Theoretical Computer Science 442 (2012) 2–12

Contents lists available at SciVerse ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Shape matching by random sampling☆

Helmut Alt, Ludmila Scharf*

Institute of Computer Science, Freie Universität Berlin, Germany

ARTICLE INFO

Keywords:
Geometric shape matching
Probabilistic algorithms

ABSTRACT

In order to determine the similarity between two planar shapes, which is an important problem in computer vision and pattern recognition, it is necessary to first match the two shapes as well as possible. As sets of allowed transformation to match shapes we consider translations, rigid motions, and similarities. We present a generic probabilistic algorithm based on random sampling for matching shapes which are modelled by sets of curves. The algorithm is applicable to the three considered classes of transformations. We analyze which similarity measure is optimized by the algorithm and give rigorous bounds on the number of samples necessary to get a prespecified approximation to the optimal match within a prespecified probability.

1. Introduction

Matching two geometric shapes under certain transformations and evaluating their similarity is one of the central problems in computer vision systems where the evaluation of the resemblance of two images is based on their geometric shape and not color or texture. Because of its significance the problem has been widely covered in the literature; see [4,16] for surveys.

We assume that shapes are modelled by sets of plane curves. As possible classes of transformations we will consider *translations*, *rigid motions* (rotation and translation) and *similarities* (rotation, scaling, and translation). Our objective is to develop an algorithm which allows an efficient implementation and whose result comes close to human similarity perception.

Several similarity measures and algorithms are known for matching two curves, especially polygonal curves. One of the most widely investigated similarity measures is the Hausdorff distance which is defined for any two compact sets A and B. Alt et al. describe in [2,4] efficient algorithms for computing the Hausdorff distance and minimizing it under translations and rigid motions for arbitrary sets of line segments. One of the drawbacks of the Hausdorff distance is that it is very sensitive to noise. A few similarity measures are defined for pairs of curves, which capture the relative course of two curves: Fréchet distance [3], turning function distance [8], and dynamic time warping distance [13]. There are few generalizations of those distances to sets of curves: In [5] a generalization of the Fréchet distance to geometric graphs is given, and in [22] Tanase et al. describe an algorithm for matching a set of polygonal curves to a single polygon. A similarity measure which is designed for sets of curves is the reflection visibility distance [15]. The reflection visibility distance is robust against different kinds of disturbances but is expensive to compute.

The method we describe and analyze in this paper is close to an intuitive notion of "matching", i.e., it finds one or more candidates for the best transformations, that when applied to the shape B map the most similar parts of the two shapes to each other. The major idea is to take random samples of points from both shapes and give a "vote" for that transformation

☆ This research was supported by the European Union under contract No. IST-511572-2, Project PROFI, and by the Priority Programme 1307 "Algorithm Engineering" of Deutsche Forschungsgemeinschaft (DFG).

* Corresponding author. Tel.: +49 30 838 75 166; fax: +49 30 838 75 166.
E-mail addresses: alt@mi.fu-berlin.de (H. Alt), scharf@mi.fu-berlin.de (L. Scharf).

0304-3975/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2010.03.023

(translation, rigid motion, or similarity) matching one sample with the other. If that experiment is repeated frequently, the distribution of votes in the transformation space allows us to approximate a certain probability function underlying the random experiment. Maxima of this function indicate which transformations give the best match between the two shapes. The matching step of our algorithm is, therefore, a voting scheme. The idea of random sampling for geometric problems with an analysis similar to ours has been used in a more general context by Cheong et al. in [11] and a similar random sampling method for symmetry detection in 3D shapes with a different clustering method was described by Mitra et al. in [17].

Related methods in the image processing community are the generalized Hough transform, also called pose clustering [1,21], the Radon transform [23] and the RANSAC algorithm [14].

In contrast to those methods we do not consider a discrete set of features that describe shapes, but work with continuous curves. Our method is independent of the choice of the parameterization and the discretization grid in the transformation space.

The algorithm itself is a quite straightforward heuristic. The challenge lies in its exact analysis. On the one hand it is not obvious what is really being computed by the algorithm, i.e., what distance or similarity function between shapes is optimized. We first clarify this question. In addition, we give rigorous bounds on the runtime (number of experiments) necessary to obtain the optimal match within a certain approximation factor with a prespecified probability. We consider these insights as the major contribution of this paper, the analysis leads to a better understanding of this kind of heuristic.

In fact, our algorithm is not meant to be directly applied to shape comparison problems arising in practice. For practical purposes it makes sense to modify our technique and enhance it with heuristic methods, which we did (see [6]) within a shape retrieval system developed in the EU-funded project PROFI. Its major application, in cooperation with the industrial PROFI-partner Thomson–Compumark in Antwerp, is to identify potentially illegal similarities between new *trademark* designs and existing trademarks of various companies in a large trademark database.

2. The probabilistic algorithm

We assume that *shapes* are modelled by finite sets of rectifiable curves in the plane, and that for each curve a random point under uniform distribution with respect to curve length can be generated in constant time. This is the case for line segments, which would be the most common representation in practice, but also for curves for which the natural parameterization, i.e., their parameterization by arc length, is explicitly given.

Given two shapes $A, B \subset \mathbb{R}^2$, a class of allowed transformations $\mathcal{T}$ (the "transformation space") and a tolerance parameter δ, we want to find a transformation $t \in \mathcal{T}$ which lets the transformed image of B, $t(B)$, "*match best*" A within a tolerance of δ. The exact definition of the quality measure by which a "best match" is defined will be given in Section 3. We follow an intuitive notion: two shapes are similar if they can be mapped to each other in such a way that large parts of them are close. We assume that the underlying metric in the plane (in the image space) is a piecewise algebraic function, e.g., an L_p metric. Commonly used distance measures are Euclidean distance (L_2), Manhattan distance (L_1), or maximum distance (L_∞). A *δ-neighborhood* of a point p is defined as $U_\delta(p) = \{x \in \mathbb{R}^2 \mid dist(x, p) \leq \delta\}$, where $dist(x, p)$ is the distance with respect to the chosen metric.

The idea of the *probabilistic approach* is quite simple. We first describe an algorithm for matching under translations:

1. Take a random point a from the shape A and a random point b from B and give one "vote" to the translation t which maps b to a, that is $t = a - b$.
2. Repeat this random experiment many times.
3. For a prespecified neighborhood size δ return the points of $\mathcal{T}$ with the highest number of votes in their δ-neighborhood as candidates for good transformations.

