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, range imaging and robotics. While there is a great variety of approaches for both, detectors, descriptors and combinations thereof for image data and 3D point clouds, there is no method readily available for 2D range data. For this reason, this paper first proposes a set of benchmark experiments based on known data sets for robot navigation to compare detector repeatability and descriptor matching performance. Secondly, the paper introduces FLIRT that stands for fast laser interest region transform, a multi-scale interest region operator for 2D range data. The transform combines a detector based on a curve approximation of the range signal with a descriptor that makes use of occupancy information in a polar region around the detected point. FLIRT combines the best detector with the best descriptor, experimentally found in a comprehensive analysis of alternative detectors and descriptor approaches. We compare the proposed method with three other multi-scale detectors on the direct range signal and a variant of the shape context descriptor. The experiments lead to very similar repeatability and matching performance values that have been found for local image features in the computer vision literature, encouraging a wide range of applications of FLIRT for 2D range data.

1 Introduction

The introduction of local image features had a great impact on many computer vision tasks such as object and scene recognition, motion tracking, stereo correspondence, or 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

Social Robotics Lab, University of Freiburg, Department of Computer Science, Georges-Koehler-Allee 106, 79110 Freiburg, Germany, e-mail: {tipaldi, arras}@informatik.uni-freiburg.de

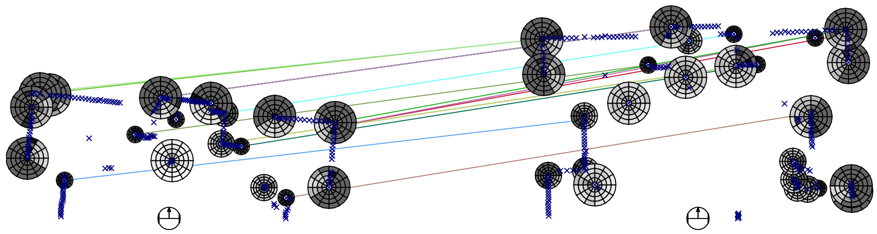

Fig. 1 Example matching of two scans of the same scene using RANSAC. The figure shows all extracted FLIRT features for both scans and the 16 inlier correspondences.

stable locations and the description to encode the image structure, there is a great variety of approaches available for image and 3D range data.

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 [1, 2], 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 the sense that do not rely on predefined models but are less efficient and operate with map representations that scale poorly with environment size and dimensionality. Interest points, on the other hand, combine the compactness of discrete features and the generality of raw range data. For other tasks such as object and place recognition with 2D range data, the same evidence accumulation and matching techniques that have been developed for visual interest points can be reused. Generally speaking, interest points for 2D range data allow for early sensor fusion with vision and an unified treatment of laser and image data.

Most related work is found in computer vision, where several detector and descriptor algorithms have been developed. A comparison of different detectors and descriptors can be found in Mikolajczyk *et al.* [3, 4]. 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 [5] 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.* [6], 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 sample point x. The scale at a point is then the neighborhood size for which the surface variation achieves a local extrema. Novatnack and Nishino [7] 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.* [8] define an integral operator that maps the input curve into its multi-scale parametrization. The operator works on the geodesic coordinate of the curve and the interest points are defined as local extrema in a geodesic neighborhood. Using data from a laser range finder, Cole *et al.* [9] 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 very 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. Bosse and Zlot [10] define entire laser scans as features and use orientation histograms to describe them. While very interesting, the descriptor is applicable only to whole scans, thus being not able to perform tasks such as object detection or partial matching for which local feature are needed.

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 lower 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, this paper proposes a method for extracting local distinctive features from 2D range data that can be used to perform matching between different scans of an object or a scene. We present a set of benchmark experiments using known data sets in robot navigation 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 finally propose FLIRT (fast laser interest region transform), that combines the most appropriate detector/descriptor pair found in the evaluation (see also Fig. 1).

