2010 IEEE International Conference on Robotics and Automation
Anchorage Convention District
May 3-8, 2010, Anchorage, Alaska, USA

FLIRT – Interest Regions for 2D Range Data

Gian Diego Tipaldi and Kai O. Arras

Abstract— **Local image features are used for a wide range of applications in computer vision and range imaging. While there is a great variety of detector-descriptor combinations for image data and 3D point clouds, there is no general method readily available for 2D range data. For this reason, the paper first proposes a set of benchmark experiments on detector repeatability and descriptor matching performance using known indoor and outdoor data sets for robot navigation. Secondly, the paper introduces FLIRT that stands for Fast Laser Interest Region Transform, a multi-scale interest region operator for 2D range data. FLIRT combines the best detector with the best descriptor, experimentally found in a comprehensive analysis of alternative detector and descriptor approaches. The analysis yields repeatability and matching performance results similar to the values found for features in the computer vision literature, encouraging a wide range of applications of FLIRT on 2D range data. We finally show how FLIRT can be used in conjunction with RANSAC to address the loop closing/global localization problem in SLAM in indoor as well as outdoor environments. The results demonstrate that FLIRT features have a great potential for robot navigation in terms of precision-recall performance, efficiency and generality.**

I. Introduction

The introduction of local image features had a large impact on many computer vision tasks such as object and scene recognition, motion tracking, stereo correspondence, or visual robot localization and SLAM. The typical strategy is to select a set of regions at locations in image space and compute a distinctive descriptor over those regions. This yields a description of the image content as a collection of local interest regions that can be used as candidates for matching. For both, the detection of stable locations and the description to encode the image structure, there is a great variety of approaches available for image and 3D range data [1], [2], [3], [4], [5].

The same reasons that make interest points attractive for the above mentioned domains also apply to 2D range data as produced by the widely employed laser scanners in robotics. For robot navigation, interest points have the potential to be an alternative to feature-based and grid-based approaches. While both paradigms have been proved to be successful under application-like conditions [6], [7], they both have strengths and weaknesses. Features allow for compact map representations and high accuracy but rely on predefined geometric models. Dense approaches using raw data or grids are general in that sense but less efficient and operate with map representations that scale less well with environment size and dimensionality.

The authors are with the Social Robotics Lab, Department of Computer Science, Albert-Ludwigs-University of Freiburg, 79110 Freiburg, Germany {tipaldi,arras}@informatik.uni-freiburg.de

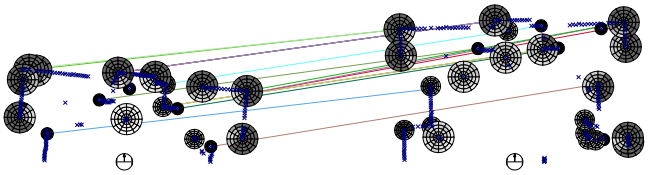

Fig. 1. Example matching of two scans from a laser range finder of the same scene using RANSAC. The figure shows the extracted FLIRT features for both scans and the 16 inlier correspondences.

Interest points, on the other hand, combine the compactness of discrete features and the generality of raw range data. They further allow for early sensor fusion with vision and an unified treatment of laser and image data.

A comparison of different detectors and descriptors for images can be found in Mikolajczyk *et al.* [8], [9]. For 3D data, several methods have been proposed to extend the scale space from images to 3D point clouds, replacing the regular image lattice with surfaces represented by a connectivity mesh. A seminal work by Taubin [10] replaced the continuous Laplacian operator ∇ from the diffusion equation by its discrete counterpart, the graph Laplacian ∇_g. A different approach has been proposed by Pauly *et al.* [11], where a surface variation quantity is computed. This quantity is formed by the eigenvalues of the sample covariance matrix computed in a local neighborhood of sampled points. The scale at a point is then the neighborhood size for which the surface variation achieves a local extrema. Novatnack and Nishino [12] detect multi-scale features using a representation of the surface geometry. This representation is encoded by the surface normals embedded in a regular and dense 2D domain. The approach, however, relies on a connectivity mesh to construct the parametrization, and on the availability of good surface normals. Unnikrishnan *et al.* [13] define an integral operator that maps the input curve into its multi-scale parametrization. The operator is defined in geodesic coordinates along the curve with interest points found as local extrema in a geodesic neighborhood. Using data from a laser range finder, Cole *et al.* [14] propose an information-theoretic measure of local saliency to find natural features in measurement space, capturing the geometry of intrinsically interesting surface patches. However, the saliency computation is expensive for an exhaustive search and the authors compute the saliency values only for randomly picked points.

