# Large Scale Place Recognition in 2D LIDAR Scans using Geometrical Landmark Relations

Marian Himstedt, Jan Frost, Sven Hellbach, Hans-Joachim Böhme, Erik Maehle

*Abstract*—**The recognition of places that have already been visited is a fundamental requirement for a mobile robot. This particularly concerns the detection of loop closures while mapping environments as well as the global localization w.r.t. to a prior map. This paper introduces a novel solution to place recognition with 2D LIDAR scans. Existing approaches utilize descriptors covering the local appearance of discriminative features within a bag-of-words (BOW) framework accompanied with approximate geometric verification. Though limiting the set of potential matches their performance crucially drops for increasing number of scans making them less appropriate for large scale environments. We present Geometrical Landmark Relations (GLARE), which transform 2D laser scans into pose invariant histogram representations. Potential matches are found in sub-linear time using an efficient Approximate Nearest Neighbour (ANN) search. Experimental results obtained from publicly available datasets demonstrate that GLARE significantly outperforms state-of-the-art approaches in place recognition for large scale outdoor environments, while achieving similar results for indoor settings. Our Approach achieves recognition rates of 93% recall at 99% precision for a dataset covering a total path of about 6.5 km.**

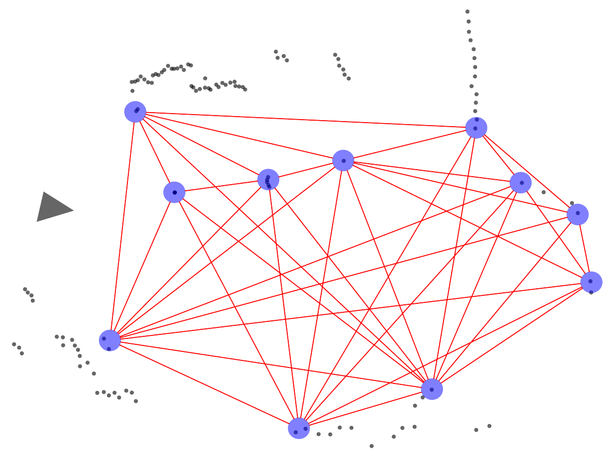

Figure 1. A graph of geometrical relations of co-occuring landmarks (blue dots) in a 2D range scan (black dots) is shown (triangle: robot pose). The relations are captured into pose-invariant scan signatures which are efficiently matched against the entire signature database.

## I. Introduction

Place recognition is an essential capability of a mobile robot in order to fulfill high-level tasks. In particular, it is necessary for detecting loop closures in SLAM as well as the global localization with respect to a prior map. In addition to these common applications, place recognition contributes to a number of further tasks, as for instance the merging of datasets collected during several surveys. Places are associated either in a pure topological or metric manner. The latter estimates a relative transformation of the query scan with respect to the matching one, whereas appearance based approaches solely match nodes of topological graphs [1]. While vision-based solutions to place recognition have been quite exhaustively investigated, only little attention has been paid to the capabilities of 2D range data. This is obviously due to the fact that, in contrast to images, 2D range scans only describe a relatively small subspace of the surrounding environment and hence provide less information to distinguish different places. Exhaustive matching of invidual laser scans is impractical in terms of time complexity when looking at large scale environments making descriptive place signatures and efficient retrieval algorithms indispensable. Similarly to vision-based methods, the extraction of local interest points serving as landmarks is beneficial. In existing approaches these landmarks are assigned descriptors capturing their local surroundings. As usual in typical bag-of-words (BOW) techniques these descriptors can be quantized to words enabling laser scans to be represented by histograms of word occurrences. While achieving promising results in visual applications [1], BOW performs rather poor in conjunction with 2D range data [2], particulary when looking at outdoor environments. Inspired by the vision community [3], Tipaldi et al. introduced Geometrical FLIRT phrases (GFP) allowing approximate geometric verification by preserving the clock-wise order of observed words [2]. Although GFP reduces the number of potential place correspondences, the number of required full geometric verfications on outdoor sequences increases with the size of the environment. This paper introduces Geometrical Landmark Relations (GLARE) explicitly taking into account relative metric displacements of landmarks observed in individual range scans. Using our approach we are able to achieve a recall of 93% at 99% precision on a 6.5 km long dataset which outperforms GFPs in both efficiency and accuracy.