The paper is structured as follows. Section 2 describes the different detectors, while Section 3 introduces the descriptors used in the comparison. The experimental setup is described in Section 4 and the results are given in Section 5. Finally, Section 6 concludes the paper.

2 Detectors

In this section we present four multi-scale detector approaches that are compared in this paper. Section 2.1 presents a detector based on the raw range signal, while Section 2.2 describes two detectors based on a normal approximation of the range data. Section 2.3 introduces a curvature based detector derived from Unnikrishnan *et al.* [8]. These detectors look for interesting points on different scales according to the discrete version of the scale space theory [11].

The scale-space theory is a framework for multi scale signal representation 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 [11] 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.

2.1 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.

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

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

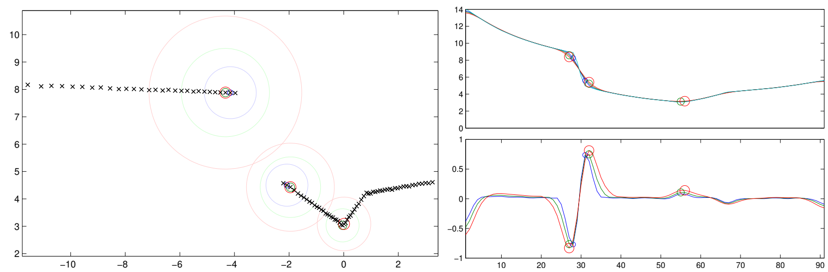

Fig. 2 Range-based detector on synthetic data taken from (0,0) with an angular resolution of 2° and a radial noise of $\sigma = 2cm$. Three scales are considered: a small scale (scale 1, blue), a medium scale (scale 2, green) and a large scale (scale 3, red). The detector is able to find the range discontinuity in the background and foreground on all scales and the convex corner on scales 2 and 3. The obtuse concave corner was not extracted as the response in the second derivative is too weak. 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 second derivative $(D_2 * S)(x;t)$ (bottom).

where

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

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

As an interest point is always a point in the scan (there are no subpixel positions), their Cartesian coordinates are given by the corresponding range and bearing measurements. The final scale in Cartesian space is obtained as the distance between the interest point position and the farthest point in its support region in signal space. The support region is the set of points within $\pm 2t$ measurements centred around the interest point.

An example of the behavior of this detector is given in Fig. 2.

2.2 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 fitting a line in a least-squares sense to a group of measurements in a sliding window centred 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 opposed to the minimization of algebraic errors. With measurements represented in polar coordinates and assuming errors in range direction only, the fit expressions have, despite being a nonlinear regression problem, a closed-form solution [12]. 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.

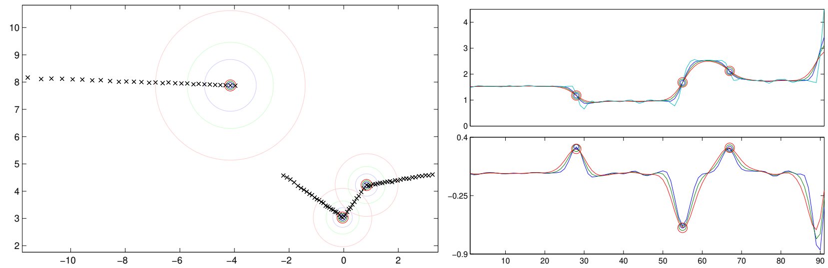

Fig. 3 Normal-based edge detector on the synthetic data set of Fig. 2. The detector finds the background part of the range discontinuity, the convex corner and the obtuse concave corner on all scales. The foreground part of the range discontinuity is not extracted as there is only one peak for the discontinuity. The two diagrams at the right show the interest points in signal space with the normal signal and its smoothed variants at each scale (top) and the gradient magnitude $\|\nabla S(x;t)\|$ (bottom).

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

$$\|\nabla S(x;t)\| = (D_1 * S)(x;t) \tag{5}$$

where

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

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).