The idea behind this algorithm is that the transformations that map large parts of shapes to each other should have significantly more votes in their δ-neighborhood than others. The size of the δ-neighborhood influences the quality of the match.

To extend this algorithm for matching under more complex transformations, we need a way to draw sample pairs and interpret them as votes for transformations. Obviously, there are infinitely many rigid motions and similarities that map one point to another. Thus, a pair of points does not give a vote for a unique rigid motion, or a unique similarity map. Therefore, *random samples* S_A, S_B of the shapes A and B, that together make a sample pair in the first step of the algorithm, should contain more than one point. The size of a random sample and the type of its elements, i.e., the number of points, or possibly a point and a direction vector, depends on the class of transformations allowed as described below. Further, we denote by $S_B \xrightarrow[\delta]{t} S_A$ the fact that the transformation t maps every element of S_B into the δ-neighborhood of the corresponding element of S_A. A "vote" generated by a pair of random samples S_A and S_B is called a *δ-region* in the transformation space, which is defined as the set of transformations t such that $S_B \xrightarrow[\delta]{t} S_A$. For transformation classes other than translations the shape of a δ-region depends on the sample pair generating it.

Before giving a generic variant of the algorithm we briefly describe the random samples and δ-regions for the basic transformation classes:

For **translations** we use a sample pair (a, b), with points $a \in A$ and $b \in B$ chosen randomly, in one random experiment to determine uniquely a translation mapping one point to the other. Thus, a random sample consists of a single randomly

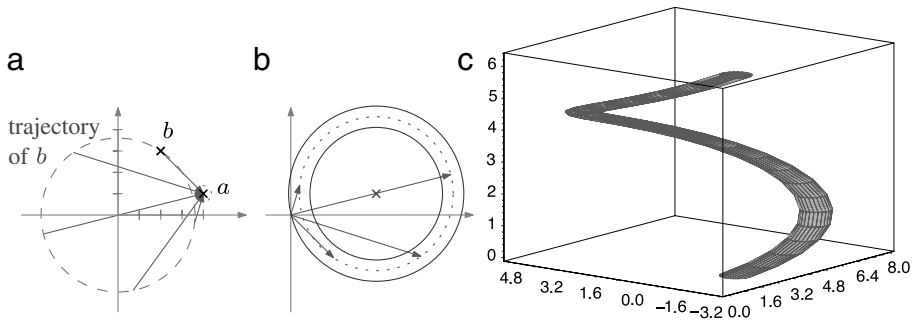


Fig. 1. δ-region in the space of rigid motions corresponding to a pair of points $a \in A$ and $b \in B$. (a) Points a and b with translation vectors corresponding to some rotated positions of b. (b) Projection of the δ-region to the translation plane for $\delta = 1$. (c) δ-region in the 3-dimensional space of rigid motions.

selected point of each shape, $S_A = a \in A$ and $S_B = b \in B$. The transformation space is two-dimensional and a δ-region in translation space corresponding to the sample pair (a, b) is a δ-neighborhood of the translation vector $t = a - b$ with respect to the same metric as used for points in image space.

A **rigid motion** $t = (\alpha, v_x, v_y)$ is defined by a rotation angle α and a translation $v = (v_x, v_y)$ and maps a point $b \in \mathbb{R}^2$ to the point $t(b) = Mb + v$, where

$$M = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} = \begin{pmatrix} m_1 & -m_2 \\ m_2 & m_1 \end{pmatrix}$$

is the rotation matrix. Thus, the transformation space for rigid motions is three-dimensional. For computational reasons we consider the four dimensional parameterization by (m_1, m_2, v_x, v_y) where $m_1 = \cos\alpha$ and $m_2 = -\sin\alpha$ and restrict it to a three dimensional algebraic variety by the constraint $m_1^2 + m_2^2 = 1$, i.e., $\det(M) = 1$. Thus, we consider the space of rigid motions as a three dimensional variety in the four dimensional space of similarity transformations.

Obviously, for any two points $a, b \in \mathbb{R}^2$ and every rotation angle α there exists a unique translation vector v_α, such that the rigid motion $t = (\alpha, v_\alpha)$ maps b to a. Therefore, we use a single random point of each shape $a \in A$ and $b \in B$ as a sample in one random experiment and record the δ-region $\{(M, v) \mid dist(M \cdot b + v, a) \leq \delta\}$, where all matrices M of the form given above are allowed.

In the three dimensional transformation space parameterized by the rotation angle and the translation vector the δ-region corresponding to a sample pair (a, b) in this approach has the shape of a spiral tube extending from 0 to 2π in the direction of the rotation axis, where for each value $\alpha \in [0, 2\pi]$ the cross-section parallel to the translation plane has the shape of the δ-neighborhood with respect to the chosen metric in image space as illustrated in Fig. 1.

For **similarity maps** the transformation space is four-dimensional. A similarity map $t = (\alpha, k, v_x, v_y)$ is defined by a rotation angle α, a scaling factor k, and a translation vector $v = (v_x, v_y)$. t maps a point $b \in \mathbb{R}^2$ to a point $t(b) = Mb + v$, where

$$M = \begin{pmatrix} k\cos\alpha & -k\sin\alpha \\ k\sin\alpha & k\cos\alpha \end{pmatrix} = \begin{pmatrix} m_1 & -m_2 \\ m_2 & m_1 \end{pmatrix}.$$

A random sample from each shape contains two points: $S_A = (a_1, a_2)$, where $a_1, a_2 \in A$, and $S_B = (b_1, b_2)$, where $b_1, b_2 \in B$. Thus, a sample pair in the first step of the algorithm is the pair of point pairs. The sample pair (S_A, S_b) determines a unique similarity transformation t mapping b_1 to a_1 and b_2 to a_2. Although a standard way to parameterize the space of similarity transformations is by (α, k, v_x, v_y), for computational reasons it is more convenient to use the parameterization (m_1, m_2, v_x, v_y) where $m_1 = k\cos\alpha$ and $m_2 = k\sin\alpha$. For a general piecewise algebraic metric a δ-region is then bounded by algebraic surfaces.

Generic probabilistic algorithm. Now we can define the shape matching problem in a generic way: Given two shapes A and B, a class of allowed transformations $\mathcal{T}$ and a certain tolerance parameter δ, we want to find a transformation $t \in \mathcal{T}$ which lets $t(B)$ in some sense match best A within a range of δ. We can solve an instance by the following probabilistic algorithm:

1. Repeat the following experiment N times: Take random samples S_A from A and S_B from B and record the corresponding δ-region in the space of transformations $\mathcal{T}$.
2. Return the points of $\mathcal{T}$ covered by the largest number of δ-regions as candidates for good transformations.