Marian Himstedt, Jan Frost and Erik Maehle are with the Institute of Computer Engineering, University of Lübeck, Germany.
Sven Hellbach and Hans-Joachim Böhme are with the Artificial Intelligence Group of the University of Applied Sciences Dresden, Germany.
E-mail: himstedt@iti.uni-luebeck.de

## II. Related work

This section presents an overview of existing work dealing with place recognition. First, it is shown how GLARE is related to the vision-based approaches. Subsequently our approach is justified against prior work in range based place recognition.

### A. Spatial relations in vision-based place recognition

There exists a large amount of work dedicated to place recognition in the computer vision literature. A survey of place recognition in visual SLAM, i.e. loop closure detection, is given by [4]. Matching of geometric constraints within visual recognition tasks has recently drawn potential interest again. In [5] Johns and Yang propose to quantize the relative orientations and distances of co-occurring features in the image coordinate frame. Also Cho et al. incorporates spatial relationships between keypoints in a graph-based object recognition system. Clemente et al. proposes to use relative distances of observed features as an evaluation step when matching submaps [6]. However, the complexity of the matching procedure limits its use for rather small numbers of submaps.

### B. Place recognition based on 2D range scans

The related work on place recognition utilizing 2D lidar scans is rather small compared to those using camera images. Bosse and Zlot analysed several detectors and descriptors for local keypoints aggregated from sequential 2D range scans into submaps. The recognition of the submaps is carried out by associating stored and observed feature descriptors using approximate nearest neighbour (ANN) search with each correspondence voting for those submaps they were observed [7]. The submaps having the highest numbers of votes are selected and checked for geometric consistency based on projection histograms and ICP.

Granström et al. concatenate a set of various features such as curvature and average range serving as input for an Adaboost based classifier that detects corresponding scans [8]. As shown in [2] this approach is computationally very expensive compared to the state-of-the-art.

Either of the mentioned approaches [7], [2], [8] extract features from raw data serving as keypoints. Different from that, authors utilized occupancy grid maps allowing to incorporate submaps built incrementally and potentially using multiple sensors. In our recent work we investigated non-negative matrix factorization (NMF) to automatically determine relevant feature types of a given dataset in terms of basis primitives constructed from adjacent grid cells [9]. By representing local grid maps as a distribution of basis primitives, the presented method also allows to reduce the storage sizes of the prior occupancy maps which scale quadratically with increasing environment sizes.

### C. Geometrical FLIRT Phrases

The most related work to our approach is given by [2], [10] and hence specifically outlined here. Tipaldi and Arras presented FLIRT, a multi-scale feature detector for 2D range scans [10]. This work was accompanied by a shape context and a beta grid desciptor. The latter comprises raw distance measurements around local interest points into polar tesselations which were shown to be superior to the well-known shape context descriptors. Tipaldi et al. extended their approach by encoding the local descriptors as words which are assigned according to an a-priori learnt vocabulary [2]. While pure appearance based place signatures based on bag-of-words (BOW) achieve rather poor results, particularly for outdoor datasets, the authors introduced Geometrical FLIRT phrases (GFP). Inspired by the visual recognition system of Zhang et al. [3], GFPs enable to match sets of features in the cyclic order of observation ensuring the preservation of geometric constraints. GFP performs better than BOW but still requires a large number of full geometric consistency checks for potential matches in order to achieve competitive recognition results.