For both detectors, the mapping of interest points and their scale into the Cartesian space is identical to the one for the range-based detector. An example of the behavior of the two detectors is given in Fig. 3 and Fig. 4.

2.3 Curvature-Based Detector

We describe here the detector introduced by Unnikrishnan and Hebert [8]. 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

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

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

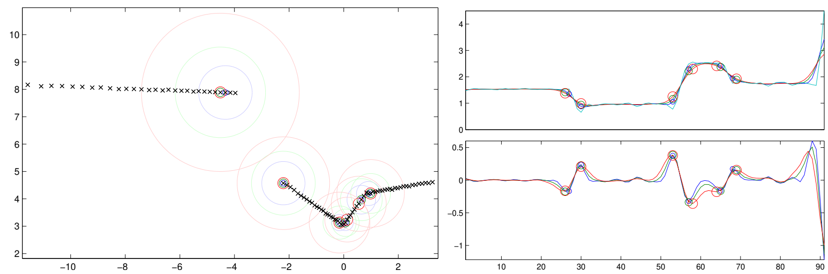

Fig. 4 Normal-based blob detector. The detector finds the range discontinuity on both sides (background and foreground) and all scales while to corners it responds by a double detection at each side. The exact interest point position varies in ± 1 point across scales due to discretization artifacts. 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 second derivative $(D_2 * S)(x;t)$ (bottom).

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 $p(s;t)$,

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

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

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,

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

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

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

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 of the behavior is provided in Fig. 5

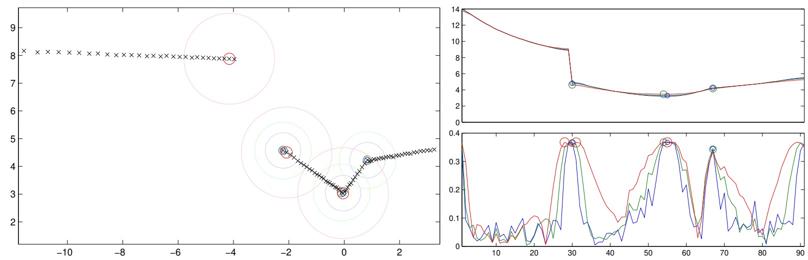

Fig. 5 Curvature-based detector. The detector finds the background part of the range discontinuity on the largest scale, the foreground part on all scales, responds on all scales for the convex corner, and on scale 1 and 2 for the the obtuse concave corner. The two diagrams at the right show the interest 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 of the signal are ignored as they are caused by the non-circularity of the data.

3 Descriptor

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.

3.1 Linear Local Shape Context

Shape context, introduced by Belongie *et al.* [13], is a shape 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 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

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

where bin_j is defined by discretizing the distance of the point and the viewing angle. Fig. 6 left shows an example descriptor.

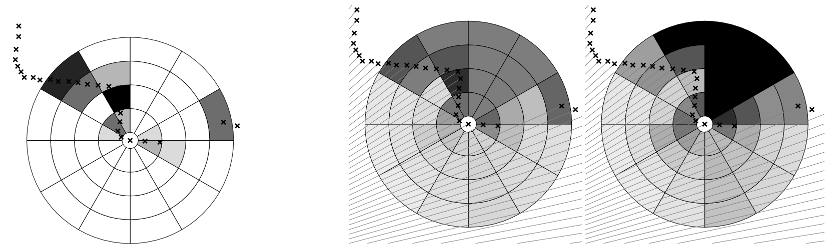

Fig. 6 Linear local shape context descriptor (left) and β-grid descriptor (center: occupancy probability, right: variance) for an example interest point in real data. The shape context descriptor captures the local structure in the histogram by counting the hits in each cell. The β-grid descriptor estimates a true occupancy probability for each cell together with a variance estimate. Thereby, the β-grid descriptor is able to encode the concepts of free space (by ray tracing) and uncertain information (darker areas in the variance histogram).