In the next section we define and analyze a suitable probability function in transformation space underlying the experiments made by the algorithm and provide the bounds on the number of experiments needed to approximate the value of the maximum of this function within a certain factor with a prespecified probability.

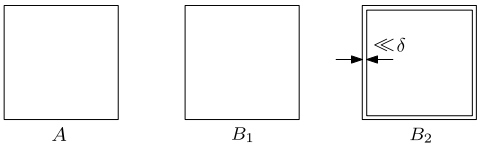


Fig. 2. The maximum of the measure of the set $M_\delta(t)$ for shapes A and B_2 is approximately twice the maximum measure of the set $M_\delta(t)$ for shapes A and B_1, whereas the maximum values of $p_\delta(t)$ are approximately equal for both pairs of shapes.

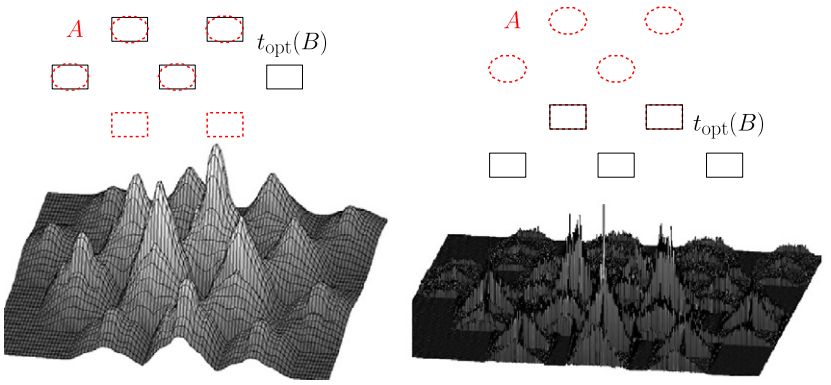


Fig. 3. Matching A (dashed lines) with B (solid lines) with large (left) and small (right) values of δ and the graphs of the corresponding functions $p_\delta(t)$ in translation space.

3. Analysis of the probabilistic algorithm

3.1. Hit probability in transformation space

In this section we determine what measure of resemblance between shapes is approximated by the algorithm and bound the number of experiments needed to get an ε-approximation of the maximum of that measure.

First we introduce some formal notation and definitions. Let Ω denote the sample space, i.e., the set of all sample pairs (S_A, S_B). By the definition of our random experiment, the samples of two shapes are drawn independently and uniformly, therefore, we have a uniform distribution on Ω.

Let $\mathcal{T} \subset \mathbb{R}^d$ denote the d-dimensional transformation space. We define a function $p_\delta : \mathcal{T} \to \mathbb{R}$ as the probability that a transformation vector t is covered by a δ-region corresponding to a randomly selected sample pair. We will call $p_\delta(t)$ the *hit probability* of transformation t. The set of sample pairs yielding a δ-region that covers a transformation t is $M_\delta(t) = \{(S_A, S_B) \in \Omega \mid S_B \xrightarrow[\delta]{t} S_A\}$, and $p_\delta(t) = \frac{|M_\delta(t)|}{|\Omega|}$, where $|\cdot|$ denotes the Lebesgue measure in the sample space. Consequently, we have

Remark 1. The hit probability $p_\delta(t)$ in the transformation space has its maximum at the transformation maximizing the Lebesgue measure of the set $M_\delta(t)$ defined as

$$M_\delta(t) = \{(S_A, S_B) \in \Omega \mid S_B \xrightarrow[\delta]{t} S_A\}.$$

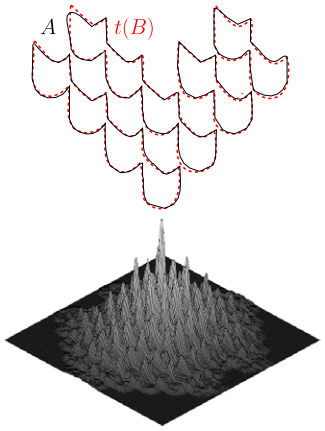
We can interpret the Lebesgue measure of the set $M_\delta(t)$ as a measure of resemblance associated with a transformation t. Intuitively, this should reflect the perceived notion of "closeness" of two shapes. However, if we want to compare several pairs of shapes, $p_\delta(t)$ appears to be a better measure of similarity. Due to normalization by the total measure of the sample space, $p_\delta(t)$ is invariant under scaling and is less sensitive to multiple occurrences of curves within one shape, as illustrated in Fig. 2.

Let us discuss the meaning of Remark 1 for the different classes of transformations.

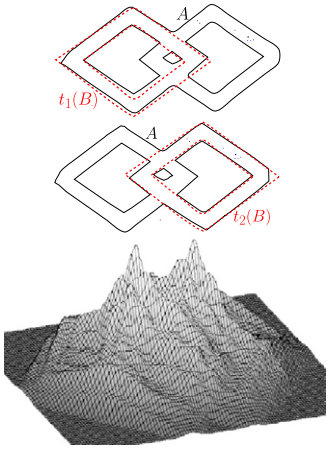
Translations and rigid motions. The sample space is in this case $\Omega = A \times B$ and $M_\delta(t) = \{(a, b) \in A \times B \mid dist(a, t(b)) \le \delta\}$. To maximize the measure of this set means to find a transformation that maps largest possible parts of the shapes into proximity of each other.

In the case of translations, it can be observed that for $\delta \to 0$ the resulting probability distribution corresponds to the normalized *generalized Radon transform* of the shape A with respect to shape B as defined in [23].

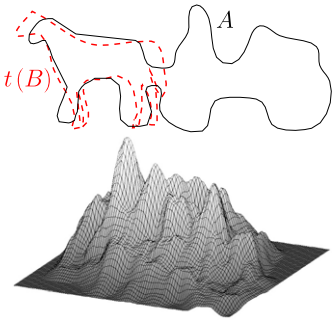
Similarity maps. In the case of similarity maps a sample taken from one shape consists of two random points, the sample space is then $\Omega = A^2 \times B^2$. By Remark 1 the similarity map with maximum coverage by the δ-regions is the one maximizing the measure of the set $M_\delta(t) = \left\{(a_1, a_2, b_1, b_2) \in A^2 \times B^2 \mid dist(t(b_1), a_1) \le \delta \text{ and } dist(t(b_2), a_2) \le \delta\right\}$. This measure is less intuitive with respect to matching shapes than the one in the previous cases. A simple consideration shows, however,



(a) Nearly congruent shapes yield a function $p_\delta(t)$ with a clear maximum.



