Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 41 (2008) 1756–1767
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Shape matching and modeling using skeletal context

Jun Xie[a,*], Pheng-Ann Heng[b], Mubarak Shah[a]

[a]School of Electrical Engineering and Computer Science, University of Central Florida, Orlando, FL 32816, USA
[b]Department of Computer Science and Engineering, Chinese University of Hong Kong, Shatin, Hong Kong

Received 3 May 2007; accepted 5 November 2007

Abstract

Shape is a significant visual clue for human perception and shape models show considerable promise as a basis for extracting objects from images. This paper proposes a novel approach for shape matching and modeling using the symmetry characterization of shape interior and the spatial relationships of shape structures. Based on the representative skeletal features, we develop a mechanism to generate a coarse segment matching between different instances of an object. Additionally, the natural correspondence of skeletal branches to sequential segments along the shape curves is employed in the matching process to avoid false correspondences across different segments. Point matches within the corresponding segments are then obtained by solving a constrained assignment problem. The validation of the proposed approach is illustrated on various data sets in the presence of considerable deformation and occlusion and the results are compared with those of popular approaches. We also demonstrate the performance of our method on biological objects for shape modeling, showing better models than those obtained by the state-of-the-art shape modeling approaches.

Keywords: Shape skeleton; Saliency structure; Shape matching; Shape modeling; Optimal matching

1. Introduction

Many applications, such as anatomical modeling and shape retrieval, require an accurate dense point correspondence across a set of shapes. One way to describe a shape is to locate a finite number of points, the landmarks, on the shape curve. Landmarks are well-defined points which are supposedly homologous from one instance to the next. Compared to boundary-based methods, shape analysis using landmarks offers many advantages such as the convenience for shape comparison, the ease of data storage and the possibility of applying powerful mathematical tools. In this representation, shape correspondence can be defined as a matching from the set of landmarks on one shape to that on the next, with the minimal shape difference measured using the corresponding landmarks.

Although manual landmarking is an intuitive way to obtain those discernable points, it is usually labor intensive in practice and studies [1] have shown that the precision of manual landmarking drifts with time and the accuracy varies among the landmarkers. This paper presents an approach to automatically establish dense matching across a set of shapes. It is a two-stage scheme involving the initial matching of shape structures and the dense point matching within corresponding curve segments.

Given a set of shapes in a plane, we first extract their symmetrical features using the medial axis transform (MAT) [2]. The benefits of applying skeleton-based methods are its natural consistency with human intuition and capability to describe the local geometrical features, allowing the performance of articulated matching. Instead of pursuing the exact matching of skeletal graphs, which could introduce extra complexity and instability into the matching process, we propose to gradually match the representative samples on shape curves. Those samples are determined by a local geometric scale function [3] based on the insight that the "best" set of samples is the smallest set which describes the shape uniquely. A set of context features are then generated to describe the spatial relationships among the sampling set.

Before exploring the dense point matching, we develop a segment matching approach to find the coarse mapping of

* Corresponding author. Tel.: +1 407 8820132; fax: +1 407 2074816.
E-mail address: jxie@cs.ucf.edu (J. Xie).

0031-3203/$30.00 © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2007.11.005

different structures of the shapes. This is enlightened by the fact that most parts of different instances of a specific object usually have comparable outlines. The structure matching strategy allows one to explore the dense point matching within a relative small space (only in corresponding parts), eliminating the erroneous matches across different structures and also offering a significant speed-up. Dense landmark matching is finally obtained by minimizing the total matching cost enforced with a spatial ordering constraint.

The rest of the paper is organized as follows. Section 2 gives a brief description of previous work on shape matching using different features. In Section 3, the proposed shape matching approach is introduced in detail. The local feature scale function is described to determine the optimal samples on the shape curve. Based on skeleton features, we illustrate how to find the coarse mapping of the shape structures. An effective solution is also presented in this section to establish the dense point matching between the corresponding shape structures. The method is then validated in Section 4 through the experiments on various databases for shape recognition and shape modeling. Finally, our conclusions are presented in Section 5.

2. Related work

2.1. Point-based shape matching

Point-based shape matching has been studied extensively in the literature and various approaches have been proposed to find the sequential point correspondence from a set of shapes. A large number of these methods (e.g. Refs. [4,5]) use local geometric features, such as curvature, angle and arc length, to identify mathematical landmarks on the shape boundary. One problem with curvature-defined landmarks is that, in some applications (e.g. biological imaging), these points are not always borne out clearly or uniquely by the individual shape, and in this case the landmarks must be inferred from a more global context.

Davies et al. [6] proposed to apply the minimum description length (MDL) as a quantitative measure of "simpleness" and considered the point correspondence problem as one of finding the parameterization for each shape. However, like other global shape correspondence methods [7], the MDL-based approach uses a very complicated and highly nonlinear cost function. Although some attempts using different cost functions [8] or other optimization techniques [9] have been proposed to improve this approach, the large optimization problem still makes it impractical for applications to large data sets.

An alternative way to describe shapes is to use arbitrary number of points. Belongie et al. [10] introduced a shape matching scheme using roughly uniform samples. They characterized the property of each sample by a local descriptor, the shape context, and the similarity of boundary points is measured by comparing their shape contexts. Although this approach has produced encouraging results for a wide variety of data sets, the employed unit-length sampling technique cannot fully capture the shape's crucial characteristics if the sampling is not dense enough. Another limitation of this approach is the absence of spatial ordering constraints, leading to additional erroneous matches. Thayananthan et al. [11] incorporated a continuity constraint into the shape context matching scheme by adding a distance measure and optimized the cost function using a dynamic programming method. Tu and Yuille [12] applied the shape context descriptor for fast shape matching. More recently, Ling and Jacobs [13] replace the Euclidean distance in the shape context framework with the inner-distance to handle articulated shapes.

2.2. Skeleton-based shape matching

The MAT is a transform of the original object which describes the object completely—the union of the interiors of all maximal disks is exactly the interior of the object. A maximal disk is a disk contained in the shape, for which there is no other disk in the shape that contains it. The medial axis highlights the underlying structures of the overall shape, while also representing important multi-scale aspects of the shape components.

Based on skeletons directly, many approaches have been developed for shape matching. Liu and Geiger [14] proposed to match shape axis tree defined by the locus of midpoints of optimally corresponding boundary points. However, their algorithm does not preserve the coherence of the shape. A variant of the medial axis, shock graph, was considered by Sharvit et al. [15] for shape recognition. Although this method leads to fairly good matches, the errors of fundamental flows can violate the hierarchical relations among parts of the shape. Siddiqi et al. [16] converted the shock graph to rooted trees and compared shapes by matching the graph trees based on subgraph isomorphism. But the choice of the oldest shock as the root of the tree is arbitrary and may lead to erroneous matches.

To reduce the sensitivity of traditional skeletons to the local deformations on the shape curves, Golland and Grimson [17] proposed a skeleton approximation method in case the topology of the skeleton is known a priori. By modeling the medial axis with an undirected graph, they first kept the critical points (endings and junctions) unchanged and drove the graph along the gradient map using a snake-like algorithm. Once the graph had been settled on the gradient ridges, the fixed points were then adjusted via a reconstruction procedure.

Sebastian et al. [18] proposed to compare two shapes using the least action path deforming one shape to another based on the shock graph, which performs well in some shape indexing problems. However, finding the optimal alignment of the shock graph is not a trivial task. In their implementation, the deformation cost of the shock edges was found through a dynamic programming method, which resulted in a slow algorithm.

