2D Figure Classification via Fusion of Shape Descriptors Based on Contours, Thickness and Skeleton-Geodesic Distance Distributions

N. A. Lomov[1], S. V. Sidyakin[2], Yu. V. Vizilter[3]

[1] *Lomonosov Moscow State University, Moscow, Russia*

[2,3] *The Federal State Unitary Enterprise "State Research Institute of Aviation System", Moscow, Russia*

Abstract

This paper is addressed to 2D shape representation and classification. Many shape descriptors usually exploit one source of information, for example, only contour or only skeleton. Mainly because it is difficult to manage these two sources of information together since they are complementary to each other. Anyway, there had been attempts to combine contour and skeleton however the result was not much satisfactory. Unlike previous complex representations we propose to combine the contour descriptor with the thickness spectrum and new skeleton-geodesic distance distributions. By definition, skeleton-geodesic distances are computed between pairs of shape regions and are routed on the object skeleton. Distance distributions are stable to flexible deformations of the object parts if deformations do not brake the topology and local thicknesses are not changed too much. Skeleton-geodesic distance distributions are more informative than histograms of geodesic distances since they contain information about local thicknesses and local directions. The overall complex descriptor is called Contour-Thickness-Distance descriptor. In the paper we explain why we need such a complex solution and how we can benefit from it. We also provide quite efficient algorithms for construction of the mentioned descriptors using continuous skeleton. Extensive experimentation have been conducted. Obtained results indicate that proposed shape representation is more successful than other modern descriptors.

Email addresses: nikita-lomov@mail.ru (N. A. Lomov[1]), sersid@gosniias.ru (S. V. Sidyakin[2]), viz@gosniias.ru (Yu. V. Vizilter[3])

Keywords: 2D Shape Classification, Feature extraction, Histograms, Computational geometry, Skeletons, Thickness spectrum, Skeleton-geodesic distances

1. Introduction

The study of 2D shapes is hampered by the fact that color and texture information contained in conventional color images is not available. Therefore traditional approaches to the object classification in color images can not be directly applied.

According to the latest research [1], [2], contour and skeleton representations are popular in the shape domain. Contour-based approaches [3], [4], [5], [6] are often good at representing detailed shape information and robust against occlusion to a certain extent, but they are sensitive to articulation and non-rigid shape deformations [7]. Skeleton-based approaches [8], [9], [10], [11], [12], [13], [14] can cope well with non-rigid deformations and have structural data as the maximum inscribed empty circles (discs), centered on the shape skeleton, carry information about the thickness of the figure parts.

At the same time, if we look closely at 2D figures in the popular benchmarks it is obvious that any class of figures can be quite diverse both locally and globally because of the big variability in poses, various distortions and occlusions. And what is more, figures of different classes can have very similar parts. In this case, the important differences between object classes can be revealed on the basis of distinctions of mutual configurations of similar parts specific to each class. In our opinion, local and global analysis needs to be considered in conjunction to account for all possible situations, because even the small contour fragment can be extremely informative and can help separate one shape from another.

Earlier some progress was made in the field of shape matching. A number of feature descriptors [15], [9], [10], [3], [4], [5], [6] has been developed. Nevertheless, many researchers admit that the matching algorithms have several major drawbacks. The fact is that the matching based classification algorithms are built upon the weighted nearest neighbor method and require a comparison between the test shape and all shapes in the dataset. If there are few examples in the dataset, matching algorithms are not capable to fully consider large differences between shapes of one class. When the number of examples for training is large, pairwise comparison with all shapes takes

an unacceptable amount of time. Therefore, matching approaches are not preferable.

That is why many scientists begin to apply the powerful learning approaches and design new convenient shape feature representations for recognition and classification tasks. Modern contour descriptors are: normalized and transformed segment of the contour ("signature") [2], [16], a set of histograms of distances from each point of the contour segment to the remaining points on this segment (context of the contour segment), bag of contour fragments (BCF) [17]. Modern skeleton descriptors are: ordered normalized set of maximal inscribed circles [2] and bag of the shortest skeleton paths (BSP) [1]. All of these descriptors can be represented in the vector form (e.g., 1D histogram). Such way of representation is convenient for comparison of shapes and for rapid search in the database because shape comparison and search are reduced to vector comparison, to which machine learning methods can be directly applied.

Bag of Contour Fragments [17] is a special one among the mentioned descriptors because BCF introduces the successful bag framework to shape recognition. It also achieved state-of-the-art results on challenging datasets when it was proposed. However, it is usually not enough to analyze contour fragments if we want to assign a correct class to an object. For the Animal dataset a typical example of such situation is the monkey image (Fig. 1b) that belongs to the monkey class (Fig. 1c) but is classified as a leopard (Fig. 1a) by BCF. For the presented case it is clear that the monkey (Fig. 1b) can be recognized correctly if we consider characteristic thicknesses that are typical for two classes. Contour nature negatively affects BCF in case of the flexible deformations inherent in many real objects due to their natural mobility. Consider another example from the Animal dataset (Fig. 2). In this case the bird contours (Fig. 2b) vary greatly however the shape structure remains unchanged. Therefore, we can trace all typical parts of the bird (Fig. 2b) in the representative (Fig. 2c) of the flying bird class and note the absence of some bird parts in the representative (Fig. 2a) of the dolphin class. Therefore, BCF can be improved by global information about the structure and thickness.

There are two major approaches [1], [2] for enriching contour descriptors by means of structural information. In both of them the structural information has skeletal nature because the shortest paths between skeleton end-points are considered and the radii of inscribed circles are calculated at certain path points. Unfortunately, sequences of radii are much less vari-

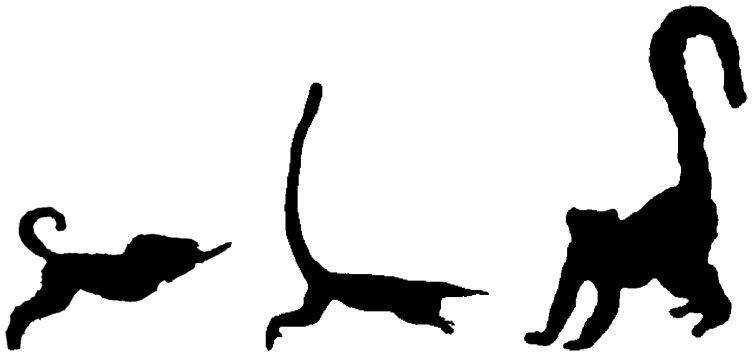


Figure 1: Shapes from the Animal dataset. BCF descriptor assigns the leopard class (a) instead of the monkey class (c) to the monkey shape (b).

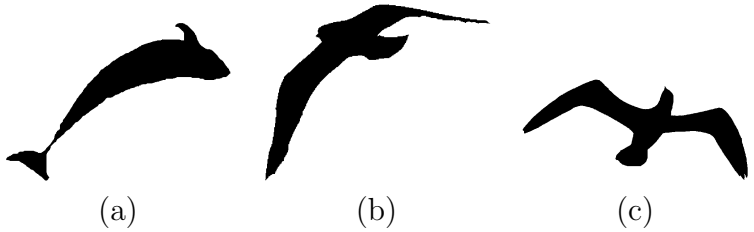


Figure 2: Shapes from the Animal dataset. BCF descriptor assigns the dolphin class (a) instead of the flying bird class (c) to the flying bird shape (b).

ous and unique when compared to the shape contexts of contour fragments. Therefore they have a weak descriptive ability. Moreover, they are not stable due to skeleton instability. Finally, a common drawback of BCF and BSP descriptors lies in the fact that they describe only local shape fragments and do not analyze global shape characteristics.

However, the shape skeleton can be useful not as an independent descriptor, but as a tool for rapid analytical computing of descriptions of some other kind. In particular, we consider two descriptors of alternative nature. We use a pattern spectrum and propose a special type of distance histograms, where distances are computed between pairs of areal shape elements.

The pattern spectrum [18] with a disc-structuring element describes the thickness distribution of the whole binary object. Hence, it is sensitive to the size of shape parts. The pattern spectrum has areal nature and therefore

it is a stable descriptor. Note that the calculation of the spectrum directly by definition occupies an unacceptable amount of time. In this paper we propose an accurate algorithm for spectrum calculation by means of a discrete thickness map [19]. In our approach values of the discrete thickness map are calculated analytically by continuous skeleton representation.

Histograms based on the statistics of pairwise distances between shape elements are also well-known in the literature [20], [21], [22], [23], [24]. Such histograms are not restricted to any specificity of the object domain, shape limitations and data dimension. Therefore they are capable of describing 2D, 3D and even multidimensional data. Classification of these histograms can be carried out on the basis of distances between shape elements (e.g., points) and on the basis of statistics that are calculated from distances. Euclidean distances, geodesic distances and diffusion distances [21] between shape points are usually applied. Full matrix of pairwise distances, shape distributions [22], shape context [23], the average distance from each point to all others [24] are often used as the descriptive statistics.

In this paper we propose 1D histograms of skeleton-geodesic distances, thicknesses and directions calculated for pairs of shape elements aiming to enrich the contour descriptor. We consider 2D and 3D combinations of these histograms too. Further, we will use a term "skeleton-geodesic histograms" for short. Skeleton-geodesic distance between two figure points is defined as the length of the shortest path along the skeleton between the projections of these points on the skeleton. Note that the proposed histograms are not skeleton descriptors. Histograms of skeleton-geodesic distances are descriptors of the areal shape elements. These areal elements are formed due to the following reasons. Obviously, several figure points may have one nearest point on the skeleton. Accordingly, a region of attraction of shape points is associated with each skeleton point. Therefore, a histogram of pairwise distances between the regions of attraction of skeleton points can be collected. Continuous skeleton representation in the form of a geometric graph allows naturally represent a skeleton-geodesic path in the form of a continuous line along skeleton edges and easily determine its length. We consider regions associated with skeleton edges since an elementary part of continuous skeleton is an edge and not a point. More specifically we use proper regions [14]. They define a partition of the figure into disjoint polygonal regions. Areas of proper regions are quickly calculated analytically by means of continuous skeleton [8], [25], [14]. The product of areas of two proper regions determines the contribution of each pair of regions in a histogram. Average thicknesses

and directions of proper regions are used too.

Thus, we propose to combine three advanced shape descriptors in one, which is called Contour-Thickness-Distance descriptor. CTD descriptor (presented as a vector) combined with SVM (Support Vector Machine) learning allows to receive the classification results surpassing the results of other modern descriptors [26], [27], [1], [28], [17], [5], [16], [2], [29], [6] on the most difficult dataset of binary figures - Animal [2]. SVM is used to allow fair comparison with the results obtained in [17], [26], [1]. CTD is superior to the listed descriptors on the other common datasets - MPEG-7 [4], Swedish leaf [30], ETH-80 [31]. It is also shown that combination of BCF and skeleton-geodesic histograms is more powerful than combination of BCF and BSP.

