

This paper has been accepted for publication in IEEE International Conference on Robotics and Automation.

DOI: 10.1109/ICRA.2017.7989618

Please cite our work as:

R. Dubé, D. Dugas, E. Stumm, J. Nieto, R. Siegwart, and C. Cadena. “SegMatch: Segment Based Place Recognition in 3D Point Clouds.” IEEE International Conference on Robotics and Automation (ICRA), 2017.

**bibtex:**

```
@inproceedings{segmatch2017,  
  title      = {{SegMatch}: Segment Based Place Recognition in 3D Point Clouds},  
  author     = {Dub{\'e}, Renaud and Dugas, Daniel and Stumm, Elena and  
              Nieto, Juan and Siegwart, Roland and Cadena, Cesar},  
  booktitle  = {IEEE International Conference on Robotics and Automation (ICRA)},  
  year       = {2017}  
}
```

# SegMatch: Segment Based Place Recognition in 3D Point Clouds

Renaud Dubé Daniel Dugas Elena Stumm Juan Nieto Roland Siegwart Cesar Cadena\*

**Abstract**—Place recognition in 3D data is a challenging task that has been commonly approached by adapting image-based solutions. Methods based on local features suffer from ambiguity and from robustness to environment changes while methods based on global features are viewpoint dependent. We propose *SegMatch*, a reliable place recognition algorithm based on the matching of 3D segments. Segments provide a good compromise between local and global descriptions, incorporating their strengths while reducing their individual drawbacks. *SegMatch* does not rely on assumptions of ‘perfect segmentation’, or on the existence of ‘objects’ in the environment, which allows for reliable execution on large scale, unstructured environments. We quantitatively demonstrate that *SegMatch* can achieve accurate localization at a frequency of 1Hz on the largest sequence of the KITTI odometry dataset. We furthermore show how this algorithm can reliably detect and close loops in real-time, during online operation. In addition, the source code for the *SegMatch* algorithm is made publicly available<sup>1</sup>.

## I. INTRODUCTION

Place recognition represents one of the key challenges of accurate Simultaneous Localization and Mapping (SLAM). As drift is inevitable when performing state estimation without global positioning information, reliable loop-closure detection is a crucial capability for many robotic platforms [1]. Many successful strategies for performing place recognition using images are proposed in the literature. However, image-based place recognition can become unreliable when strong changes in illumination occur, and in the presence of strong viewpoint variations [2]. Lidar-based localization, on the other hand, does not suffer from changes in external illumination, and since it captures geometry in a very fine resolution, does not suffer as much as vision when changes in viewpoint are present. This paper therefore considers 3D laser range-finders for their potential to provide robust localization in outdoor environments.

Current strategies for recognizing places in 3D laser data are primarily based on keypoint detection and matching [3]. In the context of performing place recognition on images, Lowry et al. [2] state that using descriptors at the level of segments or objects could provide the benefits of both local and global feature approaches. Object or segment maps also offer several advantages over their metric and topological counterparts. Among others, these maps better represent

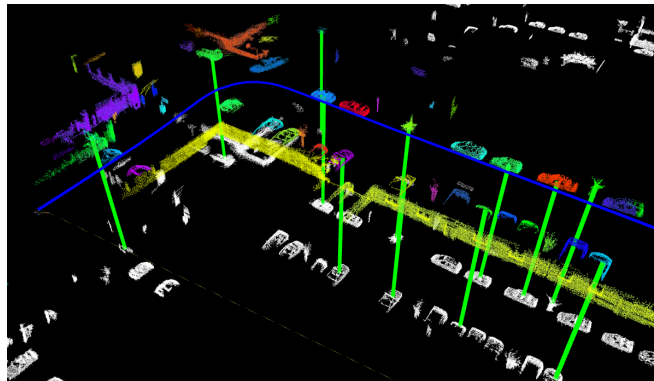


Fig. 1: An illustration of the presented place recognition framework. The reference point cloud is shown below (in white), and the local point cloud is aligned above. Colours are used to show the point cloud segmentation, and segment matches are indicated with green lines.

situations where static objects can become dynamic, and are more closely related to the way humans perceive the environment [1].

While working at the level of objects would be ideal, it also has a twofold assumption. First, that we have access to a perfect object segmentation technique, and second, that there are actual ‘objects’ in the environment, under the definition of [4]. These assumptions do not hold in general because of imperfect segmentation and because of common real-world scenarios with no distinguishable objects. This work therefore introduces *SegMatch*, a segment-based approach which takes advantage of more descriptive shapes than keypoint-based features without the aforementioned strong assumptions of object-based approaches. In other words, we recognize places by matching segments that belong to partial or full objects, or to parts of larger structures (windows, arcs, façades). Examples of such segments can be seen in Fig. 1 for data collected in an urban scenario.