### D. Summary

None of the mentioned laser based approaches makes explicit use of relative orientations and distances of co-occuring landmarks which can substantially contribute towards distinguishing places. Except for [11], [7] all methods require a prior vocabulary or training stage for feature association.

## III. Geometrical Landmark Relations

This section introduces Geometrical Landmark Relations (GLARE). At first it is explained how GLAREs are generated from the input 2D range scans. Subsequently it is shown how these signatures can be integrated into efficient retrieval algorithms.

### A. Feature detection

At first, a 2D range scan is searched for local features serving as landmarks. Similiarly to Tipaldi et al. we extract points of high curvature which were shown to be superior in the domain of 2D range data [10]. The range data of a laser is considered as a one dimensional curve mapped into a multi-scale representation. For each scale the input curve is smoothed by varying Gaussian kernels. The smoothing kernels are normalized in order to be invariant to the sampling density. The features are local extrema exceeding a certain threshold in a difference signal constructed from the input curve and the smoothed curves. The mathematical derivation of the feature detection is given by [10]. As a result we obtain a set of landmarks for each range scan.

### B. Encoding spatial relations of landmarks

For a set of $N$ landmarks $l_1, \ldots, l_N$ detected in the $k$-th range scan we estimate the Euclidean distances $\rho_{i,j}$ of each landmark $l_i = \{x_i, y_i\}$ to all others $l_j = \{x_j, y_j\}$ of the set with $i \neq j$ within the coordinate frame of the range scan. In addition to that the bearings $\theta_{i,j}$ and $\theta_{j,i}$ of co-occuring landmarks are estimated:

$$\theta_{i,j} = atan2(y_i - y_j, x_i - x_j) \quad (1)$$

Only the bearings $\theta^+_{i,j} = \max(\theta_{i,j}, \theta_{j,i})$ are considered for further processing. The reduction to one bearing $\theta^+_{i,j}$ for