In summary, the proposed CTD and skeleton-geodesic histograms have several important highlights:

- CTD provides a compact shape representation, which is a vector;
- Thickness spectrum and proposed histograms allow to enrich BCF;
- Thickness spectrum and proposed histograms are quickly and precisely calculated by continuous skeleton;
- CTD allows us to deal with shapes that have self-occluded parts and high variability in poses.

The rest of the paper is organized as follows. We review the related works in Section 2. Then we describe the details of CTD shape descriptor in Section 3 focusing on the thickness spectrum and the skeleton-geodesic histograms because the contour component is adopted from [17]. We evaluate our descriptor on several most used shape benchmarks. We do parameter selection and achieve state-of-the-art performance in shape classification in Section 4. Discussion is provided in Section 5. Section 6 concludes this paper.

2. Related Works

Let us briefly go through the recent progress in shape matching and classification over the last decades. The general principle adopted for shape representation and classification involves two basic steps: representation of an object retaining the dominant features and matching the features with a suitable distance measure. We have seen several methods in the literature, which have been developed to achieve excellent results in shape based

representation and classification task, more in particular from the field of mathematical morphology introduced by Matheron [32] and Serra [33]. It is found that, among morphology based techniques the medial axis transformation or continuous skeleton [8] is one of the popular and useful descriptor. It plays an essential role for shape representation and analysis in many applications such as content-based image retrieval systems, character recognition systems, gesture analysis and so on. Skeletons can capture shape articulation in the better form than contours [34].

For example, skeletons can be organized in the form of attribute-relation graphs (ARG) for the matching purpose. One of the examples for ARG is shock graphs introduced by Siddiqi et al. [12], [35]. Shock Graph is an abstraction of skeleton of a shape onto a Directed Acyclic Graph (DAG) and is constructed using the specialized grammar called Shock Grammar [36]. In this approach the skeleton points are labeled according to the radius function at each point. The primitives in the skeleton, called as shocks, such as branch points, end-points and skeleton segments contain both geometrical and topological information. The concept of bone graph is an extension of shock graph, that retain the non-ligature structures of the shock graph and are more stable [37]. There are several algorithms which are proposed in order to have an efficient match of these skeleton graphs. Sebastian et al. [11] proposed a matching approach for node-attributed trees which measures the edit distance between two graphs. Involvement of the complex edit operation in this method results in high computational complexity.

The techniques which involve graph matching by finding the correspondence between nodes through the conversion of skeleton graphs to skeleton trees require heuristic rules to select the root node [38], [39]. The major drawback of this method is that a small change in the shape causes the root to change resulting in the significant change in the topology of the tree representation. Apart from this, conversion from the graph to a tree structure results in loss of significant structural information and hence leads to the wrong match [10].

Bai et al. [10] proposed a method based on the path similarity between the end points of the skeleton graphs. In this method, the geodesic paths between skeleton end points are obtained and are matched using Optimal Subsequence Bijection (OSB) method. Unlike other methods, this approach does not involve the conversion of skeleton to graph and then matching the graph, which is still an open problem. This method addresses the problem of shapes having similar skeleton graphs, but different topological structures

and also the shapes having different skeletons with visually similar shapes.

Baseski et al. [40] proposed a method for shape representation namely rooted-depth-1 tree which is based on Aslan skeleton [9] This is a robust technique which deals with shape articulation. Demicri et al. [41] proposed a technique called skeletal shape abstraction from examples. In this approach, a many to many correspondence between the object parts is used to match the object skeletons. This method is invariant to shape articulation and appears to be more suitable for object detection in edge images. However, the tree abstractions are not specific enough to prevent hallucinating of target object in clutter. Earth Mover's Distance, under L^1 norm, is used as a metric during the matching phase.

Bai and et al. [42] proposed Active skeleton representation for non rigid object detection. This method provides a powerful representation for the object by forming a tree union structure, using the skeleton segments and the contour fragments of the object. The object detection phase makes use of sum and max algorithm. This method requires less training sample and is simple and effective.

Demirci et al. [43] proposed a many to many attribute matching method, presenting a powerful graph representation technique for the objects in the geometric space. In this method, the dissimilarity value between the object trees is computed by performing many to many matches between the vertices of the trees. The main highlights of this approach are its effectiveness and the robustness.

Shu and Wu [44] proposed contour points distribution histogram as the shape descriptor. This method takes into account the distribution of points on the object boundary. Earth Mover's Distance is used as a metric to compare the histograms in the matching process. This method is translation and scale invariant and is more similar to human perception. Though it is not the best one with respect to performance it gives low time complexity.

Wang et al. [45] proposed a shape descriptor which represents the object contour by a fixed number of sample point. Each sample point is associated with a height function. The height function is based on the distances of the other sample points to its tangent line. Dynamic programming algorithm is applied in the matching phase. This method is invariant to translation, rotation and scaling and it is capable of handling nonlinear deformations of the objects.

Shen et al. [46] proposed a method which finds the common structure in a cluster of the object skeleton graph. The agglomerative hierarchical cluster-

ing scheme is used for the purpose of clustering. The correspondence between end nodes is obtained by using Optimal Subsequence Bijection. The drawback of this method is the high time complexity of agglomerative hierarchical clustering.

Recently, researchers begin to apply the powerful discriminative models to shape classification [17], [26], [2], [1]. Different from the exemplar-based shape matching, in paper [17], the compact contour-based shape representation (Bag of Contour Fragments - BCF) was proposed. Wang et al. [17] were inspired by the huge progress in image classification and representation under the Bag-of-Words framework [47]. The shape is decomposed into contour fragments and contour fragments are transformed into shape codes. Then, a statistical histogram of shape codes is used to represent each shape. Similarity of shapes can be directly computed from these histograms. Matching shapes based on this new shape representation does not explicitly give correspondences between contour fragments. However using a classifier for shape classification is much more efficient than using the typical matching algorithms such as Hungarian, thin plate spline, dynamic programming. BCF has received further development in two papers [26], [1]. In paper [26], many shape vocabularies were used instead of the one in BCF. In paper [1], the BCF and skeleton feature descriptor combination was used to improve the performance of shape recognition. Skeleton was represented in the Bag of Skeleton Paths (BSP) framework - an extension of BCF for skeleton based mid-level representation. Combination of BCF and BSP was presented as concatenation of two feature vectors. Weights of the contour features and skeleton features are automatically learned by the discriminative model (linear SVM). This is an advantage of the BCF and BSP combination method compared to the Integrated Contour and Skeleton (ICS) descriptor [2], which has to fine tune the weights between contour and skeleton models.

In this paper we also address the problem of shape descriptor construction. However, unlike the complex contour-skeleton descriptors above we explore the complementary information between contour, thickness and skeleton-geodesic distances hoping to improve the performance of shape classification.

3. Contour-Thickness-Distance descriptor

To build CTD of the shape we need to compute its continuous skeleton. Because of that, the first subsection describes a concept of skeleton, important definitions and a rapid algorithm for continuous skeleton computation.

The second subsection explains how to build the thickness part of CTD and the third one - how to build the distance part of CTD. Let us proceed with continuous skeleton.

3.1. Continuous Skeletons of the Binary Shapes

Originally, the concept of skeleton was introduced for the continuous objects by Blum [48]. According to [48], skeleton of the object is the locus of the centres of the maximal disks that can be inscribed inside the object boundary. Then the skeleton concept was generalized for the discrete binary images (2D shapes) by discrete and continuous approaches. The discrete approach applies a morphological transformation based on distance maps, thinning, Voronoi diagrams of boundary points [49], [12] to the binary shape (Fig. 3a) and constructs the discrete skeleton image (Fig. 3b). The main advantage of the discrete approach is the simplicity of the algorithms and a graphic visualization of the skeleton in the source raster format. The continuous approach approximates the binary shape by the polygonal figure (Fig. 3c) and constructs the skeleton by means of analytic geometry (Fig. 3d). This skeleton is considered as a continuous skeleton of binary shape. The main advantage of the continuous approach [12] is a continuous skeleton representation of binary shape as a geometrical graph with a radial function, which determines the width of shape parts. Continuous skeleton representations allow the usage of graph theory and computational geometry algorithms for image shape analysis and recognition. The second advantage of the continuous approach is its high computational efficiency. Experiments show that the continuous skeleton computation procedure is significantly faster than the discrete skeleton computation procedure. Even taking into account the fact that the discrete approach is more suited for parallel computing [8]. Other advantages are mathematical rigor and information content.

The continuous skeleton of the binary shape can be obtained from a Voronoi diagram of the polygonal shape boundary (Fig. 4). The polygonal shape boundary is a union of linear segments and vertices (points), which are considered as the Voronoi sites. The Voronoi diagram of these sites is generated and the skeleton is extracted as a subset of the diagram. The continuous skeleton has a fairly simple structure because it is a geometrical planar graph, which edges are described by straight line segments and quadratic parabola segments. Skeleton vertices are convex vertices of the polygonal figure (one degree vertices) and also points - centers of inscribed circles, tangent to the figure boundary in three or more points (three and more degree vertices).

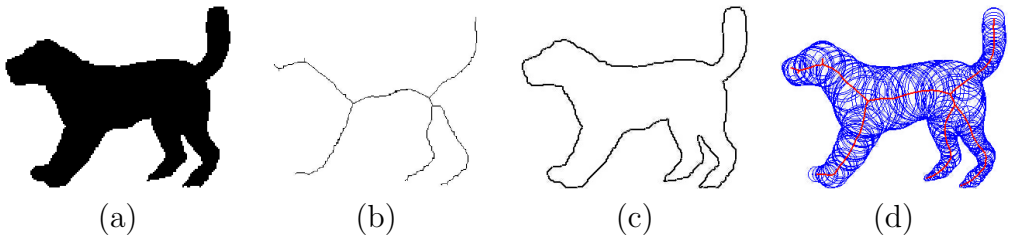


Figure 3: The concept of skeleton: the binary shape (a); the discrete skeleton (b) obtained by thinning the binary shape; the polygonal shape boundary (c); the continuous skeleton (d) of the polygonal shape boundary. Inscribed circles (d) illustrate the values of radial function at skeleton points.

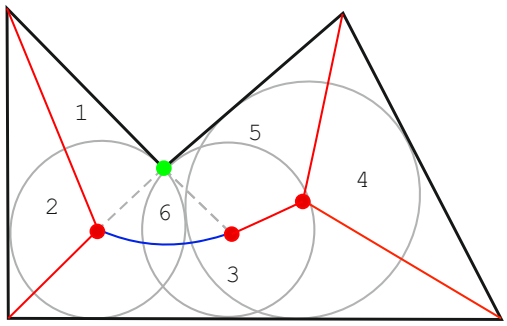


Figure 4: The continuous skeleton and the Voronoi diagram. Voronoi sites: site-segments are shown in black, the site-point is green. Voronoi regions are represented with numbers. Skeleton linear edges, internal end-points of edges are shown in red. The parabolic edge is shown in blue. Inscribed circles for end-points are grey.

