This is a preprint version of the article: Košnar, K., Vonásek, V., Kulich, M., and Přeučil, L. (2013). **Comparison of Shape Matching Techniques for Place Recognition**. *Proceedings of 6th European Conference on Mobile Robots*, which has been published in a final form at https://doi.org/10.1109/ECMR.2013.6698828.

Comparison of Shape Matching Techniques for Place Recognition

Karel Košnar, Vojtěch Vonásek, Miroslav Kulich and Libor Přeučil

***Abstract*— Place recognition is crucial for environment mapping as well as for localization. Although vision-based place recognition is well studied, where various feature-based methods can be employed, less number of works has been dedicated to laser-based place recognition. As data from laser range-finder can be represented as 2D shapes, various shape matching methods can be used to find similarities in these data. In this paper, we discuss the usage of shape matching methods for place recognition in a mapping task. Several state-of-the-art pattern recognition methods are compared on real as well as synthetic datasets. The experimental results show, that selected shape matching methods surpass the state-of-the-art robotic FLIRT feature-based algorithm in both the precision and speed.**

I. Introduction

Place recognition together with path integration are two basic principles used by humans, animals, and robots when solving navigation and mapping tasks. In robotics, the place recognition, namely the problem of decision whether a robot is revisiting an already known location or is visiting a novel one, has applications in pose initialization and localization using prior maps, closing large loops, simultaneous localization and mapping (SLAM), localization in topological maps or merging maps collected at different times or by multiple robots.

A typical utilization of place recognition arises in loop closing, where the task is to match actual sensory data with the known data stored in a map or database. When more than one match is found, hypotheses about actual position need to be built and maintained. To decrease the number of hypotheses, place recognition needs to be accurate with minimum false positives. As robots operate in a large area, the database can contain considerable amount of data, therefore, the matching has to be as fast as possible.

The majority of place recognition solutions presented in the literature focuses on vision-based techniques, which employ camera images and feature-based techniques to retrieve place signatures. The disadvantage of a camera-based perception is the inability to work under bad light conditions like in darkness or fog. The camera data needs to be preprocessed to obtain features, which increases the computational burden.

For place recognition, laser range-finders can be employed, which avoids above mentioned issues. As range finders are active sensors, they can work independently on light conditions even in total darkness. Range data processing is not computationally intensive and amount of the data to be processed and stored is low.

The authors are with Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, 166 27 Prague 6, Czech Republic {kosnar,vonasek,kulich,preucil}@labe.felk.cvut.cz

Relatively few approaches have been developed to solve place recognition problem in context of 2D range data. Although same feature-based techniques utilized in vision-based systems can be applied to 2D range data, they are less accurate, as a 2D range scan provides less information in comparison to a camera image. It is therefore hard to generate distinctive place signatures that can be applied to large datasets. As the laser data can be represented as polygons, plethora of methods known from shape matching and pattern recognition areas can be employed. Beside the matching of polygons from laser range-finder, the shape matching methods can be employed also to polygons extracted from omni-cameras. For example, using image segmentation, the polygon describing the ground can be retrieved and matched with other polygons in a database.

Many of shape matching methods have been designed to be invariant under rotation, scale, translation or affine transformation, while handling noise, occlusion or non-affine transformation are considered as additional features. In the place recognition problem, the distortion of shapes is given by placing a range-finder sensor at different places during taking 2D range scans of the same place. Therefore, the matching algorithms are highly required to cope with noise and occlusions. On the other hand, scale invariance is not desired here, as many places, especially in indoor, are different only in scale.

In this paper, the usage of state-of-the-art shape matching methods for place recognition is investigated. Several shape matching methods have been selected, implemented and their performance in place recognition has been tested on four datasets. This allows us to answer the question whether the shape matching algorithms can be used in the context of the robotic place recognition.

A. Related Work

There is a relatively little related work that is concerned with place recognition in robotic mapping based on 2D range data. Most of the approaches rely on the usage of visual information from camera images, see [1] for a survey.