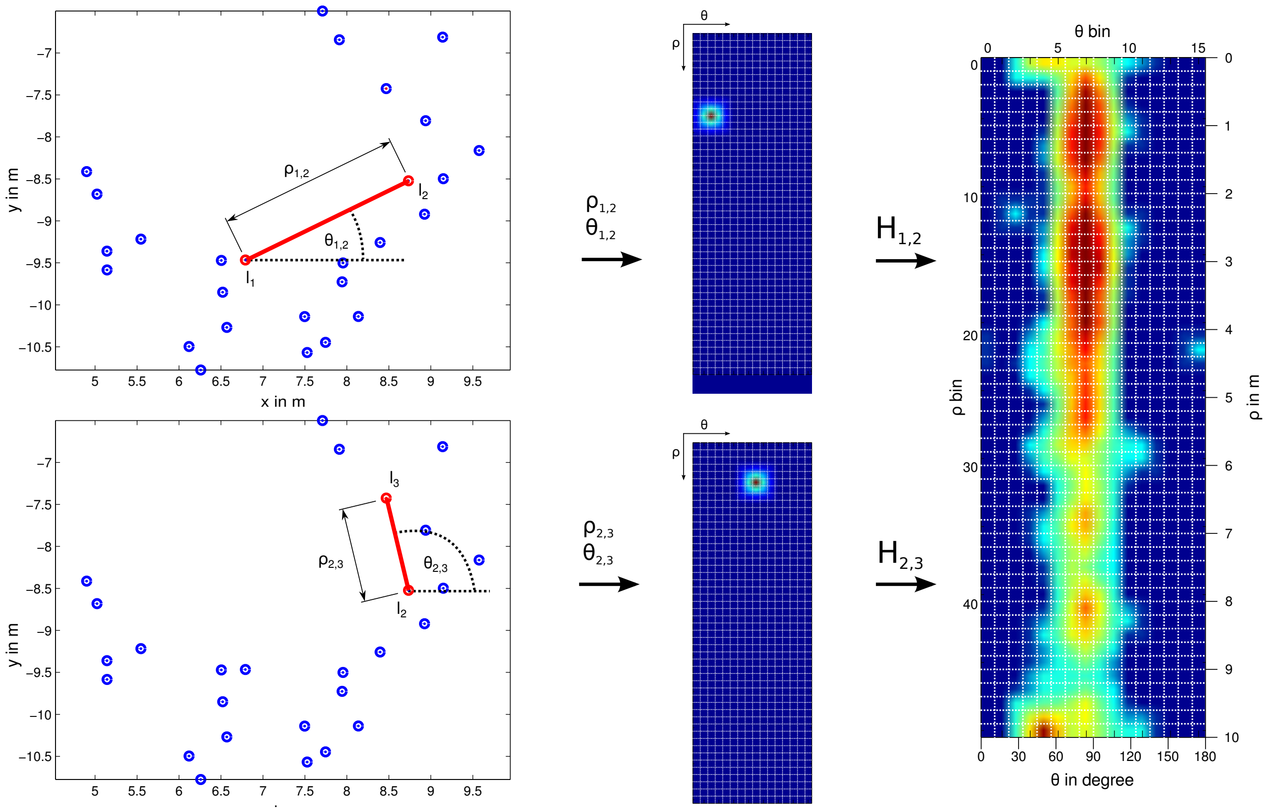


Figure 2. Generation of GLARE: Each landmark relation, given by its orientation $\theta_{i,j}$ and distance $\rho_{i,j}$, is incorporated as a multivariate Gaussian. The scan signature $S^{(k)}$ is obtained by estimating the normalized sum over all histograms $H_{i,j}$.

each relation avoids storing redundant information. Having estimated $\rho_{i,j}$ and $\theta^{+}_{i,j}$, the distribution of spatial relations is captured. For this purpose the bearings $\theta^{+}$ and distances $\rho_{i,j}$ are quantized and assigned to uniformly sized bins:

$$\left(\theta^{+}_{i,j}, \rho_{i,j}\right) \in bin\left(n_{\theta}, n_{\rho}\right) \tag{2}$$

The 2D histogram $H_{i,j}$ is constructed by a discrete multivariate Gaussian centred on the bin $n = (n_{\rho}, n_{\theta})$ given the covariance matrix $\Sigma_H$. Centering the Gaussian to the discretized bin $n$ instead of the exact position $\left(\theta^{+}_{i,j}, \rho_{i,j}\right)$ allows a precomputation of the contributions to the histogram bins covered by covariance $\Sigma_H$. The histogram position $m = (m_{\rho}, m_{\theta})$ is estimated according to:

$$H_{i,j}\left(m\right) = \mathcal{N}\left(m - n, \, \Sigma_H\right) \tag{3}$$

The signature $S^{(k_i)}$ for the landmark $l_i$ is estimated as follows:

$$S^{(k_i)} = \sum_{j} H_{i,j} \tag{4}$$

Finally, we obtain the scan signature $S^{(k)}$ for the range scan $k$ by estimating a normalized sum over all landmark signatures:

$$S^{(k)} = \eta \sum_{i} S^{(k_i)} \tag{5}$$

The factor $\eta$ is for the normalization of the signature. Figure 2 exemplarily illustrates the generation of GLARE signatures.

Our technique differs from a multitude of other place recognition approaches since the latter typically quantize the local neighbourhood of landmarks to descriptors which are quantized to words of a prior vocabulary [10], [5]. Each descriptor hence votes for exactly one word which is also referred to as hard-voting. In contrast, we use a soft-voting scheme enabling to evaluate adjacent cells depending on $\Sigma_H$. The covariance matrix $\Sigma_H$ allows to incorporate uncertainties in relative depth and orientation estimates. These are mainly driven by the sensor accuracy but can also be biased by the localization quality of the extracted features. We observed that this scheme results in smoother results which is particularly well suited for noisy range data.

