FLIRT: Interest Regions for 2D Range Data with Applications to Robot Navigation

Gian Diego Tipaldi, Manuel Braun, and Kai O. Arras

Abstract In this paper, we present FLIRT that stands for Fast Laser Interest Region Transform, a multi-scale interest region operator for 2D range data. FLIRT combines a detector based on a geodesic curve approximation of the range signal and a descriptor based on a polar histogram of occupancy probabilities. This combination was found in a set of comparative benchmark experiments on standard indoor and outdoor data sets. The analysis yields repeatability and matching performance results similar to the values found for interest points in the computer vision literature, encouraging a wide application of FLIRT on 2D range data. We show how FLIRT in conjunction with RANSAC can be applied for loop closing, global localization, incremental mapping and pose-based SLAM with particularly simple algorithms. The results demonstrate that FLIRT features have a great potential for robot navigation in terms of precision-recall performance, efficiency, generality and simplicity.

1 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 locations of interest in image space and compute a distinctive descriptor over regions centered around these locations. This yields a description of the image content as a collection of local interest regions that can be used 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 range data [13, 12].

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

Gian Diego Tipaldi · Manuel Braun · Kai O. Arras
Social Robotics Laboratory, Department of Computer Science, University of Freiburg
e-mail: lastname@informatik.uni-freiburg.de
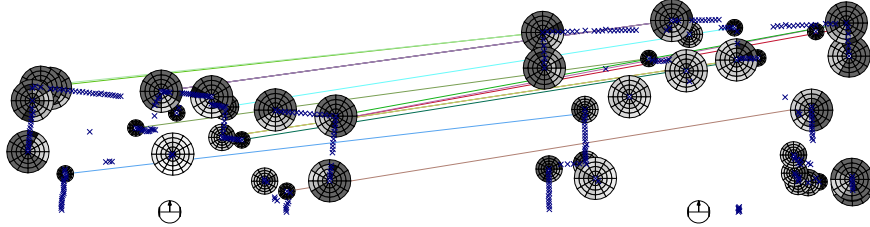
Fig. 1 Example matching of two scans from a laser range finder of the same scene using RANSAC. The figure shows the extracted FLIRT features for both scans and the 16 inlier correspondences.

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 [18, 2], they both have strengths and weaknesses. Geometric features allow for compact map representations and high accuracy but rely on predefined models. Dense approaches using raw data or grids are general in this sense but scale less well with map size and dimensionality. Interest points, on the other hand, combine the compactness of discrete features and the generality of raw range data.

A comparison of different detectors and descriptors for images can be found in Mikolajczyk *et al.* [13, 12]. 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 [17] 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.* [16], 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. Novatnack and Nishino [14] 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.* [21] 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.* [7] propose an information-theoretic measure of local saliency to find natural features in measurement space. 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 [5, 6]. In [5] the authors define entire laser scans as features and use orientation histograms to describe them. In [6], 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 descriptor support regions of $9 \times 9\,m$ defined on submaps that

are 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 [6], 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 success of visual interest points. 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. These differences motivate a specifically derived interest point transform for 2D range data.

This paper extends our previous work [19] where we compared several detector and descriptor approaches for 2D range data and proposed FLIRT as the most powerful combination in terms of stability and matching performance. Here, we report on comprehensive experiments with FLIRT applied to navigation problems such as global localization, loop closing, incremental mapping and SLAM.

The paper is structured as follows. Section 2 presents the FLIRT detector and descriptor. How FLIRT can be applied to robot navigation is described in Section 3 with the experimental results given in section 4. Section 5 concludes the paper.

2 FLIRT

In this section we present the Fast Laser Interest Point Transform (FLIRT), a multi-scale interest point operator for 2D range data. The operator consists in a detector based on a geodesic curve approximation of the range signal and a descriptor based on a polar histogram of occupancy probabilities. This particular combination is the result of a comprehensive evaluation [19], where we compared four different detectors and two descriptors in a set of benchmark experiments. We will now describe the FLIRT detector and descriptor.

