

GOSMatch: Graph-of-Semantics Matching for Detecting Loop Closures in 3D LiDAR data

Yachen Zhu¹, Yanyang Ma¹, Long Chen¹, Cong Liu¹, Maosheng Ye², and Lingxi Li³

Abstract—Detecting loop closures in 3D Light Detection and Ranging (LiDAR) data is a challenging task since point-level methods always suffer from instability. This paper presents a semantic-level approach named GOSMatch to perform reliable place recognition. Our method leverages novel descriptors, which are generated from the spatial relationship between semantics, to perform frame description and data association. We also propose a coarse-to-fine strategy to efficiently search for loop closures. Besides, GOSMatch can give an accurate 6-DOF initial pose estimation once a loop closure is confirmed. Extensive experiments have been conducted on the KITTI odometry dataset and the results show that GOSMatch can achieve robust loop closure detection performance and outperform existing methods.

I. INTRODUCTION

Loop closure detection is a problem associated with identifying places visited previously. It is a crucial part of Simultaneous Localization and Mapping (SLAM). During SLAM, pairwise scan matching odometry inevitably accumulates pose drift. Reliable loop closure detection is a key technique for SLAM systems to correct the drift error [1]. Though many vision-based methods have been proposed in recent years [2]–[4], it may draw unreliable results in the cases of dramatic changes on illumination [5] or viewpoint [6]. LiDAR, unlike cameras [7], senses the surrounding environment by generating high resolution 3D points with accurate measurements. It can work under unfavorable illumination conditions but also provide more geometric information. Therefore, the LiDAR-based loop closure detection task has attracted significant research attention.

Generally, traditional LiDAR-based methods use local keypoints [8]–[11] or other global features [12]–[18] to extract pointwise descriptors from point clouds and subsequently compare the descriptor of the query scan with the descriptors of historical scans to finally recognize the previously-visited places. These types of approaches place much emphasis on local details but ignore high-level feature constraints. Instead of defining revisited places by geometric points or other point-level features, humans understand the entire scenes from a more macro perspective [19], by recognizing objects and their relative positions in three-dimensional space.

¹Y. Zhu, Y. Ma, L. Chen, and C. Liu are with the School of Data and Computer Science, Sun Yat-sen University, China. Corresponding author: Long Chen. chen14@mail.sysu.edu.cn

²M. Ye is with the School of Geodesy and Geomatics, Wuhan University, China.

³L. Li is with Dept. of Electrical and Computer Engineering, Purdue School of Engineering and Technology, Indiana University-Purdue University Indianapolis, USA.

Inspired by the way humans identify places, we propose a novel approach that utilizes semantic objects in the scenes to detect loop closures. A couple of novel global and local descriptors formed by these semantic objects are introduced. The global descriptors are designed to efficiently seek the top similar loop candidates and the local descriptors are used to calculate the one-to-one correspondences of the semantics in two point clouds. A geometric verification step is then exploited to identify loop closures.

To the best of our knowledge, this is the first work that leverages object-level semantics to detect loop closures in 3D laser data. The main contributions of this paper are summarized as follows:

- We propose GOSMatch, an object-based approach for reliable place recognition in the urban driving environment based on LiDAR-only observations.
- New kinds of global and local descriptors are investigated for efficiently two-step loop searching, which encode spatial relationship between semantic objects.
- We perform extensive experimental validation against other state-of-the-art methods on a large public dataset and also publish the implementation of GOSMatch at <https://github.com/zhuochen/GOSMatch>.

The rest of this paper is organized as follows: Section II reviews the previous literature our work relates to. Next, we describe the presented loop closure detection framework in Section III. Section IV shows the experiment results on the KITTI odometry dataset. Finally, we conclude in Section V.

II. RELATED WORK

Although detecting loop closures from 3D point clouds has been intensively studied in the past decades, it remains an open problem in SLAM systems.

A. Traditional LiDAR-based place recognition

One kind of methods utilize local features for place recognition. Bosse and Zlot [8] extract regional shape descriptors from randomly selected keypoints in the point clouds. Subsequently, a voting strategy is used to find the nearest neighbor among the keypoints to recognize places. Steder et al. [9] extract feature description vector of each keypoint in a transformed range image. Then a kd-tree is exploited to process the high dimensional vectors efficiently. To achieve a higher recognition performance, the Normal Aligned Radial Feature (NARF) local descriptor and a bag-of-words (BoW) matching approach are employed in their extension work [10].

