

# Geometrical FLIRT Phrases for Large Scale Place Recognition in 2D Range Data

Gian Diego Tipaldi, Luciano Spinello and Wolfram Burgard

**Abstract**—Place recognition, i.e., the problem of recognizing if the robot is navigating in an already visited place, is a fundamental problem in mobile robot navigation. Efficient solutions to this problem are relevant for effectively localizing robots and for creating maps in real time. Relatively few methods have been proposed to efficiently solve this problem in very large environments using 2D range data. In this paper, we introduce geometrical FLIRT phrases (GFPs) as a novel retrieval method for very efficient and precise place recognition. GFPs perform approximate 2D range data matching, have low computational cost, can handle complicated partial matching patterns and are robust to noise. Experiments carried out with publicly available datasets demonstrate that GFPs largely outperform state-of-the-art approaches in 2D range-based place recognition in terms of efficiency and recall. We obtain retrieval performances with more than 85% recall at 99% precision in less than a second, even on data sets obtained from several kilometer long runs.

## I. INTRODUCTION

Place recognition, namely the problem of understanding if a robot is revisiting an already known area, is a fundamental problem in robot navigation tasks such as localization and SLAM. In the past, relatively few approaches have been developed for solving this problem in the context of 2D range data. The majority of solutions present in the literature focus on the use of cameras and employ feature-based techniques to compute place *signatures* for retrieval purposes. While in theory similar ideas can be applied to 2D range data, special care needs to be taken in that context. Compared to an image, a 2D range scan is a partial, scarce and sparse representation of the real 3D environment. Accordingly, a 2D range scan contains only little information for distinguishing a place with respect to another. It is therefore hard to generate appropriate place signatures that can effectively be applied also to large datasets.

In this paper, we introduce *geometrical FLIRT phrases* (GFPs) for robust and efficient retrieval in the context of 2D range scans. GFPs are inspired by the work of geometry-preserving visual phrases, introduced in [19], and make use of FLIRT [15], a state-of-the-art 2D range-data keypoint detector for building up individual scan signatures of large datasets. Our method improves classical bag of words (BoWs) retrieval approaches [1] because it encodes not only the FLIRT words distributions in a scan but it also partially captures their geometrical arrangement, in a translation and rotation invariant manner. As a result, GFPs are able to robustly capture similarities between places whose structures only

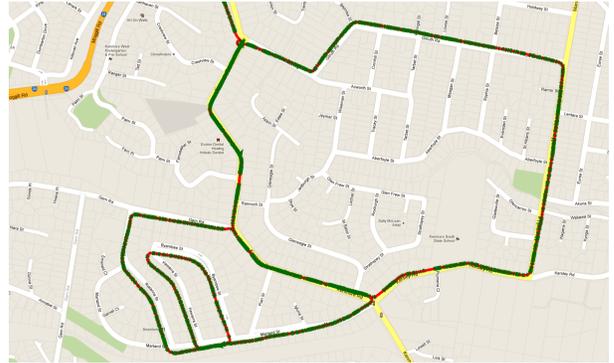


Fig. 1. Experiment on the first half of the Kenmore dataset, depicting the path of the robot (red) and the recognized places (green) at 99.9% precision. The computation time is 0.8s per query (one scan vs. 13000+).

partially match or in presence of outliers. This is seamlessly performed in the retrieval process and with a very low computational performance impact. Note that our approach does not aim to perform full geometrical verification between two scans. It rather aims at providing an effective mean for checking approximate structures between them. GFPs allow easy tuning of the structural approximate matching by a compact formulation of a special kernel function. GFPs yield a small candidate set that can effectively be checked via a RANSAC-like method to assess strong geometrical consistency and to find a valid match. The result of this is a highly effective place recognition process that can robustly operate on large-scale data sets such as the one depicted in Fig. 1.

The main contributions of this paper are the introduction of GFPs as a robust and efficient retrieval technique with low computational efforts and the presentation of practical experiments carried out with publicly available datasets that demonstrate, that GFPs outperform the state of the art both in retrieval accuracy and query time.

## II. RELATED WORK

There is relatively little related work that is concerned with place recognition using 2D range data. Most of the approaches rely on the use of visual information (see Williams *et al.* [16] for a survey). Nevertheless, few authors did focus on place recognition and loop closing using 2D laser range finders. Already in 2003, Neira *et al.* [8] presented a work on robot relocation, where geometrical features are used to find the location of the robot in a stochastic map. The authors use a combination of hard geometric constraints, feature correlation, joint compatibility, random sampling and locality to make

All authors are with the University of Freiburg, Institute of Computer Science, Germany. {tipaldi, spinello, burgard}@informatik.uni-freiburg.de.