### C. *Efficient scan retrieval*

Having generated the scan signatures, the goal is to collect those in a global repository $S$ while simultaneously managing an index structure allowing fast scan retrieval. Since $S$ consists of high-dimensional vectors, Approximate Nearest Neigbor Search (ANN) becomes indispensable in order to enable efficient retrieval [12]. Conventional kd-trees always pick that dimension of the input data having the largest variance to bisect the data at each level of the tree. This structure has shown to be less suitable in the presence of high-dimensional data [13]. Thus it is recommended to make use of multiple randomized kd-trees whose splitting bounds are randomly drawn from the top variant dimensions resulting in better representations for higher dimensions. The key difference of ANN compared to the exact nearest neighbour search is that the number of points to evaluate is bound to $\tau$. This method achieves a sufficient approximation providing the bound is set appropriately. Setting this bound is critical as it poses a trade-off of accuarcy and time complexity. In addition to that, it is affected by the balance quality of the kd tree. When using GLARE for recognizing places with respect to a prior map (global localization) we can assume to have a well-balanced tree. For loop closure detection in SLAM, the global repository $S$ is constantly updated while traversing the environment which requires the underlying kd-trees to be rebalanced after a certain amount of change. We found that incorporating the ratio of the kd-tree leafs and the number of recently appended elements (before rebalancing) indicate a promising choice of $\tau$.

Given the scan signature $S^{(k)}$, the global repository $S$ is searched for potential candidate matches using the L1-norm. For performance reasons we limit the search to the $K$ approximate nearest neighbours. Similiarly to $\tau$, $K$ needs to be set appropriately in order to obtain optimal results in terms of precision and recall. Setting $K$ to a lower value might result in missing place correspondences, whereas large values of $K$ require a potentially large number of geometric consistency checks which are computationally expensive due to the trigonometric operations involved.

We observed that the recognition of GLARE based signatures using ANN achieves optimal results in terms of time complexity and accuracy. The inverted file techniques that are commonly used for vision-based place recognition [5], [1] and also in GFP [2] are not suited for GLARE due to their dense histogram representations. The distributions of visual words for a single location are typically rather sparse representing only a small subset of the prior vocabulary. Combined with a hard-voting scheme this allows to invert the recognition by having visual words pointing towards their corresponding images or scans. Only those places being drawn are considered for further matching. Applied to GLARE, this procedure does not contribute to a faster retrieval since on the one hand the number of votes in the histogram is significantly larger ($N$ vs. $N^2/2$). Second, our soft-voting mechanism does not solely consider the closest bin but also adjacent ones.

### D. Geometric verification

Even though two places might be close to each other in the signature space, they do not necessarily describe the same place. That is why the set of $K$ putative matches for the place $S^{(k)}$, provided by the retrieval system, is checked for geometrical consistency. For this purpose, we again make use of the geometric relations of co-occurring landmarks. For each landmark $l_i$ in the $k$-th scan, the landmark signature $S^{(k_i)}$ is estimated according to Eq. 4. The landmark correspondences of place $S^{(k)}$ to the $K$ putative matches are obtained by estimating the pair-wise distances of the landmark signatures $S^{(k_i)}$. Those being above a fixed threshold are considered for a RANSAC based scan-matching which estimates a rigid transformation of the two scans. The residual error as well as the number of inlier correspondences indicate whether the two given scans satisfy our geometric consistency check and hence are considered as matching places.

## IV. Experiments

In order to evaluate the presented approach, a number of experiments were carried out. Specifically, we selected different publicly available datasets, one of an indoor environment and three captured outdoors (see Table I).