Our system presents a modular design. It first extracts and describes segments from a 3D point cloud, matches them to segments from already visited places and uses a geometric-verification step to propose place recognition candidates. One advantage of this segment-based technique is its ability to considerably compress the point cloud into a set of distinct and discriminative elements for place recognition. We show that this does not only reduce the time needed for matching, but also decreases the likelihood of obtaining false matches.

When it comes to segment description, although numerous 3D point cloud descriptors exist [5–7], there is no clear evidence of relative performance among them, such as power of generalization or robustness against symmetry in geometry for instances. Therefore, we have opted for a machine learning approach to match a variety of standard

\*Authors are with the Autonomous Systems Lab, ETH, Zurich authors@mavt.ethz.ch.

This work was supported by the European Union’s Seventh Framework Programme for Research and Technological Development under the TRADR project No. FP7-ICT-609763.

<sup>1</sup>The *SegMatch* algorithm is available at <https://github.com/ethz-asl/segmatch> and a video demonstration is available at <https://www.youtube.com/watch?v=idcCgYbqjE>.

descriptors computed over the segments. Nonetheless, due to the modular nature of the presented framework, future advances in 3D segmentation, recognition, and description can be used by replacing the respective components in our pipeline.

To the best of our knowledge, this is the first paper to present a real-time algorithm for performing loop-closure detection and localization in 3D laser data on the basis of segments. More specifically, this paper presents the following contributions:

- *SegMatch*, a segment based algorithm to perform place recognition in 3D point clouds.
- An open source implementation of *SegMatch* for online, real-time loop-closure detection and localization.
- A thorough evaluation of the algorithm performances in real-world applications.

The paper is structured as follows: Section II provides an overview of the related work in the field of place recognition in 3D point clouds. The proposed algorithm is then described in Section III and evaluated in Section IV. Section V finally concludes with a short discussion.

## II. RELATED WORK

Detecting loop-closures from 3D data is still an open problem in robot localization. The problem has been tackled with different approaches. We have identified three main trends: (i) approaches based on local features, (ii) global descriptors and (iii) based on planes or objects.

The works presented in [3, 8–11] propose to extract local features from keypoints and perform matching on the basis of these features. Bosse and Zlot [3] extract keypoints directly from the point clouds and describe them with a 3D *Gestalt* descriptor. Keypoints then vote for their nearest neighbors in a *vote matrix* which is finally thresholded for recognizing places. A similar approach has been used in [11]. Apart from such Gestalt descriptors, a number of alternative local feature descriptors exist which can be used in similar frameworks. This includes features such as Fast Point Feature Histogram (FPFH) [7] which will also be employed later in this work. Alternatively, Zhuang et al. [8] transform the local scans into bearing-angle images and extract Speeded Up Robust Features (SURFs) from these images. A strategy based on 3D spatial information is employed to order the scenes before matching the descriptors. A similar technique by Steder et al. [9] first transforms the local scans into a range image. Local features are extracted and compared to the ones stored in a database, employing the Euclidean distance for matching keypoints. This work is extended in [10] by using Normal-Aligned Radial Features (NARF) descriptors and a bag of words approach for matching. Zhang and Singh [12] are able to estimate odometry in real-time using range data. Loop-closures are mentioned but rely on an offline algorithm.

Using global descriptors of the local point cloud for place recognition is also proposed [13–15]. Rohling et al. [13] propose to describe each local point cloud with a 1D histogram of point heights, assuming that the sensor keeps a constant height above the ground. The histograms are

then compared using the *Wasserstein* metric for recognizing places. Granström et al. [14] describe point clouds with rotation invariant features such as volume, nominal range, and range histogram. Distances are computed for scalar features and cross-correlation for histogram features, and an AdaBoost classifier is trained to match places. Finally, Iterative Closest Point (ICP) is used for computing the relative pose between point clouds. In another approach, Magnusson et al. [15] split the cloud into overlapping grids and compute shape properties (spherical, linear, and several type of planar) of each cell and combine them into a matrix of surface shape histograms. Similar to other works, these descriptors are compared for recognizing places.

While local keypoint features often lack descriptive power, global descriptors can struggle with invariance. Therefore other works have also proposed to use 3D shapes or objects for the place recognition task. Fernandez-Moral et al. [16], for example, propose to perform place recognition by detecting planes in 3D environments. The planes are accumulated in a graph and an interpretation tree is used to match sub-graphs. A final geometric consistency test is conducted over the planes in the matched sub-graphs. The work is extended in [17] to use the covariance of the plane parameters instead of the number of points in planes for matching. This strategy is only applied to small, indoor environments and assumes a plane model which is no longer valid in unstructured environment. A somewhat analogous, seminal work on object-based loop-closure detection in indoor environments using RGB-D cameras is presented by Finman et al. [18]. Although presenting interesting ideas, their work can only handle a small number of well segmented objects in small scale environments.