Very popular are feature-based approaches, similar to approaches used in visual-based place recognition. Bosse and Zlot [2] evaluated several detectors and descriptors of features of 2D range data in the task of place recognition in the SLAM application. Tipaldi et al. [3] use the multi-scale interest point operator FLIRT for 2D range data and RANSAC algorithm to perform place recognition. In the later work [4], they propose geometrical FLIRT phrases, structures similar to bag of words but preserving also the geometrical

arrangement. The weak geometrical verification is used to achieve sub-linear time complexity.

Another approaches describe places with global properties of a range scan or a local map. Bosse and Zlot [5] describe local maps by an orientation histogram and a pair of projection histograms. Matching is done by cross-correlation between these histograms. Grandstom et al. [6] define a set of local properties of a scan (area, average range, circularity, etc.) and combine them using AdaBoost to build a nonlinear classifier.

On the other hand, there are many works related to shape matching problem in the context of pattern recognition and computer vision. Shape matching generally works in two steps: the description of the shape is generated for each of the compared shapes at first and these descriptors are compared consequently. Shape description techniques can be generally divided into two classes [7]: contour-based and region-based methods. While contour-based methods extract shape features from the shape boundary only, the region-based methods use the whole region of the shape to create the descriptor.

Each of these two classes can be divided into two subsequent subclasses: global and structural approaches. Contour-based global approaches derive the description from the integral boundary. This class includes techniques based on spectral transformation using Fourier [8] or wavelet [9] transformation, correspondence-based techniques [10] and scale-space method [11]. A metric distance between acquired feature vectors is usually used as a similarity measure.

Contour-based structural approaches split a boundary into segments called primitives. Methods differ in the selection of these primitives. Common methods are based on polygonal approximation [12], [13] or curvature decomposition [14].

Region-based global techniques take into account the area of the shape as a whole while creating shape descriptor. Descriptors are again often in the form of feature vectors while metric distance is used to compute similarity. Used techniques are geometric moment invariants [15], algebraic moment invariants [16] and orthogonal moments [17]. Scan-line [18] and ring-projection [19] methods use an intersection of geometrical primitives with a shape area as the descriptor. Grid based methods [20] are also studied.

Region-based structural methods decompose the shape region into parts which are used for the description. Methods like convex hull decomposition [21] or medial axis transformation [22] can be used. The graph of subregions is used as descriptor and shape similarity is then checked by graph matching. It is worth to mention that the graph matching problem is not an easy task itself.

II. Place Recognition

Place recognition is a problem of distinguishing different places from each other only by information gathered by sensors. As the sensor information is limited, two main problems arise: 1) Perceptual aliasing when two different places can be perceived as the same and 2) perceptual variability when the same place provides different sensor readings under different circumstances. The perceptual aliasing can be minimized by using as detailed place description as possible. On the other hand, the detailed description increases the perceptual variability.

The most popular sensors used in robotics are cameras and range-finder sensors. This paper focuses on range-finder sensors and mainly laser range-finders. We expect that sensor covers whole 360 degree field of view in the rest of the paper. As a sensor reading can be described by a polygon representing the place, the shape matching algorithms can be used to compare the place descriptor to decide whether the robot is located at an already known location or at a new one.

A. Shape matching methods

Various types of computer vision shape matching algorithms have been chosen trying to cover all types of them and not only the best methods are chosen for the purpose of the comparison. All the methods receive two polygons on the input and provide the measure of the polygon similarity on the output. The similarity measure is a real number in the range of $[0, 1]$ or $[0, \infty)$, where 0 means the total match.

Selected methods are outlined in the following lines to provide main ideas of each method. For detailed description, see the original papers.

1) Fourier Transformation: The range scan can be seen as a function in a polar coordinates system and this function can be described by the coefficients of the Fourier transformation [8] $X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k \frac{n}{N}} \quad k = 0, 1, \ldots, N-1$.

Only amplitudes of the Fourier coefficients are considered for the descriptor to assure a rotation invariance which permits a robot to scan the place descriptors regardless of its orientation.

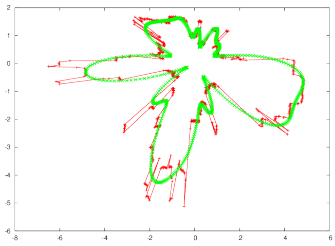


Fig. 1: The raw laser data (red) and data reconstructed using first $k = 30$ harmonic functions (green).

