Shape Matching for Robot Mapping

Diedrich Wolter[1] and Longin J. Latecki[2]

[1] University of Bremen, FB 3 — Cognitive Systems, 28359 Bremen, Germany
[2] Temple University, Dept. of Computer and Information Science, PA 19122
Philadelphia, USA

Abstract. We present a novel geometric model for robot mapping based on shape. Shape similarity measure and matching techniques originating from computer vision are specially redesigned for matching range scans. The fundamental geometric representation is a structural one, polygonal lines are ordered according to the cyclic order of visibility. This approach is an improvement of the underlying geometric models of today's SLAM implementations, where shape matching allows us to disregard pose estimations. The object-centered approach allows for compact representations that are well-suited to bridge the gap from metric information needed in path planning to more abstract, i.e. topological or qualitative spatial knowledge desired in complex navigational tasks.

1 Motivation

The problems of self-localization, i.e. localizing the robot within its internal map, and robot mapping, i.e. constructing the internal map autonomously, are of high importance to the field of mobile robotics [16]. Coping with unknown or changing environments requires to carry out both tasks simultaneously, therefore this has been termed the SLAM problem: Simultaneous Localization and Mapping [4]—it has received considerable attention [4, 6, 16]. Successful stochastical approaches have been developed that tackle representation and handling of uncertain data which is one key point in SLAM. As todays stochastical models are powerful, even linking them to a very simple geometric representation already yields impressive results. Advances in stochastical means have improved the overall performance leaving the basic spatial representation untouched. As the internal geometric representation is a foundation for these sophisticated stochastical techniques, shortcomings on the level of geometric representation affect the overall performance.

We claim that an improved geometric representation enhances the overall performance dramatically. A compact, object oriented representation based on shape is an universal yet slender one. It can outperform often-used occupancy grids in storage as well as in computational resources, since smaller sets of data need to be processed. Object-centered representations have been judged necessary to represent dynamic environments [16]. Moreover, a more comprehensive spatial representation can allow to mediate between different aspects of spatial information that are desired or even necessary in applications. We propose a shape representation of the robot's surrounding that grants access to metric information as needed in robot motion or path planning alongside with more abstract, qualitative or topological knowledge which is desired in navigational tasks and a well-suited foundation for communication.

2 Related Work

Any approach to master the SLAM problem can be decomposed into two aspects: handling of map features (extraction from sensor data and matching against the (partially) existing map) and handling of uncertainty.To address uncertainty mainly statistical techniques are used. Particle filters or the extended Kalman filter are used in most current SLAM algorithms [15, 16, 6]. As this paper focusses exclusively on the map's geometric representation, we now review related aspects in detail.

Typically, map features extracted from sensor data (esp. range finder data) are either the positions of special landmarks [4], simple geometric features like lines [10, 11, 3], or range finder data is used uninterpreted [16]. Uninterpreted use results in constructing a bitmap-like representation of the environment termed occupancy grid [5]. The simplicity of this approach causes its strength, namely universality: It may be used in unstructured, unprepared environments. However, major drawbacks also exist. First, matching a scan against the map in order to localize the robot is formulated as a minimization [10, 16, 6]. Therefore, a good estimation of the robot's position is required to prevent minimization getting stuck in local minima. Second, occupancy grids grow with the environment's size, not its complexity. As grids need to fine, it ends up in handling large data sets. This is not only a problem of storage, but, far more important, it affects run-time of algorithms as huge amounts of data need to be processed. To keep path planning in a once constructed map feasible, a topological representation can be coupled with the metric one [14].

To maintain a map at manageable size from the beginning, representations based on features or landmarks provide excellent means. These so-called object maps represent only positions of landmarks and their distinctive features. Thus, these maps grow with the environment's complexity (i.e. the number of visible landmarks), allowing for efficient processing. Using natural landmarks is of special interest as environments do not need to be prepared, like, e.g., by installing beacons [4]. For example, mapping based on line segments has been shown to improve performance in office environments [11]. A key point in feature-based approaches is a matching of perceived features against the ones represented in the map. Wrong matching result in incorrect, hence, useless maps; complex features help to prevent such mixups.As features' presence is required, application is often limited to special environments only. Choosing simple, omnipresent features can easily inhibit a reliable matching of perceived features against the map. Unreliable feature extraction, e.g. extracting line segments from round objects causes problems, too, as additional noise gets introduced.

