Intensity Scan Context: Coding Intensity and Geometry Relations for Loop Closure Detection

Han Wang, Chen Wang, and Lihua Xie

Abstract— Loop closure detection is an essential and challenging problem in simultaneous localization and mapping (SLAM). It is often tackled with light detection and ranging (LiDAR) sensor due to its view-point and illumination invariant properties. Existing works on 3D loop closure detection often leverage on matching of local or global geometrical-only descriptors which discard intensity reading. In this paper we explore the intensity property from LiDAR scan and show that it can be effective for place recognition. We propose a novel global descriptor, intensity scan context (ISC), that explores both geometry and intensity characteristics. To improve the efficiency for loop closure detection, an efficient two-stage hierarchical re-identification process is proposed, including binary-operation based fast geometric relation retrieval and intensity structure re-identification. Thorough experiments including both local experiment and public datasets test have been conducted to evaluate the performance of the proposed method. Our method achieves better recall rate and recall precision than existing geometric-only methods.

I. INTRODUCTION

Loop closure detection, aka place recognition, typically refers to the capability of identifying a re-visited place. In a SLAM system, the estimation of robot's state and trajectory often comes with inevitable drift that fails localization easily [1]. By identifying loop pairs, a robot can eliminate drifting error. More importantly, it can also prevent multiple registration of identical landmarks so that a globally consistent map can be created. A revisited place usually varies from light illumination, weather or viewing angle, which is difficult to be solved in vision based place recognition. However, LiDAR is less affected by such environmental changes and hence is widely used for place recognition in the recent years.

Existing works on place recognition often leverage on matching of 3D descriptors such as fast point feature histogram (FPFH) [2], fast laser interest region transform (FLIRT) [3] and signature of histograms of orientations (SHOT) [4]. These descriptors explore either global or local geometry information such as surface normal or neighbour points distribution, leaving the intensity information unused. A main justification is that intensity information is less straightforward than geometry reading [5], since it is affected by not only target surface characteristics (*e.g.*, roughness, surface reflectance), but also acquisition geometry (*e.g.*,
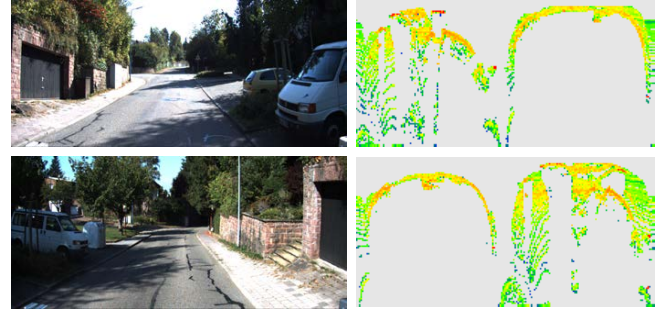
Fig. 1: Example of loop closure detected from KITTI sequence 02. The scenario is challenging due to reverse visit. Our proposed intensity scan context on the right images shows high similarity identifies loop closure.

distance) and instrument effects (*e.g.*, transmitted energy) [6]. However, the intensity channel reveals the reflectance structure of surrounding environment, *e.g.*, retro-reflective material such as metal plate usually returns high value and concrete returns low value. This information is often unique for different places. Moreover, recent works have shown some preliminary results that intensity reading can be effective for place recognition [7], [8].

In this paper, we propose a novel global descriptor, intensity scan context (ISC), that integrates both geometry and intensity characteristics for loop closure detection. We first explain how intensity information can be distinguishable to represent a place. Then we propose intensity scan context as a global signature for place recognition. To further improve the efficiency for loop closure detection, a two-stage hierarchical place re-identification strategy is proposed, including binary-operation based fast geometry retrieval and intensity structure matching. To evaluate the performance of the proposed intensity scan context, our method is tested under different scenarios including outdoor autonomous driving and indoor warehouse robot navigation. The results show that the proposed approach achieves higher recall rate and recall rate than existing geometric-only loop closure detection methods.

The main contributions of this paper are as follows:

- We propose a novel global descriptor for 3D LiDAR scan that integrates both geometry and intensity characteristics.
- An efficient loop closure detection strategy based on a two-stage hierarchical Intensity Scan Context re-

identification is proposed. It only costs 1.2 *ms* per query on average.

- A thorough evaluation on the proposed descriptor, including both local experiment and public datasets test, is conducted.