We therefore aim for an approach which does not rely on assumptions about the environment being composed of simplistic geometric primitives such as planes, or a rich library of objects. This allows for a more general, scalable solution. Inspiration is taken from Douillard et al. [19] and Nieto et al. [20] which proposed different SLAM techniques based on segments. A strategy for aligning Velodyne scans based on segments is proposed in [19] where the *Symmetric Shape Distance* is used to compare and match segments as defined in [21]. Analogously, [20] proposed an Extended Kalman Filter solution which uses segments as landmarks, rather than point features.

## III. *SegMatch* ALGORITHM

In this section we describe our approach for place recognition in 3D point clouds. The proposed system is depicted in Fig. 2 and is composed of four different modules: point cloud segmentation, feature extraction, segment matching, and geometric verification. Modularity has been a driving factor during the design phase. In the following, we propose an example implementation for every module of the system which can easily be adjusted for operating in different types of environment.

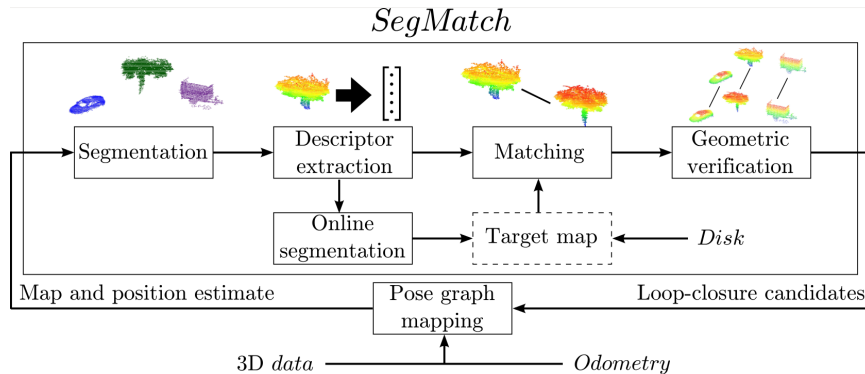


Fig. 2: Block diagram of *SegMatch*, a modular place recognition algorithm. The target map can either be loaded from disk (for localization) or computed online (for loop-closure).

### A. Segmentation

The first building block of *SegMatch* segments point clouds into distinct elements for matching. For each incoming point cloud  $P$  given in a global reference frame, we first extract a local point cloud by defining a cylindrical neighbourhood of radius  $R$ , centred around the current robot location. A voxel grid is then applied to the resulting source cloud in order to filter-out noise in voxels where there is not enough evidence for occupancy. The filtered point cloud is finally segmented into a set of point clusters  $C_i$  using the "Cluster-All Method" of [22]. This segmentation requires the ground plane to be previously removed, which can be achieved by clustering adjacent voxels based on vertical means and variances [22]. Once the ground plane is removed, Euclidean clustering is used for growing segments. For each cluster  $C_i$  the centroid  $c_i$  is computed as the average of all its points.

### B. Feature extraction

Once we have segmented the point-cloud, we extract features for each segment. This feature extraction step is used for compressing the raw data and builds segment signatures suitable for recognition and classification. As there is no clear *gold-standard* descriptor for 3D data, we use several different descriptors.

Given a cluster  $C_i$ , descriptors are computed resulting in feature vector  $f_i = [f_i^1 \ f_i^2 \ \dots \ f_i^m]$ . Whereas this feature vector could be extended to include a large quantity of descriptors, two descriptors which produced interesting results are here presented.

$f^1$  *Eigenvalue based*: In this descriptor, the eigenvalues of the segment's point cloud are computed and combined in a feature vector of dimension  $1 \times 7$ . We compute the *linearity*, *planarity*, *scattering*, *omnivariance*, *anisotropy*, *eigenentropy* and *change of curvature* measures as proposed in [23].

$f^2$  *Ensemble of shape histograms*: This feature of dimension  $1 \times 640$  is made of 10 histograms which encode the shape functions D2, D3 and A3 as described in [6]. The D2 shape function is a histogram of the distances between randomly selected point pairs while D3 encodes the area between randomly selected point triplets. The A3 shape function describes the angles between two lines which are obtained from these triplets.

### C. Segment matching

Using these features, we wish to identify matches between segments from the source and target clouds. For this operation we opted for a learning approach, as it is often difficult to select the appropriate distance metric and thresholds, especially when multiple feature types are involved. A classifier is therefore used to make the final decision about whether two segments represent the same object or object parts. In order to maintain efficiency, we first retrieve candidate matches by performing a k-d tree search in the feature space, which are then fed to the classifier.

Specifically, we employ a random forest for its classification and timing performances. The idea behind this classifier is to construct a multitude of distinct decision trees and to have them vote for the winning class. During the learning phase, each tree is trained using a bootstrapped subset of the training data set and a random subset of features. Random forests offer classification performance similar to the AdaBoost algorithm but are less sensitive to noise in the output label (such as a mis-labeled candidates) since they do not concentrate their efforts on misclassified candidates [24]. Random forests can also provide information regarding the feature's relative importance for the classification task.