To overcome these problems, we propose a representation based on shape features that is universal as shape can be extracted in any environment, but already individual features provide distinctive information as shape respects a wide spatial context. Matching of features is, thus, based on shape matching which has received much attention in the context of computer vision. The idea of applying shape matching in the context of robot mapping is not new. In the fundamental paper by Lu & Milios [10], scan matching has already been considered similar to model-based shape matching. Thrun considers this connection underexploited [16]. Recent advances in shape matching provide a good starting point to bring these fields together.

In the domain of robot mapping two key aspects dictate the applicability of shape descriptors: partial shape retrieval and the ability to deal with simple shapes. Firstly, as only partial observations of the environment can be made, any approach to shape representation that cannot handle partial shapes renders itself unemployable. This includes, for example, encoding by feature vectors like Fourier or momentum spectra. Secondly, any robot's working environment must be representable in the framework of the chosen shape descriptor. Besides these confinements, another feature is required: Much shape information perceivable often is rather poor, like for instance straightaway walls with small protrusions only. Therefore, shape recognition processes must be very distinctive, even on rather featureless shapes.

Structural approaches represent shape as a colored graph representing metric data alongside configurational information. Amongst these so-called skeleton based techniques, especially shock graphs (cp. [13]) are worth consideration[3]. Though primarily structural approaches may very well bridge from metric to more abstract qualitative or topological information (cp. [14]), recognizing shapes lacking of a rich structure of configuration has not yet proven feasible. Moreover, robust computation and matching of a skeleton in the presence of noise and occlusion has not yet been solved. Thus, we propose a boundary based approach. Considering the discrete structure provided by sensors, using polygonal lines to represent the boundaries of obstacles may be achieved easily. Related matching techniques rely on a so-called similarity measure. Various measures, often metrics, have been developed. Arkin et al. ([1]) accumulate differences in turning angle in straightforward manner; their approach fails to account for noise adequately. Basically all improvements employ a matching of boundaries to establish a correspondence prior to summing up dissimilarities of corresponding parts. Basri et al. propose a physically motivated deformation energy ([2]). More recently, an alignment-based deformation measure has been proposed by Sebastian et al. which considers the process of transforming one outline into another ([12]). However, common to these approaches is that an equal sampling rate of the outlines is required to ensure good correspondences of sampling points. Considering shape information obtained by a range sensor, scanning the same object from different positions, however, would generate this effect.

[3] Skeleton based approaches relate closely to Voronoi based spatial representations used in the field of robotics (cp. [14, 13]).

An improved performance in similarity measures for closed contours has been achieved by Latecki & Lakämper who consider a matching on basis of an a-priori decomposition into maximal arcs (cp. [8]). We will formulate the presented approach on this basis. However, it is tailored to deal with any kind of open polyline and addresses the problem of noisy data in a direct manner. The representation is complemented by a structural representation of robust ordering information. Applicability of the elementary shape similarity measure has been shown in [9].

3 Structural Shape Representation

Shape information is derived from sensor readings by a range sensor, typically a laser range finder (LRF). Shape is represented as a structure of boundaries. Polygonal lines, called *polylines*, serve as the basic entity. They represent obstacles' boundaries. Much of the spatial information represented in the map can be captured by individual polylines which form visual parts (cp. [8]). The variety of perceivable shapes in a regular indoor scenario already yields a more reliable matching than other feature-based approaches. At the same time, we are able to construct a compact representation. However, we exploit even more context information than represented by a single polyline considering shape as a structure of polylines. This allows us to cope with environments displaying mostly simple shapes with almost no extra effort. The structure captured is ordering information. For any given viewpoint, perceivable objects can be ordered in a counter-clockwise manner. A first step in the presented approach is to extract shape information from LRF data.

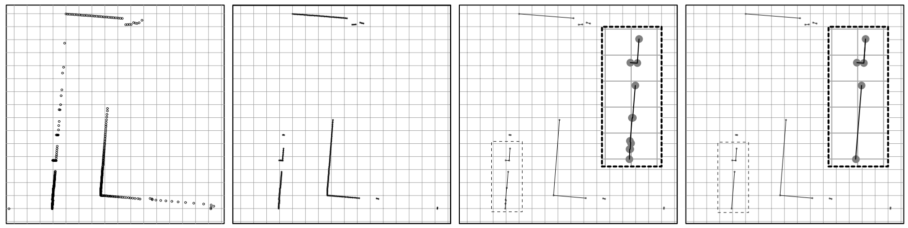
3.1 Grouping and Simplification of Polylines

Let us assume that the range data is mapped to locations of reflection points in the Euclidean plane, using a local coordinate system. Now, these points are segmented into individual polylines. For this grouping a simple heuristic may be employed: An object transition is said to be present wherever two consecutive points measured by the LRF are further apart than a given distance threshold. We used a threshold of 20cm in our experiments, however, the precise choice is not crucial and possible differences are regarded (cp. section 4.2).