3. Skeletal shape context (SSC)

In this paper, to detect feature points along shape curves, we describe shapes using the medial axis structures. Since this graph structure may change greatly due to the deformation or occlusion of shape curves, we use only its characteristic points to identify the shape components and consider an

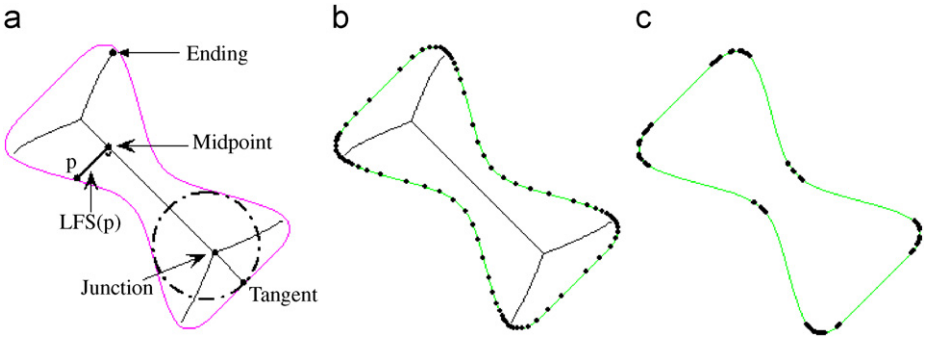

Fig. 1. An illustration of the shape sampling rule. (a) A curve (light) and its medial axis (heavy). The circle indicates a maximal disk. There are three kinds of points on the skeleton: endings, midpoints and junctions. (b) The sampling result using the γ-sampling rule. There are totally 85 samples with $\gamma = 0.4$ on this curve. (c) The sampling result using curvature maxima (top 85 points).

optimal sampling approach to capture the critical characteristics of each shape curve. A segment matching of the curves is then established based on the skeletal features and point correspondences are finally established within each pair of curve segments.

3.1. Optimal shape sampling

Given a shape, we wish to sample the shape curve with a compact set of points, which can characterize the inherent features of the shape uniquely. In other words, this set of samples shall lead to only one possible smoothed curve, the original shape. This indicates a non-uniform sampling rule.

Let $F(\xi) = (x(\xi), y(\xi))$ be a smooth curve representing a shape. A point is on the medial axis of the shape if it is equidistant from two or more points on the shape curve as shown in Fig. 1(a). There are three kinds of points on a skeleton: endings, midpoints and junctions. The endings and midpoints are the skeleton points which are equidistant from exactly two boundary points, while the junctions are equidistant from three or more points on the boundary. A skeleton point, which is equidistant from two boundary points, can be identified as an ending if there is only one skeleton point in its eight neighbors.

An intuitive way to implement the MAT is the so-called grassfire transformation. There are numerous methods for medial axis computation (see Ref. [19] for a survey of skeletal methods). In our approach, we apply a multi-scale method [20] to reduce the sensitivity of medial axes to image noise and boundary details. A multi-scale representation of shape F is achieved by computing its evolved versions as $F_\sigma(\xi) = F(\xi) \otimes g(\xi, \sigma)$, where $g(\xi, \sigma)$ denotes a 1D Gaussian of width σ, which is referred to as the scale parameter. In order to establish a skeleton-pyramid, a sequence of scales $\{\sigma_1, \cdots, \sigma_n\}$ are selected to smooth the shape curve under different resolutions. This technique is suitable for removing noise from shape curve and gradually simplifying the shape.

After getting the skeleton structure $W = \{\mathbf{w}_z\}$, $z = 1, \ldots, Z$, we can compute the local feature scale function $LFS : F \to R^2$, which is a continuous function defined as the distance of $\mathbf{p} \in F$ to W (see Fig. 1(a) for an example). Since it is based on medial axis, the LFS function is dominated by both the geometric features (e.g. curvature and orientation) and the topological features (e.g. symmetry and width). With this measure, the shape curve is sampled under the following γ-sampling condition. Let $S \in R^2$ be a set of points on shape F. Set S is defined as a γ-sample of F if, for each point $\mathbf{p} \in F$, there is a sample $\mathbf{q} \in S$ such that

$$\|\mathbf{p} - \mathbf{q}\| \leqslant \gamma LFS(\mathbf{q}). \tag{1}$$

Due to this sampling condition, the sample density varies with the local feature sizes so that areas with less detail are sampled less densely. Fig. 1(b) shows a result of this adaptive sampling method. Compared to the curvature-based method (Fig. 1(c)), it is obvious that, using the same number of samples, the LFS-based samples can describe the shape more accurately and comprehensively.

It is observed that, for $\gamma \geqslant 1$, the sample set may produce more than one curve which is the polygonal reconstruction of the smooth curve γ-sampled by S. This indicates that the sampling is not dense enough and there may be some smaller features on the shape that are not sampled adequately. Conversely, with a considerably smaller γ (e.g. $\gamma < 1$), the resulting samples are so compact that they can reconstruct the original curve uniquely [21,22].

One way to match two sets of such non-uniform samples is to consider all possible matchings among the points and choose the matching with maximal similarity. However, this is computationally expensive and prone to produce erroneous matches. It would be desirable to obtain a good correspondence in the first step. We thus propose the following two-stage matching scheme.

3.2. Segment matching

Each structural piece of the shape leads to a corresponding branch on the medial axis and the terminals of these branches, i.e. the endings, naturally correspond to the maxima of positive curvature along the shape curve [23] and thus provide a good capability for indexing these structural portions [24]. On the other hand, it is noticed that the junctions and midpoints on

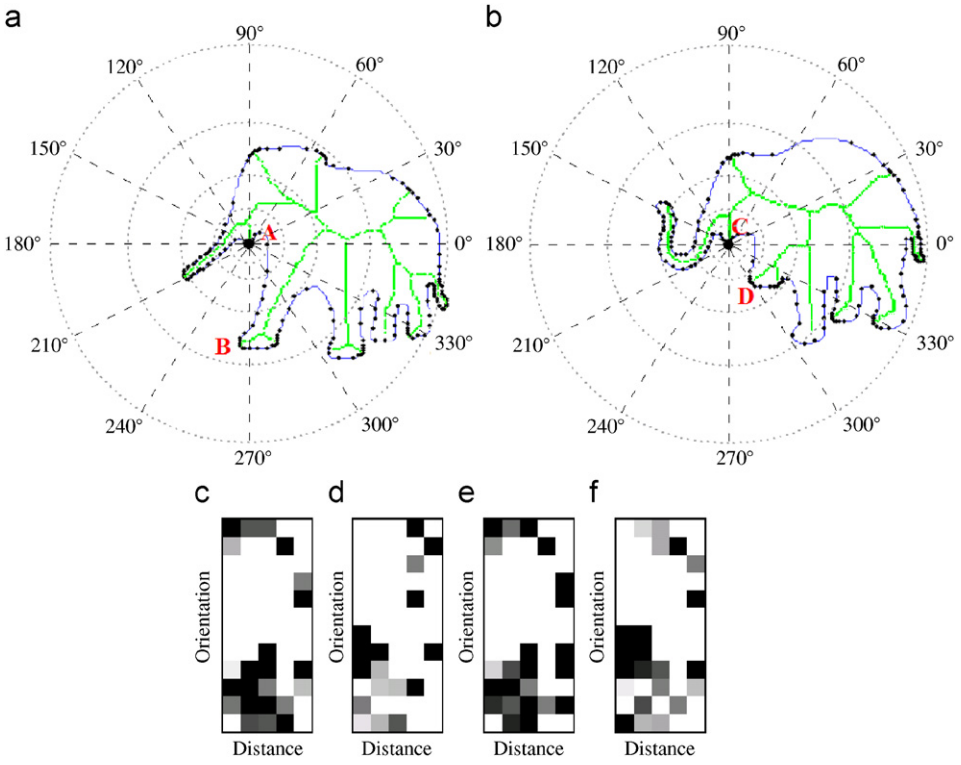

Fig. 2. Shape contexts of endings. (a) and (b) are shape outlines of two elephants. The histograms (c–f) are the computed context descriptor for four endings (A and B from (a), C, and D from (b)), respectively. It is evident that corresponding endings (e.g. endings A and C) have relatively more similar histograms. (c) For A; (d) For B; (e) For C; (f) For D.