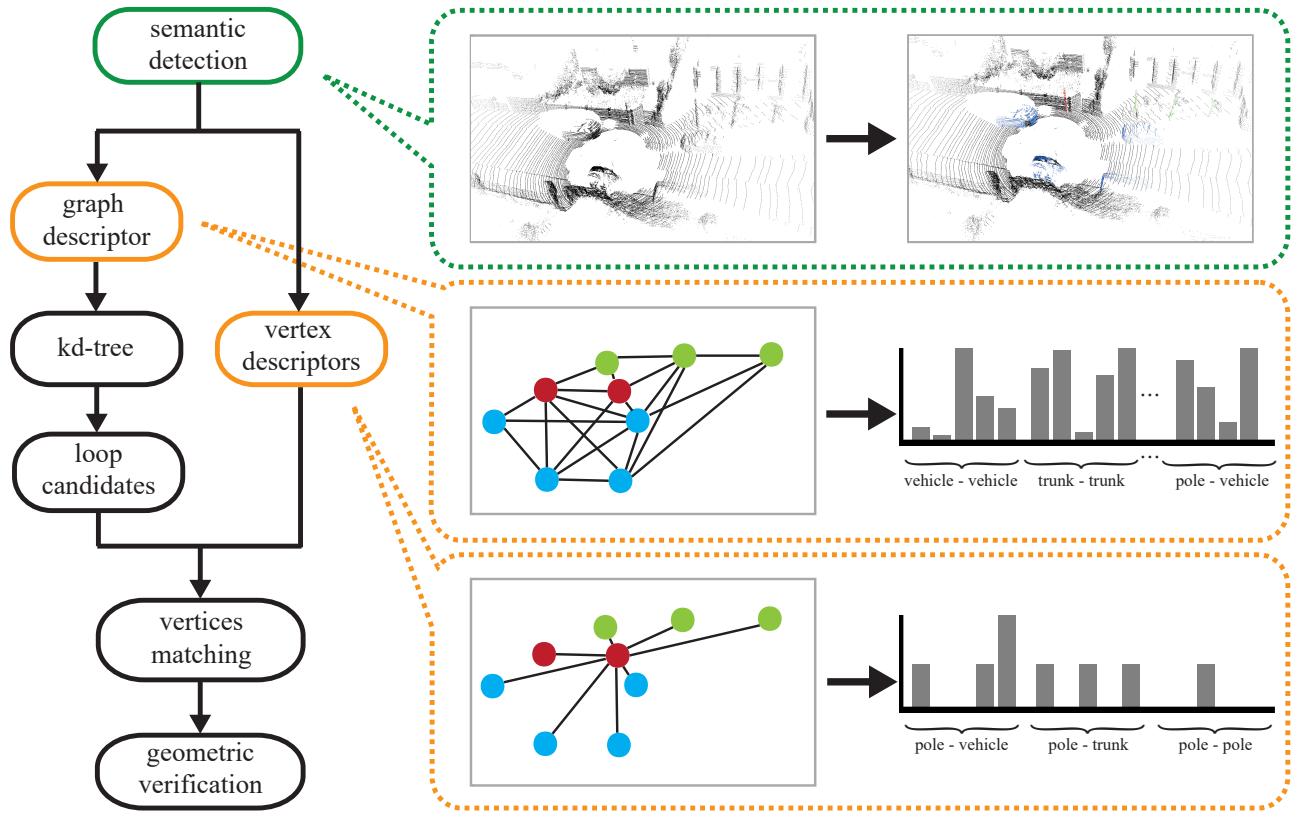


Fig. 1. The block diagram of GOSMatch. First, we segment the semantics from the raw point cloud. Then, a histogram-based *graph descriptor* is established for a fast nearest neighbor search. Note that we omit some edges in the undirected complete graph and we reduce it to a 2D representation for clarity. Next, we compare the *vertex descriptors* of the query point cloud and the retrieved candidates, a 6D initial pose estimation and a loss value are eventually derived from the geometric verification step.