The radial function is defined in each skeleton point as the radius of inscribed circle centered in this point. Note that the continuous skeleton of the polygonal figure always exists and it is unique.

There are known effective $O(n \cdot log(n))$ algorithms for the Voronoi diagram construction for the case of a simple polygon [50] and for the case of complex multiply-connected polygonal figures [8]. In our implementation we stick to more general Mestetskiy algorithm [8], [25]. According to it, the polygonal representation of the binary shape boundary is obtained by Minimal Perimeter Polygon (MPP) algorithm. It is guaranteed that the polygonal representation has no self-intersections. The idea of the skeleton computation [8] is to build a boundary adjacency tree from the polygonal description by a

sweepline method and a Delaunay graph from the boundary adjacency tree. The Voronoi diagram is constructed from the Delaunay graph. Finally, the continuous skeleton is extracted as a subset of the Voronoi diagram and each skeleton edge together with its radial function gets a parametric description for convenience and compactness.

Thus, the continuous skeleton was discussed. Let us have a look at the thickness part of CTD descriptor.

3.2. Thickness spectra and thickness maps via continuous skeletons

The thickness part of CTD is closely related to pattern spectrum [18]. The original pattern spectrum of the binary image has two parts: a figure part and a background part. In this paper we are interested only in the figure part.

Let X be a binary figure located on the frame $K : X \subseteq K$, $E(r)$ a translatable and scalable structuring element that is a compact subset of the plane $\mathbf{R}^2$ with a size $r \in \mathbf{R}, r \geq 0$. Define the pattern spectrum of X as

$$PS(r) = -\frac{\partial S(X \circ E(r))}{\partial r},$$

where $\circ$ denotes an opening, $S(X \circ E(r)) = \|X \circ E(r)\|_{L^1}$ denotes the area of $X \circ E(r)$.

Let us show that the pattern spectrum with the disc structuring element $B(r)$ can be expressed in terms of a thickness map [19].

Define the thickness map of X for each point $q \in X$ as

$$T(q) = max_{r \in \mathbf{R}}\{q \in B(r) \subseteq X\}. \quad (1)$$

Thus, the value in each point q of T is equal to the maximum size r of the structuring element B that is fully inscribed in X and is covering q.

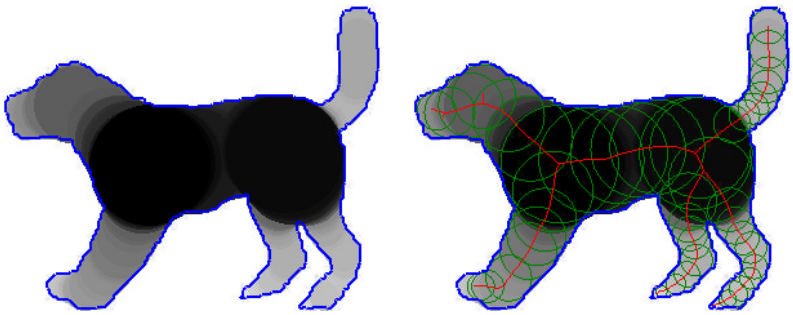
Let the relationship between the thickness map and the opening of X by B be given by:

$$X \circ B(r) = \{q : T(q) \geq r\}.$$

Consider the function $\eta(q, r)$ and the measure $\mu(r)$:

$$\eta(q, r) = \begin{cases} 1, T(q) \geq r; \\ 0, T(q) < r. \end{cases}$$

$$\mu(r) = ||\eta(q, r)||_{L^1} = \int \eta(q, r) dq.$$



(a) (b)

Figure 5: (a) the thickness map of the binary object (thickness values are coded with grey color); (b) the continuous skeleton and inscribed circles overlaid on the thickness map.

Since

$$\mu_X(r) = ||X \circ B(r)||_{L^1}, r \geq 0,$$

we see that $\mu(r)$ gives an alternative definition of the pattern spectrum:

$$PS(r) = -\partial\mu_X(r)/\partial r.$$

Therefore, the disc pattern spectrum is a distribution density function of the thickness map. In other words, the disc pattern spectrum is a thickness spectrum.

In practice it is convenient to use the discrete pattern spectrum $PS_{\Delta r}$:

$$\begin{aligned} &PS_{\Delta r}(r_i) = -\frac{S\left(X \circ B\left(r_i\right)\right) - S\left(X \circ B\left(r_{i+1}\right)\right)}{\left(r_i - r_{i+1}\right)}, \\ &r_i \in RN = \{i\Delta r\}, i = 0, ..., i_{max},\ i_{max} = \lfloor r_{max}/\Delta r \rfloor, \\ &r_{max} = max\left\{r : B(r) \subseteq X\right\}, \Delta r > 0, \end{aligned}$$

where Δr denotes a scale step, i_{max} is a maximum integer that is not greater than $r_{max}/\Delta r$.

In this case the discrete thickness map $T_{\triangle r}$ (Fig. 5) is defined similar to (1):

$$T_{\triangle r}(q) = max_{r \in RN}\{q \in B(r) \subseteq X\}. \tag{2}$$

Consider the characteristic function η_i of i-th $T_{\Delta r}$ region such that:

$$\eta_i(q) = \begin{cases} 1, T_{\Delta r}(q) = r_i; \\ 0, T_{\Delta r}(q) \neq r_i. \end{cases} \tag{3}$$

Using (3), we get the histogram h_i of $T_{\Delta r}$

$$h_i = ||\eta_i||_{L^1} = \int \eta_i(q)dq.$$

It follows that the discrete pattern spectrum is proportional to h_i:

$$\begin{aligned} &PS_{\Delta r}(r_i) = -(S(X \circ B(r_i)) - S(X \circ B(r_{i+1})))/(r_i - r_{i+1}) = \\ &= -(||X \circ B(r_i)||_{L^1} - ||X \circ B(r_{i+1})||_{L^1})/(i\Delta r - (i+1)\Delta r)) = \\ &= -(\mu_X(r_i) - \mu_X(r_{i+1}))/(-\Delta r) = ||\eta_i||_{L^1}/\Delta r = \\ &= \left(\int \eta_i(q)dq\right)/\Delta r. \end{aligned}$$

It is clear that if $\Delta r = 1$, then the discrete pattern spectrum is equal to the histogram of the discrete thickness map:

$$PS_{\Delta r}(r_i) = \int \eta_i(q)dq = h_i.$$

Therefore, a possible approach to thickness spectrum calculation is to build discrete thickness map and take its histogram.

In our case, the calculation of the thickness map is particularly justified. Since we calculate features describing the thickness in individual parts of the figure, we need quantitative and spatial distributions of thickness.

The keystone idea of fast thickness map computation is derived from the fact that centers of all inscribed empty discs belong to the figure skeleton. This allows us to define the thickness function $T(q)$ for point q of the continuous binary figure X in terms of the skeleton $Sk(X)$ and the radial function $\rho(p), p \in Sk(X)$:

$$T(q) = \max_{p \in Sk(X)} \{\rho(p) : q \in B(p)\}, \tag{4}$$

where $B(p)$ is the maximal inscribed circle centered in p.

The skeleton can be represented as a union of subgraphs:

$$Sk(X) = \bigcup_{i=1}^{k} Sk_i$$

Consider Sk_i silhouettes:

$$G(Sk_i) = \bigcup_{p \in Sk_i} B(p), i = 1, \ldots, k.$$

Let the thickness function at the point q in relation to the subgraph Sk_i be given by:

$$T_i(q) = \begin{cases} \max_{p \in Sk_i}\{\rho(p) : q \in B(p)\} & \text{if } q \in G(Sk_i); \\ -1 & \text{otherwise.} \end{cases}$$

Consider equation (4) as

$$T(q) = \max_{i=1,\ldots,k}(T_i(q)).$$

A skeleton partitioning into subgraphs can simplify the procedure of calculating the thickness value at the point due to the following theorem.

Theorem 1. *Let $Sk_i = p(t) = (x(t), y(t)), t \in [0, 1]$ be a simple parametric curve such that $Sk_i \in Sk(X)$. The radial function $\rho(p)$ increases monotonically along Sk_i. The point q belongs to the silhouette $G(Sk_i)$. Then either the point q is covered by the maximal circle $B(p(1))$ of Sk_i, or there is a point $p', p' \in Sk_i$ such that $\rho(p')$ takes the value of $T_i(q)$ and q lies on the boundary of the inscribed circle centered at p'.*

Proof. Consider the following function:

$$g(q, t) = r(t) - d(q, p(t)),$$

where $r(t) = \rho(p(t))$ and d is the Euclidean distance between the points. The function $g(q, t)$ is non-negative if the point q is covered by an inscribed circle centered at the point $p(t)$ and negative otherwise. Suppose that q does not belong to the maximal circle of the curve. Due to the continuity of the function $g(q, t)$ when q is fixed, the set $A = \{t : g(q, t) \geq 0\}$ is a compact set, and $t', t' < 1$ is a maximum number of this set. It is obvious that $r(t') = T_i(q)$, but $g(q, t') = 0$ because of the continuity of the function $g(q, t)$ with reference to t and the inconstancy of its sign. □

Note. In the proof of the theorem we can replace the function $g(q, t)$ with $f(q, t) = r^2(t) - d^2(q, p(t))$. It is more convenient to operate with $f(q, t)$.

The usage of Theorem 1 is illustrated in Fig. 6.

Thus, we need only two steps for calculating the value of $T_i(q)$:

1. If $f(q, 1) \geq 0$, assign $T_i(q) = r(1)$.

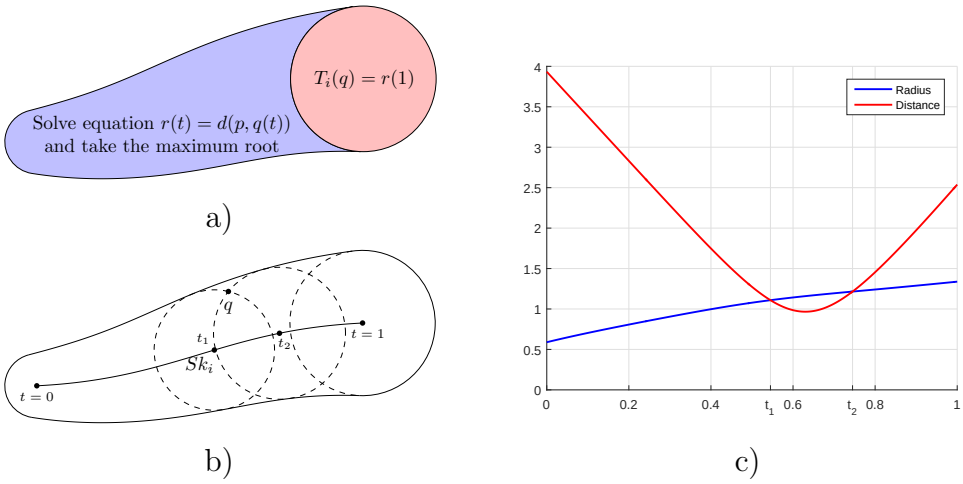


Figure 6: Search for covering circle of maximum radius in line silhouette. (a) the scheme of the approach; (b) the boundary circles for q; (c) the plots for $r(t)$ and $d(q, p(t))$. In this case $T_i(q) = r(t_2)$.