medial axes are extremely sensitive to the changes of the shape curves. Therefore, in this work, we use only the endings for segment matching to avoid mismatches of the sensitive skeletal graphs. To describe the features of these endings, we utilize the same context descriptor as in Ref. [10].

Given a set of samples $S = \{s_i\}_{i=1}^{L}$ of a shape F and the endings $E = \{e_k\}_{k=1}^{K}$ on its medial axis, the context feature of each $e_k \in E$ is defined as a log-polar histogram h_k:

$$h_k(b) = \#\{s \neq e_k : (s - e_k) \in \mathrm{bin}(b)\}, \quad b = 1, \ldots, B, \qquad (2)$$

where B denotes the number of bins used to describe the point set. The resulting coarse histogram records the relative locations of the sampling points referring to e_k. An example of this context descriptor is demonstrated in Fig. 2 where we use five bins for distance measure log r and 12 bins for angle θ. Thus, there are totally 60 bins in the log-polar space.

Using this feature, the optimal matching of two sets of endings can be estimated by minimizing the total matching cost $\sum_1^K C(e_k, e_{\pi_k})$, where π_k is the index of correspondence of e_k, and $C(e_k, e_{\pi_k})$ is the difference between the context features of e_k and e_{π_k} defined as

$$C(e_k, e_{\pi_k}) = \frac{1}{2}\sum_{b=1}^{B} \frac{[h_k(b) - h_{\pi_k}(b)]^2}{h_k(b) + h_{\pi_k}(b)}. \qquad (3)$$

If the two sets of endings have different sizes, dummy points may be added to each set. The resulting square cost matrix is then input to an assignment problem (AP), and the matching can be solved in $O(n^3)$ by using the Hungarian method [25] or a more efficient approach [26] based on the preflow concept and dynamic tree data structure [27].

Now, consider a pair of corresponding endings on the two shapes. Each ending has exactly two tangents (they collapse to one at locations with maximum positive curvature [28]). We call those tangent points ending tangent points (ETPs). Since there are only two possible mappings between the ETPs of a pair of corresponding endings, it is easy to find the correct one by selecting the matching with a small matching cost.

Fig. 3 shows an example of matching the endings between two skeletons. Notice that the structures of the two skeletal graphs are so different that, even for humans, the exact mapping of those graphs is not straightforward. With the aid of the skeletal branches and their ETPs, however, the segment matching scheme provides a correct identification of the corresponding parts of the shapes (e.g. see the enlarged views for the endings 8 and 9 at front legs) despite the apparent deformations and occlusions.

3.3. Point correspondence within segments

Once the segment matching is done, the global point-based shape matching problem can be converted to a series of local dense matching within each pair of curve segments. Given a pair of corresponding segments $\Gamma_1 =$

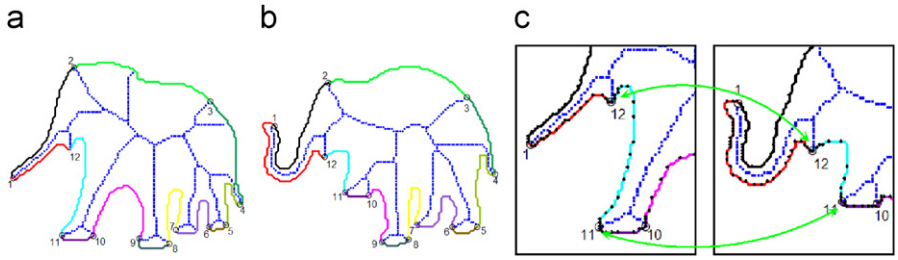

Fig. 3. Two shapes (a) and (b) with corresponding endings and segments. The blue dashed lines represent the skeletons and corresponding segments are indicated with the same color. (c) An enlarged view of the segment matching to illustrate the consecutive assignment problem for point matching within a pair of corresponding segments.

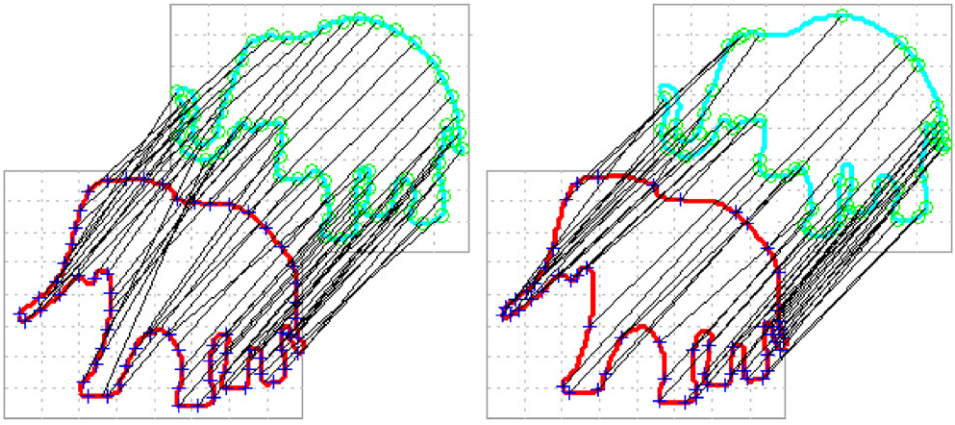

Fig. 4. The matching results of two shapes in Fig. 2. Left: the result (67 matches from 200 equally spaced samples) using the method in Ref. [10] after 10 iterations with $\lambda_0 = 1$. Right: the matching result (only 57 pairs among total 186 matches are shown for better visibility) of the proposed method with $\gamma = \frac{1}{3}$.

$\{t_{1,1}, s_{1,1}, \ldots, s_{1,i}, \ldots, s_{1,m}, t_{1,2}\}$ and $\Gamma_2 = \{t_{2,1}, s_{2,1}, \ldots, s_{2,j}, \ldots, s_{2,n}, t_{2,2}\}$, we want to find the best match across the two sets of points, where $t_{1,1}$ and $t_{1,2}$ are two ending tangents for Γ_1 and $t_{2,1}$ and $t_{2,2}$ are for Γ_2.