For the random forest classifier to determine whether clusters  $C_i$  and  $C_j$  represent the same object, we compute the absolute difference between the eigenvalue based feature vectors:  $\Delta f^1 = |f_i^1 - f_j^1|$ . The feature vectors  $f_i^1$  and  $f_j^1$  are also fed to the classifier for a total eigenvalue based feature dimension of  $1 \times 21$ . For the ten histograms of the ensemble of shape features, the histogram intersection is computed, resulting in a feature of dimension  $1 \times 10$ . Given this set of features, the random forest classifier assigns a classification score  $w$  of being a match. A threshold on  $w$  is applied for building the final list of candidate matches passed to the next module. This threshold value is dependent on the subset of features used for matching and is defined in Table I.

### D. Geometric verification

The candidate matches are fed to a geometric-verification test using random sample consensus (RANSAC) [25]. Transformations are evaluated using the segment centroids. A geometrically-consistent cluster of segments is finally accepted based on a minimum number of segments in it,

resulting in a 6DOF transformation and a list of matching segments which represent a place recognition candidate.

#### IV. EXPERIMENTS

The proposed segment based algorithm is evaluated using the KITTI odometry dataset [26]. We first illustrate how this dataset can be processed for generating segment matching samples for training and testing the classifiers (Sections IV-A and IV-B). This leads to an analysis of the performances of different classifiers’ parametrization (Section IV-C). The segment based localization strategies are then compared to a keypoint approach as a baseline (Section IV-D). We then show how the segment based loop detection framework can be used for online place recognition applications and how it can successfully operate in different environments (Sections IV-E and IV-F).

##### A. Dataset

The following three analyses are performed using sequences 00, 05 and 06 of the KITTI dataset. Sequence 06 lasts 1.2 km (114 seconds) and is only used for training the classifiers. Sequence 00 lasts 3.7 km (470 seconds) and is particularly interesting as it contains one large loop where the vehicle revisits the same environment for a stretch of 500 meters. This section with multiple traversals will therefore be used in the localization experiment. Sequence 05 lasts 2.2 km (287 seconds) and is used for presenting the online operation of the framework.

As previously described, the input of our segment based place recognition algorithm is a point cloud in a global reference frame. For generating a point cloud in real-time from the large quantity of measurements provided by a Velodyne sensor, a uniform rate sub-sampling filter is first applied for removing half of the scan’s points. These scans are added to the point cloud every time the robot drove a minimum distance of 1 meter. As the sensor configuration on-board the car is known (e.g. height of the sensor), ground plane extraction is performed by directly filtering the input scan by minimum height. This simple assumption is efficient and works very well for this driving dataset.

For extracting the source point cloud, the radius of the cylindrical neighbourhood  $R$  is set to 60 meters. The voxel grid leaf size is set to 0.1 meters, and a minimum of two points within a voxel is required to consider it as occupied. For segmentation, the maximum Euclidean distance between two occupied voxels such that they are considered to belong to the same cluster is set to 0.2 meters. We choose to consider only segments which contain a minimum of 100 points and a maximum of 15000 points.

##### B. Training and testing setup

The following procedure is performed for generating both training and testing data. During the first section of a given sequence, a target map is generated and processed by extracting and describing segments. When the vehicle revisits the same section of the environment, the ground truth information is used for storing pairs of corresponding

TABLE I: Parameters of three segment matching strategies.

Parameter	$L2$	$rf\_eigen$	$rf\_full+shapes$
Number of neighbors	200		
k-NN Feature space	Eigenvalue based		
Hard threshold value	0.0024	N/A	N/A
Number of trees	N/A	25	25
Threshold on probability	N/A	0.81	0.72
Minimum cluster size	4		
RANSAC resolution	0.4 meter		

and differing segments from the source and target clouds. For each segment in the local cloud, we perform k-Nearest Neighbors (k-NN) retrieval in feature space and identify the 200 nearest neighbors in the target map. These candidates are saved as true matches for the corresponding segments and false matches for differing segments. Using this procedure on sequence 06 of the KITTI dataset, we generate 2000 true and 800000 false segment matches. For training the random forests, we adopt a 1:50 ratio between the number of positive and negative samples which results in a training set of 102000 samples.

##### C. Segment matching performance

The goal of this first experiment is to evaluate the performances of three segment matching techniques. The first strategy titled  $L2$  applies a threshold on the Euclidean distance between two segment’s features vectors. The second strategy,  $RF\_eigen$ , is based on a random forest which relies only on the eigenvalue based features. The last strategy,  $RF\_eigen+shapes$ , uses the full set of features described in Section III-B. The parameters used for each classifier are summarized in Table I.