Polylines extracted this way still carry all the information (and noise) retrieved by the sensor. To make the representation more compact and to cancel out noise, we employ a technique called Discrete Curve Evolution (DCE) introduced by Latecki & Lakämper ([7]) to make the data more compact without loosing valuable shape information and to cancel out noise. DCE is a context-sensitive process that proceeds iteratively: *Irrelevant* vertices get removed until no irrelevant ones remain. Though the process is context-sensitive, it is based on a local relevance measure for a vertex v and its two neighbor vertices u, w[4]:

$$K(u, v, w) = |d(u, v) + d(v, w) - d(u, w)| \tag{1}$$

[4] Context is respected as in the course of simplification the vertices' neighborhood changes.



(a) (b) (c) (d)

Fig. 1. Extracting polylines from a scan. Raw scan points (a) are grouped to polylines (b), then simplified by means of DCE. The threshold used in figure (c) is 1 and 5 in (d). The two additional rectangles show magnifications of marked parts. The grid denotes 1 meter distance.

Hereby, d denotes the Euclidean distance. The process of DCE is very simple and proceeds in a straightforward manner. The least relevant vertex is removed until least relevance exceeds a given simplification threshold. Consequently, as no relevance measure is assigned to end-points, they remain fixed. The choice of a specific simplification threshold is not crucial; refer to Figure 1 for results. Proceeding this way we obtain a cyclic ordered vector of polylines.

4 Matching Shapes

To match two shapes means to match two ordered set of polylines against each other. Hence, we need to seek the *best* correspondence of individual polylines that preserves the structure, i.e. that does not violate the order. Shape similarity is the key point to quantify quality of a correspondence.

4.1 Similarity of Polylines

The similarity measure utilized in our approach is based on a measure introduced by Latecki & Lakämper; we will briefly summarize the approach and indicate changes necessary in this context—for details refer to [8]. To compute the basic similarity measure between two polygonal curves, we establish the best correspondence of maximal left- or right-arcuated arcs[5]. To achieve this, we first decompose the polygonal curves into maximal subarcs which are likewise bent. Refer to Figure 2 (c) for an illustration. Since a simple 1-to-1 comparison of maximal arcs of two polylines is of little use, due to the fact that the curves may consist of a different number of such arcs and even similar shapes may have different small features, we allow for 1-to-1, 1-to-many, and many-to-1 correspondences. The main idea here is that on at least one of the contours we have a maximal arc that corresponds to a part of the other contour that is composed

[5] The original work is based on convex and concave arcs, respectively. As we deal with open polylines here, the terms convex or concave would be meaningless.

of adjacent maximal arcs. The best correspondence can be computed using Dynamic Programming, where the similarity of the corresponding visual parts is as defined below. The similarity induced from the optimal correspondence of polylines C and D will be denoted $S(C,D)$.

Basic similarity of arcs is defined in tangent space, a multi-valued step function representing angular directions and relative lengths of line-segments only. It was previously used in computer vision, in particular, in [1]. Denoting the mapping function by T, the similarity gets defined as follows:

$$S_a(C,D) = (1 + (l(C) - l(D))^2) \cdot \int_0^1 (T_C(s) - T_D(s) + \Theta_{C,D})^2 ds \qquad (2)$$

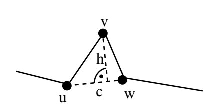
where $l(C)$ denotes the arc length of C. The constant $\Theta_{C,D}$ is chosen to minimize the integral (cp. [8]) (it respects for different orientation) and is given by

$$\Theta_{C,D} = \int_0^1 T_C(s) - T_D(s) ds. \qquad (3)$$

More appropriately, this measure should be denoted a dissimilarity measure as identical curves yield 0, the lowest possible measure. This measure differs from the original work in that it is affected by an absolute change of size rather than by a relative one. It should be noted that this measure is based on shape information only, neither the arcs' position nor their orientation are considered. This is possible due to the wide context information of polylines.