This work has been partly supported by the European Commission under the FP7-ERC-267686-LiveNav and FP7-248873-RADHAR.

the algorithm linear in the size of the map. The approach was extended by Paz *et al.* [9] to general grid maps keeping the same complexity. In [3], Bosse and Zlot evaluate several detector/descriptor-pairs for 2D range data for the task of place recognition in a graphical, submap-based SLAM application. They use a voting approach to match submaps, where each interest point votes for a candidate submap. Association between interest point is done using kd-trees. Their approach share same similarities with our work since it may be seen as “vocabulary-free” BoW, without any geometrical verification Tipaldi *et al.* [15] also used a feature-based approach combined with RANSAC to perform place recognition. The authors introduced a feature detector and descriptor for 2D range data that allows for high precision and recall in place recognition. Bosse and Zlot [2] define global features on individual scan or local maps instead of local interest points and use those descriptions to match a set of reference scans/maps to a query one. They describe local maps using projection histograms and orientation histograms as descriptors. Matching is done using cross-correlations between the histogram signatures to find the most likely match. Those histogram however, are very sensitive to the view point and showed lower recall rate of a feature based approach. Granström *et al.* [4] define a set of local features for a scan (e.g., area, average range, circularity, curvature, ...) which are used as input for AdaBoost to build a non-linear classifier capable of detecting loop closure from pairs of point clouds. In contrast to our approach, their method is not able to compute a transformation between scans and has a longer run-time. It is worth noticing that our work is orthogonal to theirs and both could be combined together.

Our method is inspired by the work of Zhang *et al.* [19] in the computer vision community. There, the authors introduce *phrases*, geometrically consistent co-occurrences of visual-words, and present an efficient technique for handling large image datasets. Also relevant to our approach is the work of Zhang and Chen [18] in which the authors introduce kernels to encode high-order spatial relations in a bag-of-visual-features model for near-duplicate image matching tasks in large datasets. Both works share the same idea of capturing local and long-range spatial visual word layouts for improving image matching with low computational cost. Their technique is able to cope with translation-only partial matching. Our approach instead provides invariance to translation and rotation. Other works [17] search repeated word co-occurrences within limited ranges and build a word lexicon by employing frequent itemset mining techniques. Our approach also relates to works that employ spatial verification in visual BoW [10, 11, 6]. However, these are hardly suitable for handling the specific nature of a 2D range scan and its low specificity for place description.

### III. PLACE RECOGNITION USING FLIRT DICTIONARIES

The Fast Laser Interest Point Transform (FLIRT) is a multi-scale interest point operator for 2D range data introduced by Tipaldi *et al.* [15]. The operator consists of a detector based on a geodesic curve approximation of the range signal and a descriptor based on a polar histogram of occupancy

probabilities. The detector models the scan as a curve in Cartesian space and applies an integral operator invariant to sampling density that smooths the input curve into its multi-scale parametrization. Interest points are then detected by finding the local maxima of the distance between the smoothed curve and the original one. For each detected interest point a linear-polar tessellation of the space around is defined for the descriptor, with a radius proportional to the scale of the interest point. For estimating the occupancy probability, a Bayesian parameter learning approach is used. More details can be found in the work of Tipaldi *et al.* [15].

Place recognition using FLIRT features can be performed in linear time w.r.t. the number of places by applying the RANSAC algorithm between the query scan and each scan present in the place database. In practice, RANSAC returns the place with the minimum reprojection error, whose inlier set is above a threshold  $n_I^{min}$ . Although the approach is very efficient, the linear complexity prevents it from being practical on very large datasets.

In order to overcome this problem, a naive solution would be to adopt a BoW approach [14]. The idea of such methods is to create a distribution of feature appearances in a certain place and to use this distribution as a description of the place itself. A place is encoded as a set of predefined words belonging to a precomputed dictionary. Each feature is associated to the closest word in the dictionary according to the distance between the word and the descriptor of the feature. Once each feature has been associated to a word, a histogram of word frequencies is used as a signature of the place. For recognition, we first sort all the places in the database according to the distance of their signature to the signature of the query path. We then perform RANSAC on the top  $H$  matches to verify the geometrical configuration of the feature and return the most likely place (or none if the retrieval criterion are not satisfied).