For 2D range data, there is little related work. Closest to this paper is the line of work by Bosse and Zlot [15], [16]. In [15] the authors define entire laser scans as features and use orientation histograms to describe them. In [16], several detector/descriptor-pairs for 2D range data are evaluated for

the task of place recognition in a graphical, submap-based SLAM application. While interesting, the main difference to our approach, is that with a descriptor support region of $9 \times 9\,m$ and being defined on submaps (collection of 20-30 scans spaced 1–2 m apart), their approach is a submap characterization technique rather than a local interest point operator. While this property was not a limitation for obtaining the good results presented in [16], we are interested in designing a general-purpose multi-scale keypoint for 2D range data that retains the important concept of locality which was key to the impact of interest points in computer vision. FLIRT features have been designed in this spirit: they are defined locally (in support regions of typically $0.5\,m$ radius) and on a single scan.

The reasons why 2D range data are different from image data and 3D point clouds are manifold. As a naive approach, one could apply the techniques from computer vision to range data, replacing the image intensity values with the range signal. While this approach leads to some results, it is not able to deal with many interesting structures since range variations around such structures can be weak (corners are an example). This is because the nature of range data is different from the nature of image data in that range data represent a manifold in a higher-dimensional space. In the case of 3D range data, this manifold is a surface in 3D, for 2D data it is a curve in Cartesian space. Further, for range data, measurement sparsity is highly non-uniform and view-point variant, partly due to the low angular resolution of range finders compared to cameras. Finally, range data can be seen to lie in between image data and 3D point clouds as they are defined over an ordered lattice (the raw data space) but also define a point cloud in Cartesian space. These differences motivate a specifically derived interest point transform for 2D range data.

Accordingly, we propose a novel set of benchmark experiments based on commonly used large-scale robot navigation data sets to define an experimental testbed for the comparison of detectors and descriptors for 2D range data. We then evaluate a number of different detectors and descriptor approaches to eventually propose FLIRT (Fast Laser Interest Region Transform) as the most appropriate detector-descriptor combination. Finally, we show how FLIRT can be used in conjunction with RANSAC to address the loop closing and global localization problem in SLAM.

The paper is structured as follows. Sections II and III describe the different detectors and descriptors, respectively, used in the comparison. The experiments are described in Section IV. Finally, Section V concludes the paper.

II. Detectors

In this section we present four multi-scale detector approaches that are compared in this paper. Section II-A presents a detector based on the raw range signal, while Section II-B describes two detectors based on a normal approximation of the range data. Section II-C introduces a curvature based detector derived from Unnikrishnan *et al.* [13]. These detectors look for interesting points on different scales according to the discrete version of the scale space theory [17].

The scale-space theory is a framework for multi-scale signal representations to handle structures in the signal that occur at different resolutions. The original signal, $s(x)$ is represented by a family of smoothed signals $S(x;t)$, parametrized by t, the size of the smoothing kernel used for suppressing fine scale structures. Formally, we have

$$S(x;t) = (K_t * s)(x) \tag{1}$$

where K_t is the smoothing kernel. In general, this kernel must not introduce new feature that do not correspond to simplification of previous features at finer scales. A typical choice in computer vision is the Gaussian kernel which has been proved to satisfy this property in the continuous case. However, Lindeberg [17] showed that in the discrete case, the filter has the form

$$K_t(x) = e^{-t} I_x(t) \tag{2}$$

where I_x are the modified Bessel functions of integer order. Loosely speaking, the kernel in Eq. (2) is a discrete equivalent of the Gaussian with t corresponding to the standard deviation.

Interest points can then be detected by considering local maxima of differential invariants computed at different scales. Typical invariants are the gradient magnitude for edge detectors and the Laplacian or the determinant of the Hessian matrix for blob detectors. With 1D data, there are no invariants for ridge and corner detectors as they are only defined for higher-dimensional signals such as image data.

Note that the scale space theory also defines a way to make image representations invariant to scale by performing automatic scale selection. The selection is based on local maxima over scales of normalized derivatives. In range data, however, there is no need of automatic scale selection or scale invariance, since the scale of features can directly be obtained from the distance information and represents discriminative information for matching.

A. Range-Based Detector

The first detector finds interest points in scale-space with a blob detector applied on the raw range information in the laser scan. This choice is inspired by the SIFT approach that uses a blob detector on the raw image data [1].

We construct the scale space using the discrete smoothing operator defined in Eq. (2) with different values for the scale t. For each t, interest points are detected by extracting the local maxima and local minima of the Laplacian of the signal. For one-dimensional signals the Laplacian operator is equivalent to the determinant of the Hessian and both are equal to the second derivative of the signal itself. This derivative is computed by convolving the signal with the discrete second derivative operator

$$\nabla^2 S(x;t) = (D_2 * S)(x;t) \tag{3}$$

$$(D_2 * S)(x;t) = S(x-1;t) - 2S(x;t) + S(x+1;t) \tag{4}$$

Interest points are then found by detecting peaks in Eq. (3).