Table I
DATASETS FOR EXPERIMENTS

| Name | Type | # Scans | # Landmarks | Path length [m] |
|---|---|---|---|---|
| Intel-lab | indoor | 2672 | 39392 | 360.7 |
| FR-Clinic | outdoor | 6917 | 190760 | 1437.6 |
| Victoria Park | outdoor | 5751 | 81795 | 4206.14 |
| Kenmore | outdoor | 13063 | 499237 | 6588.34 |

### A. Setup

For our experiments, we compared two different variants of GLARE to the state-of-the-art approach GFP [2]. GLARE-1 only uses relative distances of landmarks omitting the relative orientations. GLARE-2 incorporates distances as well as orientations. For GLARE-1 and GLARE-2 we used 100 linearly sized bins with a size of 0.5 m for the outdoor datasets and 50 bins with a size of 0.2 m for the indoor datasets. GLARE-2 additionally utilizes 8 angular bins. GFPs are used with the optimal settings, as shown in [2]. We tested GLARE-1, GLARE-2 and GFP on all datasets with three different nearest neighbour numbers, more precisely with $K = 50$, $K = 100$ and $K = 500$. The geometric verification rejects putative matches of places by thresholding based on the residual error (linear: $0.5m$, angular: $0.2rad$). The thresholds for estimating the precision/recall curves are obtained by using $a$ different numbers of inlier correspondences $c$ (here: $a = 1, 2, ..., 32$).

### B. Run time

The run times for GLARE-1, GLARE-2 and GFP are shown in Table II. GFP only gets close to GLARE's recall rates for a large number of nearest neigbors which is why $K = 500$ is used for this experiment (see Figure 3). Actually GFP requires even more putative neighbours to be considered, however, the run time for this becomes highly impracticable.

The run times for feature detection are very similar since all approaches share the same algorithms (see Section III-A). The descriptor construction (signatures) of GLARE-2 takes slightly longer than GLARE-1 but all approaches are very close to each other. It is apparent that the geometric verification for GLARE is significantly faster than GFP. This is probably due to the fact that the relative distances of landmarks allow to reject more false-positive feature correspondences than the appearance based descriptors of GFP. We observed that the majority of features on the Kenmore dataset refer to point-like features resulting in very similar GFP descriptors. Thus the number of putative correspondences passed to RANSAC is smaller for GLARE compared to GFP. Note that GLARE already achieved very accurate results for $K = 50$ (see Figure 3), which would substantially reduce the run time needed for geometric verification.

Table II
RUN TIME ON KENMORE FOR $K = 500$

| | GLARE-1 [ms] | GLARE-2 [ms] | GFP [ms] |
|---|---|---|---|
| feature detection | 7.5402 | 7.298 | 7.276 |
| feature description | 0.768 | 1.291 | 1.797 |
| scan retrieval | 18.578 | 25.312 | 81.462 |
| geometric verification | 89.596 | 88.928 | 330.586 |
| **total** | **116.482** | **122.829** | **421.121** |

### C. Results

The results for our experiments are shown in Figure 3. It is obvious that both GLARE-1 and GLARE-2 outperform GFP on all outdoor datasets. GFP gets closer to GLARE with a high number of $K$ nearest neighbours taken into consideration. The differences in precision/recall for GLARE are rather small for increasing number of $K$. The number $K$ essentially determines the run time of the place recognition since the geometric verfications that need to be carried out for $K$ scans are computationally highly expensive. Exactly due to that fact GLARE is much more efficient than GFP that requires an order of magnitude more verification steps.

Figure 3 also shows that GLARE is slightly worse than GFP on the indoor dataset (intel-lab). We observed that this is due to lower variance in relative distances of co-occuring landmarks and the high self-similarity in man-made environments in terms of spatial relations. In contrast to GFP, GLARE does not require prior training stages to learn landmark signatures from at least very similar environments.

GLARE-1 outperforms GLARE-2 in terms of run-time since the signatures of GLARE-2 are eight times as large as those of GLARE-1 which becomes apparant in the pair-wise landmark association in the geometric verification. Note that it is possible to use GLARE-2 for the scan signature generation and use GLARE-1 for the geometric verification.