To learn the dictionary we use a training database of 10000 scans randomly sampled from publicly available datasets. On each of these scans we extract FLIRT features – on average 20 per scan – and we apply the hierarchical K-means algorithm to find the number of words. We use a 3 layer hierarchy with 5 clusters per level. We compute several dictionaries starting with different random seeds and select the best dictionary using the V-measure [12]. The final vocabulary consists of 275 words.

### IV. PLACE RECOGNITION WITH FLIRT PHRASES AND WEAK GEOMETRICAL VERIFICATION

In this section, we introduce the method to recognize places by means of weak geometrical verification. First, we present the core formulation of the weak geometrical verification for matching 2D range scan data. Then, we present an efficient search structure for matching and retrieval in large datasets.

#### A. GFP-based 2D Place Recognition

Our method aims to rapidly quantify similarities between two 2D range scans, not only by considering them as unordered collections of local descriptors but also by taking into account the positional arrangement of keypoints.

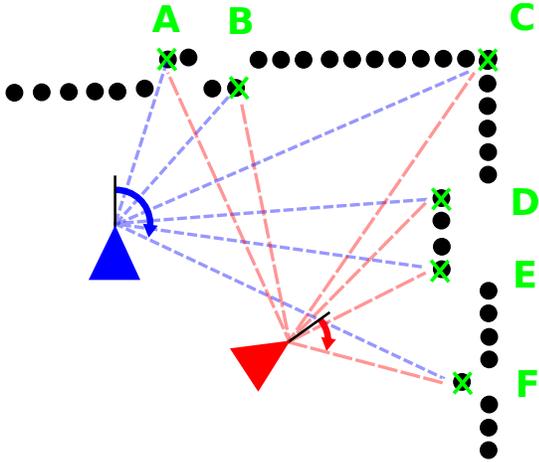


Fig. 2. Visual explanation of invariance with respect to rotation and translation of the angular order of keypoints in a scan. The figure depicts a 2D range scan (black dots), two robots (blue (upper) and red (lower) triangles) and the position of FLIRT keypoints (green crosses) with the associated FLIRT words (green letters). The angular order of the features in a scan is invariant. In the example, the blue robot observes the features (in a clockwise fashion) as  $\{A, B, C, D, E, F\}$  and the red one as  $\{E, F, A, B, C, D\}$ . Feature ordering is preserved.

A popular approach to check the geometrical similarity between two scans is by estimating the rigid transformation between them and checking the resulting reprojection error. The idea pursued in this work is to avoid this computationally expensive operation, to find an *invariant* representation of range scans and to apply simplified matching operations. Note that the transformation of the scan in the new representation should require efficient operations. The proposed solution represents the scan as a set of FLIRT keypoints with their clockwise order of appearance in the range scan as a simplified representation.

Throughout this paper, we use the term “weak geometrical verification” for the process of comparing two scans by analyzing the clockwise ordering of their FLIRT features. This process leverages the peculiarity and characteristics of 2D range data: a 2D range scan is a 1D manifold embedded in a 2D space where it is possible to define an ordering of its data, see Fig. 2. Can be proved that the set of scans with similar arrangement of keypoints contains also the ones that match with a full geometrical verification method.

Our novel theoretical framework makes use of *phrases* for a tunable formulation of the weak geometrical verification principle. The notion of phrases comes from the information retrieval community [7] and it represents a (high-order) arrangement of words. In the context of our work, we define phrases by using FLIRT words (previously generated from a standard BoW dictionary construction, see Section III). A phrase  $p^k$  of order  $k$  consists in  $k$  clockwise-ordered words (e.g.:  $p^2 = \{LS\}$ ,  $p^3 = \{GDT\}$ ,  $p^4 = \{WBFR\}$  etc). Similar to the BoW document representation, a scan that employs GFPs is encoded as a histogram  $\Phi(\cdot)$  of all possible phrases of length  $k$ . Each bin represents the occurrence of a certain phrase  $p_i^k$  in a scan. The similarity between two scans is computed by calculating the cosine distance between two histograms. The

dimensionality of the histogram  $\Phi(\cdot)$  is very large, containing  $n^k$  bins. Even with  $k = 2$  and a small vocabulary ( $n = 200$ ),  $\Phi(\cdot)$  is so large and sparse that is impractical to use it.