2.1 Curvature-Based Detector

The detector is derived from the approach by Unnikrishnan and Hebert [21] for 3D point clouds, applied onto 2D range data. The rationale behind this detector is that range data define a curve in Cartesian space and the scale space theory should be
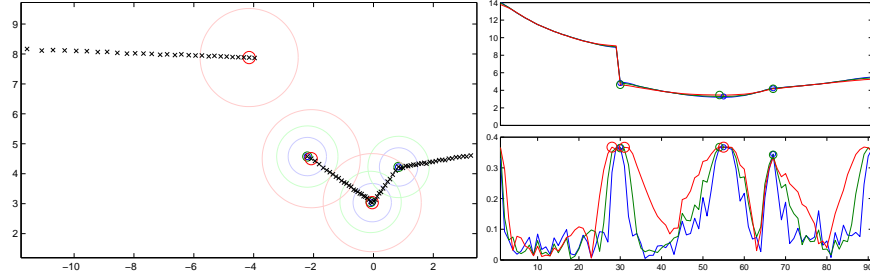
Fig. 2 Curvature-based detector on synthetic data. The detector finds the background of the range discontinuity on the largest scale (left, red circle), the foreground on all scales, responds on all scales for the convex corner, and on scale 1 and 2 for the the obtuse concave corner. The diagrams at 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.

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

$$S(\alpha(s);t) = \int_{\Gamma} k(s,u;t)\alpha(u)\mathrm{d}u \tag{1}$$

$$k(s,u;t) = \mathcal{N}(s-u;t) \tag{2}$$

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 $p(s;t)$,

$$\tilde{k}(s,u;t) = \frac{k(s,u;t)}{p(s;t)p(u;t)} \tag{3}$$

$$p(s;t) = \int k(s,u;t)p(u)\mathrm{d}u. \tag{4}$$

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,

$$\tilde{S}(\alpha(s);t) = \int_{\Gamma} \tilde{k}(s,u;t)\alpha(u)\mathrm{d}u, \tag{5}$$

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

$$F(x;t) = \frac{2\|x-\tilde{S}(\alpha(s);t)\|}{t}e^{-\frac{2\|x-\tilde{S}(\alpha(s);t)\|}{t}} \tag{6}$$

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
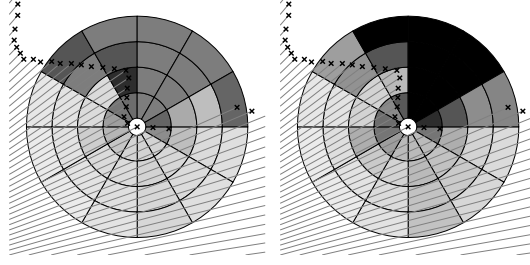
Fig. 3 The β-grid descriptor (center: occupancy probability, right: variance) for an example interest point in real data.

at a scale t correspond to places where t equals the inverse of the local curvature of the smoothed signal $\tilde{S}(\alpha(s);t)$.

Example detection results are shown in Figures 2 and 4.

2.2 β-Grid Descriptor

An important difference between image and range data is that range data encode metric distance but also directed free-space information between the sensor (emitting light or sound) and the measured object. Occupancy grids naturally deal with free-space information which is why we adopted this concept for our purpose of building meaningful descriptors for 2D range data. Concretely, for each detected interest point p_{det} we define a polar tessellation of the space around p_{det}[1]. 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 variance estimates over the occupancy values.

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 b follows a Bernoulli distribution, parametrized by the bin occupancy probability occ_j, where b 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,

$$p_{\beta}(occ_j; \alpha, \beta) = \frac{occ_j^{\alpha-1}(1-occ_j)^{\beta-1}}{\mathrm{B}(\alpha,\beta)} \tag{7}$$

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

---

[1] Note that the interest points define also a direction, given by curvature. This direction is used to obtain rotation invariance
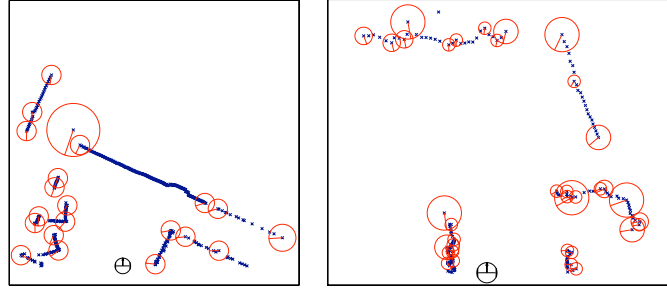
Fig. 4 Example FLIRT detection result in indoor (mit-csail, left) and outdoor (fr-clinic, right) data. The circles show the interest points and their support region (the actual descriptors are not shown).

descriptor). Over a sequence of measurements, that is, a sequence of beams $\{b_i\}_{i=1}^n$ that either hit or miss the bin, it can be shown that the update rules are

$$\alpha_i = \alpha_{i-1} + \sum 1^{b_i} \qquad \beta_i = \beta_{i-1} + \sum 1^{(1-b_i)}. \tag{8}$$

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

$$\widehat{occ}_j = \frac{\alpha}{\alpha + \beta} = \frac{\#\text{hits} + 1}{\#\text{hit} + \#\text{misses} + 2} \tag{9}$$

Accordingly, the variance of this probability is

$$\text{var}(occ_j) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}. \tag{10}$$

The collection of occupancy probabilities together with their variance estimates in the polar histogram make up the beta grid descriptor of p_{det} (see Fig. 3).