The lower order harmonics describe a coarse character of the laser data, the higher harmonics are more sensitive to noise. To minimize influence of noise in the data, the only first $k < N$ harmonic functions are considered. An example scan of a real environment and the scan reconstructed from its harmonic functions are depicted in Fig. 1.

The similarity of two descriptors $s(D_i, D_j)$ is computed as the Euclidean distance in k dimensional space.

2) Integral Invariant: The integral invariant method [23] relies on measurement of similarity between two curves that represents an integral invariant of the polygon. The discrete version of integer invariant $I(p)$ for a polygon P is defined

as $I(p) = \log \sum_{x \in P}(\|p - x\|)$ for every point $p \in P$, where $\|x - y\|$ is Euclidean distance in $\mathbb{R}^2$.

The descriptor is $D(P) = \{I(p_1), \ldots, I(p_n)\}, \quad n = |P|$, where n is the number of points in the polygon P.

The similarity is computed in two steps: The best correspondence between the points of the polygons is computed at first by minimizing the function $\sum_{\substack{k \in 1 \ldots M \\ l \in 1 \ldots N}}^{L} |I_1(k\Delta s_i) - I_2(l\Delta s_j)|^2 + \frac{(h_i + h_j)}{2} + \alpha \left| \frac{h_j - h_i}{h_i + h_j} \right|^2 \frac{h_i + h_j}{2}$, ,where (k, l) are corresponding points ,$I(k\Delta s)$ is an integer invariant in k-th point and h_i, h_j are determined by warping function $h = (h_i, h_j)$, which represents a point-wise correspondence between points of polygons (as defined in [23]).

The distance between two descriptors is defined as $s(D_i, D_j) = \sum_{k=1}^{n} |D_{iu} - D_{jv}|, \quad (u, v) = C(k)$, where $C(k)$ is k-th pair of corresponding points u and v. The integral invariant is more sensitive to global changes (the shape of the polygon) and less to changes imposed by noise in the depth data along the horizon.

3) Shape context: This approach is based on assignment of a shape context [10] to each polygon vertex, which describes a near neighborhood of the vertex in question. The shape context is defined as a two-dimensional histogram $H(p)$ of logarithmic polar distances from a particular vertex to other vertices in the polygon.

The distance between two shape contexts is given as a distance between two such histograms $C(p_i, p_j) = \frac{1}{2} \sum_{i=1}^{K} \frac{(H_i(k) - H_j(k))^2}{H_i(k) + H_j(k)}$, where K is number of bins in the histogram.

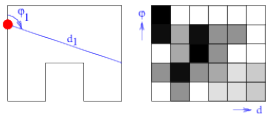


Fig. 2: Shape context of one point.

The similarity measure between two polygons can be then computed as follows: the shape context is computed for each vertex first and the shape distance between vertices of the two polygons are computed consequently. The correspondences between polygon vertices are resolved by application of the bipartite matching problem. The distance between two polygons is obtained as the sum of distances between resulting matched vertex pairs.

4) Tangent Space: Traditionally, the closed polygon can be represented as a list of vertices or by giving a list of line segments. Alternatively, a polygon can be represented using a tangent space [13] - a list of angle-length pairs, whereby the angle at a vertex is an accumulated tangent angle at this point while the length is the normalized accumulated length of polygon sides up to this point.

As the tangent space representation depends on the starting vertex, lets define a tangent space representation $t_p : [0, 1] \mapsto [0, 2\pi]$ to be a projection from a normalized length to an accumulated tangent angle starting from the point $p \in P$. Then the similarity of two polygons is $s(P, Q) =$

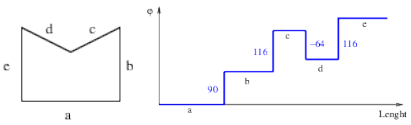


Fig. 3: Closed polygon and corresponding tangent space representation.

$min_{p \in P} \int_0^1 |t_q(x) - t_p(x)| dx$ a minimal difference between all possible variants of the tangent space representation.

