Place Recognition of 3D Landmarks based on Geometric Relations

Dario Lodi Rizzini[1]

Abstract— **Place recognition based on landmarks or features is an important problem occurring in localization, mapping, computer vision and point cloud processing. In this paper, we present GLAROT-3D, a translation and rotation invariant 3D signature based on geometric relations. The proposed method encodes into a histogram the pairwise relative positions of keypoint features extracted from 3D sensor data. Since it relies only on geometric properties and not on specific feature descriptors, it does not require any prior training or vocabulary construction and enables lightweight comparisons between landmark maps. The similarity of two point maps is computed as the distance between the corresponding rotated histograms to achieve rotation invariance. Histogram rotation is enabled by efficient orientation histogram based on sphere cubical projection. The performance of GLAROT has been assessed through experiments with standard benchmark datasets.**

I. Introduction

Robot localization and mapping rely on the recognition of already visited regions and places from sensor measurements, after travelling long paths. This important operation is known as loop closure and enables to recover consistency in environment representation. Seminal loop closure techniques have been originally developed for planar range finder measurements either in the form of occupancy grid maps [1]–[3] or, more recently, keypoint features [4]–[6]. In computer vision [7], [8] and 3D perception [9] keypoint features are commonly used to build landmark maps. Although keypoint features may fail in specific contexts like underwater environments [10], they enable compact keypoint landmark map representation.

Several SLAM systems [11]–[13] operating with cameras and RGBD sensors adopt landmark map representations. Images and point clouds have large memory footprints and their direct manipulation is computationally complex. Most SLAM tools have developed a standard structure to efficiently manage the resulting large maps. Raw sensor data can be replaced by a collection of points and, possibly, of descriptors, enabling lightweight and efficient representation of environment. The keypoint features extracted from one or more consecutive measurements are usually collected in a local map associated to a keyframe. Keyframes are selected frames of robot or sensor sampled during scene exploration. The relations among keyframes are encoded by a pose graph that is periodically optimized, when new measurements have been collected.

Given this formulation, place recognition is achieved by comparing local maps of keypoints, where each local map is associated to a keyframe. In principle each keypoint feature could be used as an independent landmark in a global map. However, direct feature-to-feature association is not reliable over large collections of points, and, moreover, is computationally demanding. The implicit hierarchical organization into submaps allows a two-step loop closure procedure. First, the candidate matching keyframes are selected on the whole map. Thereafter, point-to-point association and frame transformation estimation are performed for each candidate pair. Selection of candidate matching keyframes is important for efficient loop closure and is usually performed by computing a signature for each keyframe. Signatures enable lightweight comparison of keyframes before the more accurate association between the landmark sets. Several signatures, generally based on bag-of-word (BoW) [14], have been proposed. Some implementations of BoW achieves real-time and efficient selection of candidates using additional criteria like covisibility graph [13]. Although efficient and effective in place recognition, this approach depends on specific feature descriptors and on dictionaries obtained from training sets.

In this paper, we present GLAROT-3D, a novel rotation-invariant signature based on geometric landmark relations, which enables data association between collections of 3D landmark points. This signature is inspired by the planar signatures GLARE [15] and GLAROT [6], originally proposed to associate keypoint features extracted from planar laser scans. GLAROT-3D computes a histogram by encoding the pairwise relative positions of the keypoints belonging to a local map. The relative positions are expressed in polar form through their distances and orientations. Unlike BoW approaches, the proposed signature provides an effective description of a point map based only on geometric information instead of a dictionary. Candidate matching maps can be detected by comparing their corresponding GLAROT-3D histograms through rotation invariant metric. In our proposal, histogram bin rotations are efficiently handled through a cubical grid partition of orientations inspired by Cubemap sphere [16]. This approach has been assessed through experiments with standard benchmark datasets and compared with the state-of-the-art loop closure detection algorithm of ORB-SLAM2 [17].

The paper is organized as follows. Section II reviews the state of the art of 3D mapping systems and data association. Section III illustrates GLAROT-3D and its application to the selection of loop candidates. Section IV reports the experimental results. Section V concludes the work.

[1]Author is with RIMLab - Robotics and Intelligent Machines Laboratory, Dipartimento di Ingegneria dell'Informazione, University of Parma, Italy, dlr@ce.unipr.it

II. Related Works

A. Place Recognition in Planar Range Data

The most popular maps with range finder are occupancy grid maps and data association is performed using correlation-based techniques [1], [3]. The recent proposal of keypoint features for planar range finders enabled adaptation of pairing methods designed for features and landmarks. Tipaldi and Arras [4] proposed FLIRT, the first keypoint feature explicitly designed for planar laser scans. The FALKO algorithm [5] efficiently detects more stable keypoints from a laser scan through the selection of neighbors in the scan and the computation of a cornerness score. Moreover, such keypoint features are endowed with descriptors, which encode the local distribution of neighbor points and can be potentially exploited for data association. Geometrical FLIRT phrases (GFPs) [18] extend the BoW techniques, popular in computer vision, to planar range finders using FLIRT descriptors. Deray et al. [19] improved the BoW association and the weak geometric check using Viterbi algorithm. However, descriptors extracted from range data are often unreliable due to the limited discrimination of the descriptors. Geometric LAndmark RElations (GLARE) [15] are based on geometric invariants in a keypoint set like pairwise distances and angles. While effective in many contexts, they are not orientation invariant. GLAROT [6] is a variant of GLARE comparing all possible orientations. Techniques like GFP, GLARE and GLAROT can be classified as signatures that can be used to efficiently select loop candidates. To validate the candidates and to compute the transformation, point-to-point association methods like joint compatibility test [20] or maximum clique techniques [21] are required.