2. Otherwise, solve the equation $f(q, t) = 0$ and assign $T_i(q) = r(t_{\max})$, where $t_{\max}$ is the maximum root.

As already mentioned, the disc pattern spectrum can be approximated by the distribution of values of T_i at the points of a uniform grid N overlaid on the figure X with sufficiently small step (for example, points with integer coordinates). Computational procedure is organized as follows:

1. Represent skeleton of the figure X as a union of monotonic subgraphs $Sk_i, i = 1, \ldots, k$.
2. For all $q \in N$ take $T(q) = -1$
3. For all $i = 1, \ldots, k$
4. Change the value of $T_i(q)$ to $\max(T(q), r(1))$ for grid points that belong to the maximal circle of Sk_i
5. Find $T_i(q)$ by solving the equation $g(q, t) = 0$ for all other grid points in $G(Sk_i)$. To each $T_i(q)$ assign the value of $\max(T(q), T_i(q))$

If X is approximated by a polygon, the continuous skeleton can be efficiently obtained from the Voronoi diagram of its boundary elements (site-points and site-segments) cutting-off edges and vertices related to concave

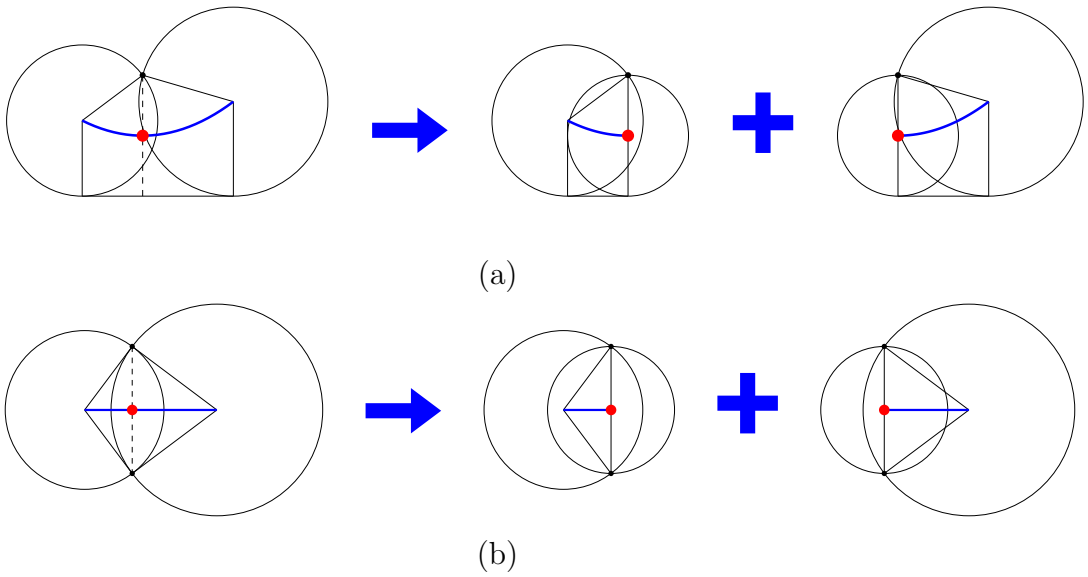


Figure 7: Subdivision of non-monotonic parabolic (a) and hyperbolic (b) edges into monotonic parts. The cut (dashed line) passes through the point of radial function minimum (shown in red), which becomes the center of the small end circle for both new edges. Edges are shown in blue. End circles for all edges are displayed.

vertices of the figure [8]. Therefore, all skeleton edges get explicit analytical description [8]. The edges of three types are distinguished depending on the types of site generators: linear edge (both sites are segments — straight lines), hyperbolic (both sites are points) and parabolic (one of the sites is a segment, the other is a site-point).

A skeleton may contain edges that are not monotonic. In the non-monotonic edge there is always only one minimum of the radial function. So we divide such edges into two monotonic sections and consider them as separate skeleton subgraphs during the computation of thickness map (Fig. 7).

It remains to clarify the question of solving the equation $f(q,t) = 0$ for each of the three types of edges. By (x_1, y_1), (x_2, y_2) denote coordinates of the edge ends. By r_1, r_2 denote radii of end circles such that $r_1 < r_2$. By (x_4, y_4) denote site-point coordinates of hyperbolic and parabolic edges.

1. The radial function $r(t)$ varies linearly during the uniform movement along the linear edge. So the linear edge is parameterized as follows:

$$x(t) = x_1 + t(x_2 - x_1);$$
$$y(t) = y_1 + t(y_2 - y_1);$$
$$r(t) = r_1 + t(r_2 - r_1).$$

The equation $f(q,t) = 0$ becomes quadratic as a result of parameterization:

$$(x_1 + t(x_2 - x_1) - x_0)^2 + (y_1 + t(y_2 - y_1) - y_0)^2 = (r_1 + t(r_2 - r_1))^2.$$

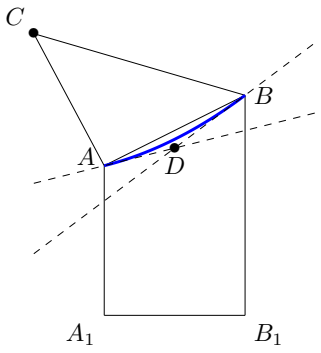


Figure 8: The control point D of a parabolic edge; C is a site-point; A_1B_1 is a site-segment; AB is a parabolic edge, shown in blue.

2. The parabolic edge is a special case of the Bezier curve of order 2. Two of its control points coincide with the ends of the edge. The third is the intersection point of tangents to the parabola at the end points (Fig. 8). We have the following parameterization:

$$x(t) = (1-t)^2 x_1 + 2tx_3 + t^2 x_2;$$
$$y(t) = (1-t)^2 y_1 + 2ty_3 + t^2 y_2;$$
$$r(t) = \sqrt{(x(t) - x_4)^2 + (y(t) - y_4)^2},$$

where (x_3, y_3) are the coordinates of the intermediate control point. The equation $f(q,t) = 0$ takes on the form:

$$(x(t) - x_0)^2 + (y(t) - y_0)^2 = (x(t) - x_4)^2 + (y(t) - y_4)^2 \Rightarrow$$
$$-2x(t)x_0 + x_0^2 - 2y(t)y_0 + y_0^2 = -2x(t)x_4 + x_4^2 - 2y(t)y_4 + y_4^2,$$

That leads to a quadratic equation.

3. The hyperbolic edge is parameterized as follows:

$$x(t) = x_1 + t(x_2 - x_1);$$
$$y(t) = y_1 + t(y_2 - y_1);$$
$$r(t) = \sqrt{(x(t) - x_4)^2 + (y(t) - y_4)^2}.$$

The equation $f(q, t) = 0$ takes on the form

$$-2x(t)x_0 + x_0^2 - 2y(t)y_0 + y_0^2 = -2x(t)x_4 + x_4^2 - 2y(t)y_4 + y_4^2$$

That leads to a linear equation.

Thus, to calculate the thickness at the point in relation to the edge it is sufficient to solve the simplest equation.

When the thickness map is computed. We calculate the thickness spectrum for the whole figure, local thickness histograms and an average thickness for each figure segment S (each edge silhouette). The thickness spectrum with l elements is calculated as follows:

$$PS(a) = \#\{p \in N \cap X : \lfloor (l - \epsilon)\frac{T(p)}{r_{max}} \rfloor = a\}, \tag{5}$$

where symbol # denotes the number of elements in the set; ϵ denotes a very small positive number; r_{max} denotes the maximal radius of the circle inscribed in the figure. The equation for the local thickness spectrum can be obtained by changing X in equation (5) to S. The average thickness of S is calculated as follows:

$$T(S) = \frac{\sum_{p \in N \cap S} T(p)}{\#\{p : p \in N \cap S\}}. \tag{6}$$

In contrast to the discrete approach [51] for thickness map construction, we do not use the rasterization of skeleton and inscribed circles by Bresenham algorithms. This is because we want to calculate the exact real values of thickness function. This is not possible with the discrete approach. In our algorithm, discretization is reflected only in the fact that we calculate the function values in a discrete set of points.

Thus, the thickness component of CTD descriptor was discussed. We can proceed to the skeleton-geodesic histograms.

3.3. Histograms of skeleton-geodesic distances, directions and thicknesses

The distance component of CTD is closely related to histograms of geodesic distances. It is known that histograms of geodesic distances are invariant to shift, rotation and scale when appropriate normalization is involved. They are also stable to bending of fat curves composing the figure if bendings do not change the topology. However, the calculation time of such histograms is extremely large, since it is necessary to calculate geodesic distance for every pair of figure points. Obviously, the complexity is proportional to the square of figure points.

Here we propose a skeleton-geodesic distance (SGD) between the points p, q of the figure X instead of the geodesic distance (GD) $d_{Geod(X)}(p,q)$. The skeleton-geodesic distance $d_{Geod(Sk(X))}(p,q)$ is equal to the geodesic distance $d_{Geod(Sk(X))}$ on the skeleton $Sk(X)$ between the projections $p_{Sk(X)}$, $q_{Sk(X)}$ of points p, q on the skeleton $Sk(X)$:

$$\forall p,q \in X : d_{Geod(Sk(X))}(p,q) = d_{Geod(Sk(X))}(p_{Sk(X)}, q_{Sk(X)}).$$

The projection $p_{Sk(X)}$ of the point p on the skeleton $Sk(X)$ is the nearest point e on the skeleton in terms of the Euclidean distance:

$$\forall p \in X : p_{Sk(X)} = \arg\min_{e \in Sk(X)} \|p-e\|,$$

$$\|p-e\|^2 = (x_p - x_e)^2 + (y_p - y_e)^2,$$

where (x_p, y_p), (x_e, y_e) are the coordinates of p, e.

In the strict sense, SGD is not a distance because

$$\forall p,q \in X : p \neq q, p_{Sk(X)} = q_{Sk(X)} \Rightarrow d_{Geod(Sk(X))}(p,q) = 0.$$

SGD is a factor-distance between regions of attraction. The region of attraction of the point e is the set of figure points $\{p \in X\}$, for which e is the nearest skeleton point.