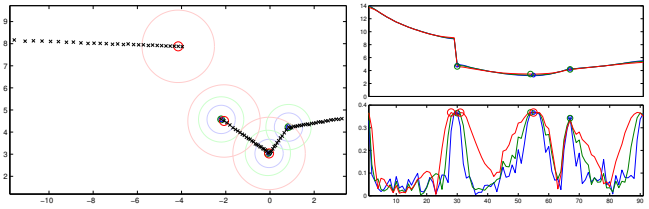

Fig. 2. Curvature-based detector on synthetic data. The detector finds the background part of the range discontinuity on the largest scale, the foreground part on all scales, responds on all scales for the convex corner, and on scale 1 and 2 for the the obtuse concave corner. The two diagrams at the right show the interest points in signal space, with the raw range signal and its smoothed variants at each scale (top) and the exponentially damped signal $F(x;t)$ on all scales (bottom). The maxima at the start and the end are ignored as they are caused by the non-circularity of the data.

B. Normal-Based Detectors

Instead of the raw range signal, we can also consider a local approximation of the normal direction at each point. The resulting *normal signal* is treated in a similar way than the range signal in the previous section. We propose two detector methods based on this signal. The first one finds interest points using an edge detector and the second one uses a blob detector.

The normal direction is estimated by least squares fitting of a line to a group of measurements in a sliding window centered around the current point. For fitting, we minimize the perpendicular error from the points onto the line so as to have geometrically meaningful estimation results. With measurements in polar coordinates, the fit expressions were shown to have a closed-form solution [18]. With the normal direction – obtained by the perpendicular to the line direction – we again construct the scale space by applying the discrete smoothing operator of Eq. (2) at different scales t.

The first detector responds to edges in the normal signal and detects points as local maxima and local minima of the gradient magnitude on every scale. For one-dimensional signals the magnitude of the gradient is equal to the first derivative of the signal itself. This derivative is computed by convolving the signal with the discrete first derivative operator,

$$\|\nabla S(x;t)\| = (D_1 * S)(x;t), \quad (5)$$

$$(D_1 * S)(x;t) = \frac{1}{2}(-S(x-1;t) + S(x+1;t)). \quad (6)$$

The second detector responds to blobs in the normal signal and detects points as local maxima and local minima of the Laplacian of the signal on every scale, according to Eq. (3).

C. Curvature-Based Detector

We describe here the detector introduced by Unnikrishnan and Hebert [13]. The rationale behind this detector is that range data define a curve in Cartesian space and the scale space theory should be applied to this curve and not to the original signal. The authors define an integral operator that maps the input curve into its multi-scale parametrization

$$S(\alpha(s);t) = \int_\Gamma k(s,u;t)\alpha(u)\mathrm{d}u \quad (7)$$

$$k(s,u;t) = \mathcal{N}(s-u;t) \quad (8)$$

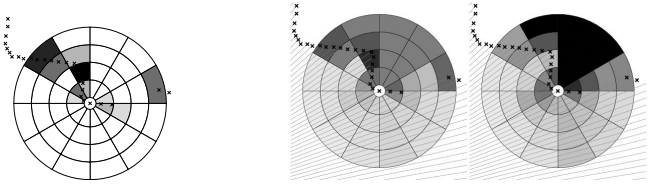

Fig. 3. Linear local shape context descriptor (left) and β-grid descriptor (center: occupancy probability, right: variance) for an example interest point in real data.

where Γ is the curve, $\alpha(s)$ the parametrization of the curve by the geodesic coordinate s and $k(s,u;t)$ is a Gaussian kernel. The operator is then made invariant to the sampling density of the curve by normalizing the smoothing kernel with the (unknown) sampling density $p(s;t)$,

$$\tilde{k}(s,u;t) = \frac{k(s,u;t)}{p(s;t)p(u;t)} \quad (9)$$

$$p(s;t) = \int k(s,u;t)p(u)\mathrm{d}u. \quad (10)$$

The sampling density $p(s;t)$ at scale t is approximated by local kernel density estimation using Gaussian kernels. This yields a curve for each scale,

$$\tilde{S}(\alpha(s);t) = \int_\Gamma \tilde{k}(s,u;t)\alpha(u)\mathrm{d}u, \quad (11)$$

of increased smoothness for increasing t's. Interest points are then detected by finding the local maxima of the exponential damping expression

$$F(x;t) = \frac{2\|x - \tilde{S}(\alpha(s);t)\|}{t} e^{-\frac{2\|x-\tilde{S}(\alpha(s);t)\|}{t}} \quad (12)$$

with the term $\|x - \tilde{S}(\alpha(s);t)\|$ being an error distance in Cartesian space between the original curve and its smoothed versions $\tilde{S}(\alpha(s);t)$. With this method, interest points at a scale t correspond to places where t equals the inverse of the local curvature of the smoothed signal $\tilde{S}(\alpha(s);t)$.

An example detection results is shown in Fig. 2

III. Descriptors

The task of the descriptor is to encode the local structure in the scan around the detected interest point with high distinctive power. We propose two approaches for this purpose, a modified shape context descriptor and a β-grid descriptor based on insights from occupancy grid mapping.

A. Linear Local Shape Context

Shape context, introduced by Belongie *et al.* [4], is a descriptor for finding correspondences between point sets. The descriptor captures the distribution of points relative to each point on the shape, represented in a log-polar histogram.