This paper is organized as follows: Section II reviews the related works on both vision based and LiDAR based approaches for loop closure detection. Section III describes the idea of using Intensity Scan Context for place recognition, followed by the two-stage hierarchical place re-identification. Section IV shows experiment results and comparison with existing works, followed by conclusion in Section V.

II. RELATED WORK

Existing works on loop closure detection consist of vision based methods and LiDAR based methods according to different perception systems used. Vision based approaches are developed for place recognition in the early stage. Those methods often leverage on bag of words model (BoW) that measures the distance of visual words according to a pre-trained visual vocabulary, *e.g.*, FAB-MAP [9] and DBoW2 [10]. They are widely used in visual SLAM such as ORB SLAM [11] and LDSO [12]. However, image stream is not resistant to light illumination or view-point so that vision based place recognition is not robust enough in practice. Recently some works target to solve environmental changes such as [13], [14], but are limited to some specific scenarios.

In comparison, LiDAR detection is more robust due to its illumination and view-point invariant properties and is subsequently introduced for loop closure detection. Existing works on LiDAR based place recognition strive for an efficient local descriptor or place signature that can accurately and concisely present a place. One of the most popular local descriptors is fast point feature histogram (FPFH) [2] which explores the local surface normal of each neighbour point. It is effective in estimating affine transform between two point clouds and is used for place recognition in the later work such as [15]. Bosse *et al.* propose a probabilistic voting approach based on Gasalt3D descriptor [16]. Despite the good performance achieved, the point cloud retrival is inefficient due to the high dimension of the proposed descriptor.

Re-identification on the local descriptor usually requires key-point extraction and massive local geometry calculation. In comparison, matching on global descriptor is more efficient in place recognition. Rizzini introduces a novel descriptor named GLAROT that encodes the relative geometric position of key-point pairs into a histogram [17]. The experiment result achieves very satisfactory recall precision and recall rate. However, building key-point relation is still computationally expensive. Kim *et al.* propose scan context which projects laser scan into global descriptor. The matching of scan context only requires element-wise multiplication so that the query speed is fast. However, the matching precision is not high enough and false positive occurs during public datasets test.
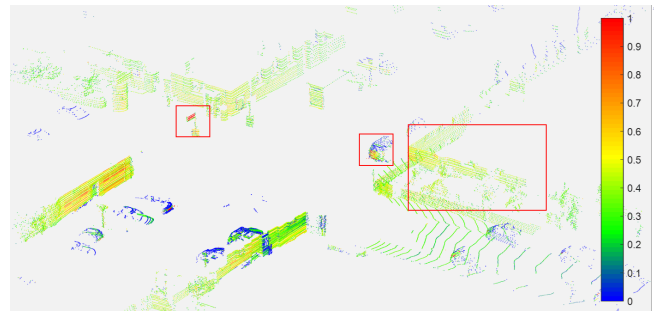
(a) Intensity scan reading of a crossroad.

(b) Camera view of the same place.

Fig. 2: An example of intensity reading from KITTI dataset. The relationship of intensity information and landmarks are highlighted with red rectangles.

The global descriptors are more efficient but the performance is not competitive. However, recent works find that the performance can be improved by integrating intensity characteristics. Guo *et al.* justify that intensity information can be distinctive for places and propose a novel local descriptor called intensity signature of histograms of orientations (ISHOT) that consists of both geometry and intensity information. The place re-identification is solved by a probabilistic voting strategy similar to [16]. Despite computationally expensive, the new descriptor outperforms geometrical-only descriptors in place recognition. This inspires our work on finding a more efficient and more accurate descriptor for loop closure detection.

III. METHODOLOGY

In this section, the proposed method is described in detail. We firstly present the concept of intensity scan and briefly explain on how intensity characteristics can be used for place recognition. Then we present the idea of using intensity scan context as a global 3D descriptor. Furthermore, an efficient two-stage hierarchical intensity scan context based retrieval is introduced, consisting of fast binary-operation based geometry retrieval and intensity structure matching.