There are some fair facts indicating that properties of the histogram of geodesic distances (HGD)

$$HGD(d) = \#\{p,q \in X : d_{Geod(X)}(p,q) = d\}$$

are similar to properties of the histogram of skeleton-geodesic distances (HSGD)

$$HSGD(d) = \#\{p,q \in X : d_{Geod(Sk(X))}(p,q) = d\}.$$

Firstly, both histograms are stable to bending of fat curves that compose the figure. If bendings do not change topological characteristics of the figure or its skeleton (the number and location of cycles in it). Secondly, if the figure X comes close to its skeleton $Sk(X)$ in the process of thinning, then HGD of the thin shape obviously comes close to HSGD. Therefore, we can conclude that for figures composed from rather thin "fat curves" (such as, for example, silhouettes of people and animals), HSGD should be close enough to HGD. At the same time, HSGD computation is fast if the calculation is carried out according to the equation:

$$HSGD(d) = \sum_{p,q \in Sk(X), d_{Geod(Sk(X))}(p,q)=d} S(p)S(q),$$

where $S(p)$ and $S(q)$ are the areas of regions of attraction for the corresponding points:

$$S(p) = \#\{q \in X : q_{Sk(X)} = p\}.$$

The computation time of HSGD is proportional to the square of skeleton points. The number of skeleton points is usually less than the number of figure points by one order of magnitude. Therefore, the difference in computations between HSGD and HGD should be two orders of magnitude.

Further, we introduce several combined histograms because information about thickness and directions can be extracted from the figure skeleton.

The histogram of directions-distances (HDD) is defined as follows

$$HDD(d, \phi) = \#\{p, q \in X : d_{Geod(Sk(X))}(p, q) = d, |\theta(p) - \theta(q)| = \phi\},$$

where $\theta(p)$ is the angle of inclination of the tangent at the point e; e is the projection $p_{Sk(X)}$ of the point p on the skeleton $Sk(X)$.

The histogram of thicknesses-distances (HTD) is defined as follows

$$HTD(d, t) = \#\{p, q \in X : d_{Geod(Sk(X))}(p, q) = d, |t(p) - t(q)| = t\},$$

where $t(p)$ is the local thickness at the point p; $t(p)$ is calculated based on the radius $r(e)$ of maximal inscribed circle centered at the point $e, e \in Sk(X)$:

$$\forall p \in X : t(p) = \max_{e \in Sk(X), \|p-e\| \le r(e)} r(e).$$

The histogram of thicknesses-directions-distances (HTDD) is defined as follows

$$HTDD(d, t, \phi) = \#\{p, q \in X : d_{Geod(Sk(X))}(p, q) = d, \\ |t(p) - t(q)| = t, |\theta(p) - \theta(q)| = \phi\}.$$

Proposed histograms can be calculated using an enriched skeleton for the same time as the histogram of skeleton-geodesic distances. The enriched skeleton is a skeleton containing information about regions of attraction of its points, directions and thickness distributions in regions. Defined histograms are invariant to shift and rotation. HTD has properties of the thickness spectrum. Note that, while the skeleton-geodesic distances are defined for pairs of figure points, rotation invariance is achieved if we consider differences of directions between pairs of points. The thickness can be regarded as a self-contained point characteristic.

So far in subsection 3.3 we have discussed the case when the skeleton description is obtained by discrete methods. In subsection 3.2 we proposed the algorithm of precise thickness map computation via continuous skeleton and learned to calculate local thickness histograms and the average thickness for figure segments. Therefore it is convenient to obtain the histogram of thicknesses-directions-distances by means of the continuous skeleton and the thickness map. Further, we reformulate HTDD and organize calculations based on an analytical description of skeleton edges.

In the case of continuous skeleton, the number of skeleton points is infinite. So we use skeleton edges as primitive elements. General view of all histograms introduced above remains the same. The thickness spectrum is still extracted from the thickness map. However this brings up the question, what is considered by the region of attraction of an edge. It is a polygon, which is cut off from the figure by line segments. These line segments connect the edge ends and their projections to the site generators. This polygon is called a proper region. Proper regions of edges cover the entire polygonal figure completely and intersect each other only by their boundary parts. Therefore, the area of the polygonal figure is equal to the sum of the areas of proper regions. The area of such region is the maximum possible medial width of its bicircle (silhouette) [14]. It can be calculated using the radii r, R of end circles and the distance l between their centers.

The area S_{lin} of the proper region of a linear edge is defined as follows:

$$S_{lin} = (r + R)t, \quad t = \sqrt{l^2 - (R^2 - r^2)},$$

The area S_{par} of the proper region of a parabolic edge is defined as follows:

$$S_{par} = \begin{cases} \Phi(R) + \Phi(r) & \text{if } t \geq 2\sqrt{r(R-r)} \\ \Phi(R) - \Phi(r) & \text{if } t < 2\sqrt{r(R-r)} \end{cases},$$

$$\Phi(z) = \frac{1}{2}(z+p)\sqrt{p(2z-p)}, \quad p = \frac{t^2}{2l^2}(R + r + \sqrt{(R+r)^2 - l^2}),$$

The area S_{hyp} of the proper region of a hyperbolic edge is defined as follows:

$$S_{hyp} = \begin{cases} \Psi(R) + \Psi(r) & \text{if } l^2 + r^2 \geq R^2 \\ \Psi(R) - \Psi(r) & \text{if } l^2 + r^2 < R^2 \end{cases},$$

$$\Psi(z) = \frac{q}{2}\sqrt{z^2 - \frac{q^2}{4}}, \quad q = \frac{1}{l}(\sqrt{[(l+r)^2 - R^2] \cdot [R^2 - (l-r)^2]}).$$

Skeleton edges have several features:

- s is an area of the proper region;
- l is an edge length; it is a length of parabola segment if an edge is parabolic, otherwise it is a distance between edge ends;
- α is an inclination angle of tangent to the axis of abscissas; the tangent is drawn in the middle of an edge, in other words, in the point located halfway from one edge end to another;
- t is an average thickness of a proper region.

To allow more efficient histogram calculation, we agree that the projection of all region points on the skeleton is the middle of the corresponding edge. In this case skeleton-geodesic distances are calculated only between the edge middle points (midpoints). To calculate thickness-related features of an edge we select the grid points belonging to its proper region (Fig. 9). It is easy to do since proper regions are unions of elementary geometric shapes taken from Voronoi diagram. Bypassing of the grid points is performed by means of the scan-line algorithm [52].

The shortest path between any pair of midpoints of edges can be represented as a sequence of transitions from one middle point to another and other midpoints are not visited under a single transition. Such transition is

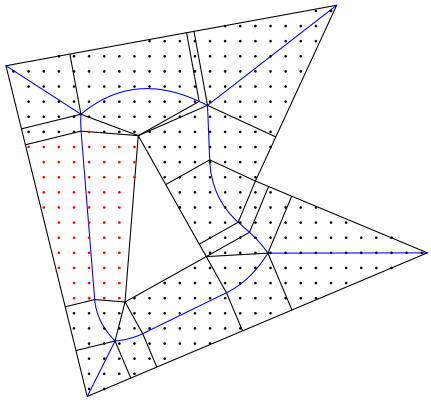


Figure 9: The partitioning of the figure into proper regions. Each proper region corresponds to one skeleton edge. Grid points of one region are shown in red.

possible only for the adjacent edges within the skeleton, where the transition length is half of the sum of the lengths of these edges. In order to reduce the calculation of histogram to the classic problem of finding the shortest paths in a graph, we construct the graph G, where each edge e_i has the associated vertex w_i. If edges e_i and e_j are adjacent in the skeleton, then w_i and w_j are connected in both directions, otherwise the vertices are non-adjacent. We assign the weight to each edge of the new graph. The weight is equal to $(l_i + l_j)/2$. In other words, the weight equals skeleton-geodesic distance between the centers of adjacent edges. An example of the skeleton and its dual graph G is shown in Fig. 10. The shortest path between the vertices of the dual graph corresponds to the shortest path between the centers of edges in the skeleton. Therefore, Johnson algorithm [53] is applied to the graph G to calculate the shortest paths between all pairs of edges.

Skeleton edges have several pairwise features:

- $t_{ij} = |t_i - t_j|$ is the difference between average thicknesses of the edges e_i and e_j;
- d_{ij} is the skeleton-geodesic distance between the edges e_i and e_j (more precise, the distance between their midpoints);
- ϕ_{ij} is the rotation angle between the edges e_i and e_j, calculated as $\max(|\alpha_i - \alpha_j|, \pi - |\alpha_i - \alpha_j|)$;

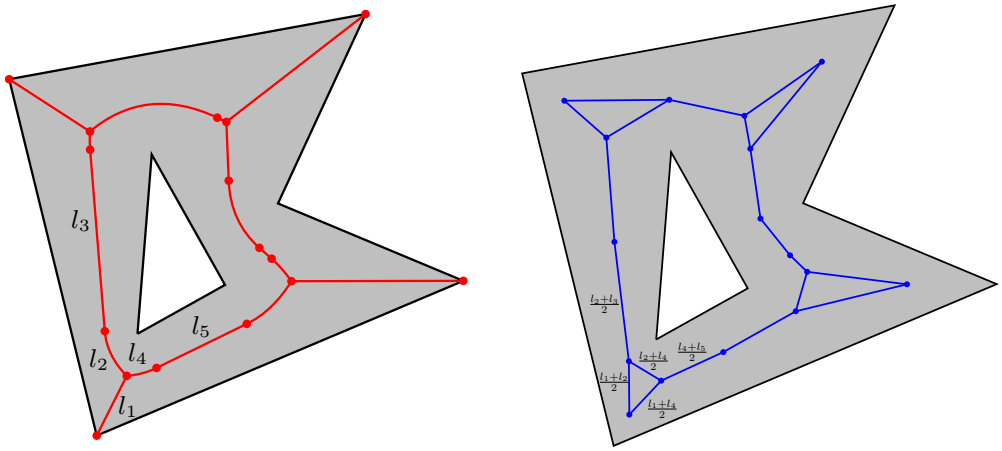


Figure 10: Skeleton-geodesic distances between pairs of skeleton edges: (a) the polygonal figure and its skeleton; (b) the dual graph which is used to calculate distances; lengths (weights) of graph edges are specified locally.