(b) Shape B occurs twice in A which results in two (almost) equally large local maxima of $p_\delta(t)$.



(c) A rough complete-partial match of horse and carriage shapes from the MPEG7-Shape B dataset results in a less distinct maximum of $p_\delta(t)$ than in the case of nearly congruent shapes.

Fig. 4. Matched shapes and the corresponding function $p_\delta(t)$ in translation space.

that maximizing the measure of $M_\delta(t)$ also means to maximize the measure of $M'_\delta(t) = \{(a, b) \in A \times B \mid dist(t(b), a) \leq \delta\}$. The measure of the set $M_\delta(t)$ is exactly $\left|M'^2_\delta(t)\right| = \left|M'_\delta(t)\right|^2$. Since the measure of a set is always non-negative, both functions have maxima at the same values of t. So the same, intuitively understandable measure, is maximized as in the cases of translations and rigid motions.

The role of the parameter δ. In the description of the algorithm we introduced a parameter δ, which defines how far apart two samples are allowed to be and still be considered close. The choice of δ, therefore, should be specified by the user and

controls the trade-off between the quality of match and the size of the parts matched. With a small value of δ our algorithm would find a transformation which maps nearly congruent parts of two shapes to each other. A large value of δ leads to a transformation which gives a rough match but for larger parts of the shapes; see Fig. 3.

For nearly congruent shapes, however, a small δ already leads to a complete matching, see Fig. 4(a). If shape B or parts of it are nearly congruent to some parts of A, then with a small value of δ we detect these occurrences as shown in Fig. 4(b) and (c).

For some applications it might be worth to consider several local maxima of the distribution, since they can give us additional information about the shapes. For example, multiple local maxima of the distribution, that are almost equally good, indicate multiple occurrences of one shape, or its parts, depending on the value of the similarity measure, within the other; see Fig. 4(b).

3.2. Approximation of the hit probability

In this section we determine how many samples are needed in order to approximate the function $p_\delta(t)$ in the transformation space within a certain accuracy ε with high probability. We also analyze the total running time of the algorithm.

In order to find a transformation covered by the highest number of δ-regions corresponding to the samples, we consider the arrangement of these δ-regions, i.e., the subdivision of the transformation space induced by the boundaries of the regions. All transformations in the same cell of the arrangement have the same region coverage. Therefore, it is sufficient to consider one point in each cell of the arrangement, which we will call a witness point. Then we can traverse the arrangement and take the witness points with the highest number of δ-regions that contain this point.

We will show that the fraction of δ-regions covering the deepest cell of the arrangement gives a good approximation to the maximum value of the hit probability $p_\delta(t)$. The number of necessary samples is expressed in terms of the allowed approximation error ε and the maximally allowed probability of failure η. Let the random variable $Z(t)$ denote the number of δ-regions produced by N random experiments that cover t. Let $\tilde{p}_\delta(t)$ denote the ratio of the number of the observed δ-regions that cover t to the total number of samples, that is $\tilde{p}_\delta(t) = \frac{Z(t)}{N}$. $\tilde{p}_\delta(t)$ is an estimate of $p_\delta(t)$.

Using Chernoff bounds, see [18], and the technique described by Cheong et al. in [11] we will bound the relative error for the estimate of the hit probability in the transformation space. The following theorem bounds the number of samples needed for an approximation with a relative error at most ε with probability of failure at most η for all transformation classes considered in this work:

Theorem 1. *Let A and B be two shapes, i.e., finite sets of rectifiable curves with total lengths L_A, L_B, respectively, and $\delta > 0$ be a given tolerance value. Assume that t_{app} is a transformation maximizing $\tilde{p}_\delta(t)$ after some number N of random experiments and t_{opt} is a transformation maximizing $p_\delta(t)$, and let $m = \max(L_A, L_B, n\delta)$, where n is the total number of curves in A and B. Then for all ε, η, $0 < \varepsilon, \eta < 1$, there exists a constant c such that for $N \geq c\frac{m^2}{\varepsilon^2\delta^2}\ln\left(\max(\frac{1}{\eta}, \frac{m^2}{\varepsilon^2\delta^2})\right)$ the probability that $\left|\tilde{p}_\delta(t_{\text{app}}) - p_\delta(t_{\text{opt}})\right| \geq \varepsilon p_\delta(t_{\text{opt}})$ is at most η.*

For the proof of the theorem we first show that for a fixed transformation t the probability of a bad estimate of $p_\delta(t)$ falls exponentially with N (Lemma 2). Then we define some experiment dependent transformations t and show that also for those t the probability of a bad estimate falls exponentially with N (Lemma 3). Finally, we argue that it is sufficient to compute the estimate $\tilde{p}_\delta$ for a finite set of experiment dependent transformations t in order to get a good approximation of the maximum of p_δ. These considerations are valid for all transformation classes considered.

Lemma 2. *For all $0 < \varepsilon, \nu < 1$, for a sample set S of size N, and any transformation represented by some vector $t \in \mathbb{R}^d$ the following holds:*

- *If $p_\delta(t) \leq \nu$ then $P(\tilde{p}_\delta(t) > (1+\varepsilon)\nu) \leq e^{-\frac{\varepsilon^2\nu N}{3}}$.*
- *If $p_\delta(t) > \nu$ then $P\left(\left|\tilde{p}_\delta(t) - p_\delta(t)\right| > \varepsilon p_\delta(t)\right) \leq 2e^{-\frac{\varepsilon^2\nu N}{4}}$.*

Proof. If $p_\delta(t) \leq \nu$:

$$\begin{aligned}
P(\tilde{p}_\delta(t) > (1+\varepsilon)\nu) &= P(Z(t) > (1+\varepsilon)\nu N) \\
&= P\left(e^{rZ(t)} \geq e^{r(1+\varepsilon)\nu N}\right) \quad \text{for all } r > 0 \\
&\leq \frac{E\left(e^{rZ(t)}\right)}{e^{r(1+\varepsilon)\nu N}} \quad \text{by the Markov inequality [18]} \\
&\leq \frac{e^{(e^r-1)p_\delta(t)N}}{e^{r(1+\varepsilon)\nu N}} \quad \text{since } r \leq e^r - 1 \\
&\leq \left(\frac{e^{(e^r-1)}}{e^{r(1+\varepsilon)}}\right)^{\nu N} = (e^{\varepsilon-(1+\varepsilon)\ln(1+\varepsilon)})^{\nu N} \quad \text{for } r = \ln(1+\varepsilon) \\
&\leq e^{-\frac{\varepsilon^2\nu N}{3}} \quad \text{for } 0 < \varepsilon < 1.
\end{aligned}$$