3.2 β-Grids

The second descriptor we consider in this paper are named *beta grids* (*β-grids* in short), a novel descriptor for range data.

An important difference between range and image data is that the former not only encodes distance information in a metric sense but also directed free-space information between the light (or sound) emitting sensor and the measured object. This is relevant extra information, not encoded in the shape context descriptor, used, for instance, in occupancy grids. We adopt this idea for the purpose of our descriptor. 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,

$$p(z; occ_j) = \begin{cases} occ_j & \text{if } z = 1 \\ 1 - occ_j & \text{if } z = 0 \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

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,

$$p_\beta(occ_j; \alpha, \beta) = \frac{{occ_j}^{\alpha-1}(1 - occ_j)^{\beta-1}}{\mathrm{B}(\alpha, \beta)} \tag{15}$$

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

$$\alpha_i = \alpha_{i-1} + \sum 1^{z_i} \qquad \beta_i = \beta_{i-1} + \sum 1^{(1-z_i)}. \tag{16}$$

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

$$\widehat{occ}_j = E[occ_j] = \frac{\alpha}{\alpha+\beta} = \frac{\#\text{hits}+1}{\#\text{hit} + \#\text{misses}+2} \tag{17}$$

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

$$\mathrm{var}(occ_j) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{(\#\text{hits}+1)(\#\text{misses}+1)}{(\#\text{hits}+\#\text{misses}+2)^2(\#\text{hits}+\#\text{misses}+3)}.$$

The collection of occupancy probabilities together with their variance estimates in the polar histogram form the beta grid descriptor of p_{det} (Fig. 6, center and right).

4 Experiments

The goal of an evaluation of detector and descriptor approaches is to study their invariance properties under typical data variations. In computer vision, a series of data sets for this purpose have been introduced in [4]. As there is no such benchmark data available for 2D range data, we follow this idea and define data sets that contain the relevant transformations of 2D range data. Three known data sets are taken for this purpose, namely the Intel data set, the MIT-CSAIL data set and the Freiburg-079 data set. All log files are freely available from the online Radish[1] repository and have been used for testing and benchmarking localization and SLAM approaches.

To evaluate the proposed detectors and descriptors, we apply four transforms: view point changes (Fig. 7), Gaussian noise (Fig. 8, top), oversampling (Fig. 8, middle), and subsampling (Fig. 8, bottom). These transforms are applied to randomly picked laser scans from all log files. In total we pick 60 locations, 20 from each data set. As ground truth, the data are corrected using a state-of-the-art SLAM algorithm [14] so as to register all laser scans into a single global reference frame. This correction is equivalent to homography for images and needed to obtain the view point transformation and to check if two interest point are found in the same location without manual data labeling.

Given the registered scan poses, view point variations are obtained by considering the overlap between a reference scan and other scans from the same log file. We define five overlap intervals: $(90\%,100\%]$, $(80\%,90\%]$, $(70\%,80\%]$, $(60\%,70\%]$,

[1] http://radish.sourceforge.net/

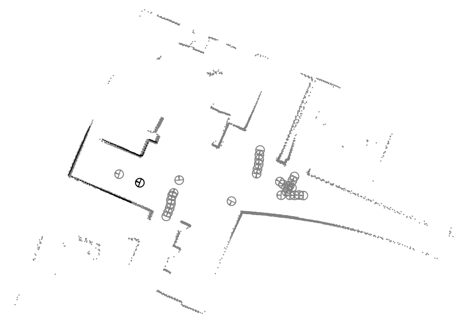

Fig. 7 Example location from the Intel data set. The reference location is indicated by the black robot and the black laser points. The figure shows the 31 locations that have a 50% overlap in the scan used for the repeatability experiments of the detectors.

