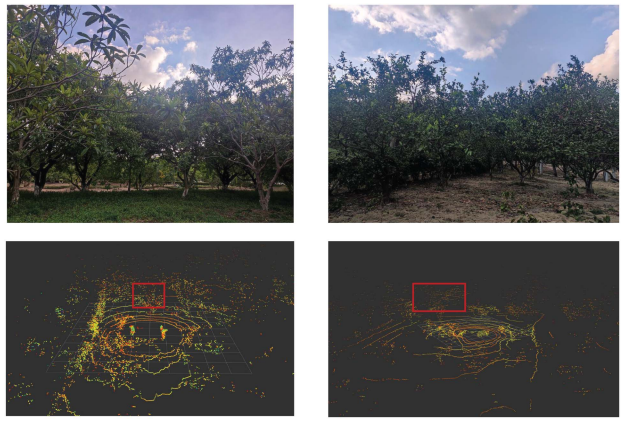


Fig. 2. An instance of an unstructured orchard, with loquats on the left and oranges on the right, dense canopy is highlighted with red rectangles.

## III. Proposed Methods

The overall algorithm framework can be summarized into three components: scan representation, contribution score maps of task-relevant features, two-stage hierarchical re-identification. Each functional step will be explained sequentially in the following subsections.

### A. Spatial Binary Pattern

LiDAR has been proven to be an effective, non-destructive scanning device for in-field canopy structure detection and estimation [32]. LiDAR scans of canopy and trunk can be directly detected and used to assist robotic tasks. In Fig. 2, we show an example from our orchard dataset for demonstration. The point cloud and image present the same place, we marked the canopy and highlight it with red rectangle. It can be observed that the scanned point cloud in densely covered vegetation (canopy) is significantly different from sparse regions (trunk). Since LiDAR perceives the environment by emitting and receiving laser beams reflected by objects, which reveals the surrounding surface reflectance structure. In this letter, a binary exponential encoding function is introduced and a place descriptor called spatial binary pattern (SBP) is proposed for orchard. The encoding values at the top canopy cover and the bottom unoccupied can be clearly distinguished by the explosive growth characteristics of the exponential calculation. In Fig. 3, the whole LiDAR scan is integrated into a bird's-eye view (BEV). The space is composed of 360-degree scan can be treated as a cylinder centered on the LiDAR, we slice the scan from the origin and divide it into $N_s \times N_r \times N_c$ voxel bins of azimuthal, radial and cylinder directions. $N_s$, $N_r$ and $N_c$ are the number of sectors, rings and cylinders respectively, the binary bin value is confirmed according to the occupancy of point cloud. In this letter, we set parameters generic to $N_c = 8$, $N_r = 20$ and $N_s = 60$, the eight-bit binary code about the voxel bins can be obtained along the vertical direction. The space binary pattern encoding function is used to project the bins onto the bird's-eye view, which can realize the conversion from point cloud to image

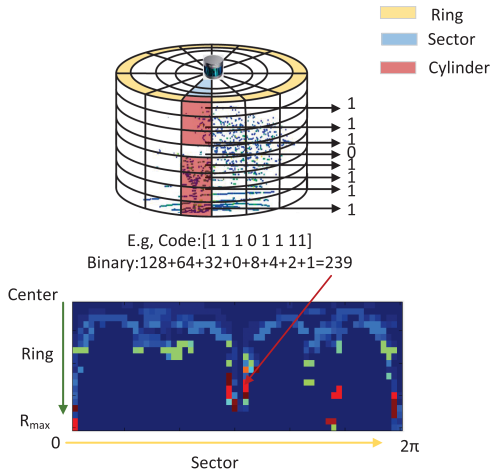



Fig. 3. Using BEV representation from a 3D scan, we partition scan sapce into 3D bins, which are split according to azimuthal (from 0 to $2\pi$ within a LiDAR frame), radial (from center to maximum sensing range $R_{max}$) and cylinder (from 0 to maximum scanning height $C_{max}$) directions. We refer to the yellow area as a ring, the blue area as a sector. Spatial binary patten (SBP) is a matrix as in the bottom. The encoding value extracted from the points located in eight-bit 3D bins is used as the corresponding pixel value of SBP.

domain. The maximum sensing valid range of a LiDAR sensor is $R_{max} = 80$ m, the maximum cylinder height $C_{max}$ needs to be set empirically. we set $C_{max}$ equal to 16 m in the loquat orchard and 8 m in the orange orchard, the circular scan azimuthal is $2\pi$ decided by inherent properties of the LiDAR. The resolutions of radial, azimuthal and cylinder are $\frac{R_{max}}{N_r}$, $\frac{2\pi}{N_s}$ and $\frac{C_{max}}{N_c}$.