In the case $p_\delta(t) > \nu$:

$$\begin{aligned} P\left(\left|\tilde{p}_\delta(t) - p_\delta(t)\right| > \varepsilon p_\delta(t)\right) &= P(|Z(t) - p_\delta(t)N| > \varepsilon p_\delta(t)N) \\ &= P(|Z(t) - E(Z(t))| > \varepsilon E(Z(t))) \\ &\leq e^{-\frac{\varepsilon^2 E(Z(t))}{2}} + e^{-\frac{\varepsilon^2 E(Z(t))}{4}} \end{aligned}$$

by the simplified Chernoff bound [18, Theorems 4.4,4.5]. Since $p_\delta(t) \geq \nu$, we get

$$P\left(\left|\tilde{p}_\delta(t) - p_\delta(t)\right| > \varepsilon p_\delta(t)\right) \leq 2e^{-\frac{\varepsilon^2 p_\delta(t) N}{4}} \leq 2e^{-\frac{\varepsilon^2 \nu N}{4}},$$

which concludes the proof. □

We associate with each cell C of the arrangement $\mathcal{A}$ of δ-regions a so-called *witness point*, i.e., a point that lies on a lowest-dimensional face F of $\mathcal{A}$ that contributes to the boundary of C. Observe, that F is in general a connected component of the intersection of k boundaries of δ-regions with $1 \leq k \leq d$. The dimension of F is $d - k$. Thus, by considering all k-subsets of δ-regions for all k, $1 \leq k \leq d$, and taking a point in each connected component of the intersection of those k region boundaries we can be sure to have at least one witness point for each cell of the arrangement.

Since we assume that the distance metric is a piecewise algebraic function of constant degree, the number of connected components of every intersection of k δ-region boundaries is bounded by a constant. Thus, the total number of witness points is at most $c\sum_{k=1}^{d}\binom{N}{k} \leq cN^d$, where c is a constant.

A witness point t is not independent of all N experiments the algorithm performs and their corresponding δ-regions, but it depends only on a few of them, namely at most d. Since t is independent of the remaining $N - d$ experiments, we can apply Lemma 2 to t and those experiments. We elaborate on this idea in the following lemma.

Lemma 3. *For all ε, ν, $0 < \varepsilon, \nu < 1$, and a sample set S of size $N \geq \frac{2d}{\varepsilon\nu} + d$, for each witness point $t \in \mathbb{R}^d$ of the arrangement of the δ-regions corresponding to the samples in S, the following holds:*

- *If $p_\delta(t) \leq \nu$ then $P(\tilde{p}_\delta(t) > (1+\varepsilon)\nu) \leq e^{-\frac{\varepsilon^2(N-d)\nu}{12}}$.*
- *If $p_\delta(t) > \nu$ then $P\left(\left|\tilde{p}_\delta(t) - p_\delta(t)\right| > \varepsilon p_\delta(t)\right) \leq 2e^{-\frac{\varepsilon^2\nu(N-d)}{16}}$.*

Proof. Observe that Lemma 2 cannot be applied to the witness points directly since they depend on the experiment, i.e., the chosen samples. However, since they depend on at most d samples, the remaining $\geq N - d$ samples are "random" for them and we can apply Lemma 2 replacing N by $N - d$. More specifically:

Let $S_1, \ldots, S_i \in S$, $1 \leq i \leq d$, be the sample pairs whose δ-regions induce the witness point t. Consider the sample set $Q = S \setminus \{S_1, \ldots, S_i\}$, $|Q| = N - i$. The point t and the sample set Q are independent. Let $Z_Q(t)$ and $Z_S(t)$ denote the number of δ-regions that cover t in sample sets Q and S, respectively, and $\tilde{p}_{\delta Q}(t)$, $\tilde{p}_{\delta S}(t)$ denote the estimate of the hit probability by sample sets Q and S, respectively, i.e., $\tilde{p}_{\delta Q}(t) = Z_Q(t)/(N-i)$ and $\tilde{p}_{\delta S}(t) = Z_S(t)/N$.

Since we consider closed regions, $Z_Q(t) = Z_S(t) - i$, $\tilde{p}_{\delta Q}(t) \leq \tilde{p}_{\delta S}(t)$ and

$$\tilde{p}_{\delta Q}(t) = \frac{Z_S(t) - i}{N - i} = \frac{Z_S(t)}{N}\frac{N}{N-i} - \frac{i}{N-i} \geq \tilde{p}_{\delta S}(t) - \frac{i}{N-i} \geq \tilde{p}_{\delta S}(t) - \frac{d}{N-d}.$$

Therefore,

$$\left|\tilde{p}_{\delta S}(t) - p_\delta(t)\right| \leq \left|\tilde{p}_{\delta Q}(t) - p_\delta(t)\right| + \left|\tilde{p}_{\delta S}(t) - \tilde{p}_{\delta Q}(t)\right| \leq \left|\tilde{p}_{\delta Q}(t) - p_\delta(t)\right| + \frac{d}{N-d}.$$

In case $p_\delta(t) \leq \nu$:

$$\begin{aligned} P(\tilde{p}_{\delta S}(t) > (1+\varepsilon)\nu) &\leq P\left(\tilde{p}_{\delta Q}(t) + \frac{d}{N-d} > (1+\varepsilon)\nu\right) \\ &= P\left(\tilde{p}_{\delta Q}(t) > (1+\varepsilon)\nu - \frac{d}{N-d}\right) \\ &\leq P\left(\tilde{p}_{\delta Q}(t) > \left(1 + \frac{\varepsilon}{2}\right)\nu\right) \quad \text{for } N \geq \frac{2d}{\varepsilon\nu} + d \\ &\leq e^{-\frac{(\varepsilon/2)^2(N-d)\nu}{3}} \quad \text{by Lemma 2} \\ &= e^{-\frac{\varepsilon^2(N-d)\nu}{12}} \end{aligned}$$