A. Intensity Calibration and Pre-processing

LiDAR perceives the environment by emitting and receiving laser beam. Generally, the distance value is measured by traveling time while the surface reflectance can be estimated by returned energy level (*i.e.*, intensity). The intensity reading reveals surrounding surface reflectance structure. Existing works on LiDAR have shown that the returned
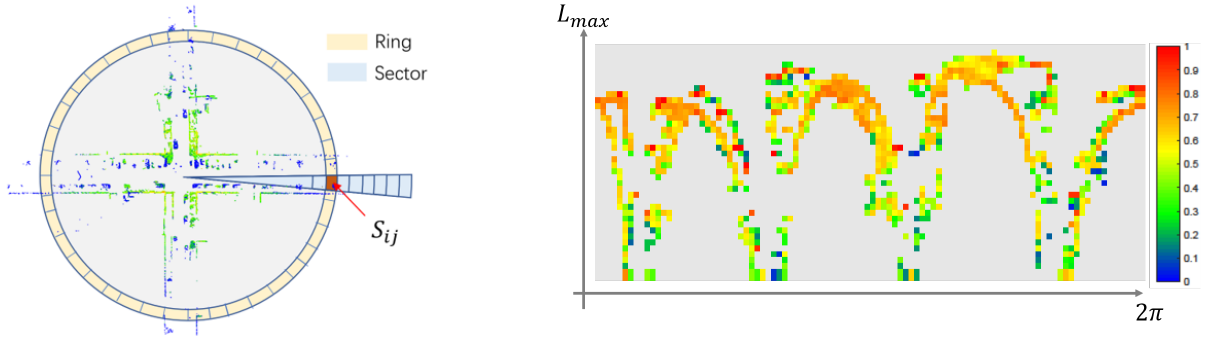
Fig. 3: A visual illustration of the proposed intensity scan context. Left figure: original point cloud is decomposed into subspace based on geometry characteristics. Right figure: derived intensity scan context by intensity projection on the subspace.

intensity readings vary for different objects [6], *e.g.*, retro-reflective material such as metal plate usually returns high value and concrete returns low value. In Fig. 2, we show an example from KITTI dataset [18] for demonstration. The point cloud and image present the same place and the intensity structure is interpreted as color in the first image. We pick 3 landmarks including car, road sign and building respectively and highlight them with red rectangles in both Fig. 2 (a) and Fig. 2 (b). It is observed that the road sign is highly distinguishable with low energy loss while the building structure (concrete) returns medium intensity. Moreover, the reflectance is consistent within object.

However, the intensity channel is noisy since it is affected by not only target surface characteristics (*e.g.*, roughness, surface reflectance), but also acquisition geometry (*e.g.*, distance) and instrument effects (*e.g.*, transmitted energy) [6]. Hence calibration is necessary in order to reduce the disturbance by other factors. Similar to [19], [8], in this work we calibrate the intensity reading η_r with a mapping function φ:

$$\eta_{cal} = \varphi(\eta_r, d), \tag{1}$$

where d is the distance reading. The mapping function φ describes the influence of distance on the received energy and can be collected based on offline experiment. Besides remapping, the intensity return from 3D LiDAR is an eight-byte integer, *e.g.*, Velodyne VLP-16. It is re-scaled into [0,1] as a float number for convenience.

In the pratical applications, some pre-processing of the LiDAR scan is necessary in order to remove the redundant information. Based on observation, the LiDAR noise increases with distance. Therefore, the LiDAR reading is firstly filtered by setting a distance threshold L_{max} to remove unreliable points. Moreover, the ground points are often not significant for place recognition so that they are optimized by the method similar to [20] in advance.

B. Intensity Scan Context

Extraction of local descriptors is often computationally expensive since we need to identify local geometry characteristics such as local norm for every single key-point.

Hence to increase the computational efficiency, the intensity information is interpreted as global descriptor in this work. Inspired by scan context [21], [22] and shape context [23], we introduce intensity scan context that can efficiently integrate both geometry and intensity characteristics into a global signature.

Denote the intensity reading η, geometry reading $[x, y, z]$, number of points n, a LiDAR scan is defined as $\mathcal{P} = \{\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_n\}$ with each point $\mathbf{p_k} = [x_k, y_k, z_k, \eta_k]$ in the local Cartesian coordinate. Each point p_k can be converted into polar coordinate, but only in x-y plane, so that

$$\begin{aligned}
\mathbf{p_k} &= [\rho_k, \theta_k, z_k, \eta_k], \\
\rho_k &= \sqrt{x_k{}^2 + y_k{}^2}, \\
\theta_k &= \arctan \frac{y_k}{x_k}.
\end{aligned} \tag{2}$$