3 Applications

We will now explore the potential of FLIRT for robot navigation. We apply FLIRT features to global localization, loop closing, incremental mapping and pose-based SLAM and show how these navigation tasks can be addressed successfully using only off-the-shelf RANSAC and a graph optimizer. Dead reckoning is only needed in the exceptional case of featureless places (e.g. structureless infinite corridors), where the system can either fall back onto odometry or a constant velocity motion model. Apart from such cases, data association with FLIRT is entirely appearance-based and does not rely on pose estimates, known to be highly error prone.

Algorithm 1: FLIRT-SLAM

Input: The scan z_t and map $m_{t-1} = \left\{ (z, \mathscr{F})_0^{t-1}, \mathscr{G}_{t-1} \right\}$ with $\mathscr{G}$ being the pose network
Output: The map $m_t = \left\{ (z, \mathscr{F})_0^t, \mathscr{G}_t \right\}$

1 $\mathscr{F}_t \leftarrow \texttt{FLIRT}(z_t)$
2 $(x_t, \Sigma_{t,t-1}, \mathscr{C}) \leftarrow \texttt{RANSAC}(\mathscr{F}_t, \mathscr{F}_{t-1})$
3 Add node x_t to $\mathscr{G}_{t-1}$
4 **foreach** $(z, \mathscr{F})_i \in m$ **do**
5 $\quad$ $(x_t, \Sigma_{t,i}, \mathscr{C}) \leftarrow \texttt{RANSAC}(\mathscr{F}_t, \mathscr{F}_i)$
6 $\quad$ **if** $|\mathscr{C}| > n_I^{min}$ and $\texttt{Error}(x_t) < E_{max}$ **then**
7 $\quad\quad$ Add constraint $(x_t \ominus x_i, \Sigma_i)$ to $\mathscr{G}_{t-1}$
8 $\quad$ **end**
9 **end**
10 $\mathscr{G}_t \leftarrow \texttt{TORO}(\mathscr{G}_{t-1})$

3.1 Global Localization and Loop Closing

Global localization is the problem of estimating the robot's pose given a sensory observation and an a priori map. This is a 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. Assuming the map *m* to be a network of nodes that each hold the corresponding robot pose, a scan *z* and the corresponding FLIRT feature set $\mathscr{F}$, global localization can be easily solved in linear time using RANSAC.

The association of two scans by matching their feature sets $\mathscr{F}_i, \mathscr{F}_j$ is as follows. After feature detection, the first step is to compute a candidate correspondence set by matching the descriptor vectors. This matching is done using the symmetric χ^2 distance and a nearest neighbor strategy with a threshold of 0.4. With the candidate correspondence set as input, RANSAC returns the displacement between the feature sets, x_{ij}, the set of inliers $\mathscr{C}$, and the residual error *e*. The inlier probability is set to 0.3. If the number of inlier correspondences is above a threshold, n_I^{min}, the solution is considered a match. The displacement estimate is then refined in a least-squares sense using the full inlier set. See Fig. 1 for an example match.

In pose tracking, the procedure delivers a set of weighted samples, with positions obtained by RANSAC and weights as the inverse of the residual error. These samples approximate the pose distribution given the observation and the map, $p(x_t | z_t, m)$, which can be used in many Bayesian filtering schemes for localization.

3.2 SLAM and Incremental Mapping

FLIRT features allow for both a feature-based SLAM approach (like [1] using visual keypoints) or a pose-based SLAM approach (like [9] based on full scans or [6] using keypoints). With data association purely based on appearance, the pose-based approach appears to be a natural choice that is taken here. As SLAM back-end, we use the TORO graph optimizer from [9].