If $p_\delta(t) > \nu$:

$$\begin{aligned} P\left(\left|\tilde{p}_{\delta S}(t) - p_\delta(t)\right| > \varepsilon p_\delta(t)\right) &\le P\left(\left|\tilde{p}_{\delta Q}(t) - p_\delta(t)\right| + \frac{d}{N-d} > \varepsilon p_\delta(t)\right) \\ &\le P\left(\left|\tilde{p}_{\delta Q}(t) - p_\delta(t)\right| > \frac{\varepsilon}{2} p_\delta(t)\right) \quad \text{for } N \ge \frac{2d}{\varepsilon\nu} + d \\ &\le 2e^{-\frac{(\varepsilon/2)^2 \nu(N-d)}{4}} \quad \text{by Lemma 2} \\ &= 2e^{-\frac{\varepsilon^2\nu(N-d)}{16}}. \quad \square \end{aligned}$$

In the above lemmata we used an additional parameter ν for the smallest value of $p_\delta(t)$ which we want to approximate well enough. Next, we eliminate this parameter:

Lemma 4. *For any two shapes A and B of total lengths L_A and L_B, respectively, and for the following classes of transformations: translations, rigid motions, and similarities, there exists a transformation t such that $p_\delta(t) \ge \frac{\delta^2}{m^2}$, where $m = \max(L_A, L_B, n\delta)$ and n is the total number of curves in A and B.*

Proof. Let s_a denote a part of one of the curves of A of length δ if there exists one, otherwise s_a denotes the longest curve of A. The length of s_a is at least $\frac{L_A}{n} \le \delta$ in the second case. Similarly s_b denotes a subcurve of length δ or the longest curve of B with length at least $\frac{L_B}{n} \le \delta$. Let v denote the translation vector that maps the center of s_b to the center of s_a. For an arbitrary point p_a of s_a and an arbitrary point p_b of s_b it holds that $dist(p_a, p_b + v) \le \delta$, therefore v is covered by the δ-region corresponding to p_a, p_b. Thus, in the case of translations $s_a \times s_b$ is a subset of $M_\delta(v)$ and

$$p_\delta(v) \ge \frac{|s_a \times s_b|}{|\Omega|} \ge \frac{\min\left(\delta^2, \delta L_A/n, \delta L_B/n, L_A L_B/n^2\right)}{L_A L_B} \ge \frac{\delta^2}{m^2}.$$

For rigid motions the same argument as above shows that the rigid motion t with rotation angle 0 and translation vector v is covered by every δ-region corresponding to an arbitrary point in s_a and an arbitrary point in s_b.

In the case of similarity maps, the shape B can be scaled by the factor $\frac{\delta}{D_B}$, where D_B is the diameter of B, so that the diameter of the scaled shape B is δ. If A contains a connected component of length at least δ, then we can place the scaled shape B in such a way that for any point of a part of A of length δ the distance to any point of the scaled B is at most δ. Therefore, the measure of the set $M_\delta(t)$ for that t is at least $L_B^2 \cdot \delta^2$. The corresponding value of p_δ is $p_\delta(t) = \frac{|M_\delta(t)|}{|\Omega|} \ge \frac{L_B^2 \cdot \delta^2}{L_A^2 L_B^2} = \frac{\delta^2}{L_A^2} \ge \frac{\delta^2}{m^2}$. Otherwise, observe that the largest connected component of A must have length at least $\frac{L_A}{n} \le \delta$. For the transformation t that maps the scaled B to the largest component of A the measure of $M_\delta(t)$ is then at least $\frac{L_A^2}{n^2} L_B^2$ and $p_\delta(t) \ge \frac{1}{n^2} \ge \frac{\delta^2}{m^2}$. $\square$

Now we can prove Theorem 1:

Proof of Theorem 1. Any witness point lies on the boundary of a k-subset of δ-regions. Furthermore, any k-subset yields a system of constantly many polynomial equations of constant degree. There are constantly many connected components of the solution set of such system of equations, and with each connected component we associate a witness point. Therefore, there are at most $c_0 \sum_{k=1}^{d} \binom{N}{k} \le c_0 N^d$ witness points, where c_0 is a constant. Then the probability that there exists a witness point t with $p_\delta(t) \ge \nu$ and $\left|\tilde{p}_\delta(t) - p_\delta(t)\right| > \varepsilon p_\delta(t)$ or with $p_\delta(t) < \nu$ and $\tilde{p}_\delta(t') > (1+\varepsilon)\nu$ is, according to Lemma 3, at most $c_0 N^d 2e^{-\frac{\varepsilon^2\nu(N-d)}{16}}$. A straightforward calculation shows that for $N \ge \frac{c_1}{\varepsilon^2\nu} \ln\left(\frac{1}{\varepsilon^2\nu}\right)$ with some suitable constant c_1 this value is at most $e^{-\frac{\varepsilon^2\nu(N-d)}{32}}$, which is less than η for $N \ge \frac{32}{\varepsilon^2\nu} \ln\frac{1}{\eta} + d$. So the probability that there exists a witness point, for which the estimate of $p_\delta(t)$ is bad in the sense described above, is at most $\eta/2$ for

$$N \ge \frac{c_2}{\varepsilon^2\nu} \ln\left(\max\left(\frac{1}{\eta}, \frac{1}{\varepsilon^2\nu}\right)\right) \tag{1}$$

for some constant c_2. Observe, that this is a combinatorial result which does not depend on the spatial position of witness points corresponding to a certain sequence of N experiments.

By Lemma 4 for any two shapes and the transformation classes considered there always exists a transformation t such that $p_\delta(t) \ge \frac{\delta^2}{m^2}$, where $m = \max(L_A, L_B, n\delta)$. The maximum of p_δ is then also greater or equal $\frac{\delta^2}{m^2}$ and we can choose the value ν as $\nu = \frac{\delta^2}{m^2}$.