Note that these features are invariant to shift and rotation. When scaling is applied, the distance d_{ij} changes proportionally to the scale factor. We calculate $d_{ij}, \phi_{ij}, t_{ij}$ for all pairs of skeleton edges, then we normalize them:

$$d^*_{ij} = \frac{d_{ij}}{\max_{i,j} d_{ij}},$$

$$\phi^*_{ij} = \frac{\phi_{ij}}{\frac{\pi}{2}},$$

$$t^*_{ij} = \frac{t_{ij}}{r_{max}}.$$

After the normalization d^*_{ij}, ϕ^*_{ij} and t^*_{ij} belong to the interval $[0, 1]$. The interval $[0, 1]$ is divided into equal parts: k parts for thicknesses, m parts for distances, n parts for directions. We associate each pair of edges with one of kmn cells of the histogram of thickness-direction-distances, according to their feature values. The contribution of the one pair of edges e_i and e_j in the corresponding cell is equal to $s_i \cdot s_j$. We normalize the histogram of thicknesses-directions-distances, so the sum of histogram values is equal to 1.

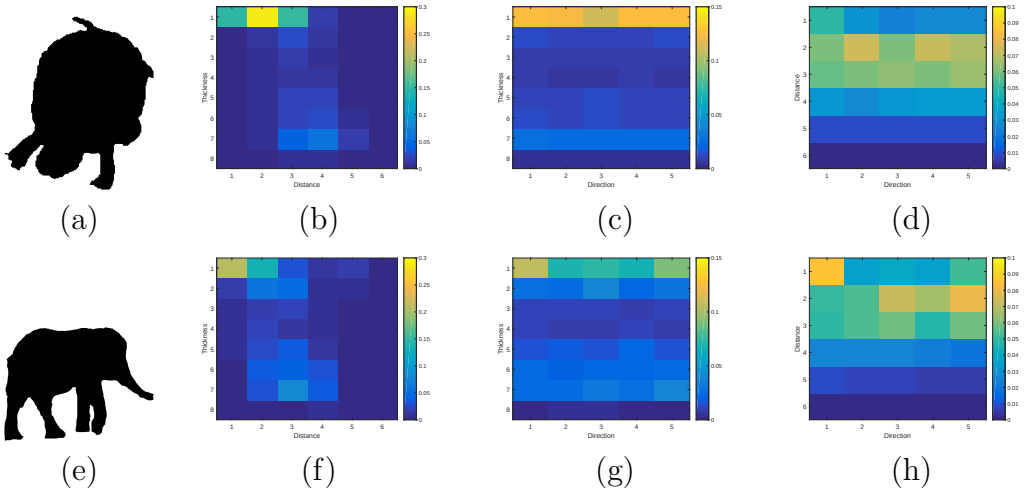


Figure 11: Images of the Animal dataset (a)(e) and their pairwise histograms of thicknesses-distances (b)(f), thicknesses-directions (c)(g) and distances-directions (d)(h)

$$HTDD(a,b,c) = \sum_{\substack{i,j:\lfloor (k-\epsilon)t^*_{ij} \rfloor = a,\\ \lfloor (m-\epsilon)d^*_{ij} \rfloor = b,\\ \lfloor (n-\epsilon)\phi^*_{ij} \rfloor = c}} s_i \cdot s_j \quad / \quad \sum_i s_i.$$

In a similar manner, we get two-dimensional (Fig. 11) and one-dimensional histograms, selecting a subset of three existing features: histogram of thicknesses-distances, histogram of thicknesses-directions, histogram of distances-directions and histogram of pairwise thicknesses, histogram of pairwise distances and histogram of pairwise directions.

We provide the following interpretation to these histograms: suppose the distance and the angle between any two points of the figure are determined by the skeleton edges whose proper regions contain these points and the thickness is averaged over the proper regions; then the histogram reflects the joint distribution of the values of the selected features if a pair of figure points is selected randomly.

Thus, the distance component of CTD descriptor is discussed. We can proceed to the final look of CTD.

3.4. Combine them all together

So far we have discussed all parts of the proposed descriptor. Among them are:

- Bag of contour fragments [17]. It contains r features and is responsible for contour information
- Skeleton-geodesic histograms:
  - histogram of pairwise thicknesses (k features)
  - histogram of pairwise distances (m features)
  - histogram of pairwise directions (n features)
  - histogram of thicknesses-distances (km features)
  - histogram of thicknesses-directions (kn features)
  - histogram of distances-directions (mn features)
  - histogram of thicknesses-distances-directions (kmn features)
- Thickness spectrum containing l features

All above mentioned components are stretched, if necessary, in the vectors, then these vectors are concatenated. Thus, CTD contains $r+(k+1)(m+1)(n+1)+l-1$ features. Which dimensions of the vectors are the best to choose and whether or not the proposed components are really informative? The answer to this question is given in the experimental section.

4. Experiments

We evaluate the proposed shape descriptor on four 2D shape benchmarks with relatively larger number of shapes (Animal [2], MPEG7 [4], Swedish leaf [30], ETH-80 [31]) and give the comparisons with the state-of-the-arts.

5.1 Datasets preprocessing

Swedish leaf dataset consists of full-color images. So the segmentation procedure has to be carried out. We applied simple segmentation by thresholding the grayscale images. The threshold value was set automatically using the Otsu's method.

Another problem for some datasets is holes. The slightest change in the object topology critically affects the medial representation. Images in Animal contain only small redundant holes appearing not due to self-occlusion. Filling the holes of less than 0.03% of the object area turned out to be sufficient. A similar option for the Swedish leaf was set to 0.6% in order to fill not only the redundant holes, but also the holes that really exist in the corrupted leaves. MPEG7 includes images with bad segmentation within the main contour. So all the holes in images of this dataset were filled. However, as a result some real holes were also filled. ETH-80 images were untouched. We adopt the default shape normalization technic reported in [1] to extract BCF shape representation for all three datasets.

5.2 Implementation details

For SVM learning we adopt the default parameters [17]. All histograms, which we add to BCF, are invariant to shift, rotation and scale by design. The elements of each histogram sum to 1.

We use two strategies for evaluating shape classification performance [26], [17]. The first is half training. We randomly select half shapes in each class for training and use the rest shapes for testing in each round. This procedure is repeated for 10 times. The second is leave-one-out. For each shape we use all shapes except one for training and use one shape for testing. Average classification accuracy, in percentage (%), is reported for both strategies.

For BCF we adopt the default parameter settings reported in [17]. The codebook size is reduced to 400 due to restrictions of computer power.

Parameters k, m and n are selected according to the results which have shown separate one-dimensional histograms in combination with the contour part of CTD descriptor. The results for individual sets were averaged taking into account the number of images in each set (Shown in Tables 1, 2, 3).

Dataset	3	4	6	8	10	15	25	50
Animal	88.49	88.52	88.75	88.78	88.81	88.76	88.73	88.51
MPEG7	97.52	97.70	97.74	97.73	97.72	97.65	97.64	97.57
Swedish leaf	97.55	97.67	97.84	98.17	98.48	98.60	98.72	98.72
Average	93.58	93.68	93.84	93.93	94.02	94.01	94.02	93.90

Table 1: Classification accuracy for the histogram of pairwise thicknesses versus the number of bins.

Dataset	3	4	6	8	10	15	25	50
Animal	87.70	87.57	87.81	87.85	87.67	87.66	87.57	87.38
MPEG7	97.58	97.61	97.65	97.61	97.54	97.48	97.40	97.29
Swedish leaf	98.08	98.15	98.09	98.04	97.96	97.84	97.68	97.41
Average	93.39	93.36	93.46	93.45	93.33	93.28	93.17	92.99

Table 2: Classification accuracy for the histogram of pairwise distances versus the number of bins.

Dataset	2	3	4	5	6	8	12
Animal	87.28	87.31	87.37	87.35	87.31	87.30	87.31
MPEG7	97.31	97.34	97.46	97.51	97.52	97.48	97.40
Swedish leaf	97.35	97.35	97.43	97.47	97.44	97.36	97.31
Average	92.94	92.96	93.04	93.06	93.04	93.00	92.97

Table 3: Classification accuracy for the histogram of pairwise directions versus the number of bins.

Let us consider various histogram combinations represented as vectors. For example, the combination of all 1D histograms, the combination of all 2D histograms, etc. The influence of each combination on the final result is reflected in Table 4. Thus, the distance component of CTD includes all listed above histograms calculated with the parameters $k = 10$, $m = 6$ and $n = 5$. When calculating thickness features grid step was set to 0.5. It is equivalent to bypass of integer-valued points for double scaled figure. We see that the histograms of all dimensions capturing both subtle and rough characteristics of the feature distribution are significant.

Dataset	1D	2D	3D	1D+2D	1D+3D	2D+3D	1D+2D+3D
Animal	89.38	89.62	88.58	89.77	89.55	89.73	89.86
MPEG7	97.35	97.88	97.28	98.16	97.97	97.94	98.18
Swedish leaf	99.03	98.95	98.03	99.21	99.12	99.05	99.25
Average	94.30	94.54	93.67	94.76	94.58	94.64	94.82

Table 4: Classification accuracy for combinations of histograms.

Finally, let us determine the number of thickness spectrum bins by Table

5. The addition of thickness spectrum to the contour and distance components of CTD slightly improves the result. Therefore, all components of CTD are important.

Dataset	4	5	7	10	15	20	30	50	100
Animal	89.85	89.91	90.02	89.97	89.89	90.03	90.06	90.02	89.94
MPEG7	98.30	98.32	98.30	98.35	98.32	98.32	98.37	98.32	98.32
Swedish leaf	99.08	99.11	99.15	99.17	99.17	99.17	99.19	99.16	99.15
Average	94.81	94.85	94.90	94.90	94.85	94.91	94.94	94.91	94.87

Table 5: Classification accuracy for CTD descriptor versus the number of thickness spectrum bins

5.3 Animal dataset

We compare CTD descriptor with other methods on the most challenging dataset for classification [2]. There are 20 animal classes in Animal. Each one comprises 100 examples. Objects are presented in various poses in each class, at the same time shapes significantly differ in size, location and articulation of parts, some parts of objects are subjected to self-occlusions.

The accuracy of the proposed descriptor with selected parameters is 90.06% (half-testing). Thus, as can be seen from Table 6 the proposed approach is superior to the known modern methods [16], [5], [17], [29], [27], [26], [28], including those that combine contour and skeleton descriptions [2], [1]. BCF combined with skeleton-geodesic histograms and thickness spectrum allows to improve the result for $6,66\%$ compared to raw BCF. When Bag of Skeleton Paths allows to improve the result for $2,1\%$ compared to raw BCF. Thus, BCF descriptor is successfully enriched by us.

To allow fair comparison with the algorithm based solely on contour features [17] we report classification accuracy for each of the 20 classes in Table 7. The proposed descriptor outperforms BCF for difficult classes that were previously bad recognized ("Monkey", "Leopard", "Cat", "Cow"). So the benefits are demonstrated.