B. Landmark Data Association in Computer Vision and 3D Processing

Several solutions to place recognition have been proposed in computer vision and visual SLAM. The most notable approach includes appearance-based and BoW methods, which represent the features extracted from an image using an a priori vocabulary. FAB-MAP [22] is one of the most popular appearance-only algorithm using SURF features [23]. DBoW2 [14] extends the BoW approach to the more efficient binary features like ORB [24]. These techniques are often integrated into complete SLAM systems like PTAM [11] and ORB-SLAM [13]. Different versions of these tools have been developed to operate with monocular cameras, but also with stereo vision and with depth cameras [17]. Stereo and depth cameras provide 3D geometric information that is exploited for keyframe registration and tracking, but not for place recognition since BoW approach exploits only visual feature descriptors detected in images. Moreover, the vocabulary and the word-based encoding used in BoW are dependent on the choice of the specific keypoint features.

An alternative approach to feature-based mapping is represented by direct methods, which perform registration, tracking and frame-to-frame association by directly comparing pixel intensities. LSD-SLAM [25] is the current state-of-the-art implementation of direct approach. Loop-closure candidates are found using OpenFABMAP [26] and, then, validated by direct image alignment. Thus, direct approach is insufficient when dealing with large scale maps and require the most robust appearance-based data association.

III. Signature for 3D Landmarks

In this section, we present GLAROT-3D, a signature for 3D landmarks based on geometric relations. We discuss the geometric properties of a point set used by the signature and how they are encoded in a histogram. The cubical mapping has been adopted for the 3D orientation histogram in order to properly compare signatures. Finally, signatures are compared using a translation and orientation invariant metric.

A. Geometric Landmark Relations in 3D Space

In place recognition problems, each point set $\mathcal{P}$ represents the landmarks detected in raw measurements. Such measurements are acquired by a sensor and, hence, the coordinates of the points are given w.r.t. the sensor frame. The same place can be observed from different viewpoints and an effective place signature must be translation and rotation invariant to allow comparison.

The original GLARE algorithm [15] computes a signature that encodes the pairwise distances and angles of a point set. This method only requires the coordinates of the input points and no other information like descriptor values often provided by keypoint features. In particular, the signature $\mathbf{G}[\mathcal{P}]$ of the point set $\mathcal{P}$ is a polar histogram that associates to each relative pairwise position an integer or real counter value. For each pair of points $\mathbf{p}_i, \mathbf{p}_j \in \mathcal{P}$, their relative position is given by the difference vector $\mathbf{r}_{ij} = \mathbf{p}_i - \mathbf{p}_j$. The difference vector can be represented in polar coordinates, which in the planar case are the angle $\theta_{ij} \in \mathbb{S}^1$ and the range $\rho_{ij} = \|\mathbf{p}_i - \mathbf{p}_j\| \in \mathbb{R}^+$. The histogram $\mathbf{G}$ is defined over a partition of 2D polar domain $\mathbb{S}^1 \times \mathbb{R}^+$. The histogram bin $\mathbf{G}_{i_\theta, i_\rho}$ corresponding to point $\mathbf{r}_{ij}$ is given by the two index functions

$$i_\theta(\mathbf{r}_{ij}) = \left\lfloor \frac{(\operatorname{atan2}(r_{ij,y}, r_{ij,x}) + 2\pi) \bmod 2\pi}{\Delta\theta} \right\rfloor \quad (1)$$

$$i_\rho(\mathbf{r}_{ij}) = \left\lfloor \frac{\|\mathbf{r}_{ij}\|}{\Delta\rho} \right\rfloor \quad (2)$$

where $\Delta\theta$ and $\Delta\rho$ are respectively the angular and range resolutions of the grid. The discussion of the signature and index function for 3D points is post-poned to focus on general aspects of geometric signatures. Since there are two possible difference vectors for each point pair, i.e. $\mathbf{r}_{ij}$ and $\mathbf{r}_{ji}$, the polar histogram can be updated either by adding both the difference vector contribution or by choosing only one between $\mathbf{r}_{ij}$ and $\mathbf{r}_{ji}$. The GLARE considers only $\mathbf{r}_{ij}$ with $p_{i,y} > p_{j,y}$ to restrict the size of polar histogram. The similarity between a source point set $\mathcal{P}_S$ and a target one $\mathcal{P}_T$ is measured by the distance between the respective signatures $\mathbf{G}_S = \mathbf{G}[\mathcal{P}_S]$ and $\mathbf{G}_T = \mathbf{G}[\mathcal{P}_T]$ according to L_1 norm.