In our case, we are interested in the *local* structure of the scan around the detected interest point and thus compute the shape context only locally, within an area proportional to the scale of the interest point. Further, we choose a linear polar histogram since the type and extent of noise in range data differ from the noise in image data. Measurement errors in range data typically occur in radial direction (which is viewpoint invariant) and can be relatively large. This

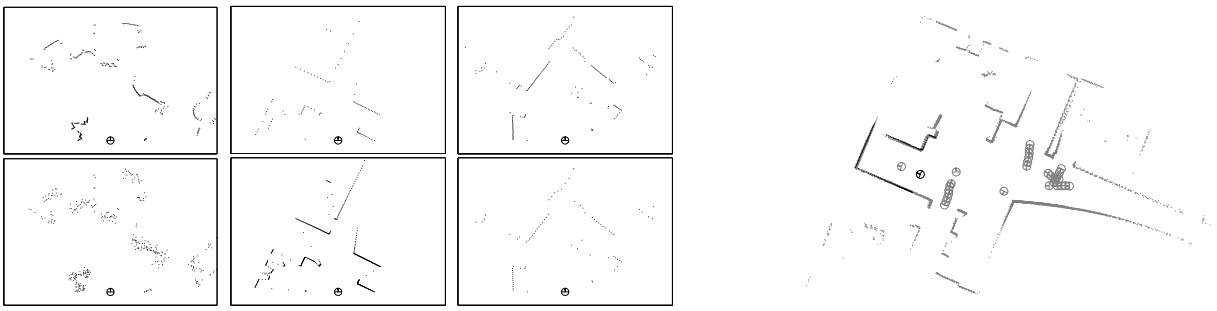

Fig. 4. Example locations and their transformations. Left, top row: reference locations, bottom row: same scans subject to extra Gaussian noise (bottom left, with $\sigma = 0.25$), oversampling (bottom center, 4 times), and subsampling (bottom right, 4 times). Right: View point changes given a reference location indicated by the black robot and the black laser points. The gray robots are map locations with a 50% overlap with the scan of the reference location.

makes that, in practice, the small bins near the log-polar histogram center tend to capture noise rather than the local scan structure. By choosing a linear tessellation in polar space, this effect is attenuated. More formally, the shape context descriptor of a detected interest point p_{det} is the histogram

$$h_{det}(j) = \#\{p_i \neq p_{det} : (p_i - p_{det}) \in bin_j\} \qquad (13)$$

where bin_j is defined by discretizing the distance of the point and the viewing angle (Fig. 3 left).

B. β-Grids

An important difference between image and range data is that the latter not only encodes metric distance information but also directed free-space information between the sensor (emitting light or sound) and the measured object. This is relevant extra information, not encoded in the shape context descriptor. Occupancy grids naturally deal with free-space information which is why we adopt this concept for the purpose of the second descriptor considered here. Concretely, for each detected interest point p_{det} we define a polar tessellation of the space around p_{det}. Again, this tessellation is linear in polar space, with a radius proportional to the scale of the interest point. For estimating the occupancy probability, we apply Bayesian parameter learning. This approach provides a sound way to initialize cell probabilities and delivers a variance estimation over the occupancy value.

We now derive the expressions for Bayesian parameter estimation for occupancy grids. Consider the j-th bin, whose likelihood to be hit by the beam z follows a Bernoulli distribution, parametrized by the bin occupancy probability occ_j, where z is equal to 1 when the laser beam is reflected inside the bin (hit) and is equal to 0 when the laser beam traverses the bin (miss). The occupancy probability is modeled using the conjugate prior of the Bernoulli which is the Beta distribution, a continuous distribution defined on the interval $[0,1]$ and parametrized by the two positive shape parameters α and β,

$$p_\beta(occ_j; \alpha, \beta) = \frac{{occ_j}^{\alpha-1}(1-occ_j)^{\beta-1}}{\mathrm{B}(\alpha,\beta)} \qquad (14)$$

with $\mathrm{B}(\alpha,\beta)$ being the Euler beta function. Learning the occupancy probability occ_j consists in estimating the parameters of a Beta distribution (hence the name of the descriptor). Over a sequence of measurements, that is, a sequence of beams $\{z_i\}_{i=1}^n$ that either hit or miss the bin, it can be shown that the update rules for the parameters are

$$\alpha_i = \alpha_{i-1} + \sum 1^{z_i} \qquad \beta_i = \beta_{i-1} + \sum 1^{(1-z_i)}. \qquad (15)$$

For $i = 0$, both parameters are set to 1 for which the Beta distribution is uniform over $[0,1]$. The point estimate $\widehat{occ}_j$ is then the expected value of the posterior Beta distribution

$$\widehat{occ}_j = E[occ_j] = \frac{\alpha}{\alpha+\beta} \qquad (16)$$

where we have substituted the update rules to get the final expression as a function of the number of hits and misses. Accordingly, the variance of this probability is

$$\mathrm{var}(occ_j) \quad = \quad \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \qquad (17)$$

The collection of occupancy probabilities together with their variance estimates in the polar histogram form the beta grid descriptor of p_{det} (Fig. 3, center and right).