The point cloud is then segmented by equally diving polar coordinate in azimuthal and radial directions into N_s sectors and N_r rings. Each segment is represented by:

$$\begin{aligned}
\mathcal{S}_{ij} = \{\mathbf{p_k} \in \mathcal{P} | \frac{i \cdot L_{max}}{N_r} &\leq \rho_k < \frac{(i+1) \cdot L_{max}}{N_r}, \\
\frac{j \cdot 2\pi}{N_s} - \pi \leq \theta_k &< \frac{(j+1) \cdot 2\pi}{N_s} - \pi\},
\end{aligned} \tag{3}$$

where $i \in [|1, N_s|]$, $j \in [|1, N_r|]$ and the symbol $[|1, N|]$ represents $\{1, 2, \cdots, N\}$. The point cloud is then divided into $N_s \times N_r$ subspace. As discussed before, the intensity reading is often consistent for the same object. Since the each subspace is much smaller than whole point cloud, we can assume that the intensity reading does not vary too much. Hence for each subspace, a coding function κ can be applied to reduce the intensity dimension. It is defined as:

$$\begin{aligned}
\eta_{ij} &= \kappa(\mathcal{S}_{ij}) \\
&= \max_{p_k \in \mathcal{S}_{ij}} \eta_k.
\end{aligned} \tag{4}$$

Note that if $\mathcal{S}_{ij} \in \varnothing$ (*i.e.*, no scan data), $\eta_{ij} = 0$. Up to this point, we can generate the intensity scan context Ω by:

$$\Omega(i, j) = \eta_{ij}. \tag{5}$$

The global signature Ω is a 2D matrix that reveals both geometry and intensity distribution of the environment. We pick a LiDAR scan from KITTI dataset [18] for illustration in Fig. 3. The left figure is the point cloud from top down view with the sub point cloud space S_{ij} marked in red color. The right figure is the coded intensity scan context with each pixel calculated by intensity coding function.

C. Place Re-identification

Place recognition targets to match current place $\mathcal{P}_n$ with the previously visited places from the historical database $\mathcal{D} = \{\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_{n-1}\}$. As more places visited, the scale of database $\mathcal{D}$ inevitably increases so that the computational cost grows accordingly. To reduce the computational cost, in this section we propose a two-stage hierarchical intensity scan context retrieval strategy that makes use of fast binary operation to speed up the process of place re-identification.

1) Fast Geometry Re-identification: Most of 3D descriptors are histogram based such as unique shape context (USC) [4], ISHOT [8], *etc*. Matching between histograms can be slow in practice since mathematical operation is inevitable, *e.g.*, multiplication of float numbers. In comparison, binary operation (or logical operation) achieves much faster speed than those mathematical operation. Inspired by [24], we introduce an efficient binary operation geometry re-identification for fast indexing. Given an intensity scan context Ω, its geometry distribution on the local coordinate can be represented as a binary matrix $\mathcal{I}$:

$$\mathcal{I}(x,y) = \begin{cases} false, & \text{if } \Omega(x,y) = 0 \\ true, & \text{otherwise} \end{cases} \tag{6}$$

For a query intensity scan context Ω^q, a candidate intensity scan context Ω^c and their binary transform $\mathcal{I}^q$, $\mathcal{I}^c$, the geometry similarity can be derived as:

$$\varphi_g(\mathcal{I}^q, \mathcal{I}^c) = \frac{XOR(\mathcal{I}^q, \mathcal{I}^c)}{|\mathcal{I}^q|}, \tag{7}$$

where $|\mathbf{x}|$ is the total number of elements in $\mathbf{x}$ and $XOR(\mathbf{x}, \mathbf{y})$ refers to element-wise exclusive OR operation between matrix $\mathbf{x}$ and $\mathbf{y}$. Because the column vector in intensity scan context represents azimuthal direction, the rotation of laser scan becomes column shift in intensity scan context [21]. Hence for place recognition, the viewing angle change can be interpreted as column-shifts of Ω. Therefore, to detect the view-point change, the final score is calculated by:

$$\Phi_g(\mathcal{I}^q, \mathcal{I}^c) = \max_{i \in [|1, N_s|]} \varphi_g(\mathcal{I}_i^q, \mathcal{I}^c), \tag{8}$$