## V. CONCLUSIONS

This paper introduced Geometrical Landmark Relations (GLARE) designed for place recognition in 2D range data. GLAREs encode relative landmark relations captured from single range scans into scan signatures. The latter reduces the scan retrieval to a very efficient approximate nearest neighbour search avoiding to evaluate each single signature of the repository for finding correspondences. It was shown that GLARE significantly outperforms the state-of-the-art approach GFP in outdoor environments, while achieving similar results indoors. In contrast to GFP, GLARE does not require any prior training stage emphasising its strengths for the application in a-priori unknown environments. Our experiments demonstrated that GLARE requires fewer candidate matches compared to GFP which substantially reduces the run time since the RANSAC based geometric verification is computationally expensive. Similarly to the state-of-the-art GLARE not only allows to recognize places but also to estimate a relative transformation and alignment error of the underlying range scans. This enables precise global localization and a convenient integration into existing graph based SLAM frameworks. In addition to the application of place recognition, GLAREs can be utilized for further tasks in the context of mobile robot navigation. For instance, GLARE signatures provide information about the spatial settings of co-occuring landmarks which is beneficial for semantic labelling of places, i.e. distinguishing narrow corridors from wide entrance halls. The contextual information about the local surroundings can be used, for example, to perform motion planning on a mobile robot.

## REFERENCES


[1] M. Cummins and P. Newman, "FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance," *The International Journal of Robotics Research*, vol. 27, no. 6, pp. 647–665, 2008.

[2] G. D. Tipaldi, L. Spinello, and W. Burgard, "Geometrical flirt phrases for large scale place recognition in 2d range data." in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Karlsruhe, Germany, 2013.

[3] Y. Zhang, Z. Jia, and T. Chen, "Image retrieval with geometry-preserving visual phrases," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011, pp. 809–816.

[4] B. Williams, M. Cummins, J. Neira, P. Newman, I. Reid, and J. Tardós, "A comparison of loop closing techniques in monocular slam," *Robotics and Autonomous Systems*, 2009.

[5] E. Johns and G.-Z. Yang, "Dynamic scene models for incremental, long-term, appearance-based localisation," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, May 2013, pp. 2731–2736.

[6] L. A. Clemente, A. J. Davison, I. D. Reid, J. Neira, and J. D. Tardós, "Mapping large loops with a single hand-held camera," in *Robotics: Science and Systems*, 2007.

[7] M. Bosse and R. Zlot, "Keypoint design and evaluation for place recognition in 2d lidar maps," *Robotics and Autonomous Systems*, vol. 57, no. 12, pp. 1211–1224, 2009.

[8] K. Granström, T. B. Schön, J. I. Nieto, and F. T. Ramos, "Learning to close loops from range data." *I. J. Robotic Res.*, vol. 30, no. 14, pp. 1728–1754, 2011.

[9] S. Hellbach, M. Himstedt, and H.-J. Boehme, "What's around me: Towards non-negative matrix factorization based localization," in *Mobile Robots (ECMR), 2013 European Conference on*, Sept 2013, pp. 228–233.

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

[11] J.-L. Blanco, J. González-Jiménez, and J.-A. Fernández-Madrigal, "A robust, multi-hypothesis approach to matching occupancy grid maps," *Robotica*, 2013.

[12] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, "An optimal algorithm for approximate nearest neighbor searching fixed dimensions," *J. ACM*, vol. 45, no. 6, pp. 891–923, Nov. 1998.

[13] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *International Conference on Computer Vision Theory and Application VISSAPP'09)*. INSTICC Press, 2009, pp. 331–340.