IV. Experiments

The goal of an evaluation of detector and descriptor approaches is to study their invariance properties under the typical variations of range data. In computer vision, a series of data sets for this purpose have been introduced in [9]. As there is no such benchmark available for 2D range data, we follow this idea and define experiments that contain the relevant transformations of 2D range data. We consider view point changes, noise, oversampling, and subsampling.

View point changes are important since recognizing places from different poses is at the basis of object recognition, global localization, or loop closing. Noise transformations are relevant for range finders whose noise properties vary with distance or incidence angle. Over- and subsampling are important since measurement sparsity is highly view-point variant causing the same scene to be sampled non-uniformly.

A. Experimental Setup

The proposed detectors and descriptors are evaluated on three data sets: fr-079, intel, and mit-csail. All log files are freely available from the online Radish repository and have been used for testing and benchmarking localization and SLAM approaches. As ground truth, the data are corrected

using a state-of-the-art SLAM algorithm [19] so as to register all laser scans into one global reference frame. The approach produces maps with a typical accuracy of several cm, rarely exceeding an error of 10 cm.

We then pick randomly 60 reference locations from all maps, 20 from each data set, and apply four transformations with increasing levels of variation on them (see also Fig. 4). For view point changes, we search all map locations whose scan has a certain overlap percentage with the reference scan. The percentage is varied from 50% to 90%. For the noise transformations, Gaussian noise with standard deviations between $\sigma = 0$ to 0.5 m is added in radial direction. For subsampling, the transformed scans are obtained by skipping each 2nd, 3rd and 4th reading of the reference scan, and for oversampling, one, two and three points are inserted using a linear interpolation in polar coordinates. We finally obtain several transformed scans $S_{T_1}, S_{T_2}, ..., S_{T_p}$ for each reference scan S_R on which compare the detectors and descriptors

More details on this experimental setup can also be found in the accompanying technical report [20].

1) Detector Performance: To quantify the stability of a detector, we use the repeatability measure introduced by Mikolajczyk *et al.* [8] and adapt it to 2D range data: the *repeatability score* of a pair of scans S_R, S_T is the ratio between the number of interest point to interest point correspondences and the total number of interest points in S_R that are also in S_T. Two interest points indexed i, j are said to correspond if the overlap error of their support regions R in Cartesian space is sufficiently small, i.e.

$$1 - \frac{R_i \cap R_j}{R_i \cup R_j} < \varepsilon_o \tag{18}$$

where ε_o is an error threshold.

With the shape of an interest point defined as the set of points in its support region, the actual presence of an interest point in both scans is checked by the modified Hausdorff distance between the two shapes,

$$H(R_i, R_j) = \frac{1}{N'} \sum_{p \in R_i} \min_{q \in R_j} \|p - q\| \tag{19}$$

where N' is the number of points in R_i.

A perfect detector will have a repeatability score of 1 for any possible transformation. However, in practice, this value is hardly obtained as structures in the environment can be hidden due to noise, extreme subsampling, view point changes, or dynamic objects. The theoretically perfect detector assumes every scan point to be an interesting point. This detector, however, is not desirable since the detected points are not distinctive and thus difficult to match.

2) Descriptor Performance: To evaluate the descriptor performances, we use a criterion similar to the one proposed in [9] based on the precision-recall curve. Recall is the number of correctly matched interest points with respect to the total number of corresponding interest points between two scans. Precision quantifies the number of correct matches relative to the total number of potential matches.

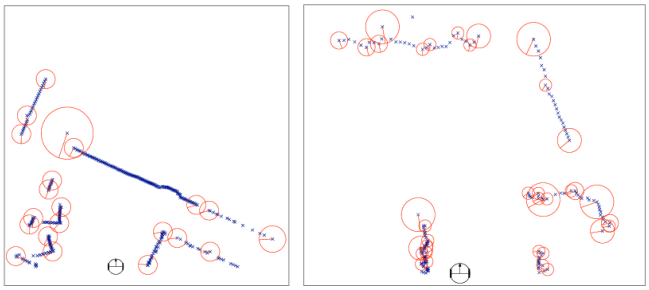

Fig. 5. Example FLIRT detection result in indoor (mit-csail, left) and outdoor (fr-clinicum, right) data. The circles show the interest points and their support region (the actual descriptors are not shown).

To obtain a matching pair, we compare each descriptor in S_R with each descriptor in S_T. We use the symmetric χ^2 distance between the descriptor histograms as a distance function. Two interest points are said to match when their distance is below a threshold. This threshold is varied in order to capture different values for precision and recall. To check if a match actually exists, we use the overlap error defined in Eq. (18).

A perfect descriptor would give a recall of 1 for every precision value. In practice, recall increases when precision decreases, since to find a correct match, many false matches have to be accepted. This is because unique features do not always exist, especially in indoor environments with symmetries, that is, similar or identical structures in different places (such as corners, columns, or doors).