Each point point $p_k = (x_k, y_k, z_k)$ can be represented as $\Gamma = (r_k, \theta_k, c_k)$ in polar coordinates.

$$\begin{aligned} r_k &= \sqrt{x_k^2 + y_k^2} \\ \theta_k &= \arctan \frac{y_k}{x_k} \\ c_k &= z_k + h_p \end{aligned} \tag{1}$$

where the actual height $c_k$ of the point cloud is related to the installation height $h_p$ of the sensor platform, and the bin attribution of the point cloud is limited by the effective range of the LiDAR:

$$\begin{aligned} P_{ij\xi} = \Bigg\{ p_k \in P \Bigg| & \frac{i \cdot R_{max}}{N_r} \le r_k < \frac{(i+1) \cdot R_{max}}{N_r}, \\ & \frac{j \cdot 2\pi}{N_s} - \pi \le \theta_k < \frac{(j+1) \cdot 2\pi}{N_s} - \pi, \\ & \frac{\xi \cdot C_{max}}{N_c} \le c_k < \frac{(\xi+1) \cdot C_{max}}{N_c} \Bigg\} \end{aligned} \tag{2}$$

where $i \in [1, N_r]$, $j \in [1, N_s]$ and $\xi \in [1, N_c]$, each binary bin $P$ contain a series of points $P_{ij\xi}$ with the same $i^{th}$ ring,$j^{th}$ sector and $\xi^{th}$ cylinder:

$$P = \bigcup_{i \in N_r, j \in N_s, \xi \in N_c} P_{ij\xi} \tag{3}$$

The point clouds falling within the bin, and an bin is considered occupied when the number is greater than 10. Otherwise the scan is considered invalid and defined as unoccupied. The voxel bins assignment function can be defined as:

$$\delta(P) = \begin{cases} 1 & \text{if } P \text{ is occupied} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where $\delta(P)$ is the function that returns a binary value, assigning zero for unoccupied bins, and one for the occupied.

All eight-bit binary bins $P$ are encoded into a so called Spatial Binary Pattern (SBP) in a $N_r \times N_s$ subspace: $\mathbb{R}^3 \longmapsto \mathbb{R}^2$ via the space binary pattern encoding function:

$$SBP(i,j) = \sum_{\xi=1}^{N_h} 2^{\xi-1} \delta(P) \tag{5}$$

where the global representation SBP is a bird's-eye view with values ranging from 0 to 255, which reveals both the point clouds geometry and density distribution. SBP segmented 3D voxel bins based on height and density information of orchard scans. It enforces the exponential property of the coding function, which means that the coding values of the exponential bits are significantly different, and can effectively distinguish low shrubs which occupy low exponential bits from tall fruit trees in higher exponential bits.

### B. Attention Score Map

*Information Entropy Weighted Strategy:* The point cloud under a specific scanning radius is projected into the bird's-eye view SBP matrix corresponding to each row, which is defined as the ring factor. Similarly, the azimuth corresponds to the column of the matrix, which is defined as the sector indicator. In this letter, the calculation of the contribution score is converted into a weighted strategy solution, which depends on the information entropy. We proposed attention score map which is a vector representation to evaluate the contribution of the sector indicator in the ring factor. In general, the ring factor of task-relevant point clouds can achieve higher contribution scores than the task-irrelevant. In our method, high-scoring point clouds are usually scanned in densely vegetated regions. Therefore, the first process of making an attention score map is to calculate the proportion $\rho_{ij}$ of the $j^{th}$ sector indicator in the $i^{th}$ ring factor, and the SBP is pixel-wise normalized to $SBP(i,j)_{norm}$:

$$\rho_{ij} = \frac{SBP(i,j)_{norm}}{\sum_{i=1}^{N_r} SBP_{ij}}, \tag{6}$$

Then the information entropy of $j^{th}$ sector indicator:

$$E_j = -K * \sum_{i=1}^{N_r} \rho_{ij} \ln \rho_{ij}, \tag{7}$$

where $K = \frac{1}{\ln N_r}$ is a constant utilized to constrain the information entropy $E_j$ value in the interval [0, 1]. Information entropy is an intuitive expression related to the point cloud information content of the observation sector indictor in each ring factor.