Fig. 3 shows the receiver operating characteristic (ROC) curves of the three methods when testing on data extracted from sequence 00. The random forest classifiers offer an improvement in performance when compared to their  $L2$  norm counterpart. Examples of corresponding segments correctly identified by the  $RF\_eigen+shapes$  strategy are illustrated in Fig. 4.

In the experiments of the following section, we illustrate how these classifiers perform during real-time localization. We define a false positive rate (FPR) of 0.2 to be the operating point of all classifiers in order to limit false segment matches and avoid false place recognitions. This parameter and the other ones summarized in Table I are used for the localization and loop-closure experiments of Sections IV-D and IV-E.

##### D. Localization performance

This section evaluates the performance of the *SegMatch* algorithm for localizing in a target segment map. The section of interest in sequence 00 (as described in Section IV-A) is used for creating the target map, and localization is performed when this section is revisited. The three segment based strategies described in section IV-C are compared to a keypoint based place recognition technique.

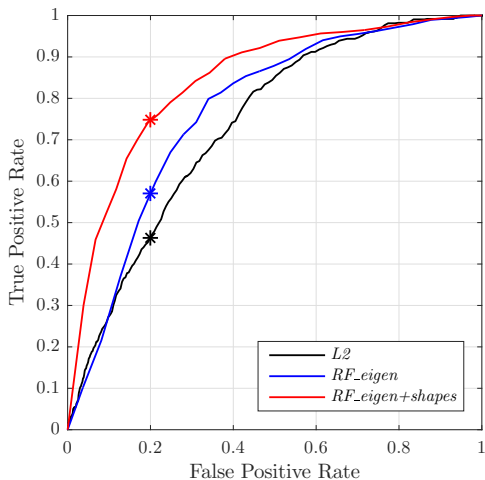


Fig. 3: ROC curves for segment matching performance using a hard threshold on the distance between segment features ( $L2$ ) compared to using random forests on two different feature sets ( $RF\_eigen$  and  $RF\_eigen+shapes$ ). The operating points of FPR = 0.2 are indicated.

1) *Keypoint baseline*: For the *keypoint* localization method, normals are first computed for every point of the filtered cloud using a radius of 0.3 meters. Keypoints are selected in the target and source clouds using the Harris 3D keypoint extractor of the PCL library [27]. These keypoints are filtered to have a minimum distance of 0.5 meters between each keypoint. This ensures that the same regions are not described twice, which in turn reduces ambiguity during the later geometrical verification step. Each keypoint is described using the Fast Point Feature Histogram (FPFH) with a radius of 0.4 meters [7]. The source keypoints are matched to their 75 closest neighbors in the target cloud and the geometric verification algorithm described in Section III-D is used to filter this list of keypoint matches and to output localizations. Parameters were chosen in an attempt to get the best performance we could find.

2) *Results*: In order to show the reproducibility of the results and because the computer load affects the locations at which localization is performed in each run, we perform 90 runs for each strategy and present the average results. The distance travelled between each localization is recorded and evaluated in a similar manner to [28]. Fig. 5 shows the probability of travelling a given distance without successful localization in the target map. Specifically, this metric is computed as follows.

$$P(x) = \frac{\text{Sum of distance travelled without localization for greater or equal to } x \text{ meters}}{\text{Total distance travelled}} \quad (1)$$

Although  $RF\_eigen+shapes$  is the most complex and computationally demanding strategy (see Table II), it never required more than 55 meters before successful localization, as compared to 67 and 88 meters for  $L2$  and  $RF\_eigen$  respectively. On the other hand, while  $L2$  is the quickest strategy, it also made 14 false localizations, which could motivate further reduction of the operating point of 0.2 FPR.

TABLE II: Timing of each of the localization modules (in ms).

Module	$L2$	$RF\_eigen$	$RF\_eigen+shapes$
Segmentation	$428.80 \pm 5.83$	$428.78 \pm 6.28$	$435.56 \pm 7.34$
Description	$1.37 \pm 0.03$	$1.37 \pm 0.03$	$103.84 \pm 2.41$
Matching	$244.52 \pm 9.76$	$289.75 \pm 10.56$	$563.23 \pm 11.96$
Geometric verification	$67.43 \pm 2.77$	$76.00 \pm 2.75$	$85.57 \pm 3.16$
Total	742.12	795.91	1188.20

For the two random forest based strategies, the vehicle can successfully localize within 35 meters 95% of the time.

Finally, all segment matching methods clearly outperform the keypoint baseline which necessitated much more work to deliver positive results. Based on keypoint matching, we were not able to obtain an interesting number of true positives without allowing for some false positives. That is, on average over a one minute localization run, the baseline detected 5.23 true positive and 3.25 false positive localizations.