B. Detector Comparison

In the evaluation of the detector approaches, the repeatability score in Eq. (18) is compared to a maximal error of $\varepsilon_o = 0.6$. We further assume that an interest point is present in a scan if the modified Hausdorff distance is less than 0.1. We used 5 scales for the detectors with increasing scale $t = t_0 \cdot (t_i)^s$, where t_0 is equal to 0.2 for the curvature detector and to 1.6 for the others. The interscale value t_i is set to 1.4 and $s \in \{0, 1, 2, 3, 4\}$ is the current scale. These parameters led to the best results in the benchmarking experiments.

Fig. 6 shows the repeatability results for the different transforms. The curvature-based detector outperforms the range-based and the two normal-based detectors in all situations. The only exception is for higher noise levels, where the range based detector shows a better behavior. This result, however, is only obtained when adding Gaussian noise with a standard deviation of more than 0.25 meters, a value far beyond the specifications of most laser scanners. The reason why the curvature-based detector is more invariant to different levels of subsampling and oversampling is due to the fact that it operates in geodesic coordinates and not on the raw range signal.

C. Descriptor Comparison

The descriptors are evaluated with 12 bins for the angle discretization and 4 bins for the distance discretization (see also Fig 3). The descriptor size is set to 0.5 m.

In Fig. 7 the precision-recall graphs are plotted for the different transforms. The β-grid descriptor performs slightly

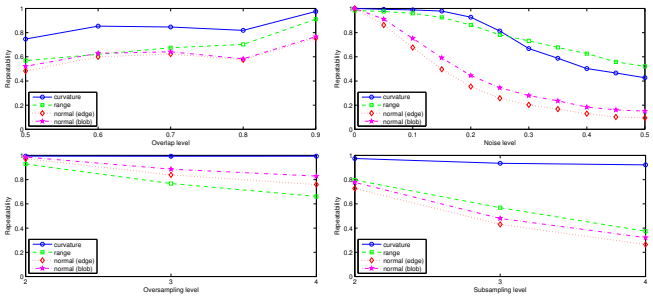

Fig. 6. Repeatability measures for different view point (top left), noise (top right), over- (bottom left) and subsampling levels (bottom right).

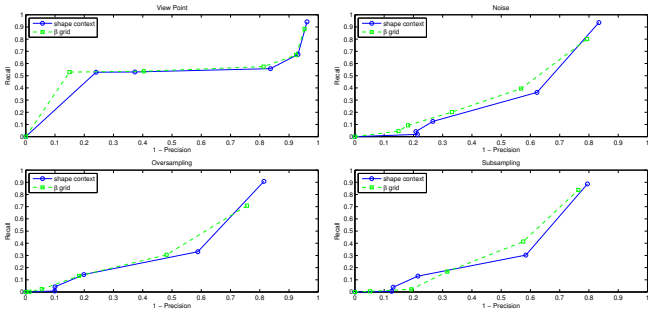

Fig. 7. Results of the matching experiments. The plots show precision vs. recall for the view point (top left), noise (top right), over- (bottom left) and subsampling transformations (bottom right).

better than the shape context-based descriptor, while the performances are not that different. One of the reasons for this result is that the β-grid descriptor is able to encode free-space information to discriminate a convex between a concave structure of the same shape.

Interestingly, compared with the corresponding plots of detectors and descriptors for image data [9], [8], the results have both, similar trends and similar values. We believe this outcome to be encouraging, given the success of local image features in computer vision.

In the light of the previous results, we finally choose FLIRT to be the combination of the curvature-based detector and the β-grid descriptor since this detector-descriptor pair shows the best performance across all other combinations.

D. Data Association with RANSAC

A key problem in localization and SLAM is global data association where the robot pose is sought given a single observation and a map. This is a highly relevant problem, e.g. for loop closing in SLAM where the robot has to decide if and where the currently observed place has already been visited. We choose four log files, three from indoor (fr-079, intel, mit-csail) and one from outdoor (fr-clinic) environments and process them with the SLAM approach in [19] to obtain a ground truth estimate.

For all data sets, we carry out the following: for each scan in the log file, the scan is first removed from the map to exclude the trivial self-match. Then, the scan is compared to all other scans using a standard RANSAC algorithm. To this end, we compute the correspondence set by matching the descriptors using the symmetric χ^2 distance and a nearest neighbour strategy with a threshold of 0.4. The correspondence set is the input to RANSAC that is applied

Data set	**Size** [m]	**Scans**	$\bar{N}_{IP}$	p_{GL}	p_{LC}	$\mathbf{t}_{sm}$	$\mathbf{t}_{ss}$
fr079 (in)	50×20	1464	27	.98	.98	0.66s	450μs
intel (in)	50×40	2672	18	.98	.98	0.52s	200μs
csail (in)	80×60	1051	23	.97	.97	0.33s	320μs
clinic (out)	550×300	1776	34	.79	.53	1.15s	650μs

TABLE I
SUMMARY OF THE RANSAC EXPERIMENT.

with an inlier probability of 0.5. If the resulting number of inlier correspondences is above a threshold, N_I^{min}, the solution is considered a candidate match and inserted in the solution set. For each candidate match, we then compute the robot pose in a least-squares sense. If the distance between this pose and the ground truth pose is within 0.5 meter and 10 degree we consider it a correct match (see Fig. 1 for an example). Notice that matching a scan against a map scales *linearly* with the map size.