where $\mathcal{I}_i^q$ is $\mathcal{I}^q$ shifted by i^{th} column. In the meantime, we can identify the best matched $\mathcal{I}_k^q$ with column-shifts of k that can be used to correct viewing angle change. The unmatched pair can be filtered out by setting an empirically determined threshold ϵ_g. In experiments we find that binary matching only costs 0.5 *ms* on a desktop computer that is very computationally efficient.
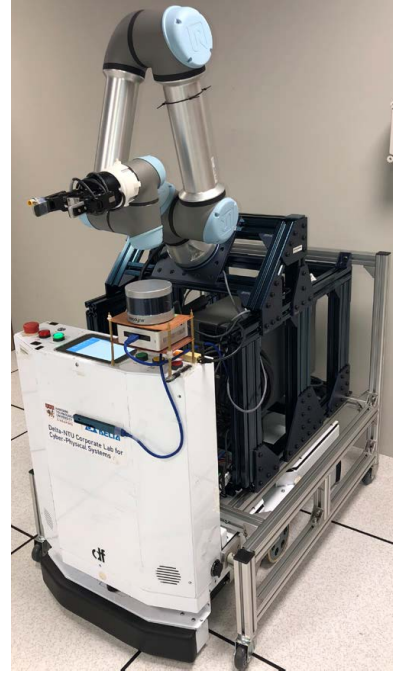

Fig. 4: Autonomous warehouse robot platform used for our experiment.

2) Intensity Structure Matching: The second stage mainly identifies the intensity similarity between two intensity scan context Ω^q and Ω^c by column-wise comparison. Let $\mathbf{v}_i^q$ and $\mathbf{v}_i^c$ be the ith column of Ω^q and Ω^c, the score can be found by taking cosine distance:

$$\varphi_i(\Omega^q, \Omega^c) = \frac{1}{N_s} \sum_{i=0}^{N_s-1} \left(\frac{\mathbf{v}_i^q \cdot \mathbf{v}_i^c}{\|\mathbf{v}_i^q\| \cdot \|\mathbf{v}_i^c\|} \right). \tag{9}$$

Similar to geometry matching, we have to correct viewing angle change. Since the viewing angle change k is already identified from geometry re-identification, we compare Ω_k^q and Ω^c where Ω_k^q is Ω^q shifted by k^{th} column. This also significantly reduces the computational cost because comparison of intensity involves mathematical operation rather than logical operation. The final score is calculated as:

$$\Phi_i(\Omega^q, \Omega^c) = \varphi_i(\Omega_k^q, \Omega^c). \tag{10}$$

The unmatched pair can be also filtered out by setting an empirically determined threshold ϵ_i.

D. Consistency Verification

As discussed in previous section, global descriptor is a highly simplified representation of original point cloud. Hence it is inevitable to have some features ignored, which can lead to false positive. Therefore it is necessary to check consistency before closing the loop.

1) Temporal consistency check: In a SLAM system, it is observed that the occurrence of single loop closure often implies high similarity on the neighbour LiDAR scans since the sensor feedback is continuous in time [25]. We can verify
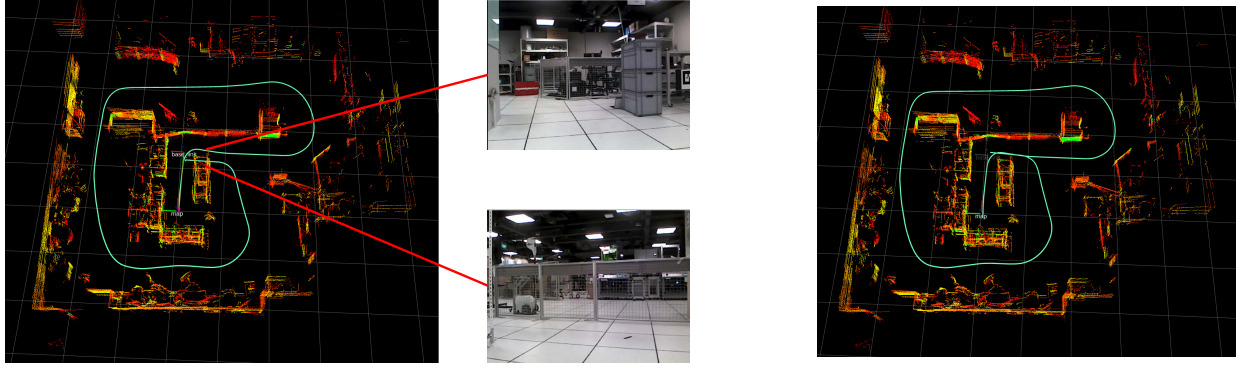
(a) Loop not detected with visual place recognition.　　　　(b) Loop identified with the proposed approach.