5) Scan Line: The scan line matching algorithm [18] computes a shape descriptor from the intersection $l \cap P$ of randomly placed lines $l \in L$ with the polygon P. All the intersecting points $x_1, x_2, \ldots, x_2 n$ are ordered and form n compact intervals. The intersection function is defined as $S_{l \cap P}(\xi) = \sum_{k=2}^{2n} \sum_{i=1}^{k-1} (-1)^{i+k+1} I\{x_k - x_i > \xi\}$ for all interval lengths $\xi > 0$, and where $I\{x_k - x_i > \xi\} = 1$ when $\|x_k - x_l\| > \xi$ and 0 otherwise; If $x_k - x_i$ represents an interval strictly interior or exterior to the polygon, the sum is incremented. If the interval represents a collection of intervals both interior and exterior, the sum is decremented.

The descriptor of the polygon is then defined as $D_{SL}(\xi) = \frac{1}{N} \sum_L S_{l \cap P}(\xi)$ and the similarity of two polygons $s(D_i, D_j) = \sum_{\xi \in \Xi} |D_i(\xi) - D_j(\xi)|$, where the Ξ is a set of interval lengths used in the descriptor.

6) Ring Projection: The ring projection [19] algorithm computes the intersection of the polygon with the growing circle placed in the center of the polygon. Lets have a function $p(x, y) : R^2 \mapsto \{0, 1\}$ which returns 1 if the point (x, y) lies inside of the polygon P and 0 otherwise.

Without loss on generality, we assume, that the origin of the frame of reference is in the center of gravity of the polygon. Now, we express the function p in a polar coordinate frame as $p(x, y) = p(\gamma \cos\theta, \gamma \sin\theta)$ where $\gamma \in [0, \infty), \quad \theta \in [0, 2\pi)$. Formally, the ring projection function for the given γ is $f(\gamma) = \int_{\theta=0}^{2\pi} p(\gamma \cos\theta, \gamma \sin\theta) d\theta$. The implementation computes intersection points $\iota \in I$ of the circle with the diameter γ with the polygon, which are ordered according to the angle. Then the point in the middle of the arc $arc_M(x, y)$ between each intersecting points is checked if it lies inside the polygon and if so, the length of the arc $arc_L(x, y)$ is added to the sum. $f(\gamma) = \sum_{i=0}^{|I|} p(arc_M(\iota_i, \iota_j)) arc_L(\iota_i, \iota_j)$ where $j = (i+1) \bmod |I|$.

7) Multi-scale Shape Representation (MRM): The multi-scale shape representation [11] stores convexity/concavity at different scale levels for each point. Different scale levels for polygon P with n points are computed by the convolution $x_\sigma(u) = \int x(u) \phi_\sigma(t - u) du$ with the Gaussian kernel $\Phi_\sigma(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}}$. The convexity and concavity of the curve is measured as a displacement of the contour between two consecutive scale levels σ, which is measured as the Euclidean distance of the corresponding contour points from two consecutive scale levels. Distinction between convex and concave parts of the contour is via a sign — positive for convex and negative for concave parts.

Similarity is based on the cost of the optimal path D_{min} found by the dynamic programming in the matrix

of mutual points distances and normalized by complexity of the compared curves. The dynamic programming solves also the pairing problem. The distances of two points u_A, u_B is the sum of the Euclidean distances of their multi-scale representations $d(U_A, \sigma)$ and $d(U_B, \sigma)$ $D(u_A, u_B) = \frac{1}{K}\sum_{\sigma=1}^{K} |d_A(u_A, \sigma) - d_B(u_B, \sigma)|$ at different levels $\sigma = 1 \dots K$, where K is number of levels.

B. Feature-based method

Fast laser interest region transform (FLIRT) method [3], which is included in the OpenSLAM repository, has been chosen as a representative of feature-based methods. It is a multi-scale interest region operator for a 2D range data which combines the curvature-based detector with the β-grid descriptor. The curve $s(x)$ is mapped onto multi-scale parametrization $s_\sigma(x)$ using a Gaussian kernel k , similarly to MRM method. Interest points at the scale σ correspond to points, where σ equals the inverse of the local curvature of the smoothed signal $s_\sigma(x)$.