Fig. 8 contains the resulting precision-recall values. The top row of the figure shows the precision, the bottom row shows the recall, both against different values for N_I^{min} and two strategies: considering the match in the solution set with the highest number of inliers (corresp) or with the lowest RANSAC error (residual). The third curve (closest) represent the (theoretically) optimal strategy to choose the closest solution to the ground truth that passed N_I^{min}. The curve demonstrates that a correct match is in the solution set although it is not the one with the minimum error.

As can be seen, the approach has both high precision and high recall values, even at small numbers of inliers. Put into words, FLIRT features enable a robot to globally self-localize from a single scan with a success probability of at least 98% within 50 cm accuracy and hundreds of milliseconds execution time. Alternatively, the figures show that, with a confidence of 0.95, FLIRT features are able to correctly identify a potential loop closure event with 98% of the scans.

Table I summarizes the RANSAC results for all four data sets. The columns are environment size in meters, number of scans in the map, the average number of interest points per scan detected ($\bar{N}_{IP}$), the probability of correct global localization from a single scan (p_{GL}), the probability of correct loop closure from a single scan given a precision of 0.95 (p_{LC}), the total time for a single scan-to-map match ($\mathbf{t}_{sm}$) and the average time for a scan-to-scan match ($\mathbf{t}_{ss}$).

In the outdoor environment of our experiment these probability are still at 79% for global localization and 53% for loop closure. This behaviour is mainly caused by the sparseness of the data that have been collected only every meter (for memory reasons) and by the lower map accuracy in the outdoor case. The numbers mean that, on average, we are still able to self-localize the robot every 1.27 scan and close a loop every second scan. Note also that all experiments have been conducted using the same set of parameters.

V. CONCLUSIONS

In this paper we addressed the problem of multi-scale interest points for 2D range data. Based on large-scale data sets for robot navigation, we proposed a set of benchmark experiments as testbed for the comparison of detectors and

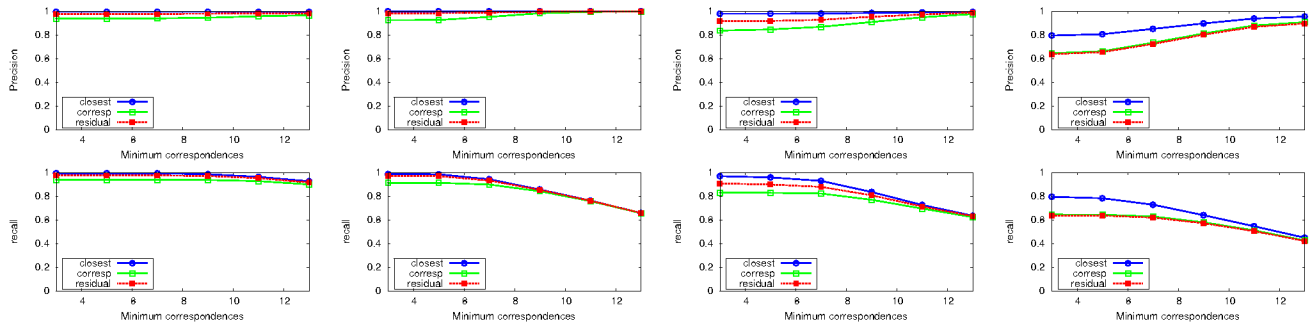

Fig. 8. Results from the data association experiment. The precision (top) and recall (bottom) values for the fr-079, intel, mit-csail and fr-clinic, respectively, are shown. They demonstrate that FLIRT features are highly appropriate for global localization or loop closing in indoor and outdoor environments.

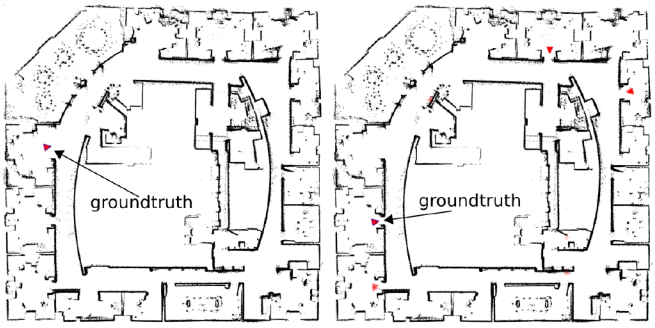

Fig. 9. Global localization result in the Intel data set (intel). Unique localization (left), ambiguous localization with more than one hypotheses (right). Red triangles indicate pose hypotheses, the black triangle stands for ground truth. See the video attachment for an animation.

descriptors for 2D range data. The data sets have been globally registered by a SLAM algorithm to serve as the ground truth. We then developed and compared a number of different detectors and descriptor approaches. We considered four multi-scale detectors on the direct range signal, a descriptor based on the idea of occupancy grids and a shape context-based descriptor.