When comparing polylines, the amount of noise and the size of shape features present are often challenging. Applying DCE to a degree that would certainly remove all noise would remove many valuable shape features as well. DCE makes vertex removal decisions in the context of a single object. A better noise identification can be made in the context of comparing corresponding polylines. We encapsulate the basic similarity measure S in another process that masks out noise in the context of corresponding polylines. It is similar to the initial curve evolution employed. When comparing two polylines C and D, we evolve each line by removing vertices if the similarity improves. Obviously, a counter weight is needed to prevent elimination of all differing shape features. This counter weight, a cost for removing a vertex from a polyline is defined on the basis of a noise model of the LRF. Vertices whose removal only results in a small contour shift can likely be caused by noise and may be removed with low cost, whereas bigger changes are inhibited by high costs. The cost function R for removing a set of vertices (respectively r for removing a single vertex v with neighbors u and w) from a polyline P is defined on the basis of area difference:

$$R_P(\{v_1,\dots,v_n\}) := \sum_{i=1}^{n} r_{P\setminus\{v_1,\dots,v_{i-1}\}} v_i, \quad r_Q(v) := \left(\frac{h}{c}\right)^2$$



The similarity measure $S^\star$ is defined on the basis of the basic similarity S considering the optimal set of vertices to mask out.

$$S^\star(C,D) := \min_{C^\star \subseteq C, D^\star \subseteq D} \{S(C \setminus C^\star, D \setminus D^\star) + R_C(C^\star) + R_D(D^\star)\} \qquad (4)$$

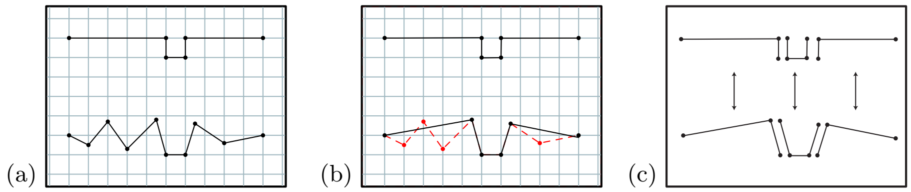


Fig. 2. (a) Two polylines from sensing an example scene with a simulated laser range finder. The upper polyline is free of noise, the lower one suffers from distortions of the magnitude of the shape features present. Using similarity measure $S^{\star}$, noise can be masked out when comparing the objects. Only the subsets shown in (b) are effective in the comparison; the determined similarity is enhanced by a factor of more than 10. Decomposition into maximal arcs and determined correspondence are shown in (c).

Computation is formulated as a greedy algorithm[6]. A prerequisite here is to use a hughly distinctive basic similarity measure. An example is depicted in Figure 2. When comparing the two polylines shown in Figure 2 (a), vertices are removed if the removal cost is lower than the gain in shape similarity (i.e. the decrease of S). This results in removing small distortions from the lower polyline, while retaining the features of both (cp. Figure 2 (b)).

4.2 Matching Polylines

The actual matching of two structural shape representations extracted from different scans is computed by finding the *best* correspondence of polylines which respects the cyclic order. Shape similarity is the key to measuring the quality of a matching. Additionally, we must take into account that (a) not all polylines may match as features' visibility changes and (b) that due to grouping differences (cp. section 3.1) not necessarily 1-to-1 correspondences exist. Noise or change of view point, for example, may lead to a different grouping. Moreover, since every correspondence of polylines induces an alignment that would align both scans involved, we demand all alignments induced to be very similar. This criterion is helpful to correctly match featureless shapes, e.g. short segments as obtained when scanning a chairs' legs. The clue in our approach is the exploitation of the correspondence of salient visual parts to correctly identify featureless parts even if no a-priori alignment is available. An estimation of the alignment is necessary to utilize an efficient matching algorithm. We will show (in Section 4.3) how to compute an estimate using shape similarity. Clearly, it can be derived from odometry if odometry data is available. Let us now assume that such an estimate exists. Further, let $\boldsymbol{B} = (B_1, B_2, \ldots, B_b)$ and $\boldsymbol{B'} = (B'_1, B'_2, \ldots, B'_{b'})$ be two cyclic ordered vectors of polylines. Denoting correspondence of B_i and B'_j[7] by

[6] Computing the true minimum may lead to combinatorial explosion, the greedy implementation avoids this problem and yields similar results.