SLAM with FLIRT features can then be carried out by the compact Algorithm 1. At start, the algorithm is called with an empty map $m = \{\}$. It adds new nodes incrementally and, for each pose, looks for valid constraints in the map. No distinction has to be made between local constraints from incremental matches and global constraints from loop closures. A constraint is valid if the inlier set is large enough, $|\mathscr{C}| > n_I^{min}$, and the reprojection error from the full scans is below a threshold $\mathtt{Error}(x_t) < E_{max}$. The second condition is to further ensure geometrical correctness of a match. The algorithm produces locally dense pose graphs where nodes are interwoven up to the point where the overlap between scans becomes too small for a valid RANSAC match.

If an environments contains symmetries (places that look alike) or featureless areas, appearance-based data association is not enough and poses need to be considered. For the former case, an additional test on the consistency of the candidate constraints is performed. To do so, we propagate the estimated covariances over the shortest path between the two edges (found by Dijkstra's algorithm) and perform a validation test at a significance level of 0.95. Like in [20], covariances of a scan-to-scan match, Σ, are computed by propagating the errors from the interest points across the least squares displacement expressions. This adds another condition into line 6. The situation where not enough interest points are extracted and the system falls back onto odometry requires two more lines, a test and the query to odometry.

Algorithm 1 also applies to the problem of incremental mapping or scan matching. Using local interest points for 2D range data, incremental mapping has the same problem statement than visual odometry and can be solved with bundle adjustment techniques. In fact, Algorithm 1 implements a pose-based variant of bundle adjustment: after extraction of FLIRT features from the current scan, they can be matched to the previous n scans from a sliding window $\mathscr{W}$. Each successful RANSAC match with a scan within $\mathscr{W}$ produces a displacement that is added as a constraint into the pose graph. A simple change to Algorithm 1 implements this: the set from which scans are picked in line 4 is changed from m to $\mathscr{W}$. In this manner, Algorithm 1 implements both, SLAM and incremental mapping.
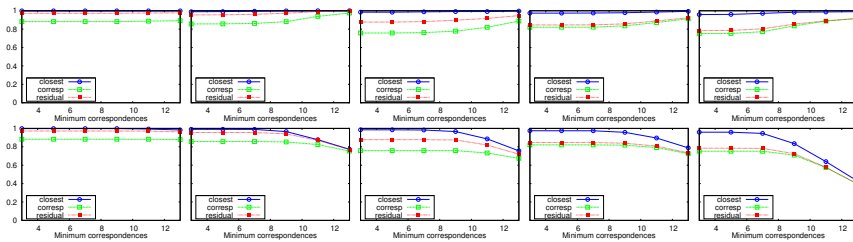
4 Experiments

The experiments have been carried out with five standard data sets, three indoor log files (fr-079, intel, mit-csail) and two outdoor log files (fr-clinic, Victoria park). These data sets have been collected by several researchers and are freely available (http://radish.sourceforge.net). For the experiments in dynamic environments, we collected additional data in the Freiburg main station.

4.1 Global Localization and Loop Closing

To perform the evaluation of FLIRT for the tasks of global localization and loop closing, we process all log files with the SLAM approach of Grisetti *et al.* [9] to

Fig. 5 Results from the global localization and loop closing experiments. The columns show the precision (top) and recall (bottom) values for the fr-079, intel, mit-csail, fr-clinic, and Victoria Park data sets, respectively. They demonstrate that FLIRT features are highly appropriate for global localization or loop detection.

obtain a ground truth map. The method typically produces maps within a precision of several centimeters. We then proceed as follows with each log file: for each scan, the scan is first removed from the map to exclude the trivial self-match and then matched to all other scans in the map by the RANSAC procedure. The set of all valid matches for a scan is said to be the *solution set* for that scan. If the solution set contains a single match, the robot is uniquely localized. In case of environment symmetries, multiple matches may occur which represent multiple pose hypotheses for global localization or loop closure candidates for SLAM. In practice, the solution set contains mostly several matches even in the unique localization case. They correspond to nearby poses that all pass the RANSAC test. The task is then to find the best match for which two strategies are considered, the highest number of inliers *(corresp)* or the lowest RANSAC error *(residual)*.

To investigate the ability of FLIRT features to perform data association we check if the match that corresponds to the true pose is in the solution set. A match is considered correct with respect to the ground truth, if the distance between its estimated pose and the ground truth pose is within 0.5 meter and 10 degrees. Fig. 5 shows 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}. We compare the two above mentioned strategies to find the best match plus a third strategy, called *closest*, that chooses the closest solution to the ground truth that passed n_I^{min}. This strategy is the theoretically optimal one and demonstrates that a correct match is in the solution set although it is not characterized by a maximum number of inliers or a minimal error.