5.4 MPEG7

MPEG7 dataset [4] is very popular shape dataset. It contains 1400 shapes of animals, artificial objects and symbols. It has 70 classes. There are 20

method	accuracy
CSS [16]	69.7%
IDSC [5]	73.6%
CS [2]	71.7%
SP [2]	67.9%
ICS [2]	78.4%
BCF [17]	83.4%
Shape Tree [29]	80.0%
Lim [27]	80.4%
Shape Vocabulary [26]	84.3%
BCF+BSP [1]	85.5%
MC-2 [28]	85.9%
CTD	90.06%

Table 6: Classification accuracy (half testing) on the Animal dataset

method	**bird**	**butterfly**	**cat**	**cow**	**crocodile**	**deer**	**dog**
BCF	87.6%	92.2%	73.8%	77.4%	76.8%	90.4%	82.6%
CTD	92.0%	97.6%	80.2%	85.6%	83.2%	92.8%	89.6%
method	**dolphin**	**duck**	**elephant**	**fish**	**flyingbird**	**hen**	**horse**
BCF	89.0%	87.0%	95.2%	79.8%	72.0%	94.2%	95.4%
CTD	93.2%	92.2%	96.2%	88.0%	83.0%	96.0%	96.6%
method	**leopard**	**monkey**	**rabbit**	**rat**	**spider**	**tortoise**	
BCF	66.4%	58.4%	85.8%	70.6%	99.2%	93.6%	
CTD	77.2%	77.6%	95.6%	87.4%	99.2%	98.0%	

Table 7: Classification accuracy (half testing) for each class on the Animal dataset

different shapes in each class. Table 8 demonstrates the classification accuracies obtained by competing methods. Classification accuracy on MPEG7 improved from 97.03% to 98.37% (half-testing) and from 98.86% to 99.00% (leave-one-out). Our method outperforms others on this dataset. So its generality is shown.

5.5 Swedish Leaf dataset

The dataset contains leaves from 15 different swedish trees with 75 leaves

method	accuracy (half testing)	accuracy (leave one out)
CSS [16]	90.9%	97.93%
CS [2]	91.1%	-
SP [2]	86.7%	-
ICS [2]	96.6%	-
PMR [54]	-	97.57%
SoS [55]	-	97.36%
RS [56]	-	98.57%
KD [57]	-	98.93%
VACF [58]	97.32	98.64%
BCF [17]	97.03%	98.86%
BCF+BSP [1]	98.35%	-
CTD	98.37%	99.00%

Table 8: Classification accuracy on MPEG7

per class. The difficulty with this dataset is that some leaves are indistinguishable to the untrained eye. We follow the experimental setting in [5]. We compare classification accuracy obtained by our descriptor with other pure shape-based recognition methods in Table 9.

method	accuracy
MAC [30]	82.00%
Fourier [5]	89.60%
CS + DP [5]	88.12%
IDSC + DP [5]	94.13%
MDM [59]	93.60%
IDSC + MS [60]	94.80%
RS [56]	95.47%
Shape Tree [29]	96.28%
BCF [17]	97.52%
CTD	99.19%

Table 9: Classification accuracy (half testing) on Swedish Leaf.

Classification accuracy on Swedish leaf has improved from 97.52% to

99.19% (one third of the dataset has been used for training).

5.6 ETH-80 dataset

The ETH-80 dataset [31] have 80 color objects from eight categories. There are 41 segmentation masks from different viewpoints for each object. The dataset has 3280 shapes. There are several classes that have meaningful holes. Leave-one-out strategy is used for evaluation. Compared with other descriptors our descriptor achieves the classification accuracy of 93.69%, outperforming the previous state-of-the-art approaches.

method	accuracy
Color histogram[31]	64.86%
PCA gray [31]	82.99%
PCA masks [31]	83.41%
SC + DP [5]	86.40%
IDSC + MS [60]	88.04%
IDSC + DP [5]	88.11%
Height function [45]	88.72%
RS [56]	90.28%
Kernel-edit [57]	91.33%
BCF [17]	91.49%
CTD	93.69%

Table 10: Classification accuracy on ETH-80.

5.7 Performance

Dataset	BCF, sec.	Thickness Spectrum, sec.	Skeleton-Geodesic Histograms, sec.	CTD, sec.
Animal	3.518	0.062	0.264	3.844
MPEG7	2.390	0.060	0.233	2.683
Swedish leaf	2.840	0.057	0.271	3.168
ETH-80	1.178	0.041	0.133	1.352

Table 11: Seconds per one shape on an average.

The experiments were carried out on a PC with Intel Core i5 750. The results shown in Table 11 confirm that thickness spectrum and skeleton-geodesic histograms are quite efficiently computed.

5. Discussion

Thus, the conducted experiments prove the superiority of the proposed descriptor over the contour descriptors [17], [26], [58] and the complex descriptors [1], [2]. The experiments underline the limitations of BCF (Fig. 12). In our opinion there are several reasons for this. To justify these reasons let

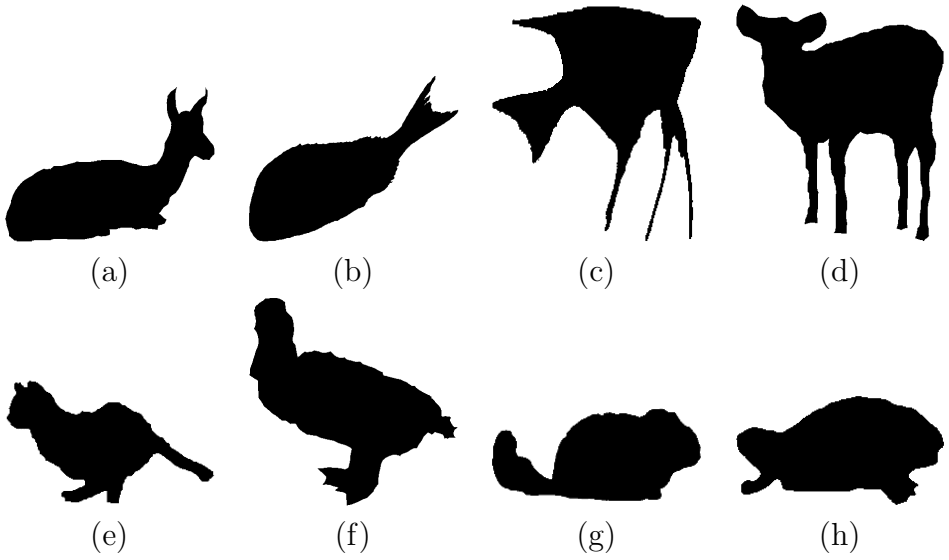


Figure 12: Shape pairs of different classes that have similar contours. Shapes (a), (c), (e), (g) are not correctly classified by BCF. In addition, the shapes (b), (d), (f), (h) are nearest to the objects (a), (c), (e), (g) in terms of L^2-metric. As can be seen from these examples the object is often recognized incorrectly if there is a general global similarity of contours with objects of other classes. Despite the presence of strong local contour bends that highlight separate segments (tail, legs, fins). Therefore, we design our complex descriptor to capture these changes in the shape structure.

us characterize the features that are used by each of the above descriptors from the viewpoint of local, global, structural and non-structural information. Contour descriptors exploit local information about the structure of various shape boundary parts. This information should be tolerant to local occlusions. Even so what if the contours are distorted or object posture

is changed globally? Due to the local nature of contour fragments classification results of BCF, Shape Vocabulary, VACF descriptors do not differ much. This suggests that local structural information is not enough. That is why Integrated Contour And Skeleton, Bag of Skeleton Path descriptors try to enrich the contour description by skeletal. However, the sequences of inscribed circles between the skeleton end points are also coding the local structural information. So the local structural information about the shape is used when the global and non-structural information is ignored. At the same time, we know that the gain from the fusion of different nature data will be considerably greater than in the case of the fusion of the same nature data. Therefore, we propose a complex descriptor and seek to use all sources of information. As other authors we also use local structural information stored in the contour fragments to provide robustness to occlusion. Nevertheless, we add global non-structural information about the object thickness by means of thickness spectrum. We add global structural information using skeleton-geodesic histograms that characterize relationships between shape regions. Note that the thickness spectrum and skeleton-geodesic histograms have areal nature. Therefore they are stable to distortions of shape boundary. Thus, our descriptor outperforms other descriptors as it contains four types of information (local, global, structural and non-structural) of different nature (contour, thickness, area) when others contain one or two information types of contour / skeleton nature. This is the first advantage. The second advantage lies in the fact that the thickness spectrum and skeleton-geodesic histograms are sufficiently stable to flexible (non-rigid) deformations (for example, the flexible deformation of the bird wings (Fig. 2)) while remaining sensitive to thickness and areal shape features. They are stable because they are based on the specific point regions (regions of constant thickness, proper regions), whereas BCF, Shape Vocabulary, VACF, ICS, BSP descriptors are based on curves (contour segments, skeleton paths) and points selected on these curves. Besides, rarity and uniqueness of regions is higher than uniqueness of lines and points. Despite the fact that proposed descriptor formally contains many diverse information, its size is slightly bigger than BCF: BCF have 31500 features in its current implementation, thickness and skeleton-geodesic histograms add 533 features. Although the number of added features is small, they have high information content that leads to the increase in the quality of classification.

6. Conclusions

In this paper we introduce the new shape descriptor CTD[1], which is suitable for 2D shape classification. It is composed of three descriptors: bag of contour fragments, thickness spectrum, skeleton-geodesic histograms. Skeleton-geodesic histograms are a kind of histograms that are based on the distance statistics between pairs of shape elements and have their useful properties. Skeleton-geodesic histograms are quickly calculated using continuous skeleton because skeleton-geodesic distances, thickness and direction differences between pairs of proper regions can be easily extracted from it. It is an advantage that CTD takes into account local and global shape features and analyzes shape structure. At the same time CTD remains a compact and informative shape feature vector that can be used for learning discriminative classifiers. The experimental results obtained on common benchmarks confirm its effectiveness.

Acknowledgement

This work was supported by Russian Science Foundation (RSF) under Grant 16-11-00082; by Russian Foundation For Basic Research (RFBR) under Grants 15-07-01323 A and 16-57-52042 MHT_a.

References

[1] W. Shen, X. Wang, C. Yao, X. Bai, Shape recognition by combining contour and skeleton into a mid-level representation, in: Pattern Recognition - 6th Chinese Conference, CCPR 2014, Changsha, China, November 17-19, 2014. Proceedings, Part I, 2014, pp. 391–400.

[2] X. Bai, W. Liu, Z. Tu, Integrating contour and skeleton for shape classification, in: Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on, 2009, pp. 360–367.

[3] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Trans. Pattern Anal. Mach. Intell. 24 (4) (2002) 509–522.

[1]MATLAB code for CTD and experiments is available at `https://goo.gl/L02Dnq`

[4] L. J. Latecki, R. Lakämper, Shape similarity measure based on correspondence of visual parts, IEEE Trans. Pattern Anal. Mach. Intell. 22 (10) (2000) 1185–1190.