Another kind of typical approaches focus on extracting the global representation of point cloud, which are usually presented in the form of histograms, such as VFH [17], ESF [18] and Z-Projection [16]. An appearance-based method [12] is proposed based on the Normal Distribution Transform (NDT) algorithm. After discretizing the clouds into cubes, the shape properties of each cell can be determined from the covariance matrix. By combining them, a surface shape histogram representing the whole cloud can be obtained. Giseop Kim [15] proposes Scan Context, a descriptor that encodes the max height of points in each divided bin into a 2D matrix. Since matching the matrix-type descriptors requires heavy computational cost, Kim extracts ring key from Scan Context to build kd-tree and speed up the searching process. He et al. [14] project 3D points on multiview 2D planes and concatenate all singular vectors derived from these planes as the point cloud descriptor. Rohling et al. [20] propose a 1D histogram of point range distribution as a global point cloud descriptor. The Wasserstein metric is then used to compare the histograms for place recognition.

B. Place recognition based on high-level descriptors

While local descriptors always lack the overall-description ability and global descriptors suffer from rotational variance problem easily, Shan et al. [21] instead extract more distinc-

tive descriptors like edge and planar features from ground points and segmented points respectively. Then, Iterative Closest Point (ICP) is performed to match feature correspondences when finding a loop closure. Dube et al. propose SegMatch [22] that match clustering segments to obtain a more general and robust solution for place recognition. However, SegMatch takes so much computing cost that the whole loop closure detection framework can only update at 1 Hz.

C. Data association with graph representation

Graph matching plays an important role in coping with the pairwise data association problem. Specifically, Graph representation is a usual approach to describe the objects as well as their topology. In this case, finding the association of the objects between two places is therefore transformed into calculating the vertices and edges correspondences between graphs. However, finding an exact solution to this problem is always NP-hard. Bailey et al. [23] generate a correspondence graph and then search maximum clique in it to obtain exact correspondences between vertices and edges in two graphs. However, this method is only suitable for matching graphs with a small-sized number of vertices.

To avoid the prohibitive computation cost, it is typical to alternatively use tolerable approximate solutions. One

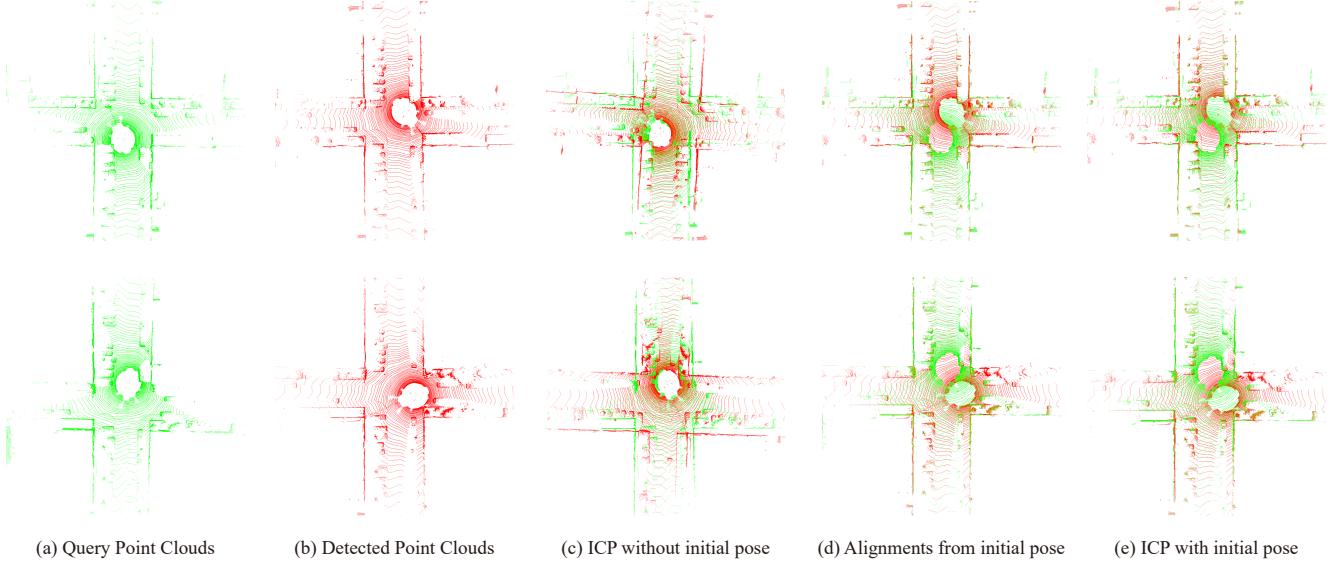


Fig. 3. The loop closure detection examples in KITTI 00 for the initial pose accuracy. (a) The frame numbers of the query point cloud are 1415 (up) and 4540 (down). (b) The frame numbers of the detected point cloud are 586 (up) and 116 (down). (c) The ICP registration results without the initial poses. (d) Align query and detected point clouds using the initial poses. (e) The ICP registration results with the initial poses.