$(50\%, 60\%]$. Overlap is defined as the percentage of points from different scans whose distance in global coordinates is smaller than 10 cm, a threshold derived from the typical accuracy of the SLAM algorithm. In the noise transformations, we take a reference scan and add zero-mean Gaussian noise in the range values. The standard deviation is varied in the range of $0.05 - 0.5$ m. For each noise level, five transformation with different random seeds are generated. Oversampling is made by linearly interpolating the range values. The oversampling level is varied for 2, 3 and 4 artificial points where a level of 1 is the reference scan. Similarly, for subsampling, readings from the reference scan are removed over a range of every 2, 3 or 4 points.

View point changes are important since recognizing places from different poses is at the basis of global localization, loop closing or object recognition. 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 making that scenes are non-uniformly sampled.

4.1 Repeatability

The key property of a good detector is its ability to stably detect the same interest point, regardless from where it is seen by the sensor. To this end, we use the repeatability measure introduced by Mikolajczyk *et al.* [3] adapted for laser data. The *repeatability score* of a pair of scans is the ratio between the number of interest point-to-interest point correspondences and the total number of interest points in the reference scan that are present in the transformed one. Two interest points are deemed to correspond if the overlap error of their support region in Cartesian space is sufficiently small:

$$1 - \frac{R_a \cap R_b}{R_a \cup R_b} < \varepsilon_o \tag{18}$$

where R is the support region of the interest points and ε_o the error threshold. The presence of an interest point is checked by using the modified Hausdorff distance between two shapes.

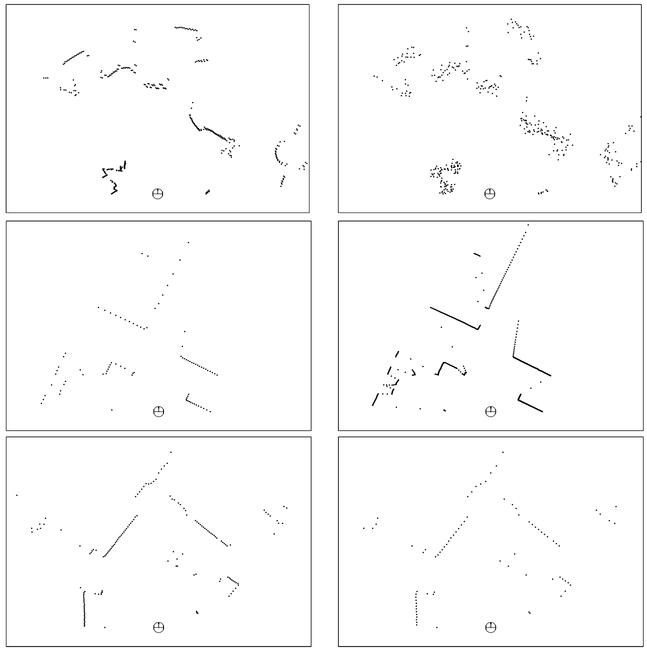

Fig. 8 Example locations with different transformations. Reference locations are shown in the left column, the right column contains the same scans subject to additional Gaussian noise (top, with $\sigma = 0.25$), oversampling (middle, 4 times), and subsampling (bottom, 4 times).

$$H(R_a, R_b) = \frac{1}{N} \sum_{a \in R_a} \min_{b \in R_b} \|a - b\| \tag{19}$$

where N is the number of points in the region R_a. The shape of an interest point is defined to be the set of points in its support region R_a. For the transformed scan, the shape R_b is obtained by translating this support region to the reference scan using the globally registered scan poses.

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, change in view point, or occlusions caused by dynamic objects. A theoretically perfect detector is the one that assume every scan point to be an interesting point. This kind of detector, however, is not preferable since the detected points are not distinctive, making matching a difficult task.

4.2 Matching Performance

We use a criterion similar to the one proposed in [4] to evaluate the descriptor performances, based on precision and recall measures. Recall is the number of correctly matched interest points with respect to the number of corresponding interest points between two scans. Precision models the number of false matches, relative to the total number of matches.