[5] H. Ling, D. W. Jacobs, Shape classification using the inner-distance, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2) (2007) 286–299.

[6] P. F. Felzenszwalb, J. Schwartz, Hierarchical matching of deformable shapes, Proc. IEEE Conf. Computer Vision and Pattern Recognition (2007) 1–8.

[7] A. M. Bronstein, M. M. Bronstein, A. M. Bruckstein, R. Kimmel, Analysis of two-dimensional non-rigid shapes, Int. J. Comput. Vision 78 (1) (2008) 67–88.

[8] L. Mestetskiy, Binary Image Skeleton Representation by Compound Bezier Curves, iConcept Press, 2013.

[9] C. Aslan, A. Erdem, E. Erdem, S. Tari, Disconnected skeleton: Shape at its absolute scale, IEEE Trans. Pattern Anal. Mach. Intell. 30 (12) (2008) 2188–2203.

[10] X. Bai, L. J. Latecki, Path similarity skeleton graph matching, IEEE Trans. Pattern Anal. Mach. Intell. 30 (7) (2008) 1282–1292.

[11] T. B. Sebastian, P. N. Klein, B. B. Kimia, Recognition of shapes by editing their shock graphs, IEEE Trans. Pattern Anal. Mach. Intell. 26 (5) (2004) 550–571.

[12] K. Siddiqi, A. Shokoufandeh, S. Dickenson, S. Zucker, Shock graphs and shape matching, Int. J. Comput. Vision 35 (1) (1999) 13–32.

[13] L. G. Domakhina, Skeleton-based segmentation and decomposition of raster pairs of shapes, Pattern Recognition and Image Analysis 20 (3) (2010) 293–302. doi:10.1134/S1054661810030053.

[14] L. Mestetskiy, Medial width of polygonal and circular figures - approach via line segment voronoi diagram, in: Proceedings of the 10th International Conference on Computer Vision Theory and Applications (VISIGRAPP 2015), 2015, pp. 379–386. doi:10.5220/0005261903790386.

[15] H. Chui, A. Rangarajan, A new point matching algorithm for non-rigid registration, Comput. Vis. Image Underst. 89 (2-3) (2003) 114–141.

[16] K. B. Sun, B. J. Super, Classification of contour shapes using class segment sets, in: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2 - Volume 02, CVPR '05, IEEE Computer Society, Washington, DC, USA, 2005, pp. 727–733.

[17] X. Wang, B. Feng, X. Bai, W. Liu, L. Jan Latecki, Bag of contour fragments for robust shape classification, Pattern Recogn. 47 (6) (2014) 2116–2125.

[18] P. Maragos, Pattern spectrum and multiscale shape representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (7) (1989) 701–716.

[19] Y. Vizilter, Y. Pytev, A. Chulichkov, L. Mestetskiy, Morphological image analysis for computer vision applications, in: M. N. Favorskaya, L. C. Jain (Eds.), Computer Vision in Control Systems-1: Mathematical Theory, Springer Publishing Company, Incorporated, 2014, Ch. 2, pp. 9–58.

[20] M. Mahmoudi, G. Sapiro, Three-dimensional point cloud recognition via distributions of geometric distances, in: Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on, 2008, pp. 1–8.

[21] D. Donoho, C. Chui, R. R. Coifman, S. Lafon, Diffusion maps, Applied and Computational Harmonic Analysis 21 (1) (2006) 5 – 30.

[22] R. Osada, T. Funkhouser, B. Chazelle, D. Dobkin, Matching 3d models with shape distributions, in: Proceedings of the International Conference on Shape Modeling & Applications, SMI '01, IEEE Computer Society, Washington, DC, USA, 2001, pp. 154–.

[23] S. Belongie, J. Malik, Matching with shape contexts, in: Content-based Access of Image and Video Libraries, 2000. Proceedings. IEEE Workshop on, 2000, pp. 20–26.

[24] A. B. Hamza, H. Krim, Probabilistic shape descriptor for triangulated surfaces., in: ICIP (1), IEEE, 2005, pp. 1041–1044.

[25] L. Mestetskiy, Skeletonization of polygonal figures based on the generalized delaunay triangulation, Programming and Computer Software 25 (3) (1999) 131–142.

[26] X. Bai, C. Rao, X. Wang, Shape vocabulary: A robust and efficient shape representation for shape matching, IEEE Transactions on Image Processing 23 (9) (2014) 3935–3949.

[27] K. L. Lim, H. K. Galoogahi, Shape classification using local and global features, in: Image and Video Technology (PSIVT), 2010 Fourth Pacific-Rim Symposium on, 2010, pp. 115–120.

[28] M. Ozay, U. R. Aktas, J. L. Wyatt, A. Leonardis, Compositional hierarchical representation of shape manifolds for classification of non-manifold shapes, in: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), ICCV '15, IEEE Computer Society, Washington, DC, USA, 2015, pp. 1662–1670.

[29] Y. Li, J. Zhu, F. L. Li, A hierarchical shape tree for shape classification, in: Image and Vision Computing New Zealand (IVCNZ), 2010 25th International Conference of, 2010, pp. 1–6.

[30] O. J. O. Soderkvist, Computer vision classification of leaves from swedish trees. Master's thesis, Linkoping University, SE-581 83 Linkoping, Sweden, 2001.

[31] B. Leibe, B. Schiele, Analyzing appearance and contour based methods for object categorization, in: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings., Vol. 2, 2003, pp. II–409–15. doi:10.1109/CVPR.2003.1211497.

[32] G. Matheron, Random Sets and Integral Geometry, Wiley, 1975.

[33] J. Serra, Image Analysis and Mathematical Morphology, Academic Press, Inc., Orlando, FL, USA, 1983.

[34] T. B. Sebastian, B. B. Kimia, Curves vs skeletons in object recognition, in: Image Processing, 2001. Proceedings. 2001 International Conference on, Vol. 3, 2001, pp. 22–25 vol.3.

[35] K. Siddiqi, B. B. Kimia, A. Tannenbaum, S. Zucker, Shapes, shocks and wiggles, Image and Vision Computing 17 (56) (1999) 365 – 373.

[36] K. Siddiqi, B. B. Kimia, A shock grammar for recognition, in: Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE Computer Society Conference on, 1996, pp. 507–513.

[37] D. Macrini, K. Siddiqi, S. Dickenson, From skeletons to bone graphs: Medial abstraction for object recognition, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), Anchorage, Alaska, USA., 2008.

[38] M. F. Demirci, A. Shokoufandeh, Y. Keselman, L. Bretzner, S. J. Dickinson, Object recognition as many-to-many feature matching, International Journal of Computer Vision 69 (2) (2006) 203–222.

[39] M. Pelillo, K. Siddiqi, S. W. Zucker, Matching hierarchical structures using association graphs, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (11) (1999) 1105–1120.

[40] E. Baseski, A. Erdem, S. Tari, Dissimilarity between two skeletal trees in a context, Pattern Recogn. 42 (3) (2009) 370–385.

[41] M. Demirci, A. Shokoufandeh, S. J. Dickinson, Skeletal shape abstraction from examples., IEEE Trans. Pattern Anal. Mach. Intell. 31 (5) (2009) 944–952.

[42] X. Bai, W. X, L. J. Latecki, W. Liu, Z. Tu, Active skeleton for non-rigid object detection, in: 2009 IEEE 12th International Conference on Computer Vision, 2009, pp. 575–582.

[43] M. Demirci, Y. Osmanlioglu, A. Shokoufandeh, S. Dickinson, Efficient many-to-many feature matching under the l1 norm, Comput. Vis. Image Underst. 115 (7) (2011) 976–983.

[44] X. Shu, X.-J. Wu, A novel contour descriptor for 2d shape matching and its application to image retrieval, Image Vision Comput. 29 (4) (2011) 286–294.

[45] J. Wang, X. Bai, X. You, W. Liu, L. J. Latecki, Shape matching and classification using height functions, Pattern Recogn. Lett. 33 (2) (2012) 134–143.

[46] W. Shen, Y. Wang, X. Bai, H. Wang, L. J. Latecki, Shape clustering: Common structure discovery, Pattern Recognition 46 (2) (2013) 539–550.

[47] J. Sivic, A. Zisserman, Video google: A text retrieval approach to object matching in videos, in: Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2, ICCV '03, IEEE Computer Society, Washington, DC, USA, 2003, pp. 1470–.

[48] H. Blum, Biological shape and visual science (part i), Journal of Theoretical Biology 38 (2) (1973) 205–287.

[49] L. d. F. D. Costa, R. M. Cesar, Jr., Shape Analysis and Classification: Theory and Practice, 1st Edition, CRC Press, Inc., Boca Raton, FL, USA, 2000.

[50] D. T. Lee, Medial axis transformation of a planar shape, IEEE Trans. Pattern Anal. Mach. Intell. 4 (4) (1982) 363–369. doi:10.1109/TPAMI.1982.4767267.
URL http://dx.doi.org/10.1109/TPAMI.1982.4767267

[51] S. Sidyakin, Morphological pattern spectra algorithm development for digital image and video sequences analysis. PhD thesis, Moscow (in Russian), 2013.

[52] D. D. Hearn, M. P. Baker, W. Carithers, Computer Graphics with Open GL, 4th Edition, Prentice Hall Press, Upper Saddle River, NJ, USA, 2010.

[53] D. B. Johnson, Efficient algorithms for shortest paths in sparse networks, J. ACM 24 (1) (1977) 1–13. doi:10.1145/321992.321993.

[54] E. Attalla, P. Siy, Robust shape similarity retrieval based on contour segmentation polygonal multiresolution and elastic matching, Pattern Recogn. 38 (12) (2005) 2229–2241.

[55] M. R. Daliri, V. Torre, Shape recognition and retrieval using string of symbols, in: 2006 5th International Conference on Machine Learning and Applications (ICMLA'06), 2006, pp. 101–108.

[56] M. R. Daliri, V. Torre, Robust symbolic representation for shape recognition and retrieval, Pattern Recogn. 41 (5) (2008) 1799–1815.

[57] M. R. Daliri, V. Torre, Shape recognition based on kernel-edit distance, Comput. Vis. Image Underst. 114 (10) (2010) 1097–1103.

[58] S. Bai, X. Wang, X. Bai, Aggregating contour fragments for shape classification, in: 2014 IEEE International Conference on Image Processing (ICIP), 2014, pp. 5252–5256. doi:10.1109/ICIP.2014.7026063.

[59] R. Hu, W. Jia, H. Ling, D. Huang, Multiscale distance matrix for fast plant leaf recognition, IEEE Transactions on Image Processing 21 (11) (2012) 4667–4672.

[60] R.-X. Hu, W. Jia, Y. Zhao, J. Gui, Perceptually motivated morphological strategies for shape retrieval, Pattern Recognition 45 (9) (2012) 3222 – 3230, best Papers of Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'2011).