Without loss of generality, suppose the two sets of points are in the following spatial order $t_{1,1} < s_{1,1} < \cdots < s_{1,i} < \cdots < s_{1,m} < t_{1,2}$ and $t_{2,1} < s_{2,1} < \cdots < s_{2,j} < \cdots < s_{2,n} < t_{2,2}$. We constrain our method to preserve this ordering in the matching process. Since the sampling rule is adaptive, the number of resulting samples at each shape may not be equal. Assume that Γ_1 contains fewer samples than Γ_2, i.e. $m < n$. Considering the fact that the samples are selected optimally, and all of them are crucial for describing the input shapes, we want to find the largest point correspondence between the pair of segments. In other words, we want to find $\min(n, m)$ pairs of correspondences across the two segments.

Fig. 3(c) gives an illustration of this matching problem. Note that the terminals on the two segments correspond to each other, i.e. $t_{1,1} \leftrightarrow t_{2,1}$ and $t_{1,2} \leftrightarrow t_{2,2}$, as described in previous section. Therefore, we can estimate the point correspondence by removing $(n - m)$ samples from Γ_2 so that both sampling sets have the same size and the remaining points lead to a natural matching. We describe each sample point using all other samples as stated in Section 3.2. Similarity of each pair of samples across Γ_1 and Γ_2 is measured using Eq. (3) and the $(n - m)$ outliers from segment Γ_2 are found by maximizing the cost function

$$H = \sum_{j}^{n-m} \sum_{i=1}^{m} C(s_{1,i}, s_{2,j}). \tag{4}$$

The search space for this consecutive AP is $C_n^{(n-m)}$, which is much smaller than that for the AP in Ref. [10]. Due to the initial segment matching, the points to be considered in this step are restricted in each pair of corresponding segments, which will greatly reduce the complexity of the global matching process. Another advantage of this two-stage matching scheme is that the mismatches from one segment to others are naturally avoided.

Fig. 4 shows the matching results of two shapes using the standard shape context method and our algorithm without any transforms. It is evident that the shape context method generated a number of erroneous matches (e.g. the incorrect correspondences from the elephant's nose to the front leg). Based on the segment matching and the ordering constraints, our approach has successfully corrected most of the mismatches.

The pseudocode in `Algorithm 1` integrates the components of our matching algorithm into a procedure which gets two shape curves and a sampling factor as input and returns a set of corresponding points along the two shape curves. The computation of context descriptor is not presented here to

increase the readability. Since the MAT is inherently affine invariant and the context descriptor is translation invariant (using relative position), scale invariant (by normalization) and robust toward small rotations, the proposed algorithm is nearly affine invariant.

Algorithm. 1. Skeletal shape context.

Input: shape curves $F_i(\mathbf{p}_j)$, $i=1, 2$, $j=1, \ldots, N$, sampling factor γ
Output: corresponding points across the two shapes
1. **Initiation**
For each shape curve:
Compute skeletons $W=\{\mathbf{w}_z\}$, $z=1, \ldots, Z$
Compute local feature size as

$$fs(\mathbf{p}) = \min_{\mathbf{w}\in W} \|\mathbf{p}-\mathbf{w}\|$$

Sampling the curve:
$P_j = r_j =$ False
Sort points $\mathbf{p}_j$ in descending order of $fs \rightarrow \{\mathbf{v}_j\}$
$P_1 = r_1 =$ True, $c=1$
while $\#\{r_j = \text{False}\} > 0$ **do**
$r_T =$ True, $T = \{j \mid \|\mathbf{v}_j - \mathbf{v}_c\| \leqslant \gamma \cdot fs(\mathbf{v}_c)\}$;
$fs(\mathbf{v}_c) = 0$;
find $\mathbf{u}_c$: $fs(\mathbf{u}_c) = \max_j fs(\mathbf{v}_j)$ and $r_c =$ False;
$P_c = r_c =$ True
end while
Samples $S = \{\mathbf{v}_j | P_j = \text{True}\}$
2. **Segment matching**
Compute the context descriptors h_{E_1} and h_{E_2}
Compute cost of ending matching

$$C(E_1, E_2) = \min \sum\nolimits_{E_1} \sum\nolimits_{E_2} C(\mathbf{e}_x, \mathbf{e}_y)$$

Compute corresponding segments: $\Gamma_k \rightarrow \Gamma_{\varpi_k}$
3. **Point match**
For each pair of segments

$$\Gamma_1 = \{t_{1,1}, \{\mathbf{s}_{1,m}\}, t_{1,2}\} \quad \text{and} \quad \Gamma_2 = \{t_{2,1}, \{\mathbf{s}_{2,n}\}, t_{2,2}\}$$

Compute the context descriptors h_{Γ_1} and h_{Γ_2}
Revise set $\mathbf{s}$ into $\hat{\mathbf{s}}$ by removing $(n-m)$ outliers found via maximizing

$$H = \sum_{j}^{n-m} \sum_{i=1}^{m} C(\mathbf{s}_{1,i}, \mathbf{s}_{2,j})$$

Return point matches $\pi : \{\mathbf{s}_{1,m}\} \rightarrow \{\hat{\mathbf{s}}_{2,m}\}$.

4. Experimental results

We have applied the algorithm to several data sets from different domains. The performances of our algorithm for shape recognition and retrieval are illustrated and compared with those of popular shape recognition approaches. In addition, we demonstrate the qualitative and quantitative results of applying the algorithm to biomedical object shapes for shape modeling. The computation complexity and the sensitivity of our algorithm against the sampling condition are also demonstrated.

For each shape, we compute the medial axis using the edge map and sample the boundary based on the local feature size. For the shape recognition and query experiments, we set $\gamma = \frac{1}{3}$ empirically. Although we have applied the multi-scale representation of shapes to compute medial axes, the resulting skeleton may still contain some spurious branches due to boundary irregularity. To improve the robustness, we made use of the parametric morphological pruning transformation [29]. Given a preferred size l, a parametric pruning consists of the removal of l pixels for each skeleton branch, starting from the endings. An alternative and faster algorithm to prune the skeletons is to eliminate only the branches with length l. The size of the pruned branches can be defined as a fixed parameter or adaptively based on the size of the skeleton, e.g. less than 3% of the skeleton length (the length of the longest branch). In the experiments reported here, we removed all the branches consisting of fewer pixels than l.

The resulting skeletons are then used to generate a piecewise matching of the shape curves. This step also captures the hierarchical relationship between the corresponding segments. To handle viewpoint changes, we measure the similarity of skeleton endings under three states: unchanged pose, vertically flipped pose and horizontally flipped pose. The initial mapping of the endings is decided under the pose with the minimal matching cost. Finally, the samples within the corresponding segments are optimally assigned with the minimal sum of matching cost.

Shape recognition: We first applied the algorithm to a small set of shapes [15] consisting of 25 shapes of six classes (Fig. 5). In the experiment, each shape in the data set was matched to all others and a measure of difference was recorded. The difference between shapes F_1 and F_2 is defined to be the overall cost of the optimal assignment, i.e.

$$D(F_1, F_2) = \frac{1}{M} \sum_{j=1}^{M} C(\mathbf{s}_{1j}, \mathbf{s}_{2\pi_j}), \tag{5}$$

where M denotes the number of correspondences established across the two shapes. For each shape, the top three most similar shapes were identified. The performance was evaluated by counting the number of first, second and third nearest neighbors that fall into the correct class.