These interest points are described by β-grids. For each interest point a polar tessellation of its surrounding with a radius proportional to the scale of the interest point is constructed. For each bin j of the polar tessellation the occupancy probability occ_j and variance $var(occ_j)$ are computed. The collection of the occupancy probabilities together with their variance estimates form the β-grid of the given interest point.

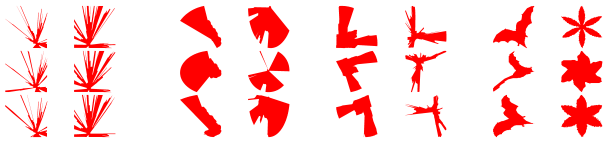
Place recognition using FLIRT is done by applying the RANSAC algorithm between two scans. The re-projection error is used as the measure. If the inlier set is not above a threshold n_I^{min}, the points in tge pair are declared as different. The FLIRTLib [24] implementation is used in this paper with default parameters.

III. Experiments

As the target of this paper is to evaluate performance of shape matching algorithms on the robotic problem of place recognition, three of the datasets are robotic. The fourth dataset is a MPEG7, which is a well-known benchmark dataset in the computer vision community. This dataset is included to provide base comparison of algorithms performance in their original field of applications. Each dataset contains a set of polygons divided into groups representing the same location. The examples from the datasets are depicted on Fig. 4 where various representations of the same place are in columns.

A. Datasets

As mentioned, the methods are compared on the four datasets. Two datasets are generated synthetically in the robotic simulator. The first synthetic dataset called Box is generated from the environment, where boxes of various orientation and size are placed randomly. Then, 11 places were chosen and for each place 21 different range scans with varying orientation and displacement were generated. This dataset simulates a cluttered unstructured environment, but with lot of significant and detectable points (like corners).



(a) Box (b) Surface (c) Robotic (d) MPEG7

Fig. 4: Example from datasets.

The second synthetic dataset called Surface is generated from the 3D undulated surface. The range scans are generated in the same manner as in the Box dataset. This dataset simulated another type of an unstructured environment without significant points.

The third dataset is collected from a real environment by a mobile robot equipped with two real laser range finders Sick LMS200, in configuration providing together full 360 degree range scan. The robot was placed to 15 different places in the building E of Czech Technical University and the robot took 12 different scans around each place. These scans are taken equidistantly on the circular trajectory with $0.6m$ diameter, which ensures various orientations of the scans.

The fourth dataset is a MPEG7 part B, which is commonly used for shape matching method comparison in pattern recognition and computer vision. MPEG7 dataset contains binary images of different shape silhouettes. The MPEG7 dataset contains 70 different shapes each in 20 variants. To get the same format as in other datasets, the images are converted to polygons and smoothed to remove redundant points of polygons.

B. Evaluation

To evaluate the performance of the methods, each polygon is compared to all others from the given dataset and each method is used to determine the similarity of the pair of polygons in question. The computation time and three different performance measures are computed: Bull's eye score, Receiver-operating characteristic and Precision-recall curve. Each of these measures focuses on a different aspect of the matching algorithms and together they give a good overview of their behavior.

1) Computation Time: The computation time is an important indicator in the robotic application. It is necessary to perform hundreds or thousands of comparisons to detect a loop closure in the mapping procedure or to globally localize a kidnapped robot.

The table I shows the computation time (i.e. mean time in milliseconds computed over all unique pairs of the robotic dataset) necessary to compare two places including creation of the descriptors and their comparison. The computation was performed on the computer with Intel Xeon processor running at 3.10GHz with 8 Gb RAM. Even the processor has 4 cores, only one core is used for each comparison.