Let t^* be a witness point of the cell of the arrangement containing t_{opt}. Plugging the value of ν in formula (1) we obtain that after

$$N = O\left(\frac{m^2}{\varepsilon^2\delta^2} \ln\left(\max\left(\frac{1}{\eta}, \frac{m^2}{\varepsilon^2\delta^2}\right)\right)\right)$$

experiments for all witness points, in particular for t^* and t_{app}, and additionally for t_{opt} it holds with probability at least $1-\eta/2$ that $\left|\tilde{p}_\delta(t)-p_\delta(t)\right| \le \varepsilon p_\delta(t)$. Combining these error bounds we get

$$\begin{aligned}
\tilde{p}_\delta(t_{\text{app}}) &\ge \tilde{p}_\delta(t^*) && \text{since } t_{\text{app}} \text{ maximizes } \tilde{p}_\delta(t)\\
&= \tilde{p}_\delta(t_{\text{opt}}) && \text{for } t_{\text{opt}} \text{ is in the cell witnessed by } t^*\\
&\ge (1-\varepsilon)p_\delta(t_{\text{opt}}) && \text{with probability } \ge 1-\eta/2
\end{aligned}$$

and

$$\begin{aligned}
\tilde{p}_\delta(t_{\text{app}}) &\le (1+\varepsilon)p_\delta(t_{\text{app}}) && \text{with probability } \ge 1-\eta/2\\
&\le (1+\varepsilon)p_\delta(t_{\text{opt}}) && \text{since } t_{\text{opt}} \text{ maximizes } p_\delta(t)
\end{aligned}$$

Therefore, $\left|\tilde{p}_\delta(t_{\text{app}})-p_\delta(t_{\text{opt}})\right| \le \varepsilon p_\delta(t_{\text{opt}})$ with probability at least $1-\eta$. □

Running time. The running time of the algorithm consists of the time needed to generate N random samples denoted by $T_{\text{gen}}(n,N)$, where n is the number of curves in the shape, and the time needed to determine the depth of the arrangement of N δ-regions denoted by $T_{\text{depth}}(N)$.

In order to generate a random point on one shape (set of curves) we proceed as follows: Let $l_1, l_2, \ldots$ be the lengths of the single curves and $L = \sum_i l_i$ the total length of the set of curves. We consider the interval $[0, L]$ and partition it into subintervals of lengths $l_1, l_2, \ldots$. Then after choosing a random value a in $[0, L]$ we can determine by binary search in which subinterval it is contained, which identifies one of the curves. The position of a within the subinterval gives a unique point on this curve. Obviously, this procedure requires $O(n)$ preprocessing time and $O(\log n)$ time to generate a random point. Therefore, $T_{\text{gen}}(n,N) = O(n + N\log n)$.

For general metrics L_p and the considered classes of transformations the boundaries of δ-regions are algebraic hypersurfaces. In order to determine the depth of the arrangement of N δ-regions we can generate a point in each cell of the arrangement using the $O(N^{d+1})$-time algorithm by Basu et al. [10], which computes a sample point from each cell of a set of N hypersurfaces in R^d. As we have seen above, there are at most $O(N^d)$ such sample points. Then we can determine the depth of each of these points and thus the depth of the arrangement in $T_{\text{depth}}(N) = O(N^{d+1})$ time.

Summarizing these results and using Theorem 1 we obtain the following theorem for all three classes of transformations considered:

Theorem 5. *Let A and B be two shapes represented by finite sets of rectifiable curves in the plane and $\delta > 0$ be the given tolerance value. Let t_{opt} denote the transformation maximizing $p_\delta(t)$, L_A, L_B the total lengths of the curves in A and B, respectively, n the total number of curves in both shapes, and $m = \max(L_A, L_B, n\delta)$. Then for all ε, η, $0 < \varepsilon, \eta < 1$, there exists a constant c, such that for $N \ge c\frac{m^2}{\varepsilon^2\delta^2}\ln\left(\max(\frac{1}{\eta}, \frac{m^2}{\varepsilon^2\delta^2})\right)$ the generic probabilistic algorithm with probability at least $1-\eta$ computes a transformation t_{app} such that $\left|\tilde{p}_\delta(t_{\text{app}})-p_\delta(t_{\text{opt}})\right| \le \varepsilon p_\delta(t_{\text{opt}})$. Its runtime is $O(n + N\log n + N^{d+1})$, where d is the dimension of the transformation space.*

The theorem states only that with high probability the numerical value obtained by $\tilde{p}_\delta(t_{\text{app}})$ is close to $p_\delta(t_{\text{opt}})$ which is a measure for the closeness of the two shapes. t_{app} and t_{opt} need not be close in transformation space. But it is also easily possible to derive that the transformation t_{app} is "good" in the sense that $p_\delta(t_{\text{app}})$ is close to the optimum, since $\tilde{p}_\delta(t_{\text{app}})$ is close to $p_\delta(t_{\text{app}})$.

Observe that, at least for sufficiently small values of δ, the runtime of the algorithm depends much more on the parameters ε and η than on the combinatorial input size n, which is needed only in the preprocessing and the drawing of random samples.

For translations in combination with convex algebraic distance functions and for similarities in combination with the L_1 and L_∞ metrics the running time of the algorithm is actually better than that stated in Theorem 5: In the case of translations, the arrangement of δ-regions is an arrangement of pseudo-disks, which can be constructed straightforwardly in time $O(N^2)$. During the construction of the arrangement we can keep record of the depth of the cells. Then at the end of the construction algorithm we know the depth of the deepest cell. For similarities in combination with the L_1 or L_∞ metric the δ-regions in transformation space are bounded by a constant number of 3-dimensional hyperplanes. Using the algorithm of Edelsbrunner et al. [12] the arrangement of N such δ-regions can be constructed in $O(N^4)$ time.

For translations, further speed-up can be achieved in combination with the depth approximation algorithm by Aronov and Har-Peled [9] resulting in running time $T_{\text{depth}}(N) = O(N\varepsilon^{-2}\log N)$.

4. Problems, variants, and concluding remarks

Downscaling problem. The algorithm that we described for similarity transformations has a minor problem that could be called "downscaling problem". On the one hand, as we observed in Lemma 4 a transformation t_1 that scales the shape B down to a shape of diameter δ and maps the scaled B to some position on the shape A has a measure of resemblance of $\frac{\delta^2}{m^2}$. On the other hand, if the shapes A and B are similar, then the transformation t_2 that matches best the shape B to the

shape A has approximately the same measure of resemblance, since for every point b of B there is a segment $s_a(b)$ of A of length (approximately) δ such that b is in the δ-neighborhood of every point in $s_a(b)$. Because of this overrating of shrinking transformations some other, reasonable transformations are likely to be missed.

This problem can easily be avoided by setting a lower bound for the allowed scaling factor to some constant times $\frac{\delta}{D_B}$, where D_B is the diameter of the shape B. Samples yielding a smaller scaling factor will be discarded. By this restriction of the scaling factor we could achieve good experimental results. A similar problem arises in matching under affine transformations, where the analogue to a lower bound on the scaling factor is a lower bound on the value of the determinant of the linear transformation matrix.

Variants for rigid motions and similarities. We also considered some variants of random sample generation for rigid motions and similarities, which we partially analyzed and tested experimentally.