$$W_j = \frac{1 - E_j}{N_s - \sum E_j}. \tag{8}$$

where $W_j$ denotes the weight of the sector indicator. Next, the attention score map can be constructed by aggregating the ring factor.

*Feature Aggregation:* Since our goal is to apply the attention map on an efficient point cloud retrieval pipeline, thus the global descriptors SBP is aggregated into a discriminative and compact vector representation. Each row of a SBP, r, is encoded into a single real value in the vector. The elements in the vector reflect the contribution of observed objects on the current scanning radius. We apply the attention weight matrix $W$ to ring factors, which reweights the global feature SBP as follows:

$$\underset{i\in N_r}{s_i} = \sum_{j=1}^{N_s} W_j * SBP(i,j), \tag{9}$$

$s_i$ is the comprehensive score of the ring factor. $S$ denotes the final attention score map which concerns multi-channel sector entropy information:

$$S = (s(r_1), \ldots, s(r_{N_r})), \text{where } s : r_i \mapsto \mathbb{R}, \tag{10}$$

The output attention map size is $N_r \times 1$, the example of attention score map is presented in the upper right of Fig. 1. In this letter, the attention score is calculated based on the information entropy to obtain the weight. Since it is only related to column elements in the SBP matrix, which is not associated with the azimuth value in the current scan. Therefore, the attention score map is independent of the viewpoint.

### C. Two-Phase Search Algorithm

Place recognition aims to match the current place with the previously visited places from the historical database. As more scenes are visited, the scale of the database inevitably increases in large-scale orchards so that the computational cost grows accordingly. In order to improve the retrieval efficiency while ensuring accuracy in the intricate orchard. This letter proposes a two-stage step-by-step retrieval scheme including a fast attention score map relation retrieval and the overlap re-identification of global descriptors.

*Fast Attention Score Map Candidates Search:* As explained in the previous section, each ring factor of the descriptor is encoded into a single attention score value via information entropy. Although the $N_r \times 1$ attention score map obtains less information than $N_r \times N_s$ global descriptors, can quickly find a revisiting place and more directly represent the contribution of LiDAR scan. The historical attention score map is utilized to construct a KD-Tree. In the KD-Tree, the $K$-Nearest neighbors of the current query point cloud are retrieved as possible visited places (loop closure candidates). In comparison, KD-Tree retrieval of low-dimensional vectors can achieve faster computation speed than global descriptor registration directly. These constant number of candidates spatial binary patterns are compared against the query spatial binary pattern by taking cosine distance.

$$d(SBP^q, SBP^c) = \frac{1}{N_s}\sum_{j=1}^{N_s}\left(1 - \frac{v_j^q \cdot v_j^c}{\left\|v_j^q\right\|\left\|v_j^c\right\|}\right), \tag{11}$$

where $SBP^q$ and $SBP^c$ are spatial binary patterns acquired from a query point cloud and loop closure candidates respectively. $v_j^q$ and $v_j^c$ are two column vectors at the same index.

$$D(SBP^q, SBP^c) = \min_{n\in[N_s]} d(SBP^q, {SBP_n}^c), \tag{12}$$

when $SBP^c$ is shifted $n$ column from the original, the minimum distance is found. Note that the candidates SBP need to satisfy an acceptance threshold that can be selected as the revisited place:

$$c^* = \min_{c_k\in C} D(DBP^q, DBP^{c_k}), \text{s.t} D < \tau, \tag{13}$$

where $c_k$ is a set of indexes of candidates and $c^*$ is the top similarity candidate as preliminary loop closure index.

*Overlap Estimations Re-identification:* The concept of overlap has been utilized in photogrammetry to estimate the similarity, it can also be a flexible tool for LiDAR-based place recognition [33], [34], [35].