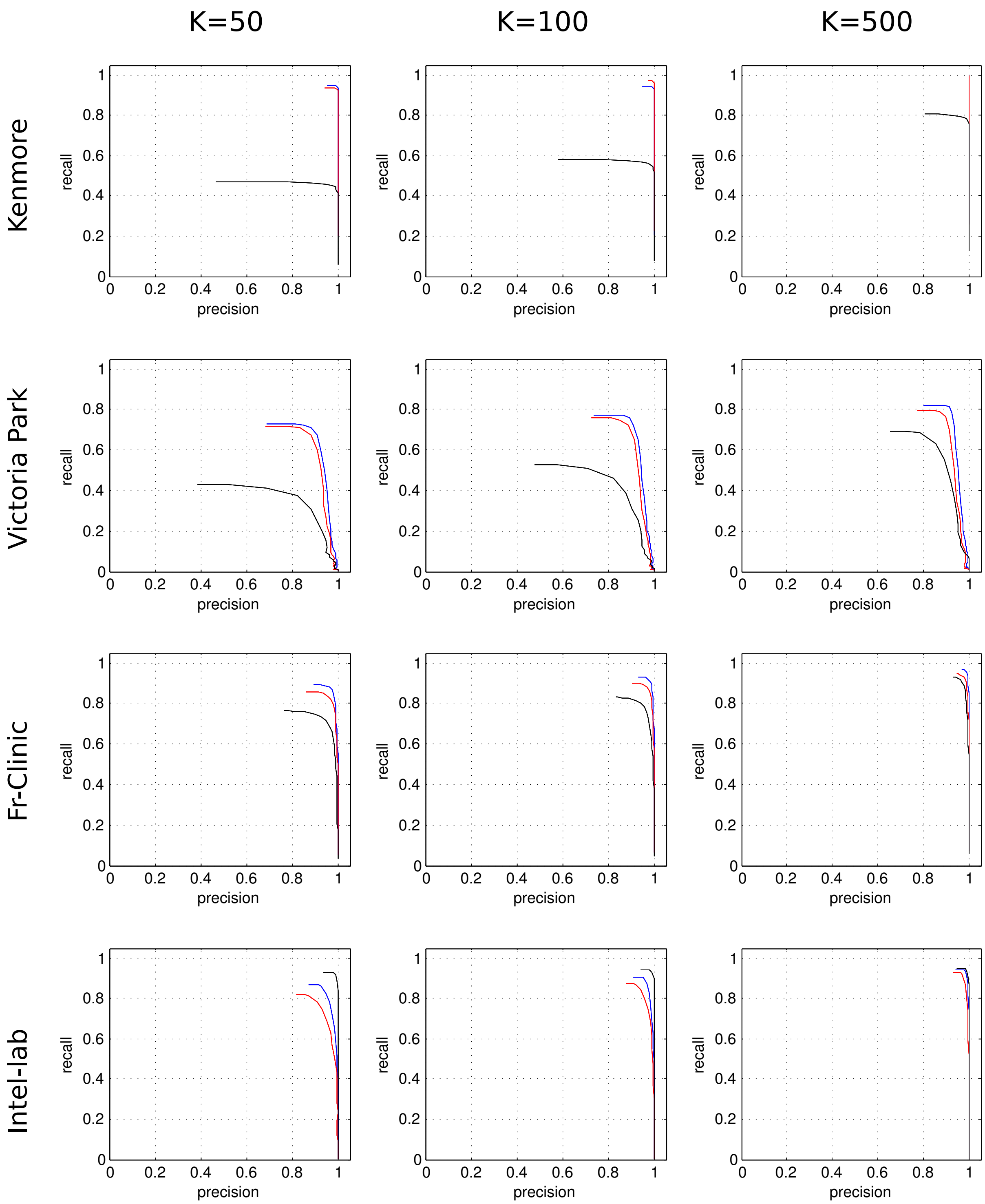


Figure 3. Experimental results obtained using GLARE-1 (red), GLARE-2 (blue) and GFP (black) are shown for different number of nearest neighbours $K$ taken into consideration. The results for GLARE-1/2 on the Kenmore dataset ($K = 500$) are so close that only one is visible in the plot.