[7] To be more precise: correspondences of either B_i and $\{B'_j, B'_{j+1}, \ldots, B'_{j'}\}$ or $\{B_i, B'_{i+1}, \ldots, B'_{i'}\}$ and B'_j since we consider correspondences of types 1-to-many and many-to-1, too.

the relation $\sim$, the task can be formulated as minimization.

$$\sum_{(\boldsymbol{B}_i,\boldsymbol{B'}_j)\in\sim} (S^\star(\boldsymbol{B}_i,\boldsymbol{B'}_j)+D(\boldsymbol{B}_i,\boldsymbol{B'}_j)) + \sum_{B\in\tilde{B}} P(B) + \sum_{B'\in\tilde{B}'} P(B') \overset{!}{=} \min \quad (5)$$

Hereby, $\tilde{B}$ (rsp. $\tilde{B}'$) denotes the set of unmatched polylines. P is a penalty function for not matching a polyline. This is necessary, as not establishing any correspondences would yield the lowest possible value 0 suggesting maximum similarity. The penalty function is chosen to linearly grow with the polyline's size modeling a higher likelihood for smaller polylines to appear or disappear[8]. D denotes the aforementioned alignment measure quantifying the deviation of the estimated alignment from the one induced by the correspondence $\boldsymbol{B}_i \sim \boldsymbol{B}'_j$. The best correspondence can so be computed by applying an extended Dynamic Programming scheme. The extension regards the ability to detect 1-to-many and many-to-1 correspondences and results in a linear extra effort such that the overall complexity is $O(n^3)$. The basic idea here is to consider in each step of the computation if it is advantageous to establish a grouping with the latest correspondence determined so far, if the summed up (dis-)similarity values and skipping penalties can be decreased.

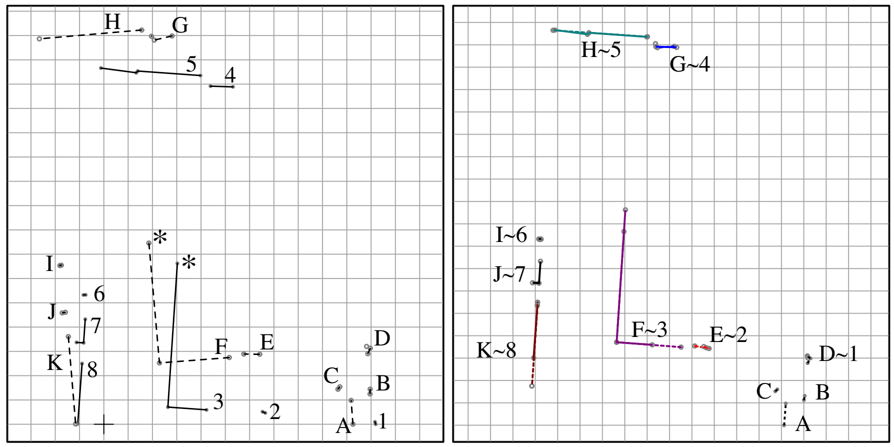
4.3 Matching in the Absence of Odometry

The outlined matching is capable of tracking *complex* shapes even if no estimate of the induced alignment is available.We will detail now how to obtain an alignment estimate purely by shape similarity. If we had two corresponding polylines, hence, the induced alignment, we could use this as the estimation in the matching. Observing that many shapes can be matched only in consideration of shape similarity, the matching can be employed to obtain this correspondence. Thus, the matching can be computed in a two pass process. Within the first matching pass the consideration of induced alignments' similarity is ineffective. Then, the *most reliable* correspondence is selected. Finally, the actual matching is computed using the alignment induced by the selected matching. To quantify reliability, a measure based on shape similarity and shape complexity has been proposed [9]. A polyline's shape complexity may be expressed by summing up inner points' relevance measures (cp. equation 1). If a polyline has no inner points, complexity is given by half its length. Terming this complexity measure C, the reliability is defined as

$$Q(P,Q) = C(P) + C(Q) - S^\star(P,Q). \quad (6)$$

The idea is to express reliability as high similarity of complex shapes (cp. [9] for details). An exemplary result is presented in Figure 3 where two scans are matched against each other only concerning shape (a). Based on the most reliable correspondence the estimated alignment is computed. Accordingly aligned

[8] When comparing polylines affected by similar noise, similarity values grow linearly with the polylines' size, too.