We define the set of FLIRT words associated to the scan  $\mathbf{x}_i$  as  $\mathcal{X}_i^W$ . Similar to Zhang and Chen [18], it is possible to formulate the GFP histogram comparison with a function

$$K(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathcal{X}_i^W), \Phi(\mathcal{X}_j^W) \rangle \quad (1)$$

that computes the inner product of two GFP histograms. The dot product is an operator that is symmetric and positive semi-definite, so that  $K(\cdot)$  satisfies the Mercer condition and defines a kernel function. Eq. (1) can be normalized by dividing each histogram by its norm. This kernel has an interesting property:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \sum_{l=1}^{n^k} \frac{U_{p_l}(\mathcal{X}_i^W)}{\sum_{m=1}^{n^k} U_{p_l}(\mathcal{X}_i^W)} \frac{U_{p_l}(\mathcal{X}_j^W)}{\sum_{q=1}^{n^k} U_{p_l}(\mathcal{X}_j^W)} \quad (2)$$

Here,  $U_{p_l}(\mathcal{X}^W)$  is a function that returns the number of phrases  $p_l$  in the set  $\mathcal{X}^W$ . In practice, the dot product of the histograms  $K(\cdot)$  is equal to counting the total number of co-occurring phrases of the same kind in the two scans. This information can be exploited for a highly efficient solution of Eq. (1). The intuition is that when a GFP phrase of order  $k$  is present in two scans there must be a constant shift in the clockwise-order arrangement of  $k$  FLIRT words. Let us consider the clockwise ordered set  $\tilde{\mathcal{X}}_i^W$  of words in the scan. For a word  $w \in \tilde{\mathcal{X}}_i^W$  placed at position  $m$  in the ordered set, we consider its position within the ordered set  $\tilde{\mathcal{X}}_j^W$  associated with the scan  $\mathbf{x}_j$  and we compute the following:

$$\mathcal{J}_w = \{\text{indices of } (w \in \tilde{\mathcal{X}}_j^W)\} \quad (3)$$

$$\mathcal{O} = m - \mathcal{J}_w \quad (4)$$

$$\Omega(\mathcal{O}) = \Omega(\mathcal{O}) + 1 \quad (5)$$

The term  $\Omega(\cdot)$  is the histogram of offsets that collects word co-occurrences and it has  $O = \max(Z, T) \ll n^k$  bins, where  $Z = \#\tilde{\mathcal{X}}_j^W$  and  $T = \#\tilde{\mathcal{X}}_i^W$ . In practice, in case that co-occurring words are present with consistent clockwise-ordered layout in two scans, their difference in index falls into the same bin of the offset histogram  $\Omega(\cdot)$  (see Fig. 3 for an illustration). Thanks to the offset space, it is possible to compare scans that share complicated interleaved word patterns or large displacements in the ordered word-sets, and to rewrite Eq. (2):

$$K(\mathbf{x}_i, \mathbf{x}_j) = \sum_{l=1}^O \binom{\Omega(l)}{k} \quad (6)$$

The binomial coefficient is used to count the phrases, constituted by the combinations of  $m$  words taken  $k$  times, co-occurring in each offset bin. In this manner, we can easily compute scan similarities with phrases of all orders (1 to  $\infty$ ) with the same computational efforts.

### B. GFP-enabled Efficient Retrieval

An efficient technique used for searching documents in large databases in a BoW context, is the *Term Frequency-Inverse Document Frequency* (TFIDF) with inverted file index [13].

In our case, we consider only IDF phrase weights, that is the frequency of a certain phrase appearing in the entire corpus. For efficiency reasons, we define the IDF weight of a phrase as the sum of the words' IDF composing it. It is possible to rewrite the kernel in a way that takes into account IDF weighting. The function  $K_{\text{IDF}}(\mathbf{x}_i, \mathbf{x}_j)$  is equal to the summation of the IDF weights of the co-occurring phrases.

For matching two scans, we store the number of clockwise-arranged co-occurring words in the offset histogram  $\Omega$  and the summation of their IDF weights in the weight histogram  $\Xi$ . To clarify, the co-occurrence of the word  $w$  in two scans with offset position  $m$  is used to update Eq. (5) and  $\Xi(\cdot)$  :

$$\Xi(m) = \Xi(m) + v_w \quad (7)$$

$$K_{\text{IDF}}(\mathbf{x}_i, \mathbf{x}_j) = \sum_{l=1}^O \Xi(l) \binom{\Omega(l)-1}{k-1} \quad (8)$$