2) Bull's eye score: The retrieval rate is measured by so called Bull's eye score. This score expects that the dataset D can be divided into a set of m disjunctive classes $C = \{c_1, \dots, c_m\}$; $\bigcup_{c \in C} c = D$; $\bigcap_{c \in C} c = \emptyset$ and each class

Method	Datasets				Time [ms]
	Box	Surface	Mpeg7	Robot	
Fft	78.54	57.43	65.40	72.22	0.52
Integral	83.24	50.38	48.21	51.99	1.60
Ring	**98.80**	52.34	34.84	81.20	4.46
Shapecontext	92.35	**70.95**	63.72	35.37	10.47
Scanline	67.59	50.48	41.42	38.98	70.79
MRM	97.20	60.52	**77.95**	**81.76**	87.63
Tangent	39.08	58.94	18.38	30.37	15.33
Flirt	97.84	31.99	16.12	43.01	191.23

TABLE I: Bull's eye score and processing time.

Method	Datasets			
	Box	Surface	Mpeg7	Robot
Fft	0.89	0.73	0.89	0.90
Integral	0.91	0.72	0.85	0.85
Ring	0.97	0.74	0.80	0.89
Shapecontext	0.93	**0.81**	0.85	0.68
Scanline	0.86	0.67	0.76	0.71
MRM	0.98	0.75	**0.91**	**0.93**
Tangent	0.65	0.77	0.60	0.65
Flirt	**0.98**	0.59	0.59	0.70

TABLE II: Area under ROC.

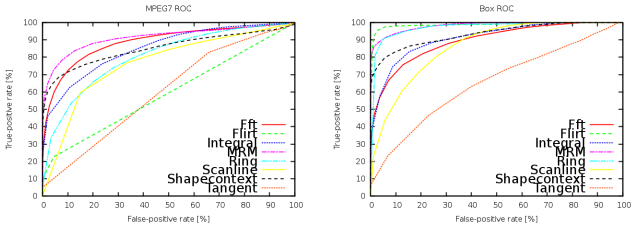
$\forall c \in C, |c| = n$ has the same number n of shapes. Every shape is compared to all other shapes and the $k = 2n$ best matches are considered. The number of true positive hits h_i (both shapes are from same class) from the k best matches is computed for each shape $i \in D$. The Bulls eye score B is then the ratio of the total number of shapes from the same class to the highest possible number $B = \frac{\sum_{i \in D} h_i}{|D|n}$. Thus, the best possible rate is 100%.

The results for the selected methods and datasets are summarized in the table I. The best score for each dataset is displayed in a bold font. The Bull's eye scores for MPEG7 dataset can differ from the values listed in other papers as the shapes are converted to polygons and smoothed. This procedure can influence performance of some methods.

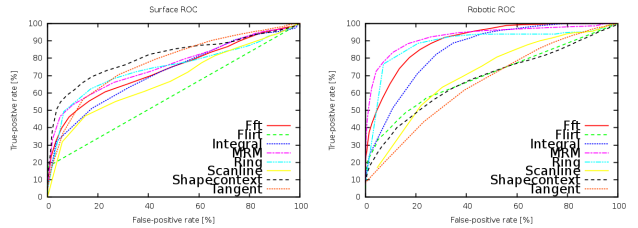
3) Receiver Operating Characteristic: Receiver operating characteristic [25] (ROC) is a graphical plot oftrue positive rate (called also recall or sensitivity), vs. false positive rate (also fall-out) for a binary classifier system as its discrimination threshold ϑ is varied. The true positive rate $TPR(\vartheta) = \frac{\sum_{x \in P} c_\vartheta(x) \geq \vartheta}{|P|}$ is a fraction of true positive hits of a classifier $c : \{P \cup N\} \mapsto \{0, 1\}$ out of all positive P cases in a dataset with a given threshold ϑ and false positive rate $FPR(\vartheta) = \frac{\sum_{x \in N} c_\vartheta(x) \geq \vartheta}{|N|}$ is the fraction of false positive hits (false alarms) out of all negative N cases. ROC curves for all the datasets are depicted on Fig. 5.

The quality of the classifier can be measured also by the area under ROC curve (AUC). The value 1 means an optimal classifier and value 0.5 is equal to random guess. Table II summarizes the values for all methods and all datasets. It can be seen that it differs from the Bull's eye score.