As can be seen, the approach has both high precision and high recall values, even at small numbers of inliers. To further summarize the results, we define the prob-

Data set	**Size** [m]	**Scans**	$\bar{n}_{IP}$	p_{GL}^c	p_{LC}^c	p_{GL}^r	p_{LC}^r	$\mathbf{t}_{sm}$	$\mathbf{t}_{ss}$
fr079 (in)	50×20	1464	27	.98	.98	.98	.98	0.66s	450μs
intel (in)	50×40	2672	18	.98	.98	.96	.96	0.52s	200μs
csail (in)	80×60	1051	23	.97	.97	.89	.73	0.33s	320μs
fr-clinic (out)	550×300	6917	34	.96	.96	.85	.65	6.45s	930μs
Victoria (out)	500×500	5751	20	.97	.97	.79	.38	2.05s	357μs

Table 1 Summary of the global localization/loop closing experiment.
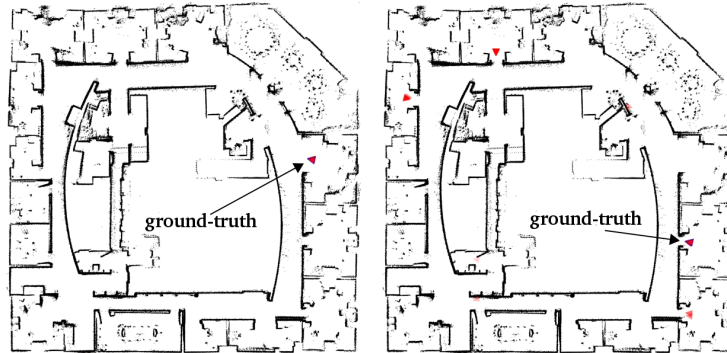
Fig. 6 Global localization result in the Intel data set. Unique localization (left), ambiguous localization with more than one hypotheses (right). Red triangles indicate pose hypotheses, the black triangle show the ground truth.

ability of correct global localization from a single scan p_{GL} and the probability of correct loop closure from a single scan p_{CL}. p_{GL} is defined to be the precision value at maximum recall which is the recall at the minimum number of inlier correspondences (three). p_{CL} is defined to be the recall value at a precision of 0.95. To demand a high precision is crucial for loop detection since incorrect loop closures can lead to severe map inconsistencies.

Table 1 summarizes the RANSAC results for all five data sets. The columns are environment size, number of scans in the map, the average number of interest points per scan detected $\bar{n}_{IP}$, the probabilities of correct global localization for the closest and the residual strategy, p_{GL}^c, p_{GL}^r, the probabilities of correct loop closure for both strategies, p_{CL}^c, p_{CL}^r, 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 a C++ implementation. All experiments have been conducted using the same set of parameters.

The global localization and the loop closure probabilities are shown separately for the strategies residual and closest. The probabilities of the closest strategy show the ability of FLIRT to produce a solution set that approximates the distribution over poses given observations and the map, $p(x_t|z_t, m)$. This is relevant in a Bayesian filtering schemes for pose tracking, for instance. With values higher than 0.96 across all data sets, we conclude that the solution sets produced by FLIRT are highly appropriate for this task.

The residual strategy is applied in a maximum likelihood estimation scheme where, given a map, the robot pose is sought from a single scan. In this case, FLIRT features enable an indoor robot to globally self-localize from a *single* scan with a success probability of at least 89% within 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 from a single scan with at least 73% probability. In the two outdoor environments these probabilities are smaller. However, even for the Victoria park data set, the numbers mean that, on average, we are still able to self-localize the robot every 1.3 scan and close a

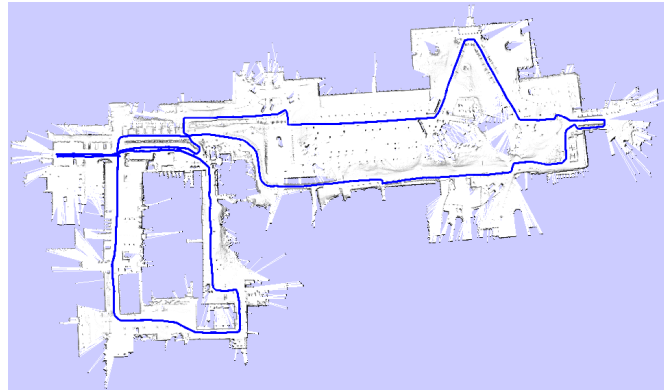**Fig. 7** SLAM experiment on the intel data set.