The original signature GLARE satisfies invariance only to translation, but not to rotation. The solutions proposed to

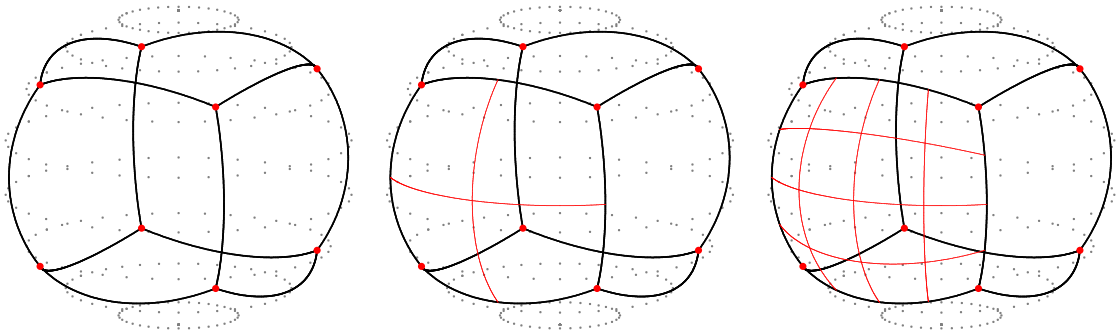

Fig. 1: Cubical grids on sphere with different number of face cells l.

f	label	$\mathbf{d}_f$	$\mathbf{u}_f$	$\mathbf{v}_f$
0	x^+	$\mathbf{x}^+$	$\mathbf{y}^+$	$\mathbf{z}^+$
1	x^-	$\mathbf{x}^-$	$\mathbf{z}^-$	$\mathbf{y}^-$
2	y^+	$\mathbf{y}^+$	$\mathbf{z}^+$	$\mathbf{y}^+$
3	y^-	$\mathbf{y}^-$	$\mathbf{x}^-$	$\mathbf{z}^-$
4	z^+	$\mathbf{z}^+$	$\mathbf{x}^+$	$\mathbf{y}^+$
5	z^-	$\mathbf{z}^-$	$\mathbf{y}^-$	$\mathbf{x}^-$

TABLE I: Faces of cubemap with their index f, label, orthogonal axis $\mathbf{d}_f$ and parallel axes $\mathbf{u}_f$ and $\mathbf{v}_f$.

achieve complete invariance to isometry include GSR [27] and GLAROT [6]. The first approach assesses the relative orientation between the two point sets using the original raw measurements to estimate the normal directions to the points. While rotational invariant, the computation of GSR is expensive due to the processing of raw measurement of a grid and is less general. GLAROT achieves same or better results by comparing the target signature $\mathbf{G}_T$ to signature $\mathbf{R}\mathbf{G}_S$, where $\mathbf{R} \in \mathcal{R} \subseteq SO(2)$ is one of the $n_\theta = 2\pi/\Delta\theta$ rotations shifting an orientation bin to another one. This approach requires n_θ evaluations of L_1 norm for each shifted histogram.

The discussion of geometry-based signature shows the main issues to be addressed to define GLAROT-3D. The first issue is the structure of the GLAROT-3D histogram $\mathbf{G}$ implicitly defined by index functions like eq. (1)-(2). While the range index i_ρ is trivially extended to 3D domain, a different orientation index function $i_{\mathbb{S}^2}$ must be used instead of i_θ. The subdivision of sphere $\mathbb{S}^2$ induced by $i_{\mathbb{S}^2}$ must correspond to bins with equal or nearly equal areas. The second issue is the comparison procedure, which must be independent from the reference frames of the point sets. In particular, GLAROT-3D must handle histogram rotations to define a metric invariant to the frame orientation. These problems are discussed in the following.

B. Cubical Orientation Grid

The definition of a signature for 3D landmarks requires a proper histogram for the polar values of pairwise relative position of landmarks. The main issue is the orientation histogram formally defined by angular index function $i_{\mathbb{S}^2}$ mapping a point of $\mathbb{R}^3$ to the histogram bin indices. The angular bin $\int_i \subseteq \mathbb{S}^2$ corresponding to index i is a subset of unit sphere surface s.t.

$$\int_i = \{\mathbf{u} \in \mathbb{S}^2 | i_{\mathbb{S}^2}(\mathbf{u}) = i\} \tag{3}$$

The bins constitute a partition of the solid angle which must satisfy some convenient requirements:

1) the solid angle is partitioned into equal angular bins, i.e. the areas $|\int_i| = |\int_j|$ for all i, j;
2) for all bins $\int_i$ and $\int_j$, there is a rotation $\mathbf{R} \in \mathcal{R} \subseteq SO(3)$ s.t.

$$\mathbf{R}\int_i = \{\mathbf{R}\mathbf{u} | \mathbf{u} \in \int_i\} = \int_j \tag{4}$$

and $\mathbf{R}\int_i = \mathbf{R}\int_j$ implies $i = j$.