4) Precision-Recall: Precision-Recall graph (P-R) is a graphical plot of precision vs. recall or true positive rate as a threshold ϑ is varied. The recall is defined in same manner as

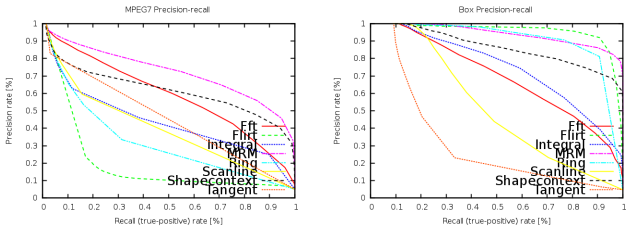


(a) MPEG7 (b) Box

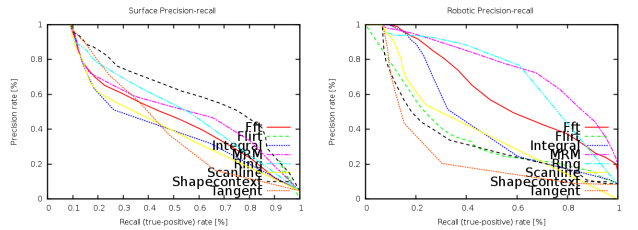


(c) Surface (d) Robotic

Fig. 5: ROC from the datasets.



(a) MPEG7 (b) Box



(c) Surface (d) Robotic

Fig. 6: Precision-Recall for all datasets.

for ROC curve. The precision $Pr(\vartheta) = \frac{\sum_{x \in P} c_\vartheta(x) \geq \vartheta}{\sum_{x \in P \cup N} c_\vartheta(x) \geq \vartheta}$ is a fraction of true positive hits out of all instances classified as true (true positive and false positive hits). The P-R graph is tightly correlated with the ROC curve but it is more suitable for the datasets, where the numbers of positive and negative instances are very different like all the here used.

F-Score combines the precision and the recall using a harmonic mean $F = 2\frac{precision \cdot recall}{precision + recall}$. The table III summarizes the balanced F-Score of the all methods and datasets.

The tables I, II and III show that there is no superior method that is best for all datasets. The ring projection method gets very low score on the MPEG7 dataset, and will not be chosen, if only the performance in shape matching field will be considered, but behaves very well on the robotic dataset and the Box dataset. The shape context method reaches a good score on all the synthetic datasets but the score is low on the real robotic data. On the other hand, the

Method	Datasets			
	Box	**Surface**	**Mpeg7**	**Robot**
Fft	0.62	0.49	0.57	0.54
Integral	0.66	0.43	0.42	0.40
Ring	0.86	0.53	0.32	0.68
Shapecontext	0.80	**0.61**	0.62	0.36
Scanline	0.47	0.43	0.25	0.33
MRM	0.89	0.54	**0.68**	**0.70**
Tangent	0.28	0.44	0.10	0.24
Flirt	**0.91**	0.35	0.20	0.37

TABLE III: F-Score

tangent space method performs bad on all datasets.

The MRM method is near to be the best matcher as it gets a good score in all cases but with the payoff of long processing time. If the speed is the main criterion, the FFT is the fastest method and the overall performance is over the average.

The FLIRT method is best for the synthetic box dataset where lot of significant points exists and therefore the FLIRT features are easy to compute. On the other hand, the performance is not better than the average for the other datasets, and it is the worst method on the MPEG7 dataset as expected, because it is not designed for arbitrary polygons.

IV. Conclusion

The results presented in this paper show that methods developed for shape matching are suitable for place recognition in the context of robotic mapping with 2D range data. The multi-scale representation method (MRM) achieves very good results for all the datasets but with payoff of high computation time. Even the very simple method as the ring projection, which has poor results on the 'standard' MPEG7 dataset, has very good performance on the both real and synthetic robotic datasets.

Half of the tested shape matching methods outperforms the state-of-the-art method based on the FLIRT features in both the speed and the performance. It is not a surprise, as the shape matching methods use similar approaches as feature-based methods but they compare all the points and not only a small subset.