Our results for this data set are $(\frac{25}{25}, \frac{25}{25}, \frac{24}{25})$ while the approach in Ref. [10] reported $(\frac{25}{25}, \frac{24}{25}, \frac{22}{25})$ by using 100 samples. Our method gets only one erroneous match ($15 \rightarrow 10$) in the top three ranked matches and corrects three mismatches of the shape context method: $8 \rightarrow 4$, $7 \rightarrow 4$ and $4 \rightarrow 8$. The same experiment was also performed by Sharvit et al. [15] and Gdalyahu et al. [4] and their reported results are $(\frac{23}{25}, \frac{21}{25}, \frac{20}{25})$ and $(\frac{25}{25}, \frac{21}{25}, \frac{19}{25})$, respectively.

Shape query: A larger data set [18] tested in our experiments is constructed from the shape database created for testing the compression rates for MPEG7. This data set consists of 216

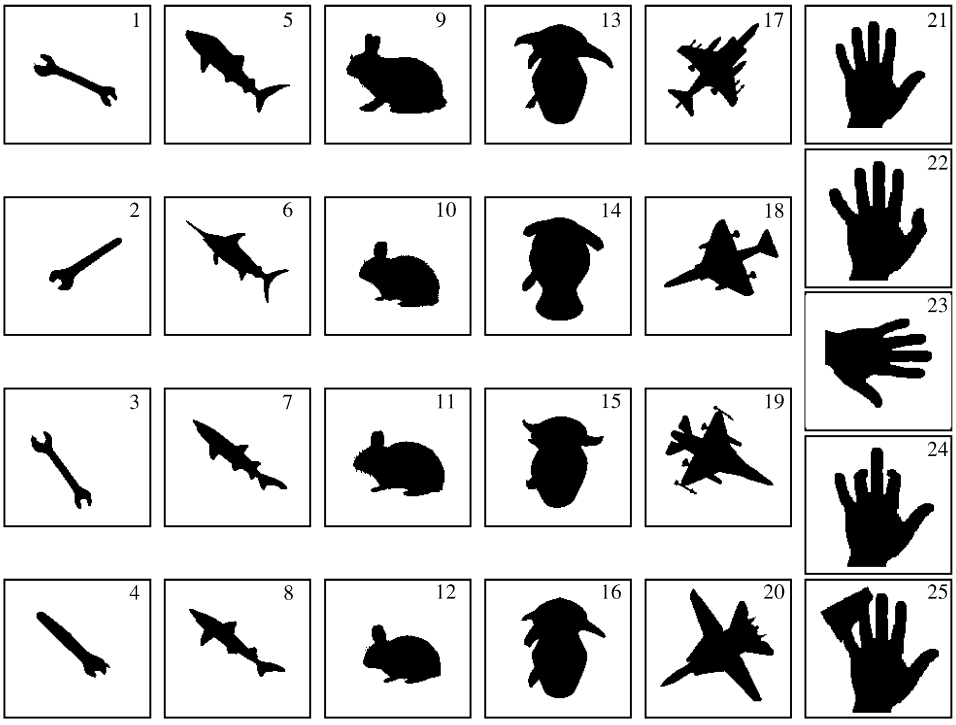

Fig. 5. Shape recognition result on a small data set [15]. Each row shows instances of a different object category.

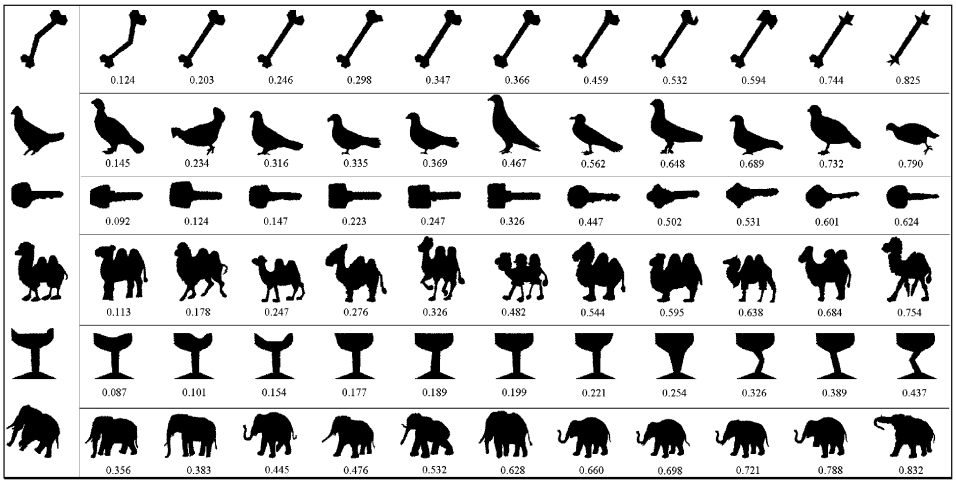

Fig. 6. Several query results on the Latecki data set [18]. Left column shows query shapes and the right rows show the first 10 ranked nearest neighbors for each query shape, respectively.

images including 18 categories, and each category has 12 shapes with variations in form and occlusion of parts. Each shape was matched against all others. Excluding the query shape, there are 11 possible correct matches for each test. The matches were ordered by the shape difference, and the quality of the performance was measured by a vector representing the proportion of correct matches. Several query results of our approach are depicted in Fig. 6. Our ranked recognition rates in % of this data set are (100,100,100,99,99,99,99,98,97,92). A same test was conducted by Sebastian using the shock graph approach [18] and their results are (100,100,100,99,99,99,97,96,95,87).

Furthermore, the whole MPEG7 CE-Shape-1 database [30] was used for shape query. This data set consists of 1400 silhouette images from 70 classes and each class has 20 shapes. For each shape, we match it with all other shapes in the database and the top 40 matched shapes are counted. The recognition score is the ratio of the number of correct hits of all shapes to the complete matches hits. Table 1 shows the reported results from different approaches on this database. The inner-distance-based method [13] reported the best performance followed by Tu's method [12] and our approach. Our algorithm is implemented and interpreted in Matlab on a regular

Pentium IV, 1.5 GHz PC. A single pairwise match with $\gamma = \frac{1}{3}$, including the MAT, optimal sampling, mapping of skeleton endings and point matching in the segments, takes 80 ms on average. Due to the two-step matching strategy, our algorithm is much faster than the shape context algorithm and those using statistical optimization approaches. For example, with 100 landmarks, the shape context algorithm takes about 1.2 s for a single comparison of two shapes and Ling and Jacobs [13] reports 0.31 s on a 2.8 GB PC.

Another database used in our experiments is a leaf data set constructed from the plant species database collected by the Smithsonian Institution [32] that consists of the largest, verified plant-type specimen database in the world. The database contains 912 leaf images from 38 different species. For each species, there are 24 instances with different orientations, deformations and outlines. Fig. 7 shows some silhouettes of those leaves. For each image, five distance bins and 12 orientation bins are used for both SC and SSC. We sampled each silhouette with 300 evenly spaced points for the SC matching and set $\gamma = 0.2$ for our algorithm to handle this high-resolution database. The classification rate is 78.45% using the SC matching and 89.12% with the proposed SSC approach.

Table 1
Retrieval rates of different approaches for the MPEG7 database

Algorithm	Visual parts [30]	SC + TPS [10]	Curve edit [31]
Score %	76.45	76.51	78.17
Algorithm	Gen. model [12]	IDSC + DP[13]	SSC
Score%	80.03	85.40	79.92

Shape modeling: We further tested the proposed approach on a set of biomedical objects for shape modeling. For a specific object, each instance in the training set $\{F_i\}_{i=1}^{T}$ was matched to all others. The shape with the minimal total difference was selected to find the overall correspondence of the whole set by determining the overlapped correspondences among its matchings with other shapes.