The first requirement implies that there is no privileged orientation. Clearly, the standard altitude and azimuth parameters do not lead to uniform partition of sphere. The second requirement enables efficient comparison of signatures by reducing all possible rotations to a discrete set of bin-to-bin comparisons. Planar rotations of GLAROT signatures are reduced to circular shifts preserving the adjacency relations among bins.

Unfortunately, there is no trivial and general way to partition the solid angle into regular regions. The only regular sphere tessellations with congruent regions and equal arc lengths are modelled on the five inscribed platonic solids. The symmetry groups of platonic solids enable correct face-to-face and, thus, bin-to-bin comparisons. Since there are only five tessellations with these properties, the bin number cannot be adapted to arbitrary resolution of orientation. *Cubemap* [16] is a popular projection used in computer graphics for visualization and represents a trade-off between such requirements, simplicity and efficiency. The idea is to divide the sphere $\mathbb{S}^2$ into six regions according to the projection of cube faces onto the inscribed sphere. These spherical regions are also called faces and each face is further divided into l^2 parts, where l is the number of face grid cells. The total number of bins of this orientation histogram is $6l^2$. The cube satisfies the octahedral symmetry, which allows comparisons between face grids. Figure 1 shows examples of this partition of sphere.

The six faces of cubemap are orthogonal to an axis and are identified either by the axis label (e.g. x^+, x^-, etc.) or by an index f from 0 to 5. The orthogonal axis to face is

n_r	rotation					
	x^+	x^-	y^+	y^-	z^+	z^-
1	x^+	x^-	y^+	y^-	z^+	z^-
2	x^-	x^+	y^-	y^+	z^+	z^-
3	y^-	y^+	x^+	x^-	z^+	z^-
4	y^+	y^-	x^-	x^+	z^+	z^-
5	x^-	x^+	y^+	y^-	z^-	z^+
6	x^+	x^-	y^-	y^+	z^-	z^+
7	y^+	y^-	x^+	x^-	z^-	z^+
8	y^-	y^+	x^-	x^+	z^-	z^+
9	x^-	x^+	z^+	z^-	y^+	y^-
10	x^+	x^-	z^-	z^+	y^+	y^-
11	z^+	z^-	x^+	x^-	y^+	y^-
12	z^-	z^+	x^-	x^+	y^+	y^-
13	x^+	x^-	z^+	z^-	y^-	y^+
14	x^-	x^+	z^-	z^+	y^-	y^+
15	z^-	z^+	x^+	x^-	y^-	y^+
16	z^+	z^-	x^-	x^+	y^-	y^+
17	y^+	y^-	z^+	z^-	x^+	x^-
18	y^-	y^+	z^-	z^+	x^+	x^-
19	z^-	z^+	y^+	y^-	x^+	x^-
20	z^+	z^-	y^-	y^+	x^+	x^-
21	y^-	y^+	z^+	z^-	x^-	x^+
22	y^+	y^-	z^-	z^+	x^-	x^+
23	z^+	z^-	y^+	y^-	x^-	x^+
24	z^-	z^+	y^-	y^+	x^-	x^+

TABLE II: List of 24 discrete rotations $\mathcal{R}$ corresponding to the orientation-preserving symmetries of cube.

referred as $\mathbf{d}_f$, while the other two axes parallel to the face are called $\mathbf{u}_f$ and $\mathbf{v}_f$. The vectors $\mathbf{u}_f$ and $\mathbf{v}_f$ are also used to compute the indices of a point projection into the face grid. Table I shows the list of faces with their respective indices f, labels, orthogonal axis $\mathbf{d}_f$ and parallel axes $\mathbf{u}_f$ and $\mathbf{v}_f$. The same symbol is used both for labels and, in bold font, for the corresponding vectors, e.g. $\mathbf{x}^+$ is the axis vector orthogonal to face x^+. The orientation of a point $\mathbf{r}_{ij} \in \mathbb{R}^3$ is described on the cubical orientation histogram through its projection on the grid. The face corresponding to such point is found as

$$f = \underset{f}{\operatorname{argmax}}\, \mathbf{d}_f^\top \mathbf{r}_{ij} \tag{5}$$

The indices of the grid cell on face f are computed w.r.t. the axes of the face $\mathbf{u}_f$ and $\mathbf{v}_f$ as

$$u = \left\lfloor l \left(\frac{2}{\pi} \arctan \left(\frac{\mathbf{u}_f^\top \mathbf{r}_{ij}}{\mathbf{d}_f^\top \mathbf{r}_{ij}} \right) + \frac{1}{2} \right) \right\rfloor \tag{6}$$

$$v = \left\lfloor l \left(\frac{2}{\pi} \arctan \left(\frac{\mathbf{v}_f^\top \mathbf{r}_{ij}}{\mathbf{d}_f^\top \mathbf{r}_{ij}} \right) + \frac{1}{2} \right) \right\rfloor \tag{7}$$