We finally proposed FLIRT as the most powerful detector-descriptor pair, combining a detector based on a curve approximation of the range signal and a descriptor that encodes occupancy information in a polar region around the interest point. It was found that the repeatability and precision diagrams have both, similar trends and values than the corresponding plots of detectors and descriptors for image data. We further showed how a naive application of RANSAC already leads to state-of-the-art results for the loop closing and global localization problem in terms of matching accuracy, efficiency and simplicity.

We conclude that FLIRT features have a great potential for robot navigation based on 2D range data. In future work we will analyze the benefit of more advanced RANSAC variants with guided data association schemes and the variance information in the β-grid descriptors. We will also explore the application of FLIRT to navigation tasks such as SLAM, multi-hypothesis pose tracking and object recognition.

The data sets and the source code of the detector and descriptor implementations are available on the author websites.

References

[1] D. Lowe, "Object recognition from local scale-invariant features," in *Proc. of the 7th IEEE Int. Conf. on Computer Vision (ICCV'99)*, 1999.

[2] C.-K. Tang and G. Medioni, "Robust estimation of curvature information from noisy 3d data for shape description," in *Proc. of the 7th IEEE International Conference on Computer Vision (ICCV'99)*, 1999.

[3] S. Lazebnik, C. Schmid, and J. Ponce, "Affine-invariant local descriptors and neighborhood statistics for texture recognition," in *Proc. of the 9th IEEE Int. Conf. on Computer Vision*, Washington DC, 2003.

[4] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. on Patt. Anal. and Mach. Intell.*, vol. 24, 2002.

[5] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded-Up Robust Features," in *Proc. of the 9th European Conf. on Computer Vision (ECCV'06)*, Graz, Austria, 2006.

[6] S. Thrun, M. Bennewitz, W. Burgard, A. Cremers, F. Dellaert, D. Fox, D. Haehnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, "Minerva: A second generation mobile tour-guide robot," in *Proc. IEEE Int. Conf. on Robotics and Automation*, Detroit, USA, 1999.

[7] K. O. Arras, R. Philippsen, N. Tomatis, M. de Battista, M. Schilt, and R. Siegwart, "A navigation framework for multiple mobile robots and its application at the Expo.02 exhibition," in *Proc. IEEE Int. Conf. on Robotics and Automation*, Taipei, Taiwan, 2003.

[8] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *Int. J. Comp. Vis.*, vol. 65, 2005.

[9] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 27, 2005.

[10] G. Taubin, "A signal processing approach to fair surface design," in *Int. Conf. on Comp. Graph. and Interact. Techn.*, New York, NY, 1995.

[11] M. Pauly, R. Keiser, and M. Gross, "Multi scale feature extraction on point sampled surfaces," in *Proc. of Conf. of European Association for Computer Graphics*, 2003.

[12] J. Novatnack and K. Nishino, "Scale-dependent 3d geometric features." in *Proc. of Int. Conf. on Comp. Vis.*, 2007.

[13] R. Unnikrishnan and M. Hebert, "Multi-scale interest regions from unorganized point clouds," in *Workshop on Search in 3D, IEEE Conf. on Comp. Vis. and Patt. Rec.*, 2008.

[14] P. N. David Cole, Alastair Harrison, "Using naturally salient regions for slam with 3d laser data." in *Workshop on SLAM, IEEE Int. Conf. on Robotics and Automation*, Barcelona, 2005.

[15] M. Bosse and R. Zlot, "Map matching and data association for large-scale two-dimensional laser scan-based slam," *IJRR*, vol. 27, 2008.

[16] ——, "Keypoint design and evaluation for place recognition in 2d lidar maps," *Robotics and Autonomous Systems*, vol. 75, 2009.

[17] T. Lindeberg, *Scale Space Theory in Computer Vision*. Norwell, MA, USA: Kluwer Academic Publishers, 1994.

[18] K. O. Arras and R. Siegwart, "Feature extraction and scene interpretation for map-based navigation and map building," in *Proc. of SPIE, vol. 3210, Mobile Robotics XII*, Pittsburgh, USA, 1997.

[19] G. Grisetti, C. Stachniss, S. Grzonka, and W. Burgard, "A tree parameterization for efficiently computing maximum likelihood maps using gradient descent," in *Proc. of Robotics: Science and Systems*, Atlanta, GA, USA, June 2007.

[20] G. D. Tipaldi and K. O. Arras, "FLIRT – Interest Regions for 2D Range Data," Institut für Informatik, University of Freiburg, Tech. Rep. 249, 2009, http://www.informatik.uni-freiburg.de/tr.