The computational requirements of this algorithm on an Intel i7-4900MQ CPU @ 2.80GHz are depicted in Table II. Note that all operations including ensemble of shape feature extraction, histogram intersection, and random forest classification could benefit of parallelization.

#### E. Loop-closure performance

We now show how our segment based loop detection algorithm can be used online and how it can easily be integrated with a pose-graph trajectory estimation system. In this scenario, the target map is built online by accumulating segments extracted from the source clouds, as opposed to being loaded before the experiment as performed in Section IV-D. Special care is taken to avoid cluttering the target map with ‘duplicate segments’, i.e. segments resulting from the same object part, but segmented at different times and overlapping in the target map.

The results of applying this strategy on sequence 05 of the KITTI dataset is illustrated in Fig. 6. For this sequence, the global map is created using ICP for adding constraints between Velodyne scans. This introduces a drift over time, as expected in GPS-free state estimation solutions. On this sequence, our real-time algorithm successfully detected 12 true positive and no false positive loop-closures. Once loops are detected, they are fed in a pose-graph optimization system similar to the one described in [29]<sup>2</sup>. The result of this optimization is used to update the target segment positions and remove duplicate segments from the target map as aforementioned.

#### F. Demonstration with more complex data

To conclude the experiment section, we briefly show that the proposed place recognition algorithm can be applied to other environments and sensor modalities by simply replacing sub-modules of the pipeline. As an example, in situations where the segmentation algorithm described in Section III-B cannot be applied, this module can be replaced by a different algorithm. Fig. 7 shows an example of a correct loop detection by matching segments obtained from segmenting

<sup>2</sup>This separate contribution is available at [https://github.com/ethz-asl/laser\\_slam](https://github.com/ethz-asl/laser_slam).



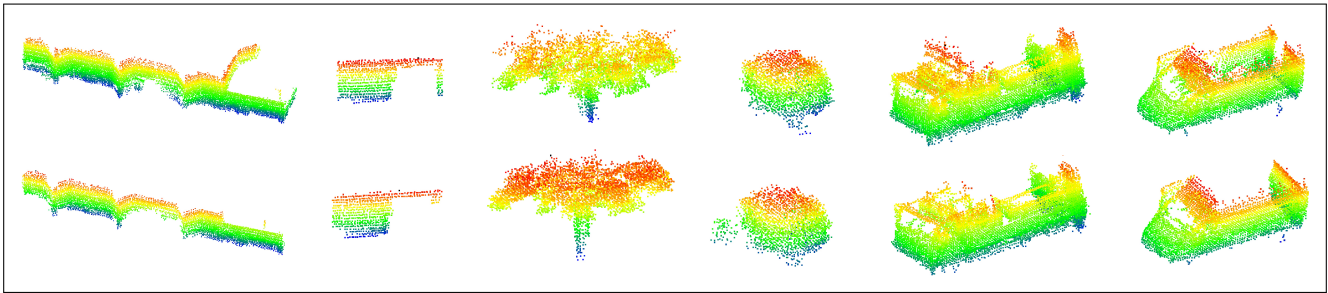


Fig. 4: Corresponding segments successfully detected by the *SegMatch* algorithm. The top and bottom rows illustrate segments from the target and the source clouds respectively.

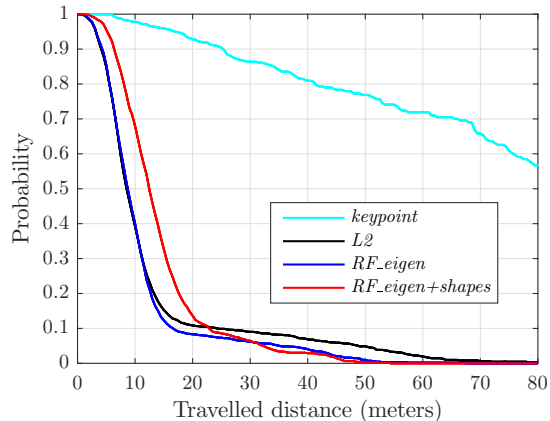


Fig. 5: Probability of travelling a given distance before localizing in the target segment map. Data is obtained from 90 localization runs for each strategy on drive 00 of the KITTI dataset. Over these 90 runs, the *keypoint* and *L2* strategies respectively detected 292 and 14 false localizations while *RF\_eigen* and *RF\_eigen+shapes* made no false detections.

the point cloud based on region growing with smoothness constraints [30]. Although these types of segments may appear to be less meaningful for humans, they provide discriminative features for the loop-closure algorithm, as illustrated by the matches shown in Fig. 7.

## V. CONCLUSION

This paper presented *SegMatch*, an algorithm for performing place recognition in 3D laser data based on the concept of segment matching. Compared to a keypoint approach, acting at the level of segments offers several advantages without making any assumptions about perfect segmentation or on the presence of ‘objects’ in the environment. Our modular approach first extracts segments from a source point cloud, which are then described and matched to previously mapped target segments. A geometric-verification step is finally applied to turn these candidate matches into place recognition candidates.