Indices u and v have values between 0 and $l-1$ and locate the specific bin on the square grid of the face. Thus, the orientation index function is the combination of the above indices into a unique identifier

$$i_{\mathbb{S}^2}(\mathbf{u}) = f \cdot l^2 + u \cdot l + v \tag{8}$$

number of face cells l

The proposed cubical histogram encodes the angular distribution of pairwise relative positions of landmarks. As anticipated before, it does not perfectly satisfy the second

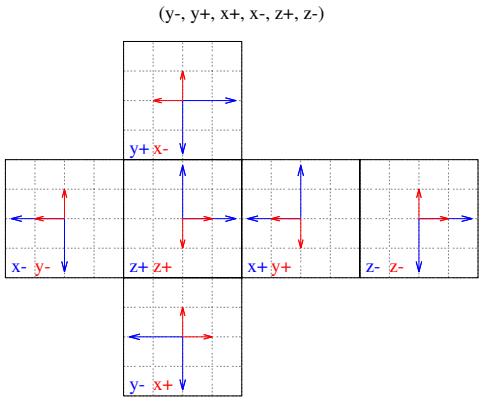

Fig. 2: Example of cubemap rotation ($n_r = 3$ in Table II): the reference face labels and axes are in blue, the rotated faces and axes are in red.

requirement, since comparisons are performed on faces instead of bins. Table II shows the discrete rotation set $\mathcal{R}$ corresponding to the 24 orientation-preserving symmetries of cube. The notation for rotations is similar to the standard notation used for permutation. For example, rotation $n_r = 3$ maps face x^+ to y^-, x^- to y^+, y^+ to x^+, etc. Figure 2 illustrates the face mapping after rotation. Observe that the order of adjacent faces w.r.t. a given face is unchanged by the rotation, since the transformation is orientation-preserving. However, while comparing two overlapping face grids, the orientation grid should be rotated according to its axes $\mathbf{u}_f$ and $\mathbf{v}_f$. For example, face y^- lies on face x^- (respectively red and blue in Figure 2) and its axes are rotated by $3 \times 90°$ counterclockwise w.r.t. x^- axes. These rotations can be precomputed for each discrete rotation of $\mathcal{R}$.

C. Computation and Comparison of Signatures

The histogram $\mathbf{G}$ of GLAROT-3D can be computed using the previously illustrated index functions $i_{\mathbb{S}^2}(\mathbf{r}_{ij})$ and $i_\rho(\mathbf{r}_{ij})$. The total number of $\mathbf{G}$ bins is equal to $6l^2 n_\rho$, where l is the number of face grid cells and n_ρ is the number of range grid cells. The range size n_ρ is chosen according to the range resolution $\Delta\rho$ and the expected maximum pairwise distance of two landmarks. If the signature histogram is vectorized, the signature bin index for a points $\mathbf{r}_{ij}$ is $i_{\mathbb{S}^2}(\mathbf{r}_{ij})\, n_\rho + i_\rho(\mathbf{r}_{ij})$.

The computation of the histogram of a landmark set $\mathcal{P}$ is straightforward. For all the point pairs $\mathbf{p}_i, \mathbf{p}_j \in \mathcal{P}$ with $i \neq j$, the pairwise positions $\mathbf{r}_{ij} = \mathbf{p}_i - \mathbf{p}_j$ and $\mathbf{r}_{ji} = -\mathbf{r}_{ij}$ are computed. Then, the bins corresponding to $\mathbf{r}_{ij}$ and $\mathbf{r}_{ji}$ are incremented by one unit. Contrary to GLARE [15], GLAROT-3D considers both the two pairwise relative positions of landmarks. Moreover, the counter values of $\mathbf{G}$ are integer instead of real values sampled from Gaussian functions. These straightforward design solutions are more suitable to the complex structure of 3D domain.

As anticipated, the distance between two GLAROT-3D signatures $\mathbf{G}_T$ and $\mathbf{G}_S$ is computed according to a variant of L_1 norm. The cubical orientation grid efficiently handles

the rotations $\mathbf{R}$ belonging to discrete set $\mathcal{R}$ of octahedral symmetries. The rotated histogram $\mathbf{RG}_S$ is obtained by the order-preserving permutation of its face grids and the relative rotation of face axes as previously illustrated. The *rotated* L_1 *norm* is defined as

$$RL_1(\mathbf{G}_S, \mathbf{G}_T) = \min_{\mathbf{R}\in\mathcal{R}} \|\mathbf{G}_T - \mathbf{RG}_S\|_1 \tag{9}$$

Although the computation of $RL_1(\cdot,\cdot)$ requires 24 evaluations of L_1 norm, all the operations involved are lightweight, and consists of integer arithmetic operations and histogram access.