where  $v_w$  is the IDF weight of the word  $w$ . We modified the inverted file index technique for the GFPs approach by modifying the indexing structure. The novel inverted file index  $\mathcal{G}$  contains the indices of the range scans of the corpus in which a certain word  $w_m$  appears and it caches the position of the word in those scans. In particular:

$$(w_i, v_{w_i}, \mathcal{S}_{w_i}) \in \mathcal{G}, \quad (q, \mathbf{l}_q) \in \mathcal{S}_{w_i} \quad (9)$$

where  $\mathbf{l}_q$  is the vector containing the indices of the position in clockwise-order of the word  $w_i$  in the set  $\tilde{\mathcal{X}}_q^W$  relative to the scan  $q$ . The query of a scan  $\mathbf{x}_i$  with a 2D range scan corpus is very fast. Each word  $w$  in  $\tilde{\mathcal{X}}_i^W$  is used to select the associated entry in  $\mathcal{G}$  and the offsets  $\mathcal{O}$  are computed for each  $\mathcal{S}_w$  by subtracting the index of  $w$  with  $\mathbf{l}_q$ . Then, the histograms  $\Omega$  and  $\Xi$  are updated for each of the scans relevant to all the words in  $\tilde{\mathcal{X}}_i^W$ . Eventually,  $K_{\text{IDF}}$  scores all the relevant scans by using Eq. (8). All these scans are ranked with their score to provide the best matches. The RANSAC scan-matching technique is then employed to compute the rigid transformation on the top  $H$  GFP-matches, similar to the BoW explained in Sec. III.

Note that the computational effort in the retrieval is bounded by the number of words in the query scan. The geometrical verification is very fast because it consists only in a simple difference between vector indices, and no trigonometric transformations are run. The binomial coefficient is implemented as a pre-computed look-up table.

## V. EXPERIMENTS

In this section we describe our experimental set-up and show the performances of the proposed approach. For the experiments, we used six publicly available datasets of 2D scans consisting of three indoor datasets and three outdoor datasets. Tab. I lists the datasets used and their size. Each dataset has been corrected using the SLAM algorithm of Grisetti *et al.* [5] to obtain the ground-truth trajectory and evaluate if the solution retrieved by the system is correct.

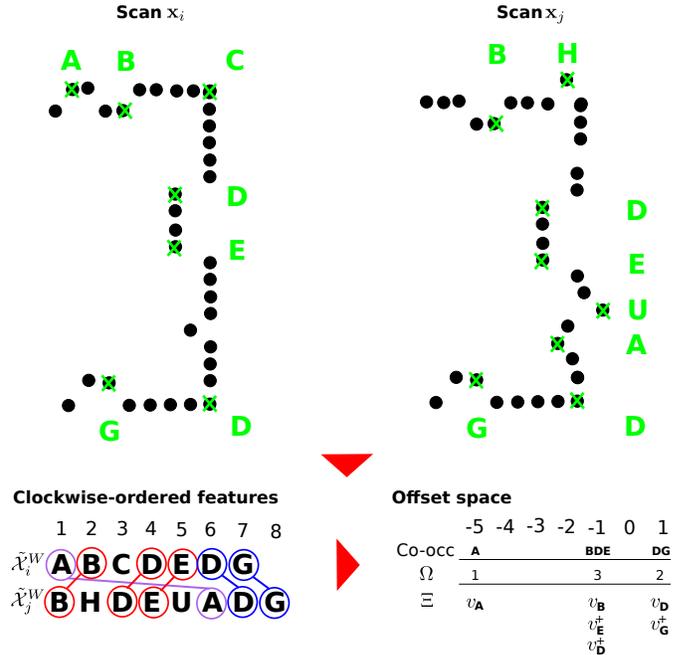


Fig. 3. Efficient solution to GFP matching. In the figure, two similar scans are being compared. The clockwise feature arrangement is computed, the histogram of words co-occurrence  $\Omega$  and IDF weights  $\Xi$  is filled. Note that with this method, we can successfully match phrases encoded with complicated interleaved word patterns (eg: BDE). Given the two histograms, we can compute distance kernels of different orders at the same computational cost (eg: matching duplets, triplets, etc), see Eqn. 8.

TABLE I  
STATISTICS OF THE DATASETS USED IN THE EXPERIMENTS.