This framework has been exhaustively evaluated on the KITTI dataset. We first analysed the impact of using a random forest classifier to learn an adequate feature distance metric for the purpose of matching segments. We have then shown that the algorithm is able to accurately localize at a frequency higher than 1Hz in the largest map of the KITTI dataset. We also demonstrated how it is possible to robustly detect loops in an online fashion, and how these

can be fed to a pose-graph trajectory estimator. Thanks to the framework’s modular approach, we have furthermore illustrated that it can easily be applied to different scenarios by simply changing building blocks of the algorithm. The source code for the entire framework is available online, offering real-time segmentation and loop-closure detection for streams of 3D point clouds.

Based on this segment matching technique, we foresee several possible advantages in systems which do more than mapping - using segments for both matching and describing the environment. We will pursue supervised learning techniques to interpret these *segment*-based maps into structural and object semantic classes.

## REFERENCES

- [1] S. Thrun *et al.*, “Robotic mapping: A survey,” *Exploring artificial intelligence in the new millennium*, vol. 1, pp. 1–35, 2002.
- [2] S. Lowry, N. Sunderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford, “Visual place recognition: A survey,” *IEEE Trans. on Robotics*, 2016.
- [3] M. Bosse and R. Zlot, “Place recognition using keypoint voting in large 3D lidar datasets,” in *IEEE Int. Conf. on Robotics and Automation*, 2013.
- [4] B. Alexe, T. Deselaers, and V. Ferrari, “What is an object?” in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2010.
- [5] P. Scovanner, S. Ali, and M. Shah, “A 3-dimensional sift descriptor and its application to action recognition,” in *ACM Int. Conf. on Multimedia*, 2007.
- [6] W. Wohlkinger and M. Vincze, “Ensemble of shape functions for 3d object classification,” in *IEEE Int. Conf. on Robotics and Biomimetics*, 2011.
- [7] R. B. Rusu, N. Blodow, and M. Beetz, “Fast point feature histograms (fpfh) for 3d registration,” in *IEEE Int. Conf. on Robotics and Automation*, 2009, pp. 3212–3217.
- [8] Y. Zhuang, N. Jiang, H. Hu, and F. Yan, “3-d-laser-based scene measurement and place recognition for mobile robots in dynamic indoor environments,” *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 2, pp. 438–450, 2013.
- [9] B. Steder, G. Grisetti, and W. Burgard, “Robust place recognition for 3D range data based on point features,” in *IEEE Int. Conf. on Robotics and Automation*, 2010.
- [10] B. Steder, M. Ruhnke, S. Grzonka, and W. Burgard, “Place recognition in 3d scans using a combination of bag of words and point feature based relative pose estimation,” in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2011.
- [11] A. Gawel, T. Cieslewski, R. Dubé, M. Bosse, R. Siegwart, and J. Nieto, “Structure-based Vision-Laser Matching,” in

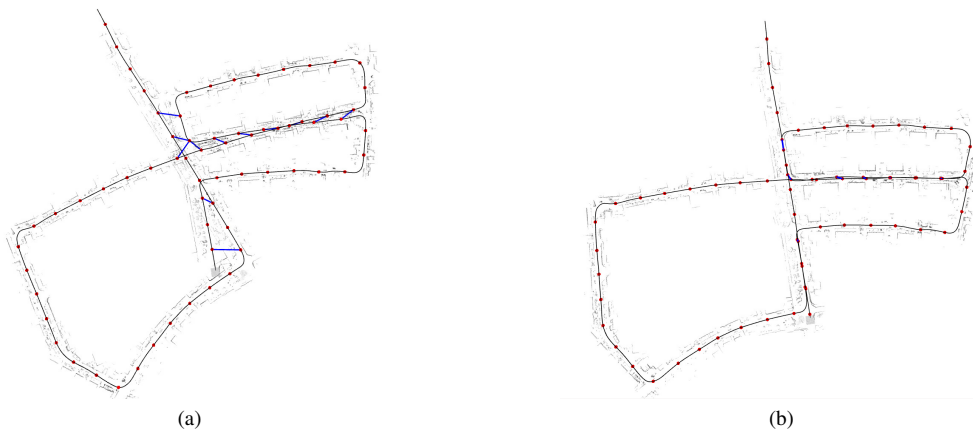


Fig. 6: Illustration of loop-closure with *SegMatch*: (a) Shows loops detected in real time by the segment based strategy *RF\_eigen+shapes* during sequence 05 of the KITTI dataset. The red dots represent locations where segmentation and loop-closure detection were performed and the blue lines indicate the detected loops (with no false positives). For illustration purposes, we show the odometry trajectory in (a) whereas (b) illustrates the result of feeding these loops to an online pose-graph trajectory estimator.