IV. Experiments

In this section we present the experimental setup and the results obtained in loop candidate detection using GLAROT-3D and DBoW2 algorithm [14] designed for binary descriptors like ORB. Six sequences have been used: *fr3_structure_texture_far*, *fr2_desk* and *fr3_long_office* from from TUM RGB-D Benchmark [28], and stereo vision sequences *kitti_00*, *kitti_05* and *kitti_06* from KITTI visual odometry dataset [29]. Sequences *fr3_structure_texture_far*, *fr2_desk* and *fr3_long_office* consist of RGB and depth images that have been acquired indoor by a hand-held RGB-D camera. In the first sequence *fr3_structure_texture_far* the operator performs limited translation and significant rotation motions, whereas in *fr2_desk* and *fr3_long_office* the depth camera performs a loop about 20 m long in an office environment. Sequences *kitti_00*, *kitti_05* and *kitti_06* are collected by stereo cameras mounted on a car executing one or more closed $2-3$ km long paths. ORB-SLAM2 [17] has been used to build a global map consisting of smaller ORB features maps.

The coordinates of features are referred to the keyframe of the local map, i.e. the reference frame of the sensor from where the features were observed. The detected ORB features along with the BoW weight vector of each features have been exported to file. We adopted the ORB vocabulary distributed with ORB-SLAM2.

GLAROT-3D signature is computed by exploiting only the coordinates of ORB keypoints. The corresponding descriptors are used to compute the DBoW2 vector according to the adopted dictionary. Table III presents the number of keyframes n_{frames}, the average number of points per keyframe n_{points} and the size of GLAROT-3D histograms. The average and maximum pairwise point distances are respectively about 0.5 m and 5.0 m in TUM sequences, whereas about 10.0 m and 17.0 m in KITTI sequences. The parameters n_ρ and $\Delta\rho$ must be chosen according to the keypoint distribution obtained by perception in a given enviroment. Trials with different reasonable values of such range parameters achieve similar results, when the range domain is covered ($n_\rho \cdot \Delta\rho$ greater than maximum pairwise landmark distance) and $\Delta\rho$ enables accurate distance distribution. The number of face grid cells l is less sensitive to scale differences of environments.

The loop closure procedure used in the experiments is the following.

dataset	n_{frames}	n_{points}	l	n_ρ	$\Delta\rho$ [m]
fr3_structure_texture_far	24	409.9	2	70	0.05
fr2_desk	179	273.1	2	70	0.05
fr3_long_office	191	344.4	2	70	0.05
kitti_00	1515	471.5	2	200	0.10
kitti_05	898	461.7	2	200	0.10
kitti_06	465	461.6	2	200	0.10

TABLE III: Parameters of the datasets (number of keyframes n_{frames}, average number of points per keyframe n_{points}) and of GLAROT-3D signature (number of face cells l, number of range cells n_ρ and range resolution $\Delta\rho$) used in the experiments.

1) Given a query local map $\mathcal{P}_q$, ORB-SLAM2 provides an initial list of potentially matching local maps $\mathcal{P}_m$ based on co-visibility between them and the query one. To avoid trivial matches between consecutive maps, we exclude from the candidate list all the $\mathcal{P}_m$ with $|m-q| < 5$.
2) A subset of $k = 10$ candidates $\mathcal{P}_m$ are selected according either to GLAROT-3D or DBoW2. GLAROT-3D scores each $\mathcal{P}_m$ by comparing the corresponding signature $\mathbf{G}_m$ and the signature of query map $\mathbf{G}_q$. In particular, the k selected maps are the k-nearest neighbors measured according to the rotated signature metric $RL_1(\mathbf{G}_q, \mathbf{G}_m)$. Similarly, DBoW2 scores the candidate according to the BoW vectors and L_1 norm.
3) Point-to-point association and registration between $\mathcal{P}_q$ and each $\mathcal{P}_m$. The candidate loop closure is accepted, if a sufficiently high number of correspondences is found after registration. In particular, the number of mutual nearest point pairs of the transformed $\mathcal{P}_q$ and $\mathcal{P}_m$ must be greater then the acceptance threshold.

The correctness of each association is assessed by comparing the estimated relative poses after the registration. If the relative location between the two frames is close to the groundtruth value, the closure is marked as a true positive, otherwise as a false positive. The maximum allowed position and angular errors are respectively 0.5 m and 10° for TUM and 4.0 m and 10° for KITTI datasets.

Figure 3 illustrates the precision-recall curves of *fr3_structure_texture_far*, *fr2_desk*, *fr3_long_office*, *kitti_00*, *kitti_05* and *kitti_06* obtained from the procedure described above. Each point of the curve corresponds to a different threshold on the number of point-to-point correspondences. GLAROT-3D performs clearly better than DBoW2 in all the scenarios except for *kitti_06*. In this sequence, the bag-of-word approach discriminates more accurately the k nearest candidates, when the acceptance threshold on the number of corresponding points is low. The average time for comparing two GLAROT-3D signatures is 2 ms for TUM and 32 ms for KITTI sequences, whereas DBoW2 requires less than 1 ms in all cases using a Intel i7-3630QM CPU @ 2.40GHz, 8 GB RAM. Thus, DBoW2 is faster than GLAROT-3D, but seems to achieve less accurate loop closure detection.