After the automatic landmarking process, a set of landmarks were obtained on each shape in the training set. To build shape models automatically from a number of shapes with different orientations, sizes and locations, one needs to estimate a transformation which may transfer the shapes into a common coordinate frame. In our experiments, we applied the affine transformation and its parameters were determined by minimizing the total distance between each set of corresponding landmarks. With M representing the number of landmarks on each shape curve, the coordinates of those landmarks were arranged in a $2M$ dimensional vector. The mean shape and covariance matrix were then computed using the T shape vectors in a standard procedure for statistical shape modeling (see Ref. [1] for details).

Fig. 8 shows two shape models generated automatically using the proposed approach. The hand training set contains

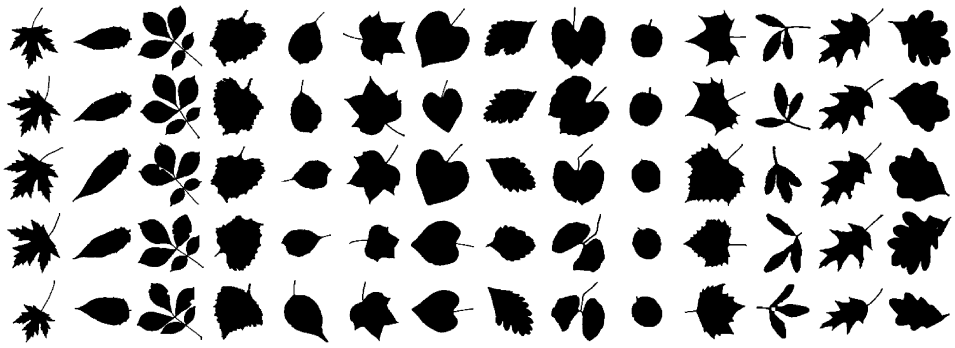

Fig. 7. Typical images from the leaf database. Each column shows five images from a particular species.

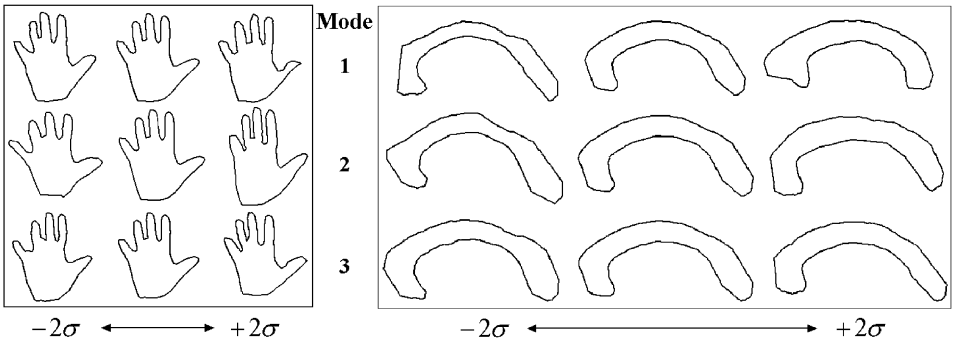

Fig. 8. Two shape models automatically generated using the proposed approach. Left shows the first three modes of the hand model generated from the hand posture database in Ref. [33] and right shows the modes of a callosum model.

Table 2
Evaluation of the three kidney models

Kidney model	Number of points	Variance			
		Mode 1	Mode 2	Mode 3	Total
Automatic landmarking	81	46.89	34.92	13.08	94.89
Manual landmarking	54	60.13	34.91	14.58	109.62
Evenly-spaced sampling	81	172.86	35.44	30.82	239.12

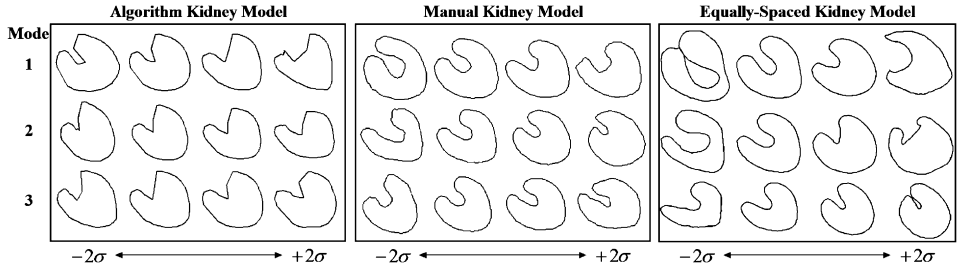

Fig. 9. Three kidney models generated using the proposed method, manual landmarks and evenly spaced samples.

15 hand outlines extracted from the Jochen Triesch static hand posture database [33] and the callosum set contains 14 callosum shapes extracted from MRI brain slices. Using the proposed automatic landmarking algorithm, we obtained 128 landmarks on each shape for the hand data set and 112 for the callosum data set. The three rows for each modeling approach illustrate that the first three modes of the shape models varied by $\pm 2\delta$, where δ denotes the standard deviation along the principal directions. As seen, compared with the manual models, our approach can generate much better automatic results than the even-space samples-based method.

To evaluate the effectiveness of our approach for shape modeling, we compared the models generated using different methods in terms of variance of the three largest modes of the models. The training data consist of 20 2D left kidney shapes from different patients. The recorded quantitative results (Table 2) show that our model has the smallest variances in all three principal directions, indicating a more compact statistical shape model, compared with methods using manual landmarking and even-space sampling (Fig. 9).

Finally, we performed leave-one-out tests on the callosum data set to access the generalization capability of the proposed approach. Given a set of the callosum outlines, a model $\Phi = \{\mathbf{u}_m\}_{m=1}^{M}$ was built from all but one. We then fitted the model to the example missed out $\mathscr{F} = \{\mathbf{p}_j\}_{j=1}^{N}$ and recorded the mean of the distances between each point $\mathbf{u}_m \in \Phi$ and its closest point in $\mathscr{F}$:

$$MPD(\Phi, \mathscr{F}) = \frac{1}{M} \sum_{\mathbf{u}_m \in \Phi} \min_{\mathbf{p}_j \in \mathscr{F}} \|\mathbf{u}_m - \mathbf{p}_j\|. \quad (6)$$

The mean and deviation of pixel distance measure the generalization capability of a model to represent unseen shape instances of the object.

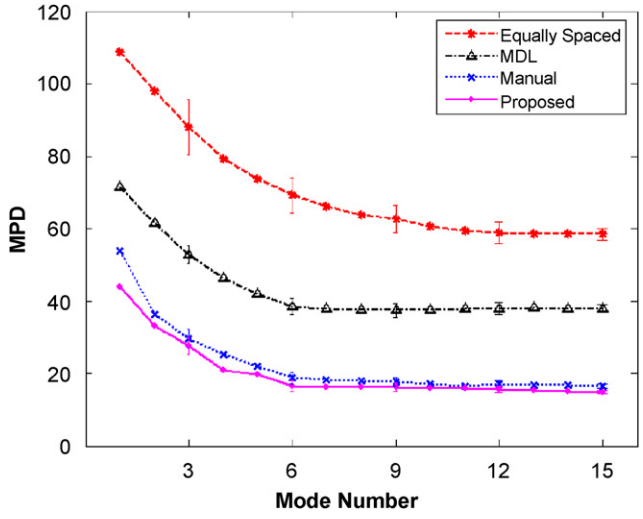

Fig. 10. The results of the leave-one-out experiment on each model of the callosum data set. The plot shows the mean pixel distances against the number of modes used for shape fit test. The error bars indicate the standard deviation of the pixel distances of each model to the test shape.