Two interest points are matched if the distance between their descriptors is below a threshold. Each descriptor of the reference laser scan is compared with each descriptor in the transformed scan. The threshold is varied to capture the different values for precision and recall. We use the symmetric χ^2 distance between histograms as a distance function. The number of correct matches and correspondences is defined with the overlap error of Eq. (18).

A perfect descriptor would give a recall equal to 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 phenomenon is due to the fact that unique features do not always exist, especially in indoor environments that can contain similar or even the same structures in different places (e.g. corners, columns, doors).

4.3 Efficiency

The computational efficiency of the different detector and descriptor approaches is as follows. Let N be the number of points in a laser scan, S the number of scales, k the size of the differentiation kernel, and $k(s)$ the size of the smoothing kernel at scale s, then the range-based detector requires $N \cdot k(s)$ operations to create a smooth signal and $N \cdot k$ for the convolution, yielding a worst case complexity of $2 \cdot N \cdot K \cdot S$ operations where $K = max(k, k(s))$ for all s. The normal-based detectors differs from the range-based detector by the local normal approximation which has a $N \cdot K$ complexity, making a total of $3 \cdot N \cdot K \cdot S$ operations per laser scan. The curvature-based detector computes a minimum spanning tree which, for sparse graphs with N vertexes, has a $N \cdot log(N)$ complexity, giving a total of $N \cdot log(N) + 2 \cdot N \cdot K \cdot S$ operations.

For the descriptors, let M be the number of interest points extracted with a detector, N' the number of beams that intersect the disk of a descriptor and B the number of descriptor bins. Then, the local linear shape-context descriptor requires N' operations to count the number of hits and B operations for normalization, making a total of $M \cdot (N' + B)$ operations. The β-grid descriptor performs ray tracing which is a $N' \cdot B$ operation, times two for the variance, making a total of $2 \cdot M \cdot N' \cdot B$ operations per laser scan. In practice, as N' is relatively small, $N' \ll N$, the difference between the two descriptors is sensible but not extensive.

5 Results

In this section we present and discuss the experimental results of the evaluation. In the detector experiments we average the repeatability values over all 60 sampled locations. For the descriptor experiment, averaging of the precision and recall values is done over the 60 sampled locations and all transformation levels.

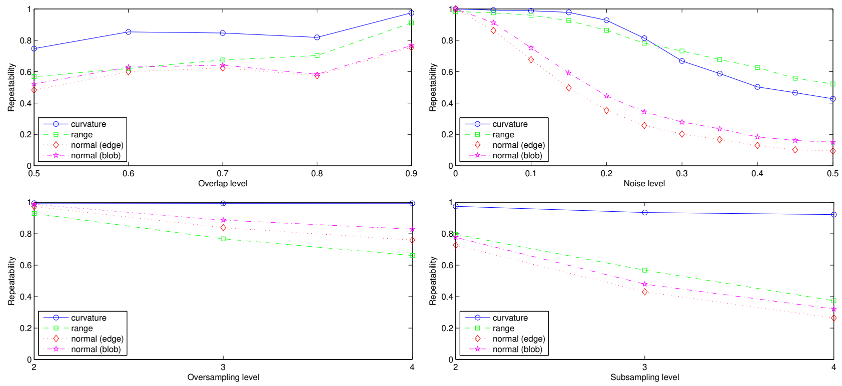

Fig. 9 Results of the repeatability experiments. The repeatability measures for different view point (top left), noise (top right), over- (bottom left) and subsampling levels (bottom right) are shown.

5.1 Detector Comparison

For the performance evaluation of the detector approaches, the repeatability score in Eq. (18) is compared to a threshold of $\varepsilon_o = 0.6$. We assume that an interest point is present in a scan if the modified Hausdorff distance Eq. (19) is less than 0.1.