Data set	Size [m]	Scans	Path length [m]
FR-079 (in)	50×20	1464	390.803
INTEL-Lab (in)	50×40	2672	360.695
MIT-CSAIL (in)	80×60	1051	382.942
FR-Clinic (out)	550×300	6917	1437.6
Victoria Park (out)	500×500	5751	4206.14
Kenmore (out)	1000×1000	13063	6588.34

### A. Experimental Set-Up

We compared place recognition strategies based on BoW and GFPs. We employ TFIDF scaling for reweighting the bins of the BoW histograms [13]. In particular we tested standard TFIDF (BoW-1) and robust variants called “maximum TF normalization” (BoW-2) and “sublinear TF scaling” (BoW-3). For GFPs, we used a kernel size of 2 (GFP-2) and 3 (GFP-3). No particular difference was found for bigger kernels.

To evaluate the performances of the different approaches, we used the following procedure for each of the dataset. At first, features are extracted from every scan in the dataset, are associated to the words of the dictionary and the corresponding signature is computed. Then, for each scan of the dataset, we compute the similarity of that scan with all the others in the dataset using the BoW or the GFP histograms.

To assess if two scans represent a valid match we check if the number of inlier correspondences returned by RANSAC is above a threshold,  $n_I^{\text{min}}$ . Once we obtain the set of valid

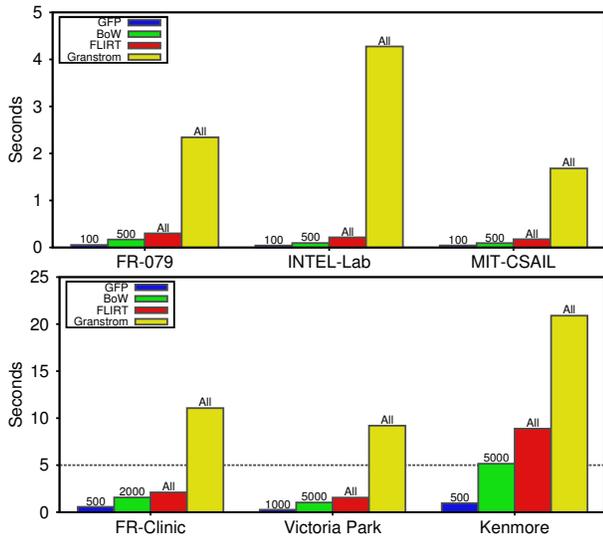


Fig. 4. Run-time in seconds using respectively GFPs, BoW, FLIRT [15] and Granström [4] for the indoor (top) and outdoor (bottom) datasets at maximum precision/recall values. The numbers on top of the boxes indicate the top- $H$ . The indoor timings are given below the dashed line.

matches, the one with the minimum residual error is returned. A match is finally considered correct with respect to the ground-truth, if the distance between its estimated pose and the ground truth pose is within  $0.5m$  and 10 degrees.

## B. Results

Fig. 5 shows the precision/recall curve for all datasets and strategies. The curves have been obtained by changing the inlier threshold  $n_l^{min}$ . As can be seen, GFPs outperform the BoW technique. If we analyze the behavior on the smaller indoor dataset, we notice that GFPs are able to obtain the same precision/recall values of BoWs but require only one fifth of RANSAC operations. This is far more evident on the larger outdoor datasets where BoW is not able to obtain recall values above 50% when using the same numbers of RANSAC operations of GFPs. We found that on average BoW needs to perform RANSAC on half of the scans of the entire dataset to reach similar results. The reason of this behavior is that 2D range data offer limited variability in the descriptor space due to the limited appearance information present in the data. The most meaningful information is indeed the spatial arrangement of the points, which is discarded in BoW techniques. Geometrical phrases, on the contrary, bring back the lost information by considering the rigid ordering of the words in a scan.

Comparing the three TFIDF schemes for BoW, we can see that BoW-3 tends to perform slightly better than the others, although showing the same general behavior. We found out that using a kernel of dimension 3 for GFPs has a slight decrease in performances. This is due to the fact that it imposes harder rigidity constraints and is less robust to outliers and noise in the detected features.

The experiments on the Kenmore dataset show that the approach also outperforms state-of-the-art techniques [2, 3], with a recall rate of 80% at 99% accuracy and 68% at

100% precision, and a retrieval time of only 0.8s. Similar performances have also been obtained in [4], showing a recall rate of 84% at 99% accuracy and 30% at 100% precision, but with a retrieval time of about 20s, more than 20 times our performances. Finally, the approach also obtains the same performance as the naive matching approach of Tipaldi *et al.* [15] but with a relative speed-up of 10 (see Fig. 4).