(a) (b)

Fig. 3. The two scans depicted in (a) (numbered 1–8 and A–K) are matched only wrt. shape, the most reliable match (marked $\star$) is selected. The induced alignment helps to determine the final matching. The correspondences found and the two scans aligned according to the estimation are shown in (b). Observe that the scans' origins are farer apart than 1m (grid denotes 1m distance) and no odometry has been used.

scans and the matching is shown in (b). The presented technique can cope with differences in the scans' position of more than 1m without the help of any means of estimating the robot's current position. Observe, that this is a dramatical improvement compared to the precision required by standard scan matching approaches which typically rely on a hill climbing strategy [6].

5 Conclusion and Outlook

We have presented a comprehensive geometric model for robot mapping based on shape information. Shape matching has been tailored to the domain of scan matching. The matching is powerful enough to disregard pose information and cope with significantly differing scans. This improves performance of today's scan matching approaches dramatically. Based on the presented shape processing, we plan to propose a complete robot mapping architecture. This is the topic of a forthcoming paper. We believe mapping based on shape to be particularly promising. For example, shape matching can also be exploited to map alignment. Equation 3 already provides the rotational difference. We are aware that statistical methods are needed to guarantee robust performance, but did not include any as we concentrated on geometric models exclusively. So, future work comprises also the coupling with a state-of-the-art stochastical model besides attacking the problem of cycle detection.

Acknowledgment

This work was carried out in the framework of the SFB/TR 8 Spatial Cognition, project R3 [Q-Shape]. Financial support by the Deutsche Forschungsgemeinschaft is gratefully acknowledged. Additionally, this work was supported in part by the National Science Foundation under grant INT-0331786 and the grant 16 1811 705 from Temple University Office of the Vice President for Research and Graduate Studies. Thomas Röfer is acknowledged for providing scan data.

References

1. M. Arkin, L. Chew, D. Huttenlocher, K. Kedem, and J. S. B. Mitchell. An efficiently computable metric for comparing polygonal shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, 1991.
2. R. Basri, L. Costa, D. Geiger, and D. Jacobs. Determining the similarity of deformable shapes. *Vision Research*, 38, 1998.
3. I. J. Cox. Blanche: Position estimation for an autonomous robot vehicle. In I. J. Cox and G. Wilfong, editors, *Autonomous Robot Vehicles*, pages 221–228. Springer-Verlag, 1990.
4. G. Dissanayake, P. Newman, S. Clark, H. Durrant-Whyte, and M. Csorba. A solution to the simultaneous localization and map building (SLAM) problem. *IEEE Transactions of Robotics and Automation*, 2001.
5. A. Elfes. *Occupancy Grids: A Probabilistic Framework for Robot Perception and Navigation*. PhD thesis, Department of Electrical and Computer Engineering, Carnegie Mellon University, 1989.
6. D. Hähnel, D. Schulz, and W. Burgard. Map building with mobile robots in populated environments. In *Proceedings of International Conference on Intelligent Robots and Systems (IROS'02)*, 2002.
7. L. J. Latecki and R. Lakämper. Convexity rule for shape decomposition based on discrete contour evolution. *Computer Vision and Image Understanding*, 73, 1999.
8. L. J. Latecki and R. Lakämper. Shape similarity measure based on correspondence of visual parts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(10), 2000.
9. L. J. Latecki, R. Lakämper, and D. Wolter. Shape similarity and visual parts. In *Proceedings of the 11th International Conference on Disrecte Geometry for Computer Imagery (DGCI), Naples, Italy*, November 2003.
10. F. Lu and E. Milios. Robot pose estimation in unknown environments by matching 2D range scans. *Journal of Intelligent and Robotic Systems*, 1997.
11. T. Röfer. Using histogram correlation to create consistent laser scan maps. In *Proceedings of the IEEE International Conference on Robotics Systems (IROS-2002)*, 2002.
12. T. B. Sebastian, P. N. Klein, and B. B. Kimia. On aligning curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(1):116–125, 2003.
13. K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, and S. W. Zucker. Shock graphs and shape matching. *International Journal of Computer Vision*, 35(1):13–32, 1999.
14. S. Thrun. Learning metric-topological maps for indoor mobile robot navigation. *Artificial Intelligence*, 99(1):21–71, 1998.
15. S. Thrun. Probabilistic algorithms in robotics. *AI Magazine*, 21(4):93–109, 2000.
16. S. Thrun. Robotic mapping: A survey. In G. Lakemeyer and B. Nebel, editors, *Exploring Artificial Intelligence in the New Millenium*. Morgan Kaufmann, 2002.