Fig. 5: Local experiment result of the proposed algorithm.

the loop closure by measuring the temporal consistency:

$$P(\mathcal{P}_m, \mathcal{P}_n) = \frac{1}{N} \sum_{k=1}^{N} (\Phi_g(\mathcal{I}_{m-k}, \mathcal{I}_{n-k}) + \Phi_i(\Omega_{m-k}, \Omega_{n-k})),$$
(11)

where N is the number of frames included for temporal consistency verification. Note that in case of reverse visit (viewing angle change detected in geometry re-identification), $\mathcal{I}_{m-i}$ becomes $\mathcal{I}_{m+i}$ accordingly. The loop candidate can be accepted by taking threshold ξ on the final temporal consistency score.

2) Geometrical consistency: The geometrical consistency verifies the raw scan-to-scan similarity. Similar to [8], FPFH features are extracted and matched to get an initial guess of the rigid transform matrix. Starting from this initial estimate, iterative closest point (ICP) [26] is applied to find minimal distance error between the query scan and candidate scan.

IV. EXPERIMENT EVALUATION

A. Experiment Setup

In this section we present preliminary results from our proposed method, including both indoor warehouse robot navigation and outdoor autonomous driving experiments. The proposed method is implemented in C++ and is integrated to robot operating system (ROS) on an Intel NUC mini computer. For indoor warehouse navigation, an autonomous guided vehicle (AGV) equipped with Velodyne VLP-16 and Intel Realsense r200 is used for the experiment. As shown in Fig. 4, the robot is developed for warehouse manipulation tasks such as packaging and transportation which require high accuracy on long-term localization. The maximum speed of AGV is 1 *m/s*. For outdoor autonomous driving, KITTI dataset [18] is used for evaluation. The proposed place recognition approach is used to reduce localization drifts and improve the mapping accuracy. A list of parameters used in the experiment is shown in Table I.

B. Experiment on Autonomous Robot

The robot is tested in real warehouse environment consisting of operating machines, shelves, human, *etc*. LiDAR odometer is estimated based on feature points matching via

point cloud library (PCL) [27] and the robot trajectory (front-end SLAM) are collected from the fusion of both wheel odometer and LiDAR odometer. To illustrate, we simulate a common task where the robot leaves the docking station to fetch materials and come back in a reverse direction. Our approach is compared with vision based approach tested based on a front-mounted camera. In particular, we use DBoW2 [10] which is implemented in ORB SLAM [11]. The result is shown in Fig. 5 with the estimated trajectory plotted in green color. For DBoW2, the reverse visit changes the view angle significantly so that there is not enough similarity to conclude loop closure. Hence the trajectory collides with shelves according to the final result, which is incorrect. In comparison, our proposed method is able to identify the revisited place. By applying global optimization, the resultant trajectory and the mapping result is much more reasonable. This is due to the view angle invariant property of our proposed two-stage hierarchical Intensity Scan Context based retrieval.

C. Evaluation on Public Dataset

To further demonstrate the performance of the proposed method, we test the algorithm on KITTI dataset which is commonly used for place recognition. KITTI dataset is collected from an autonomous car equipped with various perception systems including front-mounted cameras, Velodyne HDL-64E LiDAR, GPS, *etc*. The recorded scenarios are challenging for loop closure detection due to similar architectures and dynamic environment. The proposed method is tested with multiple recordings such as sequence 00, 02