Fig. 4 shows the time spent to match one scan with the entire dataset. We are able to run the approach at 5 to 10 Hz in typical indoor environments and at 2 to 1 Hz in outdoor datasets. The reason the algorithm is slower outdoor stems from the presence of more self-similar places due to the increased size. The only exception is the Victoria Park dataset, where the number of detected features is quite low, making the approach faster at the cost of lower precision/recall values. Note that the approach is always faster than BoW, extensive FLIRT search and Granström’s approach [4]. This difference is more evident in outdoor datasets, where GFPs scales sublinearly.

## VI. CONCLUSION

In this paper we presented an efficient and robust method for place recognition in large scale environments using 2D range data. Our geometrical FLIRT phrases (GFPs) are based on the previously developed FLIRT features. We realized a dedicated retrieval scheme that also performs a weak geometrical verification. Based on this, our method not only computes similar places but also calculates a relative pose transformations between the retrieved and the query places.

Experiments show that our method largely outperforms other state-of-the-art approaches in retrieval performance and metric precision. GFPs yield higher recall rates compared to the state of the art obtaining retrieval performances with 99% precision at more than 85% recall in less than a second, even data sets recorded over several kilometers. This corresponds to a tenfold improvement of the retrieval time. Thanks to the high performances of this method and the usage of 2D scan data, we open the door for place recognition lidar data which are less sensitive to variations of illumination and weather conditions. In future work, we will explore the use of hashing techniques and parallel processing to further reduce the retrieval time. The source code is available on the authors webpages.

## REFERENCES

- [1] R. A. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., 1999.
- [2] M. Bosse and R. Zlot. Map matching and data association for large-scale two-dimensional laser scan-based SLAM. *Int. J. on Rob. Res.*, 27, 2008.
- [3] M. Bosse and R. Zlot. Keypoint design and evaluation for place recognition in 2d lidar maps. *J. Rob. Aut. Sys.*, 75, 2009.
- [4] K. Granström, T. Schön, J. Nieto, and F. Ramos. Learning to close loops from range data. *Int. J. Rob. Res.*, 30, 2011.
- [5] G. Grisetti, R. Kümmerle, C. Stachniss, U. Frese, and C. Hertzberg. Hierarchical optimization on manifolds for online 2D and 3D mapping. In *IEEE Int. Conf. on Rob. & Aut.*, 2010.
- [6] H. Jegou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In *Eur. Conf. on Comp. Vis. (ECCV)*, 2008.
- [7] K. Jones. *Natural language information retrieval*, chapter What is the role of NLP in text retrieval? Kluwer, 1999.
- [8] J. Neira, J. Tardis, and J. Castellanos. Linear time vehicle relocation in SLAM. In *IEEE Int. Conf. on Rob. & Aut.*, 2003.