For rigid motions, in addition to a point of the shape we take the (interpolated) direction of the tangent line at that point and restrict the corresponding δ-region so that not only the points but also the sample directions are close. The idea behind this approach is to reduce the search space.

For similarity transformations, a sample of one shape can alternatively consist of one point, the direction of the tangent line at that point and the (possibly interpolated) curvature at that point instead of a sample consisting of two points. This approach is not affected by the downscaling problem described above.

Both alternative approaches, for rigid motions and for similarities, are best suitable for shapes where the tangent slopes actually contribute to the shape characterization, as opposed to shapes with noisy contours or shapes composed of many sparse and small parts.

Higher dimensional matching problems. Although we did not carry it out in detail, we expect that the techniques used here can be applied for matching higher dimensional objects, as well. In fact, analysis of an analogous variant of our method has been carried out for two-dimensional *regions* in two-dimensional space, see [7], and an extension to objects of arbitrary dimension in arbitrary-dimensional space should be possible. Practical experiments using the idea for matching triangulated surfaces in three-dimensional space in the framework of a M.Sc. thesis [19] in our work group showed promising results.

Experimental evaluation. As was mentioned before, for practical purposes we enhanced this algorithm with various heuristics (see [6]) but even with an implementation of the simple form presented here we observed reasonable matching results in experiments with the MPEG-7 Core Experiment CE-Shape-1 dataset and a selection of trademark images. The experiments show that the theoretical bounds on the number of experiments N are rather pessimistic. For simple images, like most of the MPEG-7 Core Experiment CE-Shape-1 images, reasonable matching results were achieved with N between 1000 and 2000 for translations and $N = 50\,000$ to $100\,000$ for rigid motions. For similarity transformations, after restricting the range of the scaling factor for the reasons described above, good matching results could be obtained with 50 000 to 500 000 experiments depending on the structure of the shapes.

Concluding remarks. We presented a probabilistic approach for matching two shapes which comes close to the human notion of match and is easy to implement. In this paper, we considered only three classes of transformations: translations, rigid motions, and similarities, since those are the ones most commonly used for shape matching. However, our approach is much more general. An elaborate description of the details of the algorithm and the analysis for homotheties, shear transformations and affine transformations can be found in [20].

In general the probabilistic algorithm presented here is robust to noise, deformations and cracks in the representation of shapes and does not require shapes to be modelled by a single contour line. It is applicable to the problem of complete and partial matching.

References

[1] A.S. Aguado, E. Montiel, M.S. Nixon, Invariant characterisation of the Hough transform for pose estimation of arbitrary shapes, Pattern Recognit. 35 (2002) 1083–1097.
[2] H. Alt, B. Behrends, J. Blömer, Approximate matching of polygonal shapes, Ann. Math. Artif. Intell. 13 (1995) 251–265.
[3] H. Alt, M. Godau, Computing the Fréchet distance between two polygonal curves, Internat. J. Comput. Geom. Appl. 5 (1995) 75–91.
[4] H. Alt, L.J. Guibas, Discrete geometric shapes: matching, interpolation, and approximation, in: Handbook of Computational Geometry, Elsevier, 1999.
[5] H. Alt, G. Rote, C. Wenk, A. Efrat, Matching planar maps, J. Algorithms (2003) 262–283.
[6] H. Alt, L. Scharf, S. Scholz, Probabilistic matching and resemblance evaluation of shapes in trademark images, in: Proc. of the 6th ACM International Conference on Image and Video Retrieval, 2007, pp. 533–540.
[7] H. Alt, L. Scharf, D. Schymura, Probabilistic matching of planar regions, Comput. Geom. 43 (2) (2010) 99–114.
[8] E.M. Arkin, L.P. Chew, D.P. Huttenlocher, K. Kedem, J.S.B. Mitchell, An efficiently computable metric for comparing polygonal shapes, IEEE Trans. Pattern Anal. Mach. Intell. 13 (3) (1991) 209–216.
[9] B. Aronov, S. Har-Peled, On approximating the depth and related problems, in: Proc. 16th ACM–SIAM Symposium on Discrete Algorithms, 2005, pp. 886–894.
[10] S. Basu, R. Pollack, M.-F. Roy, On computing a set of points meeting every semi-algebraically connected component of a family of polynomials on a variety, J. Complexity 13 (1997) 28–37.
[11] O. Cheong, A. Efrat, S. Har-Peled, On finding a guard that sees most and a shop that sells most, in: Proc. 15th ACM–SIAM Symposium on Discrete Algorithms, 2004, pp. 1091–1100.
[12] H. Edelsbrunner, J. O'Rourke, R. Seidel, Constructing arrangements of lines and hyperplanes with applications, SIAM J. Comput. 15 (2) (1986) 341–363.
[13] A. Efrat, Q. Fan, S. Venkatasubramanian, Curve matching, time warping, and light fields: new algorithms for computing similarity between curves, J. Math. Imaging Vision 27 (3) (2007) 203–216.

[14] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395.
[15] M. Hagedoorn, M. Overmars, R. Veltkamp, A new visibility partition for affine pattern matching, in: Proc. Discrete Geometry for Computer Imagery Conference, DGCI 2000, Springer-Verlag, Berlin, 2000, pp. 358–370.
[16] L.J. Latecki, R.C. Veltkamp, Properties and performances of shape similarity measures, in: Proc. of Int. Conf. on Data Science and Classification, 2006.
[17] N.J. Mitra, L.J. Guibas, M. Pauly, Partial and approximate symmetry detection for 3d geometry, ACM Trans. Graph. 25 (3) (2006) 560–568.
[18] M. Mitzenmacher, E. Upfal, Probability and Computing: Randomized Algorithms and Probabilistic Analysis, Cambridge University Press, 2005.
[19] M. Mohajer, A random algorihm to determine resemblance between two 3D-surfaces. Master's thesis, Freie Universität Berlin, Institut für Informatik, January 2008.
[20] L. Scharf, Probabilistic matching of planar shapes, Ph.D. Thesis, Freie Universität Berlin, Institut für Informatik, March 2009. http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000010520.
[21] G. Stockman, Object recognition and localization via pose clustering, Comput. Vis. Graph. Image Process. 40 (1987) 361–387.
[22] M. Tanase, R.C. Veltkamp, H. Haverkort, Multiple polyline to polygon matching, in: Proc. of the 16th Annual Symposium on Algorithms and Computation, LNCS 3827, 2005, pp. 60–70.
[23] P. Toft, The radon transform — theory and implementation, Ph.D. Thesis, Department of Mathematical Modelling, Technical University of Denmark, 1996.