TABLE I
RECALL AT 100% PRECISION WITH DIFFERENT n VALUES

	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 4$
KITTI 00	95.36%	94.49%	92.87%	92.61%	89.13%
KITTI 05	74.51%	71.14%	67.74%	55.58%	50.61%
KITTI 07	92.01%	89.21%	87.84%	88.61%	86.14%
KITTI 08	78.44%	77.91%	76.49%	68.64%	66.90%

Note that GOSMatch-labeled uses the point cloud data with labels as input while the other five methods use the raw point cloud data as input. Methods that are not object-based, especially Z-Projection and VFH, have shown poor performance. The reason is that these descriptors heavily depends on the accuracy of the normal estimation step. Besides, these types of descriptors only care about the angles formed by the normal with the centroid of the point cloud or with z-axis direction vectors but ignore the geometric space information. In this case, if the probability distribution of normal vectors is similar between two different places, these methods easily derive non-distinctive descriptors.

Scan Context and M2DP perform a comparatively good result in our experiment. However, they are unable to handle the challenging KITTI 08 dataset where all loops are generated when the car revisits the same place in the opposite direction. Only the rotational invariant algorithms can detect loop closures in this sequence with high performance. M2DP can not successfully detect loop closures in KITTI 08 and the precision-recall curve of Scan Context also decreases dramatically.

Overall, GOSMatch performs slightly worse than GOSMatch-labeled as the effect of the point classification errors from RangeNet++. Even so, both of them exhibit significant robustness and competitive rotation-invariance performance comparing with other methods. Our approach

benefits from stable descriptors formed by valuable semantics. Such descriptors will not be greatly affected even if the viewpoint changes a lot, thus we can perform reliable place recognition.

C. Noise sensitivity

This section focus on evaluating the robustness of GOSMatch against noise. In this experiment, we choose to extract segments directly from the labeled point cloud data provided by SemanticKITTI for avoiding classification noise interference. We randomly remove X objects in each scan to simulate changes in the vehicle positions (e.g. the previously parked vehicles have driven away) and some misidentifications, X is a random variable and $X \sim B(n, \frac{1}{2})$, where n means the maximum number of objects that can be removed in each scan. We test the effect of different n values on the four sequences and we perform 100 runs for each n value in each sequence. The average recall at 100% precision are presented in TABLE I.

D. Initial pose accuracy

The goal of this section is to evaluate the initial pose obtained from GOSMatch. To get more reliable registration results, the initial pose can be provided to ICP. Some examples from KITTI 00 are depicted in Fig. 3. In Fig. 3 (e), the results show that query and detected point clouds can be