Fig. 8 SLAM experiment on the fr-clinic data set.

loop every 2.6 scan. Note also that the definition of p_{GL} and p_{CL} is conservative in the sense that true positives are only counted in the $0.5m/10°$-neighborhood of the ground truth pose. In case of environment symmetries causing multiple modes in the solution set, we only count matches in the mode of the ground truth which further decreasing these resulting probabilities.

4.2 Simultaneous Localization and Mapping

Given the ability to globally localize or detect loops, we can address the SLAM problem. We apply our SLAM algorithm onto two indoor data sets (intel and fr-079) and an outdoor data set (fr-clinic). We further evaluate the robustness of FLIRT with respect to environment dynamics using a data set collected at the Freiburg main station during rush hour. The experiments have been conducted using the same set of parameters for all data sets, where a number of inlier of 6 and a maximum average

error per inlier of $E_{max} = 0.05$ are taken to assess a valid RANSAC match. For speed reasons, we limit the number of constraints added at each time step to the best 15, in order to not overly strain the optimization backbone. The resulting maps are shown in Fig. 7 and 8. The high level of local connectivity in the pose graph illustrate the view point invariance property of FLIRT. Even pairs of scans with marginal overlaps from relatively far apart vantage points still produce valid constraints.

We repeated the experiments using the displacements from RANSAC matches as initial estimates for scan matching. Employing ICP for this purpose, the resulting indoor maps were even more accurate. However, in the outdoor data set, the use of ICP produced sometimes worse results at places with no structure, very noisy clutter, or 3D objects such as ramps. Then, RANSAC could not find valid matches, and without ICP, our SLAM procedure fell back onto odometry. With scan matching switched on, ICP attempted to find a match and failed, inserting incorrect constraints into the pose graph.

This brings up the role of FLIRT as a detector of structuredness: a low number of interest points is indication of a featureless place, a small inlier set at a high number of interest points is indication of random clutter or 3D structures. In such situations – that are straightforward to detect with a feature-based approach –, scan matching is likely to fail and should be ignored. These results have been obtained with the ICP method from [5]. Plain ICP has also been evaluated with worse results.

To further quantify the performance of FLIRT for the task of simultaneous localization and mapping, we compare our algorithm (Graph FLIRT) with the state-of-the-art in dense grid mapping (RBPF) [8] and graph-based SLAM [9] with constraints computed by the method from Olson [15] (Graph Olson). For the comparison, we employ the metric defined by Kümmerle *et al.* [10]. The metric is based on a set of relative relations between poses that are manually extracted from ground truth trajectories. These relations are compared with their corresponding relations from the trajectory estimated by the algorithm under consideration. For the fr-clinic data set, two types of relations are considered: local and global relations. The reason is to separately analyze the local and global consistency of the map.

The results, given in Table 2, show that our approach has similar performance than the approach in [15] and largely outperforms the RBPF approach. The similarity in performance with [15] is due to the fact that both methods use the same

Data set	RBPF (50 particles)	Graph Olson	Graph FLIRT
Translation error in m			
fr079	0.06 ± 0.04	0.06 ± 0.04	0.06 ± 0.09
intel	0.07 ± 0.08	0.03 ± 0.03	0.02 ± 0.02
fr-clinic (local)	0.64 ± 2.64	0.14 ± 0.18	0.18 ± 0.27
fr-clinic (global)	12.3 ± 11.7	11.6 ± 11.9	8.3 ± 8.6
Rotation error in degree			
fr079	0.6 ± 0.6	0.6 ± 0.6	0.8 ± 1.1
intel	3.0 ± 5.3	1.3 ± 4.7	0.3 ± 0.3
fr-clinic (local)	1.3 ± 2.3	0.9 ± 2.2	0.9 ± 2.0
fr-clinic (global)	5.5 ± 5.9	6.3 ± 6.2	5.0 ± 5.3

**Table 2** Residual errors of our approach and the state-of-the-art.
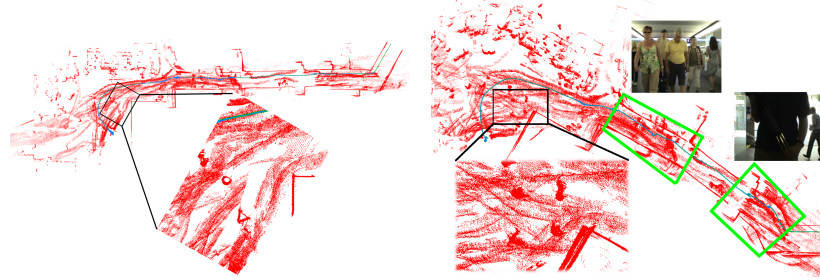
Fig. 9 Results from odometry-free incremental mapping in the underground tunnel of the Freiburg main station during rush hour (left: FLIRT, right: ICP). Despite the amount of dynamics, the FLIRT-based method produces accurate maps. Both enlarged areas show a convex corner, a round column and a triangular trash bin at the end of the exploration path. The two highlighted areas on the right figure visualize typical situations during the experiment.