V. Conclusion

In this paper, we have presented GLAROT-3D, a novel rotation-invariant geometric signature for loop closure candi-

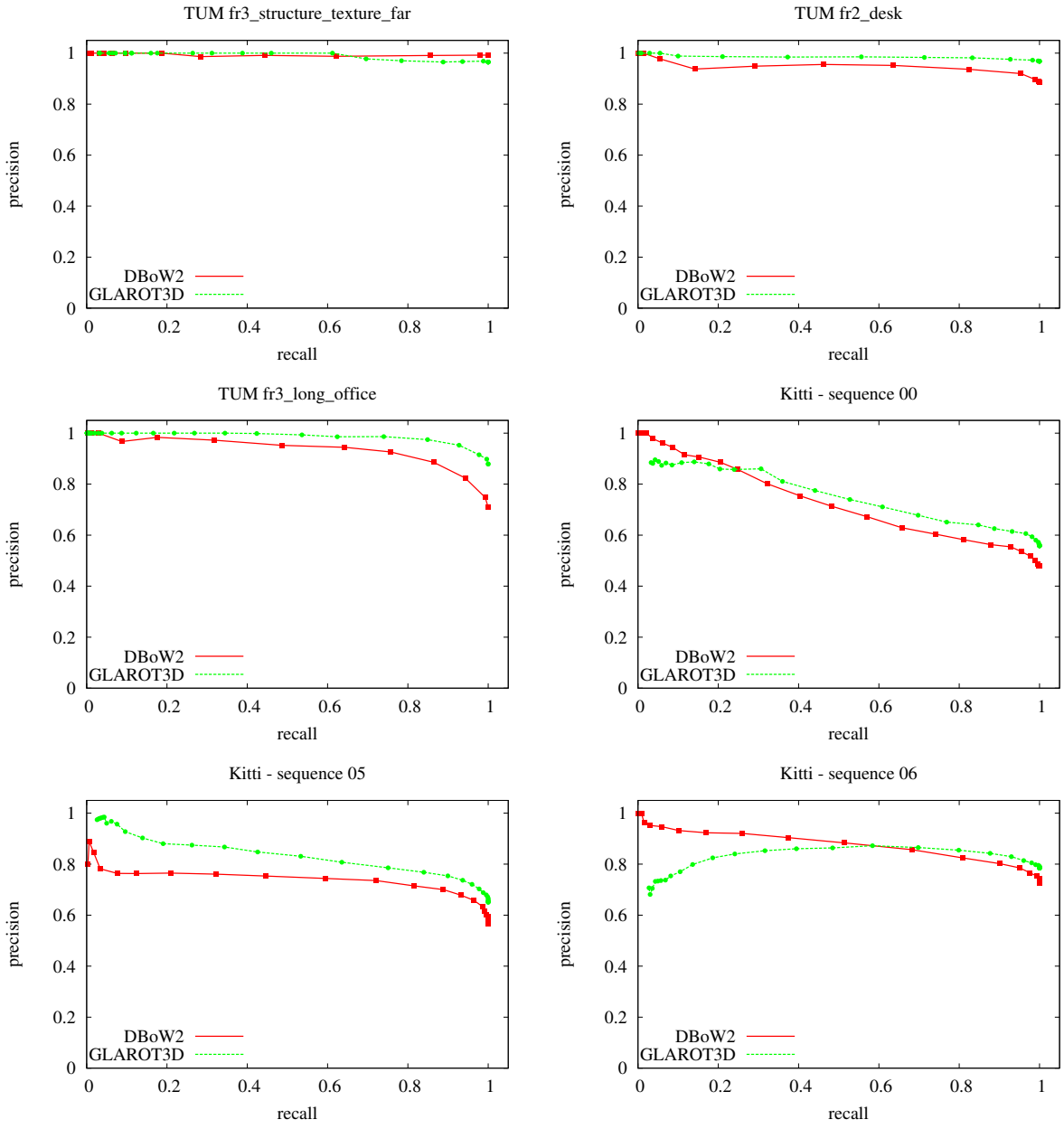

Fig. 3: Precision-recall curves of loop closure based on GLAROT-3D and DBoW2 in datasets *fr3_structure_texture_far*, *fr2_desk*, *fr3_long_office*, *kitti_00*, *kitti_05* and *kitti_06*.

date selection. The signature is a histogram encoding length and orientation of the pairwise relative positions obtained from a landmark map. It is easy to implement and does not require any prior training stage like bag-of-words and other approaches, since it is based only on geometric relations. GLAROT-3D is invariant to translation and rotation of the reference frame of the local map. To achieve invariance, the algorithm is able to efficiently rotate the signature histogram by exploiting the support of a cubical grid partition of 3D orientations. In experiments with standard benchmark datasets, GLAROT-3D has been able to find loop closure candidates with similar or better accuracy than state-of-the-art approaches. In future works, we expect to integrate the proposed algorithm into a state-of-the-art 3D SLAM system in order to better assess its performance on large scale datasets and to compare with other state-of-the-art techniques.

VI. ACKNOWLEDGEMENTS

This research has been supported by the University of Parma in the FIL 2016 *R3D-MAN* (Robust 3D Mapping and Navigation) project. I thank Prof. Stefano Caselli for his support in the discussion and the review of this work.