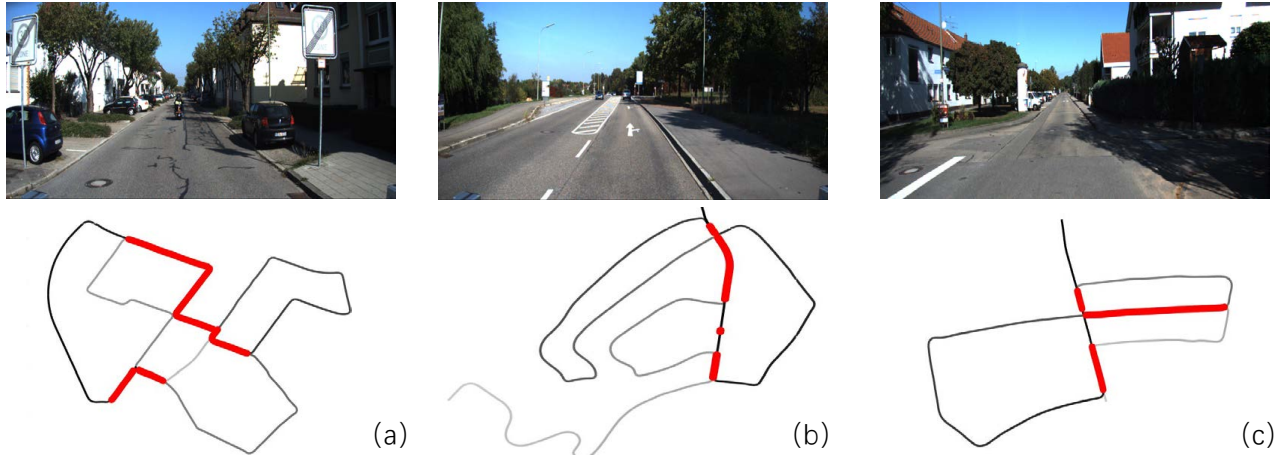| Parameter	Description	Value
L_{max}	Maximum radius	50
N_s	Number of sectors	20
N_r	Number of rings	60
ϵ_g	Geometry matching threshold	0.9
ϵ_i	Intensity matching threshold	0.92
N	Frames used in temporal consistency check	5
ξ	Temporal consistency checking threshold	1.8

TABLE I: Parameter List.

Fig. 6: Loop closure detection result. (a) KITTI sequence 00. (b) KITTI sequence 02. (c) KITTI sequence 05.

and 05. Sequence 00 and 05 are commonly used for place recognition since most of loop closure places are forward visited. Sequence 02 contains both forward and reverse visit and is considered more challenging for place recognition.

Recall rate and precision are collected from each test. Precision is the percentage of successful pairing and recall rate is the reported loop pairs against total loop pairs. Higher recall rate can significantly reduce drifting error and higher recall precision can prevent wrong registration of points into map. The number of total loop closure is collected based on GPS result since the test-field is large in scale. The results of our proposed approach can be found in Fig. 6, with the GPS trajectory plotted from light to dark with time going. Each of the loop closure place detected is marked with red color.

The results are also compared with existing works such as [21], [17], [24] that are commonly used in SLAM applications. We also include the state-of-the-art vision based place recognition methods such as [10] for comparison. The experiment results of [21], [17], [24] are collected from respective paper due to unavailability of source code and [10]

is tested with local experiment. The results are listed in Table II. Compare to vision based approach, our approach achieves competitive precision and recall rate on both sequence 00 and sequence 05. On more challenging dataset sequence 02, our approach achieves much higher recall rate. This is because in sequence 02 vision based approach fails to identify the reverse visit so that the recall rate drops significantly. However, false positive is reported for LiDAR based approach such as our method and [21]. The failure case comes from non-residential area where both sides of roads are trees so that the geometry and intensity characteristics are very limited for place recognition. Compared to Lidar based approach such as [21], [17], [24], we achieve both higher recall precision and recall rate across all three datasets. In the meanwhile, the proposed intensity scan context only costs 1.2 *ms* for each query which is very efficient in practice.

V. CONCLUSION

In this paper, we presented a robust loop closure detection approach by integrating both geometry and intensity information. Existing works on LiDAR based loop closure detection mainly leverage on the geometric-only descriptor and ignore intensity reading. Inspired by the recent researches on LiDAR intensity, we argue that the intensity information can be effective for place recognition and propose a global 3D descriptor named intensity scan context. To reduce the computational cost, an efficient two-stage hierarchical intensity scan context retrieval is proposed, consisting of fast binary-operation based geometry indexing and intensity structure re-identification. It costs only 1.2 *ms* for each query on a normal computer in practice. Thorough experiments have been conducted including local run test with autonomous warehouse robot and public dataset test to evaluate our proposed method. The results show that our proposed method achieves competitive recall precision and recall rate compared to existing state-of-the-art methods.