A. Acknowledgment

This work has been supported by by the Czech Science Foundation under research project No. 13-30155P and by the Technology Agency of the Czech Republic under the project no. TE01020197 "Centre for Applied Cybernetics". The access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme "Projects of Large Infrastructure for Research, Development, and Innovations" (LM2010005) is highly appreciated.

References

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

[2] M. Bosse and R. Zlot, "Keypoint design and evaluation for place recognition in 2d lidar maps," *Robot. Auton. Syst.*, vol. 57, no. 12, pp. 1211–1224, Dec. 2009.

[3] G. D. Tipaldi, M. Braun, and K. O. Arras, "FLIRT: Interest regions for 2d range data with applications to robot navigation," in *Proceedings of the International Symposium on Experimental Robotics (ISER)*, Dehli, India, 2010.

[4] G. D. Tipaldi, L. Spinello, and W. Burgard, "Geometrical FLIRT phrases for large scale place recognition in 2d range data," in *IEEE International Conference on Robotics and Automation*, 2013.

[5] 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, no. 6, pp. 667–691, 2008.

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

[7] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognition*, pp. 1–19, 2004.

[8] H. Kauppinen, T. Seppänen, and M. Pietikäinen, "An experimental comparison of autoregressive and fourier-based descriptors in 2d shape classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 2, pp. 201–207, Feb. 1995.

[9] A. Q. Faouzi, F. A. Cheikh, and M. Gabbouj, "Wavelet-based multi-level object retrieval in contour images," in *Proc Very Low Bit Rate Video Coding (VLBV99) Workshop*, 1999.

[10] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.

[11] T. Adamek and N. O'Connor, "A multiscale representation method for nonrigid shapes with a single closed contour," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, no. 5, pp. 742–753, 2004.

[12] R. Mehrotra and J. E. Gary, "Similar-shape retrieval in shape data management," *Computer*, vol. 28, no. 9, pp. 57–62, Sept. 1995.

[13] P. Van Otterloo, *A Contour-Oriented Approach to Shape Analysis*. Prentice Hall International, 1991.

[14] S. Berretti, A. Del Bimbo, and P. Pala, "Retrieval by shape similarity with perceptual distance and effective indexing," *Trans. Multi.*, vol. 2, no. 4, pp. 225–239, Dec. 2000.

[15] S. X. Liao and M. Pawlak, "On image analysis by moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 3, pp. 254–266, 1996.

[16] G. Taubin and D. B. Cooper, "Geometric invariance in computer vision," J. L. Mundy and A. Zisserman, Eds. Cambridge, MA, USA: MIT Press, 1992, ch. Object recognition based on moment (or algebraic) invariants, pp. 375–397.

[17] S. Li, M.-C. Lee, and C.-M. Pun, "Complex zernike moments features for shape-based image retrieval," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 39, no. 1, pp. 227–237, 2009.

[18] M. Cannon and T. Warnock, "A shape descriptor based on the line scan transform," 2004.

[19] Y. Tang, B. Li, H. Ma, and J. Lin, "Ring-projection-wavelet-fractal signatures: a novel approach to feature extraction," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 45, no. 8, pp. 1130–1134, 1998.

[20] G. Lu and A. Sajjanhar, "Region-based shape representation and similarity measure suitable for content-based image retrieval," *Multimedia Syst.*, vol. 7, no. 2, pp. 165–174, Mar. 1999.

[21] J.-M. Lien and N. M. Amato, "Approximate convex decomposition of polygons," *Comput. Geom. Theory Appl.*, vol. 35, no. 1, pp. 100–123, Aug. 2006.

[22] H. Blum, "A Transformation for Extracting New Descriptors of Shape," in *Models for the Perception of Speech and Visual Form*, W. W. Dunn, Ed. Cambridge: MIT Press, 1967, pp. 362–380.

[23] S. Manay and B.-W. Hong, "Integral invariants for shape matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1602–1618, 2006.

[24] G. D. Tipaldi and K. O. Arras. (2010) Flirtlib. [Online]. Available: http://http://srl.informatik.uni-freiburg.de/ tipaldi/FLIRTLib/

[25] T. Fawcett, "An introduction to roc analysis," *Pattern Recogn. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.