We repeated this procedure by missing out each of the examples in turn. The performances of four models generated using different methods are shown in Fig. 10. As seen from the plots, for any number of modes used to fit the unseen shape, the proposed algorithm could obtain a comparable result with that using manual landmarks and performed considerably better than both the MDL approach and evenly spaced landmarks-based method.

The performances of our algorithm with different settings of the sampling factor are illustrated in Fig. 11 including the size of the resulting correspondences (NC) across the training set and the pixel distances measured using five modes. It is clear that a small sampling factor can lead to a good result due to

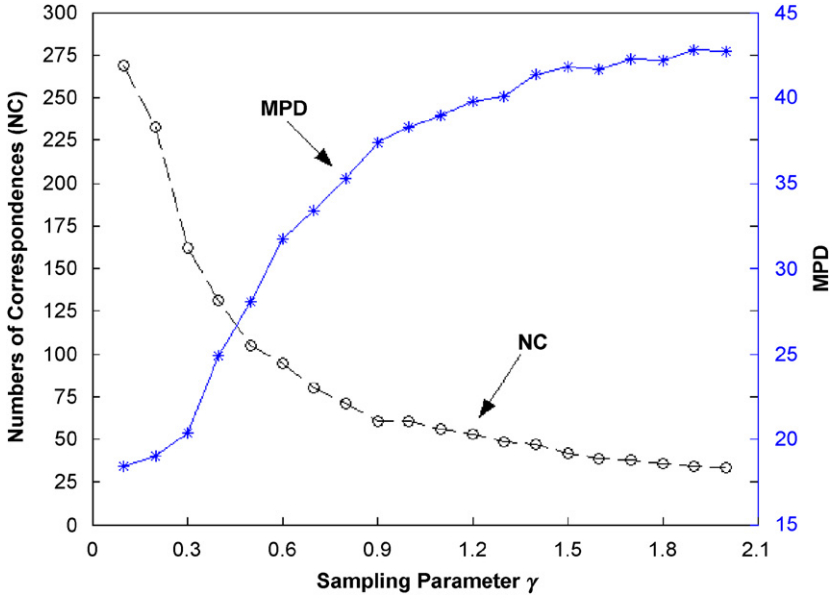

Fig. 11. The overall number of correspondences and the mean pixel distance using five modes for the callosum data set against the value of sampling parameter γ.

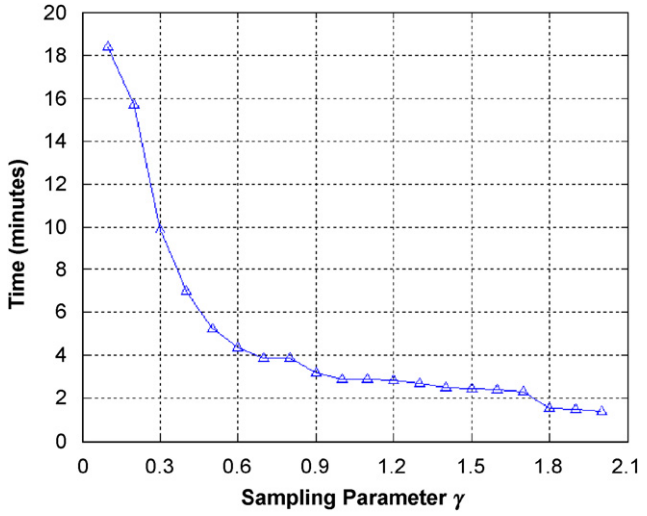

Fig. 12. The computational time for the leave-one-out experiment against different sampling factors.

the large set of correspondences. However, increasing the size of samples would introduce more computational complexity in the matching algorithm as demonstrated in Fig. 12 which shows the computational time of our algorithm for the leave-one-out tests on the callosum data set with different values of γ. This indicates that parameter γ should be chosen small enough to capture the crucial characteristic of shapes (with considerable dense of landmarks), but not too small such that the shape matching process becomes too complex. For the same training data, another automatic modeling algorithm, the MDL approach, took about 1 h using four levels of refinement.

5. Conclusions and discussions

In this paper, we have proposed and illustrated a method for determining point correspondences among different shapes. Instead of using uniform samples as in Refs. [10,13], we describe a shape curve using a set of optimal samples extracted using both geometric and topological features to capture the crucial characteristic of the shape. Then the shape correspondence is obtained through a two-stage matching approach, which stands for the major contribution of this paper. The coarse mapping of the skeleton endings provides an initial mapping of the major parts of the shapes, yet allows the following local point matching so that the computational complexity can be significantly reduced. Meanwhile, the consecutive enforcement employed in the point matching process enables the proposed approach to avoid the mismatching raised in other landmark-based methods (e.g. Ref. [10]).

Compared with some methods which require the precise matching of the skeleton structures (e.g. Ref. [18]), our approach is more flexible and efficient. The introduced coarse-to-fine matching strategy enables it to handle occlusions and deformations naturally. The shape retrieval experiments show the validity of the proposed shape matching approach. The inherent flexibility of our technique makes it possible to establish correspondences in variable density, which is a particular admirable feature in applications where the information of a large number of objects is available in various levels. The quantitative results for shape modeling illustrate that the proposed method may generate better models than the traditional approaches and obtain comparable outputs with the MDL algorithm [6] at a considerably lower computational cost.

To further improve the robustness of our approach for complex shapes, the initial segment matching may be pursued in a multi-scale manner. The generality of this approach to tasks on larger data sets with complex background will also be explored in the future. The presented algorithm considers only objects represented by closed boundaries. The extension of this approach to open curves is straightforward by connecting the end-points with a straight line. However, it

is difficult for our approach to handle self-occluding boundaries due to the challenges of finding the accurate symmetry structures and the complex topology embedded in the resulting graphs. Addressing this problem is done in our current work, together with extending the correspondence algorithm to voxellated surfaces in 3D with a medial surfaces computation technique.

References

[1] T. Cootes, C. Beeston, G. Edwards, C. Taylor, A unified framework for atlas matching using active appearance models, in: Proceedings of the International Conference on Image Processing in Medical Imaging, Springer, Berlin, 1999, pp. 322–333.

[2] H. Blum, A transformation for extracting new descriptors of shape, in: Proceedings of the Models for the Perception of Speech and Visual Form, Cambridge, MA, 1967, pp. 362–380.

[3] J. Ruppert, A new and simple algorithm for quality 2-dimensional mesh generation, in: Proceedings of the ACM–SIAM Symposium on Discrete algorithms, 1993, pp. 83–92.

[4] Y. Gdalyahu, D. Weinshall, Flexible syntactic matching of curves and its application to automatic hierarchical classification of silhouettes, IEEE Trans. Pattern Anal. Mach. Intell. 21 (12) (1999) 1312–1328.

[5] E.G.M. Petrakis, A. Diplaros, E. Milios, Matching and retrieval of distorted and occluded shapes using dynamic programming, IEEE Trans. Pattern Anal. Mach. Intell. 24 (11) (2002) 1501–1516.