REFERENCES

[1] J.-S. Gutmann and K. Konolige, "Incremental Mapping of Large Cyclic Environments," in *Proc. of the IEEE Int. Symposium on*

Computational Intelligence in Robotics and Automation (CIRA), 1999, pp. 318–325.

[2] G. Grisetti, C. Stachniss, and W. Burgard, "Improved Techniques for Grid Mapping with Rao-Blackwellized Particle Filters," *IEEE Trans. on Robotics*, vol. 23, no. 1, pp. 34–46, 2007.

[3] D. Lodi Rizzini and S. Caselli, "Metric-topological maps from laser scans adjusted with incremental tree network optimizer," *Robotics & Autonomous Systems*, vol. 57, no. 10, pp. 1036 – 1041, 2009.

[4] G. D. Tipaldi and K. O. Arras, "Flirt-interest regions for 2d range data," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2010, pp. 3616–3622.

[5] F. Kallasi, D. Lodi Rizzini, and S. Caselli, "Fast keypoint features from laser scanner for robot localization and mapping," *RAL*, vol. 1, no. 1, pp. 176–183, jan 2016, dOI 10.1109/LRA.2016.2517210.

[6] F. Kallasi and D. Lodi Rizzini, "Efficient Loop Closure based on FALKO LIDAR Features for Online Robot Localization and Mapping," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2016, pp. 1–8.

[7] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," vol. 27, no. 10, pp. 1615–1630, 2005.

[8] S. Krig, "Interest point detector and feature descriptor survey," in *Computer Vision Metrics*. Springer, 2014, pp. 217–282.

[9] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, and J. Wan, "3D Object Recognition in Cluttered Scenes with Local Surface Features: A Survey," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2270–2287, Nov 2014.

[10] D. Lodi Rizzini, F. Kallasi, F. Oleari, and S. Caselli, "Investigation of vision-based underwater object detection with multiple datasets," *International Journal of Advanced Robotic Systems (IJARS)*, vol. 12, no. 77, pp. 1–13, may 2015.

[11] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *Proc. IEEE ACM Int. Symp. Mixed Augmented Reality*, 2007, pp. 225–234.

[12] H. Strasdat, A. Davison, J. Montiel, and K. Konolige, "Double window optimisation for constant time visual slam," in *Proc. IEEE Int. Conf. Comput. Vision*, 2011, pp. 2352–2359.

[13] R. Mur-Artal, J. Montiel, and J. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," *IEEE Trans. on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[14] D. Galvez-Lopez and J. Tardos, "Bags of Binary Words for Fast Place Recognition in Image Sequences," *IEEE Trans. on Robotics*, vol. 28, no. 5, pp. 1188–1197, Oct 2012.

[15] M. Himstedt, J. Frost, S. Hellbach, H.-J. Boehme, and E. Maehle, "Large scale place recognition in 2D lidar scans using geometrical landmark relations," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2014, pp. 5030–5035.

[16] N. Greene, "Environment mapping and other applications of world projections," *IEEE Computer Graphics and Applications*, vol. 6, no. 11, pp. 21–29, Nov 1986.

[17] R. Mur-Artal and J. Tardos, "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras," *IEEE Trans. on Robotics*, vol. PP, no. 99, 2017.

[18] G. D. Tipaldi, L. Spinello, and K. O. Arras, "Geometrical flirt phrases for large scale place recognition in 2d range data," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2013, pp. 2693–2698.

[19] J. Deray, J. Solà, and J. Andrade-Cetto, "Word ordering and document adjacency for large loop closure detection in 2-d laser maps," *IEEE Robotics and Automation Letters (RA-L)*, vol. 2, no. 3, pp. 1532–1539, 2017.

[20] J. Neira and J. Tardós, "Data Association in Stochastic Mapping Using the Joint Compatibility Test," *IEEE Trans. on Robotics*, vol. 17, no. 6, pp. 890–897, 2001.

[21] T. Bailey, E. Nebot, J. Rosenblatt, and H. Durrant-Whyte, "Data association for mobile robot navigation: a graph theoretic approach," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2000, pp. 2512–2517.

[22] M. Cummins and P. Newman, "Appearance-only slam at large scale with fab-map 2.0," *Int. Journal of Robotics Research*, vol. 30, no. 9, pp. 1100–1123, 2011.

[23] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[24] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. IEEE Int. Conf. Comput. Vision*.

[25] J. Engel, J. Stueckler, and D. Cremers, "Large-Scale Direct SLAM with Stereo Cameras," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2015, pp. 1935–1942.

[26] A. Glover, W. Maddern, M. Warren, R. Stephanie, M. Milford, and G. Wyeth, "OpenFABMAP: an open source toolbox for appearance-based loop closure detection," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2012, pp. 4730–4735.

[27] M. Himstedt and E. Maehle, "Geometry matters: Place recognition in 2D range scans using Geometrical Surface Relations," in *Proc. of the European Conference on Mobile Robots (ECMR)*, 2015, pp. 1–6.

[28] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2012, pp. 573–580.

[29] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.