The tractate judges the possibility of a pair of SBP originating from the same scene based on the overlap. For a pair of LiDAR scans, the overlap of SBP can be described in the following details. The absolute differences of all corresponding binary pattern value in $SBP_q$ and $SBP_{c^*}$ are calculated and considering that only the occupied bins are valid scanning. The overlap is then calculated as the percentage of all differences in a certain threshold $\varepsilon$, the overlap is formulated as follows:

$$O_{qc^*}^j = \max_{j\in N_s} \frac{\sum_{(i,j)} \Phi\left\{\| SBP_q(i,j) - SBP_{c^*}^j(i,j) \| < \varepsilon\right\}}{\min\left(valid\left(SBP_q\right), valid\left(SBP_{c^*}\right)\right)}, \tag{14}$$

where $\Phi\{a\} = 1$ if $a$ is true and 0 otherwise. $valid(SBP)$ refers to the non-zero elements in the global description, which means valid scanning of the region of interest $(i^{th}ring, j^{th}sector)$, since not all pixels might have a positive LiDAR mapping associated after the projection. $SBP_{c^*}^j$ is a $j$ column-shifts operation of the matrix $SBP_{c^*}$. Because the column vector in a spatial binary pattern represents azimuthal direction, corresponding to the rotation change [3]. Therefore, to identify the best matching pair, we calculate overlap with all possible column-shifted, only when the overlap exceeds the minimum limit $\mu$, the corresponding scan $c*$ will be finally determined as the loop closure.

## IV. Experiments

The experimental evaluation is designed to support the performance of our approach and to evaluate the claims that our approach is able to: i) Detect loop closure candidates for large-Scale unstructured orchards using only LiDAR data without any other information. ii) Generalize to long-term outdoor scenarios by fine-tuning the parameters. iii) Compared to algorithms that directly use point cloud intensity or height maximum, the encoding-based descriptor extraction method does not require more time consumption due to the presence of attention score map. iv) Introduce an information entropy weighting method to highlight task-relevant features and improve the accuracy of loop closure detection.

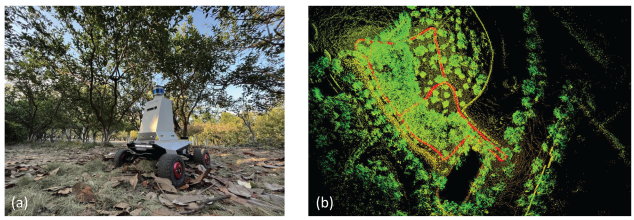



Fig. 4. Localization and mapping result in orchard1. (a) AGV platform for experiment. (b) Localization and mapping result of the proposed method. The robots trajectory is plotted in red color.

TABLE I
COMPARISON OF RUNTIME AND TRANSLATION ERROR WITH STATE-OF-THE-ART METHODS

| Dataset (Pose) | Approach | Runtime [s] | Translation RMSE [m] |
|---|---|---|---|
| Orchard1 (4590) | Scan Context [3] | 120 | 0.29 |
| | ISC [9] | **107** | 0.22 |
| | LeGO-LOAM [37] | 145 | 0.41 |
| | Ours | 128 | **0.15** |
| Orchard2 (5460) | Scan Context [3] | 170 | 0.41 |
| | ISC [9] | **159** | 0.56 |
| | LeGO-LOAM [37] | 209 | 1.86 |
| | Ours | 185 | **0.27** |
| Orchard3 (2432) | Scan Context [3] | 69.3 | 0.26 |
| | ISC [9] | **66.3** | 0.20 |
| | LeGO-LOAM [37] | 77.1 | 0.43 |
| | Ours | 84.1 | **0.18** |

The performance of our method will be evaluated by orchard and KITTI dataset [36] experiments. Both experiments are carried out on Intel NUC mini computer with i7 memory 500 GB, the proposed method is implemented in C++ and is integrated into the robot operating system (ROS). The outdoor long-term experiment on KITTI 00 and KITTI 05 are used to verify the generality of the algorithm.