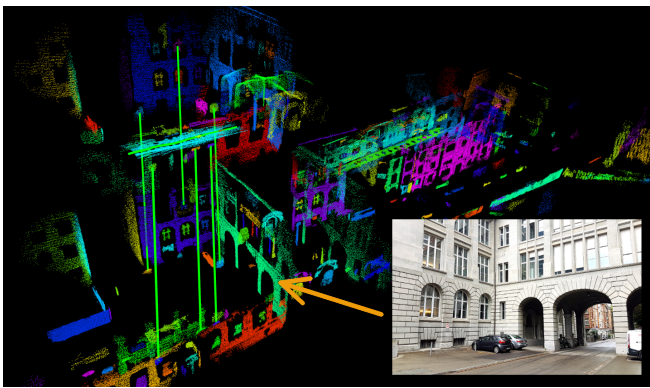


Fig. 7: An illustration of a correct loop-closure detection based on region growing segmentation with smoothness constraints on data from the Clausiusstrasse in Zurich. The reference point cloud is shown below, and the local point cloud is aligned above. Colours are used to show the point cloud segmentation, and segment matches are indicated with green lines. Note the parts of larger structures (windows, arcs, façades).

*IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Daejeon, 2016.

- [12] J. Zhang and S. Singh, “Loam: Lidar odometry and mapping in real-time,” in *Robotics: Science and Systems*, 2014.
- [13] T. Rohling, J. Mack, and D. Schulz, “A fast histogram-based similarity measure for detecting loop closures in 3-d lidar data,” in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2015.
- [14] K. Granström, T. B. Schön, J. I. Nieto, and F. T. Ramos, “Learning to close loops from range data,” *The Int. Journal of Robotics Research*, vol. 30, no. 14, pp. 1728–1754, 2011.
- [15] M. Magnusson, H. Andreasson, A. Nüchter, and A. J. Lilienthal, “Automatic appearance-based loop detection from three-dimensional laser data using the normal distributions transform,” *Journal of Field Robotics*, vol. 26, no. 11-12, pp. 892–914, 2009.
- [16] E. Fernandez-Moral, W. Mayol-Cuevas, V. Arevalo, and J. Gonzalez-Jimenez, “Fast place recognition with plane-based maps,” in *IEEE Int. Conf. on Robotics and Automation*, 2013.
- [17] E. Fernández-Moral, P. Rives, V. Arévalo, and J. González-Jiménez, “Scene structure registration for localization and mapping,” *Robotics and Autonomous Systems*, vol. 75, pp. 649–660, 2016.
- [18] R. Finman, L. Paull, and J. J. Leonard, “Toward object-based place recognition in dense rgb-d maps,” in *ICRA workshop on visual place recognition in changing environments*, 2015.
- [19] B. Douillard, A. Quadros, P. Morton, J. P. Underwood, M. De Deuge, S. Hugosson, M. Hallström, and T. Bailey, “Scan segments matching for pairwise 3d alignment,” in *IEEE Int. Conf. on Robotics and Automation*, 2012.
- [20] J. Nieto, T. Bailey, and E. Nebot, “Scan-slam: Combining ekf-slam and scan correlation.” Springer, 2006, pp. 167–178.
- [21] B. Douillard, J. Underwood, V. Vlaskine, A. Quadros, and S. Singh, “A pipeline for the segmentation and classification of 3d point clouds,” in *Experimental Robotics*. Springer, 2014, pp. 585–600.
- [22] B. Douillard, J. Underwood, N. Kuntz, V. Vlaskine, A. Quadros, P. Morton, and A. Frenkel, “On the segmentation of 3d lidar point clouds,” in *IEEE Int. Conf. on Robotics and Automation*, 2011.
- [23] M. Weinmann, B. Jutzi, and C. Mallet, “Semantic 3d scene interpretation: a framework combining optimal neighborhood size selection with relevant features,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 2, no. 3, p. 181, 2014.
- [24] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [25] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [26] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2012.
- [27] R. B. Rusu and S. Cousins, “3D is here: Point Cloud Library (PCL),” in *IEEE Int. Conf. on Robotics and Automation*, 2011.
- [28] C. Linegar, W. Churchill, and P. Newman, “Work smart, not hard: Recalling relevant experiences for vast-scale but time-constrained localisation,” in *IEEE Int. Conf. on Robotics and Automation*, 2015.
- [29] R. Dubé, H. Sommer, A. Gawel, M. Bosse, and R. Siegwart, “Non-uniform sampling strategies for continuous correction based trajectory estimation,” in *IEEE Int. Conf. on Robotics and Automation*, 2016.
- [30] T. Rabbani, F. Van Den Heuvel, and G. Vosselmann, “Segmentation of point clouds using smoothness constraint,” *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 36, no. 5, pp. 248–253, 2006.