Dataset	Approaches	Precision (%)	Recall Rate (%)
sequence 00	Kim [21]	100	87
	GLAROT3D [17]	86	40
	Cieslewski [24]	92	80
	Gálvez-López [10]	100	92
	Proposed	**100**	**90.2**
sequence 02	Kim [21]	90	73
	Gálvez-López [10]	100	80.6
	Proposed	**98**	**91**
sequence 05	Kim [21]	100	90
	Cieslewski [24]	93	60
	GLAROT3D [17]	80	80
	Gálvez-López [10]	100	87.6
	Proposed	**100**	**91.2**

TABLE II: Comparison with existing methods.

REFERENCES

[1] A. Angeli, D. Filliat, S. Doncieux, and J.-A. Meyer, "A fast and incremental method for loop-closure detection using bags of visual words," *IEEE Transactions on Robotics*, pp. 1027–1037, 2008.

[2] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (fpfh) for 3d registration," in *2009 IEEE International Conference on Robotics and Automation*, 2009, pp. 3212–3217.

[3] G. D. Tipaldi and K. O. Arras, "Flirt-interest regions for 2d range data," in *2010 IEEE International Conference on Robotics and Automation*, 2010, pp. 3616–3622.

[4] F. Tombari, S. Salti, and L. Di Stefano, "Unique shape context for 3d data description," in *Proceedings of the ACM workshop on 3D object retrieval*, 2010, pp. 57–62.

[5] C. Wang, J. Yuan, and L. Xie, "Non-iterative slam," in *Advanced Robotics (ICAR), 2017 18th International Conference on.* IEEE, 2017, pp. 83–90.

[6] A. Kashani, M. Olsen, C. Parrish, and N. Wilson, "A review of lidar radiometric processing: From ad hoc intensity correction to rigorous radiometric calibration," *Sensors*, vol. 15, no. 11, pp. 28 099–28 128, 2015.

[7] K. P. Cop, P. V. Borges, and R. Dubé, "Delight: An efficient descriptor for global localisation using lidar intensities," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 3653–3660.

[8] J. Guo, P. V. Borges, C. Park, and A. Gawel, "Local descriptor for robust place recognition using lidar intensity," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1470–1477, 2019.

[9] A. Glover, W. Maddern, M. Warren, S. Reid, M. Milford, and G. Wyeth, "Openfabmap: An open source toolbox for appearance-based loop closure detection," in *2012 IEEE international conference on Robotics and automation (ICRA)*, 2012, pp. 4730–4735.

[10] D. Gálvez-López and J. D. Tardos, "Bags of binary words for fast place recognition in image sequences," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1188–1197, 2012.

[11] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[12] X. Gao, R. Wang, N. Demmel, and D. Cremers, "Ldso: Direct sparse odometry with loop closure," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 2198–2204.

[13] A. Anoosheh, T. Sattler, R. Timofte, M. Pollefeys, and L. Van Gool, "Night-to-day image translation for retrieval-based localization," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 5958–5964.

[14] T. Naseer, M. Ruhnke, C. Stachniss, L. Spinello, and W. Burgard, "Robust visual slam across seasons," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 2529–2535.

[15] R. Dubé, D. Dugas, E. Stumm, J. Nieto, R. Siegwart, and C. Cadena, "Segmatch: Segment based place recognition in 3d point clouds," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 5266–5272.

[16] M. Bosse and R. Zlot, "Place recognition using keypoint voting in large 3d lidar datasets," in *2013 IEEE International Conference on Robotics and Automation*, 2013, pp. 2677–2684.

[17] D. L. Rizzini, "Place recognition of 3d landmarks based on geometric relations," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 648–654.

[18] 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.

[19] J. Levinson and S. Thrun, "Unsupervised calibration for multi-beam lasers," in *Experimental Robotics*. Springer, 2014, pp. 179–193.

[20] T. Shan and B. Englot, "Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 4758–4765.

[21] G. Kim and A. Kim, "Scan context: Egocentric spatial descriptor for place recognition within 3d point cloud map," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 4802–4809.

[22] G. Kim, B. Park, and A. Kim, "1-day learning, 1-year localization: Long-term lidar localization using scan context image," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1948–1955, 2019.

[23] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 4, pp. 509–522, 2002.

[24] T. Cieslewski, E. Stumm, A. Gawel, M. Bosse, S. Lynen, and R. Siegwart, "Point cloud descriptors for place recognition using sparse visual information," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 4830–4836.

[25] M. J. Milford and G. F. Wyeth, "Seqslam: Visual route-based navigation for sunny summer days and stormy winter nights," in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 1643–1649.

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

[27] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 9-13 2011.