### A. *Experiment on Orchard Robot*

*Experimental Settings:* We conduct field experiments in three scales of orchards, the proposed place recognition approach is utilized as the loop closure detection in the SLAM system to reduce localization drifts and improve the mapping accuracy. we implement the algorithm on an autonomous guided vehicle (AGV) equipped with RoboSense LiDAR (RS-LiDAR-16) with a maximum speed is 1.5 m/s, as shown in Fig. 4(a). In this experiment, the algorithm superiority is verified on the well-known LeGO-LOAM [37], which is a lightweight LiDAR SLAM system with ground optimization. The localization and mapping result as shown in Fig. 4(b), with the trajectory of the robot plotted in red. The existing method SC-LeGO-LOAM replaces the loop closure detection with scan context [3] to improve localization accuracy and reduce drift. ISCLOAM performs loop closure detection based on the spatial intensity descriptor ISC [9].

*Experimental Analysis:* We firstly set the parameter default to $R_{max} = 80$ m, $N_s = 60$, and $N_r = 20$ by referring to the state-of-the-art method used for comparison. As shown in Table I, the performance and time consumption of our representation is compared to Scan Context [3], ISC [9], and original LeGO-LOAM. In orchards with dense canopy, GNSS is always blocked resulting in ground truth not being available for evaluation. Also, the unavailability of GNSS in orchard is one of the motivations for this letter. An alternative solution is utilized when evaluating the accuracy of the algorithm, we mark the starting point as the origin. When the experimental robot completes the final loop closure event back to the origin, the root mean squared error (RMSE) between the starting and ending segments of the trajectory during 50 meters is calculated. As can be seen from Table I, our method always has an advantage in accuracy. However, the introduction of voxel bins and descriptor encoding functions lead to an increase in time consumption. The common orchard robots' map building and localization are mainly used to assist in tasks such as weeding, picking, and spraying, which are the basis of unmanned multi-collaborative operation of orchard robots and high requirements for accuracy. Our approach has the advantage of reliable localization correction at an acceptable time cost in seconds. In particular, in three orchard scenes, our loop closure method shows significant development. Our method achieves 0.26 m, 1.59 m and 0.25 m improvement in orchards respectively, relative to the RMSE of the original LeGO-LOAM trajectory. There are even significant advantages compared to typical SC-LeGO-LOAM methods.

The superiority of our method can be explained as follows: the point cloud is transformed into a binary projection of the SBP representation in the image domain. Because of the exponential growth property of the coding function, the descriptor SBP can effectively distinguish different vegetation such as fruit trees, weeds and shrubs by height. Unlike other scenes that can be distinguished by other significant objects. In the single-object vegetation orchard, the descriptors are highly distinguishable due to the different coding values of the vegetation and non-vegetation cover. Furthermore, aiming at highly similar scenes in the orchard, the attention score map is proposed to highlight task-relevant features. The attention score map reflects the contribution of features for initial fast retrieval of the scene refines features to reduce interference from irrelevant distributions and finally determines the effectiveness of loop closure by overlapping. The whole process is realized from the point cloud spatial geometric distribution to the appearance of the projection descriptor.

In addition, a study on $R_{max}$ (Maximum valid radius) utilizing time consumption and RMSE is shown in Fig. 5. In our approach, in addition to the default parameters, $R_{max}$ can usually be fine-tuned for different scenarios to achieve better performance. As can be seen, as the value of $R_{max}$ increases, the computational cost usually increases, but the performance of the algorithm is better. This is because setting a larger scan radius can encode more point cloud contextual information, making the descriptors more representative. However, the error does not always decrease as the effective radius increases, because although the scan radius is large enough, point clouds beyond a certain range usually contain more noise, which will increase the interference of useless information and bring more

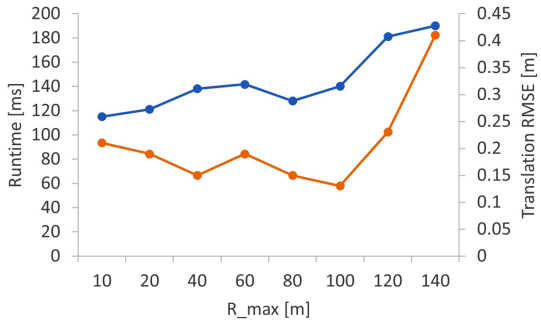



Fig. 5. Study on $R_{max}$ utilizing time consumption and RMSE.