optimization back-end (TORO from [9]) but differ only on the front-end, that is how the constraints are computed. Interestingly, our approach achieves better results for global relations which means that it can better deal with global data association which is generally harder than local data association where robot poses are approximately known. Note that our approach is simply based on FLIRT and RANSAC to generate constraints, whereas the method in [15] is much more complex.

In a final set of experiments, we evaluate the robustness of FLIRT for navigation in highly dynamic environments. To this end, we collected two data sets in the Freiburg main station during rush hour. A first log files in a narrow underground tunnel that connects the platforms, and a second log file around a loop through the main hall and an adjacent platform. The first data set is particularly challenging as the walls are feature-less and people can fill the entire width of the tunnel. Our setup did not provide any dead reckoning sensor such as odometry. We therefore test the ability of FLIRT to create maps from odometry-free incremental mapping. As described in section 3.2, this is implemented by Algorithm 1 with $\mathcal{W}$ as the sliding window from which scan are drawn for matching.

Fig. 9 and 10 show the results. Fig. 9 shows the results from the underground tunnel both using FLIRT (left) and the ICP algorithm from [5] (right). The robot (a manually pushed cart carrying a SICK LMS range finder) starts at the right end in Fig. 9 (left) and enters the tunnel to the left. The enlarged portions of the map in both subfigures shows the accuracy of the algorithms at the end of the trajectory. In particular, it shows a convex corner, a triangular trash bin and a circular column. The results show clearly that the FLIRT-based approach is much better in dealing with environment dynamics than ICP, the objects are accurately registered. People and other dynamic objects are shown by their blurred traces.

Fig. 10 shows the map from the second data set using both FLIRT (top) and ICP [5] (bottom). The map contains a big loop of around $500m$ length that goes from the main hall to the outdoor railway platform and reenters the building through a long hallway with shops and restaurants. As can be seen from the enlarged part of the images, FLIRT is able to create a clearly more precise incremental map.
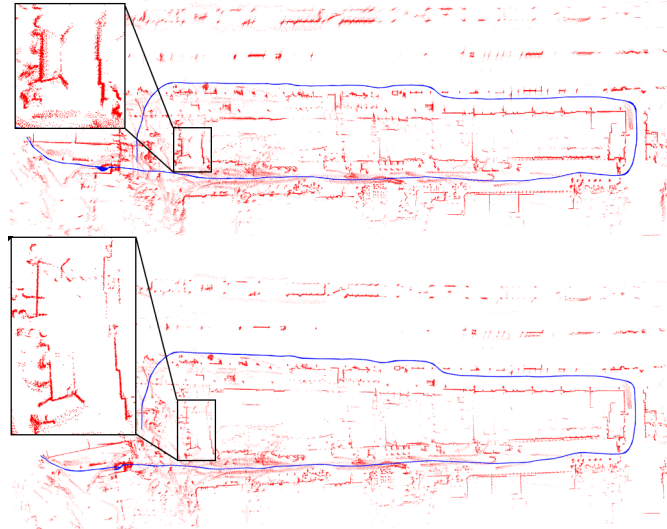
Fig. 10 Odometry-free incremental mapping (top: FLIRT, bottom: ICP). Results of the main hall data set at the Freiburg main station during rush hour. The enlarged areas show the registration error after completion of the 500m loop. Note that the map has been acquired incrementally, no loop closure is performed. The map learned with FLIRT features is largely consistent.

5 Conclusions

In this paper we introduced the Fast Laser Interest Region Transform, a multi-scale interest point operator for 2D range data. FLIRT features are generalized features that make no assumptions about the structure of world. We applied FLIRT to robot navigation tasks such as global localization, loop detection, incremental mapping and SLAM using five standard data sets plus newly collected log files in dynamic environments. We demonstrated that, to solve these tasks, not more than FLIRT, RANSAC and an open source graph optimizer is needed. Accordingly, the related algorithms become very compact.