[6] R. Davies, C. Twining, T. Cootes, J. Waterton, C. Taylor, A minimum description length approach to statistical shape modeling, IEEE Trans. Med. Imag. 21 (5) (2002) 525–537.

[7] A. Hill, C. Taylor, A. Brett, A framework for automatic landmark identification using a new method of nonrigid correspondence, IEEE Trans. Pattern Anal. Mach. Intell. 22 (3) (2000) 241–251.

[8] H. H. Thodberg, H. Olafsdottir, Adding curvature to minimum description length shape models, in: Proceedings of the British Machine Vision Conference, vol. 2, 2003, pp. 251–260.

[9] A. Ericsson, K. Astrom, Minimizing the description length using steepest descent, in: Proceedings of the British Machine Vision Conference, vol. 2, Norwich, UK, 2003, pp. 93–102.

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

[11] A. Thayananthan, B. Stenger, P.H.S. Torr, R. Cipolla, Shape context and chamfer matching in cluttered scenes, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, vol. I, Madison, USA, 2003, pp. 127–133.

[12] Z. Tu, A.L. Yuille, Shape matching and recognition—using generative models and informative features, in: European Conference on Computer Vision, vol. 3, 2004, pp. 195–209.

[13] H. Ling, D.W. Jacobs, Using the inner-distance for classification of articulated shapes, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, 2005, pp. 719–726.

[14] T.-L. Liu, D. Geiger, Approximate tree matching and shape similarity, in: Proceedings of the IEEE International Conference on Computer Vision, Corfu, Greece, 1999, pp. 456–462.

[15] D. Sharvit, J. Chan, H. Tek, B.B. Kimia, Symmetry-based indexing of image databases, J. Visual Commun. Image Representation 9 (4) (1998) 366–380.

[16] K. Siddiqi, S. Bouix, A. Tannenbaum, S.W. Zucker, The Hamilton–Jacobi skeleton, in: Proceedings of the IEEE International Conference on Computer Vision, Corfu, Greece, 1999, pp. 828–834.

[17] P. Golland, W. Grimson, Fixed topology skeletons, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, vol. I, Hilton Head Island, SC, 2001, pp. 10–17.

[18] 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.

[19] S.M. Pizer, K. Siddiqi, G. Szekely, J.N. Damon, S.W. Zucker, Multiscale medial loci and their properties, Int. J. Comput. Vision 55 (2–3) (2003) 155–179.

[20] R.L. Ogniewicz, Skeleton-space: a multiscale shape description combining region and boundary information, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 1994, pp. 746–751.

[21] N. Amenta, M. Bern, M. Kamvysselis, A new Voronoi based surface reconstruction algorithm, in: Computer Graphics (SIGGRAPH'98 Proceedings), vol. 32, 1998, pp. 415–421.

[22] T. Dey, J. Giesen, S. Goswami, J. Hudson, R. Wenger, W. Zhao, Undersampling and oversampling in sample based shape modeling, in: Proceedings of the IEEE Visualization, vol. I, San Diego, CA, 2001, pp. 83–90.

[23] F. Leymarie, M.D. Levine, Simulating the grassfire transform using an active contour model, IEEE Trans. Pattern Anal. Mach. Intell. 14 (1) (1992) 56–75.

[24] R. Kimmel, D. Shaked, N. Kiryati, Skeletonization via distance maps and level sets, Comput. Vision Image Understanding 62 (3) (1995) 382–391.

[25] C.H. Papadimitriou, et al., Combinatorial Optimization: Algorithms and Complexity, Prentice-Hall, Englewood Cliffs, NJ, 1982.

[26] A. Goldberg, et al., A new approach to the maximum flow problem, J. Assoc. Comput. Mach. 35 (1988) 921–940.

[27] D.D. Sleator, R.E. Tarjan, A data structure for dynamic trees, J. Comput. Syst. Sci. 26 (3) (1983) 362–391.

[28] M. Leyton, Symmetry-curvature duality, Comput. Vision Graph. Image Process. 38 (3) (1987) 327–341.

[29] P. Soille, Morphological Image Analysis—Principles and Applications, Springer, Berlin, New York, 2004.

[30] L.J. Latecki, R. Lakamper, U. Eckhardt, Shape descriptors for non-rigid shapes with a single closed contour, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, Madison, USA, 2000, pp. 424–429.

[31] T.B. Sebastian, P.N. Klein, B.B. Kimia, On aligning curves, IEEE Trans. Pattern Anal. Mach. Intell. 25 (1) (2003) 116–125.

[32] Electronic field guide, ⟨http://herbarium.cs.columbia.edu/index.php⟩.

[33] J. Triesch, C. Malsburg, A system for person-independent hand posture recognition against complex backgrounds, IEEE Trans. Pattern Anal. Mach. Intell. 23 (12) (2001) 1449–1453.

About the Author—JUN XIE received the M.S.E degree in computer science from the Nanjing University of Science and Technology, Nanjing, China, in 2001, and the Ph.D. degree in electronic engineering from Chinese University of Hong Kong, Shatin, in 2004. During 2005, he worked as a Postdoctoral Fellow in the Department of Computer Science and Engineering, the Chinese University of Hong Kong. At present, he is a Research Associate at the Vision Lab of University of Central Florida. His current research focuses on ultrasound image segmentation, deformable shape modeling and human perception.

About the Author—PHEN-GANN HENG (S'90–M'92) received the B.Sc. degree from the National University of Singapore, Singapore, in 1985, and the M.Sc. degree in computer science, the M.A. degree in applied mathematics and the Ph.D. degree in computer science, all from Indiana University, Bloomington, in 1987, 1988 and 1992, respectively. Currently, he is a Professor in the Department of Computer Science and Engineering, the Chinese University of Hong Kong (CUHK), Shatin. In 1999, he set up the Virtual Reality, Visualization and Imaging Research Centre at CUHK and serves as the Director of the centre. He is also the Director of the CUHK Strategic Research Area in Computer Assisted Medicine, established jointly by the Faculty of Engineering and the Faculty of Medicine in 2000. His research interests include virtual reality applications in medicine, visualization, 3D medical imaging, user interface, rendering and modeling and interactive graphics and animation.

About the Author—MUBARAK SHAH is a Professor of computer science, and the Founding Director of the Computer Vision Laboratory at University of Central Florida (UCF). He is a co-author of two books "Video Registration" (2003) and "Motion-Based Recognition" (1997), both by Kluwer Academic Publishers. He has published close to 150 papers in leading journals and conferences on topics including activity and gesture recognition, violence detection, event ontology, object tracking (fixed camera, moving camera, multiple overlapping and non-overlapping cameras), video segmentation, story and scene segmentation, view morphing, ATR, wide-baseline matching and video registration. He is a Fellow of IEEE and was an IEEE Distinguished Visitor Speaker for 1997–2000. He received the Harris Corporation Engineering Achievement Award in 1999, the TOKTEN Awards from UNDP in 1995, 1997 and 2000, Research Incentive Award in 2003 and IEEE Outstanding Engineering Educator Award in 1997. He is an Editor of international book series on "Video Computing", Editor in Chief of Machine Vision and Applications journal and an Associate Editor of Pattern Recognition journal. He was an Associate Editor of the IEEE Transactions on Pattern Analysis and Machine Intelligence and a Guest Editor of the special issue of International Journal of Computer Vision on Video Computing.