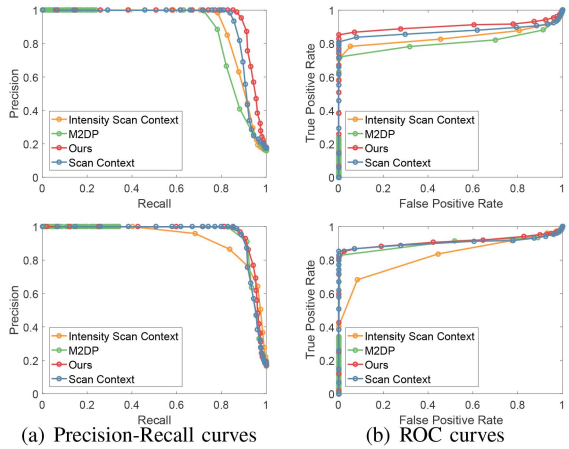
TABLE II
COMPARISON WITH STATE OF THE ART

| Dataset (Pose) | Approach | AUC | F1 |
|---|---|---|---|
| KITTI 00 (4541) | ISC [9] | 0.82 | 0.85 |
| | M2DP [22] | 0.83 | 0.87 |
| | Scan Context [3] | 0.84 | 0.84 |
| | OverlapNet [34] | 0.87 | 0.87 |
| | OverlapTransformer [33] | **0.91** | 0.88 |
| | Ours | **0.91** | **0.92** |
| KITTI 05 (2761) | ISC [9] | 0.84 | 0.87 |
| | M2DP [22] | 0.81 | 0.84 |
| | Scan Context [3] | 0.87 | 0.90 |
| | Ours | **0.91** | **0.93** |

computational consumption. Taking into account the time cost and performance requirements, our algorithm finally takes the optimal value of $R_{max} = 80$ m.

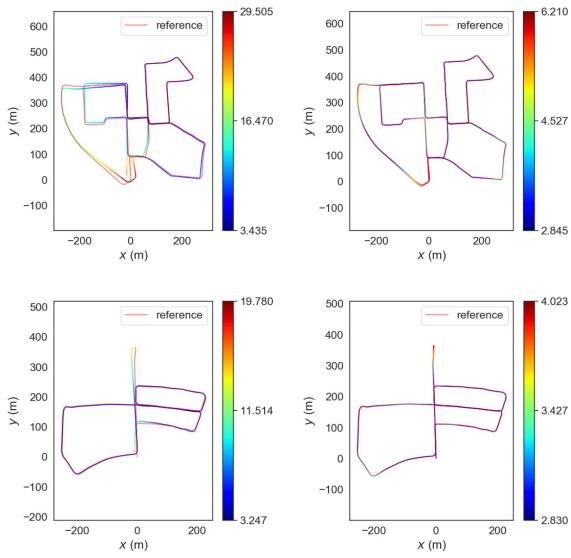
### B. Evaluation on KITTI Dataset

*Experimental Settings:* A Robustness verification experiment to support the claim that our method applies to the general outdoor scenarios, the performance of our method is analyzed using the precision-recall evaluation on KITTI which is an outdoor autonomous driving dataset [38] collected in Germany with a 64-beam LiDAR sensor. The recorded scenarios include various complex scenes such as city, residential, and campuses which are challenging for loop closure detection due to similar architectures and dynamic environments. We select two representative sequences 00 and 05 which contain loop-closure events for algorithm performance evaluation. We compare our method with several state-of-the-art algorithms, including ISC [9], M2DP [22], Scan-Context [3], OverlapNet [34] and OverlapTransformer [33]. We utilize the area under the ROC curve (AUC) and F1 scores for evaluation and the results are shown in Table II. To ensure the fairness of the experiment, we follow the default parameters in Scan Context and ISC as $N_s = 60$, $N_r = 20$ and $L_{max} = 80$ m. The metric method requires the sampling of positive pairs and negative pairs. If a ground truth poses distance between the query and the matched node is less than 4 m which is positive pair, the pair is considered as true-positive only if the two-phase search meets the threshold condition. We examine the results by precision-recall (P-R) and ROC curves in Fig. 6.




(a) Precision-Recall curves (b) ROC curves

Fig. 6. Precision-Recall and ROC curves of different methods on KITTI 05 (the first row) and KITTI 00 (the second row).