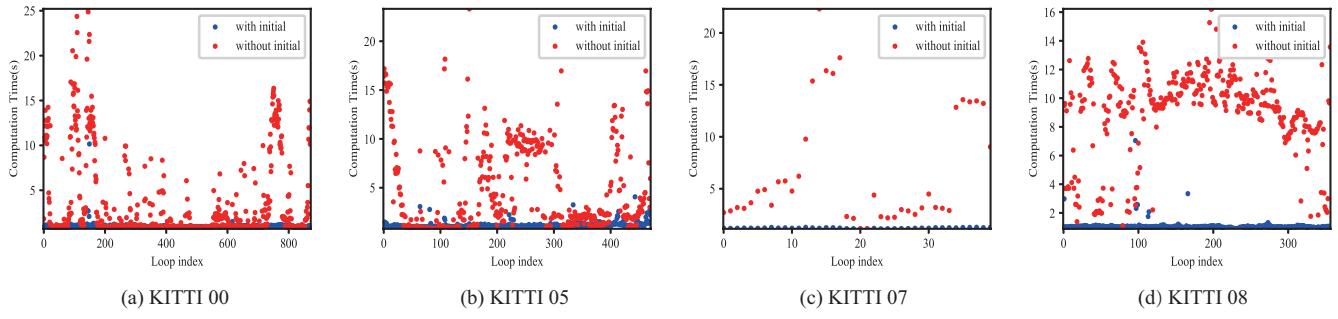


Fig. 4. Computation time for each KITTI sequence

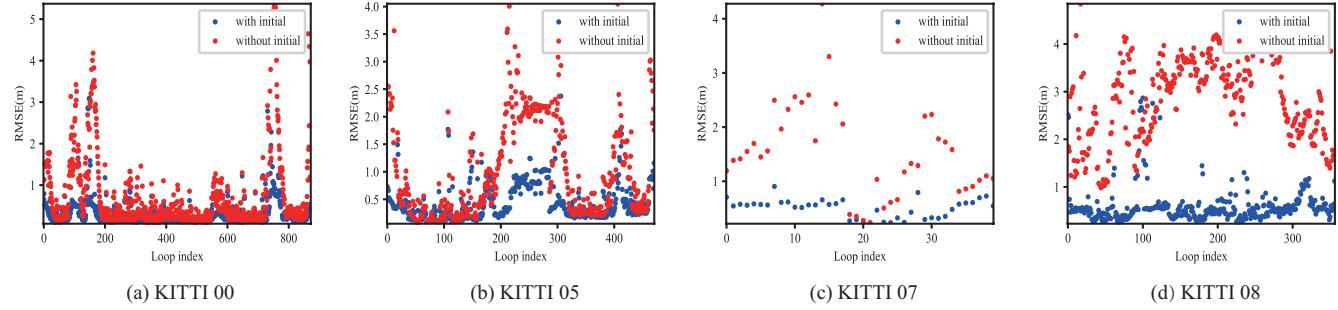


Fig. 5. ICP point-to-point RMSE for each KITTI sequence

TABLE II
AVERAGE TIME COSTS ON KITTI 00 IN SECONDS

	Descriptors calculation	Loop searching	Total time per scan
Scan Context	0.0424	0.0514	0.0938
M2DP	0.2461	0.0220	0.2681
Z-Projection	0.4045	0.0201	0.4246
VFH	0.5625	0.0379	0.6004
GOSMatch	0.0183 +0.1892*	0.0278	0.2353

*semantic detection execution time

successfully registered using the initial poses. However, in Fig. 3 (c), we can see that without the provided initial poses, the registration procedures of ICP can easily fail. Besides, the initial poses obtained from GOSMatch are shown in Fig. 3 (d). The alignment results using the initial poses are close to results which are optimized by the ICP algorithm, revealing the high accuracy of the estimated initial pose. The experimental results indicate that GOSMatch is able to predict a precise initial pose when detecting a loop closure.

In our experiment, we also test and verify the effect of the initial pose on the performance of ICP. The computation time and point-to-point RMSE of ICP with and without the initial pose is given in Fig. 4 and Fig. 5 respectively. The ICP registration procedure benefits a lot from the initial pose GOSMatch provided.

E. Computational complexity

The computational requirements for different methods are evaluated on KITTI 00 and the concrete results are shown in Table II. Because our method is different from traditional methods in that GOSMatch requires a semantic

detection step, so we separately show the time of calculating the two graphical descriptors and the time of semantic detection. In addition, only the semantic detection module in our method is performed on the GPU, all the other experiments are conducted on the CPU. The searching time of GOSMatch consists of searching k-nearest neighbors, matching *vertex descriptors* and verifying the geometric consistency. It is important to note that the time GOSMatch takes depends heavily on the semantic detection module, therefore a faster object detection algorithm can significantly improve the execution efficiency of GOSMatch.

V. CONCLUSIONS

In this paper, we propose GOSMatch, an algorithm for detecting loop closures in 3D point clouds based on two graphical descriptors. One is designed for efficiently searching the similar loop candidates from the historical point clouds, the other one is for further meticulous matching to give a solution to vertex-to-vertex correspondences between two places. Unlike the previous works, our method performs at a semantic level which is not only robust to environment changes but also able to give an accurate 6D initial pose estimation. The results of the exhaustive evaluation experiments show the potential of using relative position relationship between objects, which is significantly helpful for reliable loop closure detection.

VI. ACKNOWLEDGMENT

This work is supported by National Key R&D Program of China under Grant 2018YFB1305002.