Fig. 9 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 present after adding Gaussian noise with a standard deviation of more than 0.25 meters, a value far beyond actual laser scanner specifications. The reason why the curvature-based detector is highly invariant to different levels of subsampling and oversampling is due to the fact that it works in geodesic coordinates and not on the raw range signal.

5.2 Descriptor Comparison

The descriptors are evaluated with 12 bins for the angle discretization and 4 bins for the distance discretization (see also Fig 6). The actual descriptor size is defined by the detector scales.

In Fig. 10 the precision-recall graphs are plotted for the different transforms. The β-grid descriptor performs slightly 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 which enables it to, e.g., discriminate a convex from a concave structure of the same shape.

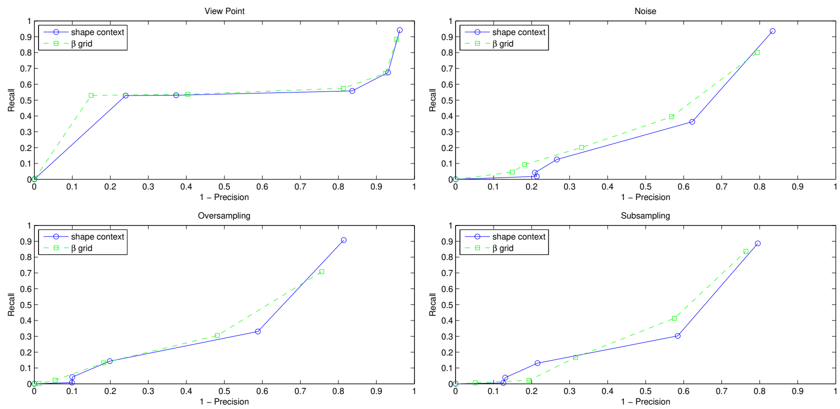

Fig. 10 Results of the matching experiments. The plots show precision vs. recall for the view point (top left), noise (top right), over- (bottom left) and subsampling transformations (bottom right).

5.3 FLIRT

In the light of these results, we 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. Interestingly, the repeatability and precision recall graphs in Figs. 9 and 10 have both, similar trends than the corresponding plots of detectors and descriptors for image data and similar values. Repeatability values obtained in [3] are in the range of $50-80\%$, recall and precision values reported in [4] are in the range of $30-50\%$ and $40-60\%$ respectively. This encourages the use of FLIRT for a similar range of applications than in computer vision. Finally, Fig. 1 shows an example match between two scans using RANSAC.

6 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 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. The experiments showed performance values similar to results that have been reported for local image features in computer vision, encouraging the use of FLIRT in a wide range of applications with 2D range data.

In the future we will analyze the benefit of the variance information in the β-grids for partial occlusions and explore the use of FLIRT for applications such as loop closing, global localization, place and object recognition. The data sets and the Matlab source code of the detector and descriptor implementations will be made available on the author websites.

7 Acknowledgments

This work has partly been supported by the German Research Foundation (DFG) under contract number SFB/TR-8.

References

1. 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 (ICRA)*, Detroit, USA, 1999.
2. 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.
3. 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.
4. K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 27, 2005.
5. G. Taubin, "A signal processing approach to fair surface design," in *Proc. of Int. Conf. on Comp. graph. and Interactive Techn.* New York, NY, USA: ACM, 1995.
6. M. Pauly, K. R, and M. Gross, "Multi scale feature extraction on point sampled surfaces," in *conf. of European Ass. for Computer Graphics*, 2003.
7. J. Novatnack and K. Nishino, "Scale-dependent 3d geometric features." in *Proc. of Int. Conf. on Comp. Vis.*, 2007.
8. 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.
9. 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.
10. M. Bosse and R. Zlot, "Map matching and data association for large-scale two-dimensional laser scan-based slam," *Int. J. Rob. Res.*, vol. 27, 2008.
11. T. Lindeberg, *Scale Space Theory in Computer Vision*. Norwell, MA, USA: Kluwer Academic Publishers, 1994.
12. 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.
13. 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.
14. 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.