(a) LeGO-LOAM (b) Ours (SBP-LeGO-LOAM)

Fig. 7. Transaction and Rotation error on KITTI sequence 00 (The first row) and KITTI sequence 05 (The second row).

*Experimental Analysis:* In Fig. 6, the first row and the second row respectively reflect the performance of the algorithms in KITTI 05 and KITTI 00. From left to right, there are precision-recall (P-R) and ROC curves, our method has the largest area under the curve. As can be seen AUC and F1 scores in Table II, our approach achieves the best performance on KITTI 05 and is also very competitive on KITTI 00. There is a reason for the worse performance of M2DP relies on the histogram and distinguishes places only when the structure of the visible region is substantially different. Evaluation on KITTI datasets shows that the SBP descriptor can be used in other non-single vegetation scenarios and our approach can more easily distinguish positive match pairs from negative match pairs. Because that our encoding method of cylindrical segmentation bins based on LiDAR scan is also suitable for highlighting any objects with geometric differences in space. Fig. 7 shows the full Absolute

Pose Error (APE) error (including translation and rotation error) of the estimated poses with respect to the ground truth (reference). The left figure shows the LeGO-LOAM and the right figure shows ours SBP descriptor integrating to construct SBP-LeGO-LOAM. The colors bar represents the full APE and error amplitude. We can see that after integrating our method, the odometry is more accurate. Since we extract distinctive descriptors based on the height value of the objects in the scene, and in order to solve the problem of high similarity in multiple scans, we highlight task-relavant point clouds with an attention score map.

## V. Conclusion

In this letter, for the specific unstructured, large-scale, long-term orchard, a novel descriptor SBP containing the geometric and density distribution is proposed and an attention score map is introduced to reduce the influence of task-irrelevant features. This idea is incorporated into LiDAR SLAM as a loop closure detection and verified on an actual orchard robot. It is proved that the algorithm can realize the automatic localization requirements of the multi-task robot in the complex orchard. At the same time, the robustness is verified by experiments on public datasets. The future work will conduct more experiments to evaluate the proposed method in more application scenarios.

## References


[1] N. Shalal, T. Low, C. McCarthy, and N. Hancock, "Orchard mapping and mobile robot localisation using on-board camera and laser scanner data fusion–Part A: Tree detection," *Comput. Electron. Agriculture*, vol. 119, pp. 254–266, 2015.

[2] P. M. Blok, K. van Boheemen, F. K. van Evert, J. IJsselmuiden, and G.-H. Kim, "Robot navigation in orchards with localization based on particle filter and Kalman filter," *Comput. Electron. Agriculture*, vol. 157, pp. 261–269, 2019.

[3] G. Kim and A. Kim, "Scan context: Egocentric spatial descriptor for place recognition within 3D point cloud map," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 4802–4809.

[4] A. Papadimitriou et al., "Loop closure detection and SLAM in vineyards with deep semantic cues," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2022, pp. 2251–2258.

[5] S. Saeedi, M. Trentini, M. Seto, and H. Li, "Multiple-robot simultaneous localization and mapping: A review," *J. Field Robot.*, vol. 33, no. 1, pp. 3–46, 2016.

[6] L. C. Santos, A. S. Aguiar, F. N. Santos, A. Valente, J. B. Ventura, and A. J. Sousa, "Navigation stack for robots working in steep slope vineyard," in *Proc. SAI Intell. Syst. Conf.*, 2020, pp. 264–285.

[7] A. S. P. de Aguiar, F. B. N. dos Santos, L. C. F. dos Santos, V. M. de Jesus Filipe, and A. J. M. de Sousa, "Vineyard trunk detection using deep learning–An experimental device benchmark," *Comput. Electron. Agriculture*, vol. 175, 2020, Art. no. 105535.

[8] M. Chen, Y. Tang, X. Zou, Z. Huang, H. Zhou, and S. Chen, "3D global mapping of large-scale unstructured orchard integrating eye-in-hand stereo vision and SLAM," *Comput. Electron. Agriculture*, vol. 187, 2021, Art. no. 106237.

[9] H. Wang, C. Wang, and L. Xie, "Intensity scan context: Coding intensity and geometry relations for loop closure detection," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 2095–2101.