From the experiments carried out in structured, unstructured, indoor, outdoor, static and highly dynamic environments, we conclude that FLIRT is able to robustly capture the invariant structures in the data. We achieved very high global localization and loop detection probabilities from a single scan. As data association scales linearly with map size, the method is also fast. The maps built with our FLIRT approach are consistent and accurate. Using a recently introduced SLAM metric, FLIRT maps are better or on par with the state-of-the-art while being algorithmically simpler. In a final set of odometry-free incremental mapping experiments using data from our main station during rush hour, we analyzed the ability of the approach to cope with highly dynamic environments. Compared to a recent ICP variant, FLIRT produced clearly more consistent and accurate maps.

Besides being a front-end for robot navigation that provides robust constraints, FLIRT opens up new ways to process 2D range data including the uniform treatment

of range and image data. This includes early fusion with vision, for instance, in voting-based detection schemes.

The data sets and the source code of FLIRT is available on the authors website at `http://srl.informatik.uni-freiburg.de/~tipaldi/FLIRTLib`.

Acknowledgments

References

1. Agrawal, M., Konolige, K.: FrameSLAM: From bundle adjustment to real-time visual mapping. IEEE Transactions on Robotics **24**(5) (2008)
2. Arras, K.O., Philippsen, R., Tomatis, N., de Battista, M., Schilt, M., Siegwart, R.: 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. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded-Up Robust Features. In: Proc. of the 9th European Conf. on Computer Vision (ECCV'06). Graz, Austria (2006)
4. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Trans. on Patt. Anal. and Mach. Intell. **24** (2002)
5. Bosse, M., Zlot, R.: Map matching and data association for large-scale two-dimensional laser scan-based slam. Int. J. on Robotics Rresearch **27** (2008)
6. Bosse, M., Zlot, R.: Keypoint design and evaluation for place recognition in 2d lidar maps. Robotics and Autonomous Systems **75** (2009)
7. David Cole Alastair Harrison, P.N.: Using naturally salient regions for slam with 3d laser data. In: Workshop on SLAM, IEEE Int. Conf. on Robotics and Automation. Barcelona (2005)
8. Grisetti, G., Stachniss, C., Burgard, W.: Improved techniques for grid mapping with rao-blackwellized particle filters. Transactions on Robotics **23**, 34–36 (2007)
9. Grisetti, G., Stachniss, C., Burgard, W.: Nonlinear constraint network optimization for efficient map learning. Trans. Intell. Transport. Sys. **10**(3), 428–439 (2009)
10. Kümmerle, R., Steder, B., Dornhege, C., Ruhnke, M., Grisetti, G., Stachniss, C., Kleiner, A.: On measuring the accuracy of slam algorithms. Autonomous Robots (2009)
11. Lowe, D.: Object recognition from local scale-invariant features. In: Proc. of the 7th IEEE Int. Conf. on Computer Vision (ICCV'99) (1999)
12. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Patt. Anal. Mach. Intell. **27** (2005)
13. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.V.: A comparison of affine region detectors. Int. J. Comp. Vis. **65** (2005)
14. Novatnack, J., Nishino, K.: Scale-dependent 3d geometric features. In: Proc. of Int. Conf. on Comp. Vis. (2007)
15. Olson, E.: Robust and efficient robotic mapping. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA (2008)
16. Pauly, M., Keiser, R., Gross, M.: Multi scale feature extraction on point sampled surfaces. In: Proc. of Conf. of European Association for Computer Graphics (2003)
17. Taubin, G.: A signal processing approach to fair surface design. In: Int. Conf. on Comp. Graph. and Interact. Techn. New York, NY (1995)
18. Thrun, S., Bennewitz, M., Burgard, W., Cremers, A., Dellaert, F., Fox, D., Haehnel, D., Rosenberg, C., Roy, N., Schulte, J., Schulz, D.: Minerva: A second generation mobile tour-guide robot. In: Proc. IEEE Int. Conf. on Robotics and Automation. Detroit, USA (1999)

19. Tipaldi, G.D., Arras, K.O.: Flirt – interest regions for 2d range data. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). Anchorage, USA (2010)
20. Tipaldi, G.D., Grisetti, G., Burgard, W.: Approximated covariance estimation in graphical approaches to slam. In: IEEE Int. Conf. on Intelligent Robots and Systems. San Diego (2007)
21. Unnikrishnan, R., Hebert, M.: Multi-scale interest regions from unorganized point clouds. In: Workshop on Search in 3D, IEEE Conf. on Comp. Vis. and Patt. Rec. (2008)