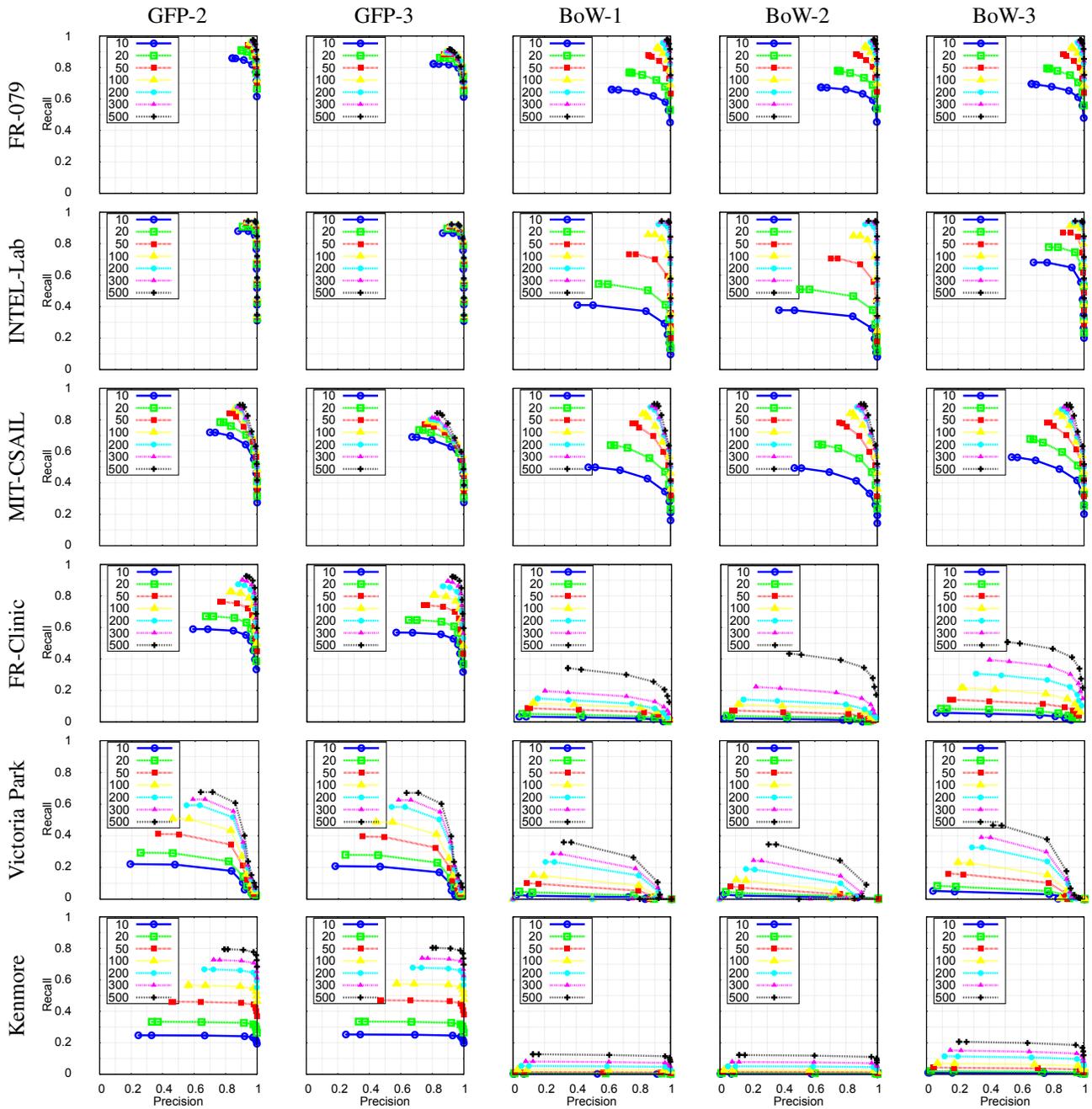


Fig. 5. Results using GFP-2, GFP-3, BoW-1, BoW-2, BoW-3 with different top- $H$ . GFPs always outperforms BoWs at same  $H$ , with GFP-2 being the best.

[9] L. M. Paz, P. Piniés, J. Neira, and J. Tardós. Global localization in SLAM in bilinear time. In *IEEE/RSJ Int. Conf. on Intel. Rob. and Sys.*, 2005.

[10] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *IEEE Conf. on Comp. Vis. and Patt. Rec.*, 2007.

[11] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Lost in quantization: improving particular object retrieval in large scale image databases. In *IEEE Conf. on Comp. Vis. and Patt. Rec.*, 2008.

[12] A. Rosenberg and J. Hirschberg. V-measure: A conditional entropy-based external cluster evaluation measure. In *Proc. of Joint Conf. on Emp. Methods in Nat. Lang. Process.*, 2007.

[13] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Inf. Processing and Management*, 24(5):513–523, 1988.

[14] J. Sivic and A. Zisserman. Video Google: A Text Retrieval Approach to Object Matching in Videos. In *Int. Conf. on Computer Vision*, 2003.

[15] G. D. Tipaldi, M. Braun, and K. O. Arras. FLIRT: Interest regions for 2D range data with applications to robot navigation. In *Proc. of the Int. Symp. on Exp. Rob. (ISER)*, 2010.

[16] B. Williams, M. Cummins, J. Neira, P. Newman, I. Reid, and J. Tardós. A comparison of loop closing techniques in monocular SLAM. *J. Rob. Aut. Sys.*, 57, 2009.

[17] J. Yuan, Y. Wu, and M. Yang. Discovery of collocation patterns: from visual words to visual phrases. In *IEEE Conf. on Comp. Vis. and Patt. Rec.*, 2007.

[18] Y. Zhang and T. Chen. Efficient kernels for identifying unbounded-order spatial features. In *IEEE Conf. on Comp. Vis. and Patt. Rec.*, 2009.

[19] Y. Zhang, Z. Jia, and T. Chen. Image retrieval with geometry-preserving visual phrases. In *IEEE Conf. on Comp. Vis. and Patt. Rec.*, 2011.