[10] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2009, pp. 3212–3217.

[11] F. Tombari, S. Salti, and L. Di Stefano, "Unique shape context for 3D data description," in *Proc. ACM Workshop 3D Object Retrieval*, 2010, pp. 57–62.

[12] H. J. Kim, E. Dunn, and J.-M. Frahm, "Predicting good features for image geo-localization using per-bundle VLAD," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1170–1178.

[13] H. Jin Kim, E. Dunn, and J.-M. Frahm, "Learned contextual feature reweighting for image geo-localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2136–2145.

[14] S. Xie, S. Liu, Z. Chen, and Z. Tu, "Attentional shapecontextnet for point cloud recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4606–4615.

[15] G. D. Tipaldi and K. O. Arras, "Flirt-interest regions for 2D range data," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2010, pp. 3616–3622.

[16] G. D. Tipaldi, L. Spinello, and W. Burgard, "Geometrical flirt phrases for large scale place recognition in 2D range data," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2013, pp. 2693–2698.

[17] R. B. Rusu, N. Blodow, Z. C. Marton, and M. Beetz, "Aligning point cloud views using persistent feature histograms," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2008, pp. 3384–3391.

[18] S. Salti, F. Tombari, and L. Di Stefano, "SHOT: Unique signatures of histograms for surface and texture description," *Comput. Vis. Image Understanding*, vol. 125, pp. 251–264, 2014.

[19] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.

[20] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 433–449, May 1999.

[21] Y. Wang, Z. Sun, C.-Z. Xu, S. E. Sarma, J. Yang, and H. Kong, "LiDAR iris for loop-closure detection," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 5769–5775.

[22] L. He, X. Wang, and H. Zhang, "M2DP: A novel 3D point cloud descriptor and its application in loop closure detection," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2016, pp. 231–237.

[23] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 652–660.

[24] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet: Deep hierarchical feature learning on point sets in a metric space," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5105–5114.

[25] M. A. Uy and G. H. Lee, "PointNetVLAD: Deep point cloud based retrieval for large-scale place recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4470–4479.

[26] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "NetVLAD: CNN architecture for weakly supervised place recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 5297–5307.

[27] H. J. Kim, E. Dunn, and J.-M. Frahm, "Learned contextual feature reweighting for image geo-localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3251–3260.

[28] Y. Zhu, J. Wang, L. Xie, and L. Zheng, "Attention-based pyramid aggregation network for visual place recognition," in *Proc. 26th ACM Int. Conf. Multimedia*, 2018, pp. 99–107.

[29] T. Naseer, G. L. Oliveira, T. Brox, and W. Burgard, "Semantics-aware visual localization under challenging perceptual conditions," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2017, pp. 2614–2620.

[30] N. Piasco, D. Sidibé, V. Gouet-Brunet, and C. Demonceaux, "Learning scene geometry for visual localization in challenging conditions," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2019, pp. 9094–9100.

[31] G. Peng, Y. Yue, J. Zhang, Z. Wu, X. Tang, and D. Wang, "Semantic reinforced attention learning for visual place recognition," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2021, pp. 13415–13422.

[32] T. Lowe, P. Moghadam, E. Edwards, and J. Williams, "Canopy density estimation in perennial horticulture crops using 3D spinning LiDAR SLAM," *J. Field Robot.*, vol. 38, no. 4, pp. 598–618, 2021.

[33] 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 Robot. Automat. Lett.*, to be published, doi: 10.1109/LRA.2022.3178797.

[34] X. Chen, A. Milioto, E. Palazzolo, P. Giguere, J. Behley, and C. Stachniss, "Suma++: Efficient lidar-based semantic SLAM," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2019, pp. 4530–4537.

[35] X. Chen, A. Milioto, E. Palazzolo, P. Giguere, J. Behley, and C. Stachniss, "SuMa: Efficient LiDAR-based semantic SLAM," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2019, pp. 4530–4537.

[36] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, 2013.

[37] T. Shan and B. Englot, "LeGO-LOAM: Lightweight and ground-optimized LiDAR odometry and mapping on variable terrain," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 4758–4765.

[38] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 3354–3361.