Shape Retrieval Using Triangle-Area Representation and Dynamic Space Warping

Naif Alajlan, Ibrahim El Rube, Mohamed S. Kamel
and George Freeman

August, 2006

Abstract

We present a shape representation and a matching method for nonrigid shapes with closed contours. The representation utilizes the areas of the triangles formed by the boundary points to measure the convexity/concavity of each point at different scales (or triangle side lengths). This representation is effective in capturing both local and global characteristics of a shape, invariant to translation, rotation, and scaling, and robust against noise and moderate amounts of articulation and occlusion. In the matching stage, we employ a dynamic space warping (DSW) algorithm to search efficiently for the optimal (least cost) correspondence between the points of two shapes. Then, we derive a dissimilarity distance based on the optimal correspondence. We demonstrate the performance of our method using four standard tests on two well-known shape databases. The results show the superiority of our method over other recent methods in the literature.

1 Introduction

The ever growing number of images generated everyday has motivated researchers to develop, evaluate and implement sophisticated algorithms for the retrieval of images from large databases based on their content rather than their textual annotations alone. Among other generic image features that are used to achieve this objective, like color and texture, shape is considered the most promising for the identification of entities in an image. It can be argued

that most real subjects are identified easily using only their silhouettes. A user survey in [10] indicated that 71% of the users were interested in retrieval by shape.

The representation of shapes requires a number of criteria to be satisfied for reliable shape matching and retrieval. It should be invariant to geometrical transformations, such as rotation, scale, translation, and skew. This requirement arises from the problem of projecting the 3D real world objects into 2D images. In addition, a shape representation should satisfy the following criteria: high discrimination capability, computational efficiency, robustness to distortion and noise, compactness, generality of the application, and handling large image databases without heavy degradation in the performance. These criteria are also required by the MPEG-7 standard for measuring the similarity between shapes [16].

Global shape descriptors are generally robust to moderate amounts of noise. However, they face major difficulties in capturing fine details of shape boundaries. On the other hand, local shape descriptors are superior in describing fine details, but they are usually sensitive to noise. Therefore, in our opinion, the two requirements (the robustness to noise and the discrimination of fine details) conflict with each other and the choice between them is context-dependent unless another semantic-based measure is employed to distinguish between noise and fine details information.

In this report, triangle-area representation (TAR) is utilized in order to derive multi-scale, affine-invariant descriptors for 2D closed-boundary shapes where the triangle areas at each boundary point are used in the matching via dynamic programming (DP). Many researchers adopted the areas of the triangles between boundary points for shape representation [20, 7, 26, 25]. However, the generation of multi-scale (i.e. of different triangle side lengths) signatures at all boundary points proposed is novel. TAR provides useful information about shape features such as the convexity/concavity at each boundary point; therefore, TAR provides high discrimination capability. For shape matching, a dynamic programming (DP) algorithm, which is called Dynamic Space Warping (DSW), is employed to find the best alignment between two shape representations.

The main contribution of the technique presented in this report is its ability to provide higher retrieval accuracy than all published methods based on the MPEG-7 CE-shape-1 part B retrieval test [18], which constitutes the most comprehensive shape retrieval test in the literature so far. Besides, TAR exhibits high robustness to affine transformations which enables it to

be suitable for 3D applications. TAR is also robust against noise and moderate amounts of deformations. Regarding the computational complexity, the matching complexity of our algorithm is $O(N^2)$, where N is the number of the boundary points.

The remainder of this report is organized as follows. Section 2 reviews the related work in the literature. The TAR representation is presented in Section 3. Then, Section 4 describes our DSW matching algorithm. Section 5 discuses the computational complexity of our approach and Section 6 includes the experimental results. Finally, Section 7 concludes the report.

2 Related Work

The literature on 2D shape representation and matching is relatively huge. Good review papers can be found in [15, 28]. However, we focus our review here on selected methods that are based on dynamic programming (DP) for two main reasons. The first is their close relation with our work. The second reason is that DP-based methods generally offer better performance than other methods which do not use DP.

Bartolini *et al* proposed a method for shape matching and retrieval based on the Fourier descriptors that is called *WARP* [4]. They chose to use the phase of the Fourier descriptors and claimed to outperform the state-of-the-art Fourier-based methods. At first, the low-frequency coefficients are normalized in terms of translation, scale and rotation. For matching, the inverse DFT is used to obtain normalized versions of the original contours in the spatial domain. Then, a DP method is employed to find the similarity distance between the two transformed contours. Although this technique outperformed other Fourier-based methods, the authors reported less retrieval accuracy than the curvature scale-space (CSS) method [1]. Besides, WARP is not invariant to the general affine transformation which makes it not suitable for 3D applications.

Ling and Jacobs were mainly motivated by computing a distance for articulated shapes [14]. The inner-distance, which is the length of the shortest path between two boundary points within the shape boundary, was derived to be invariant to shape articulation. The authors claim that the inner-distance is the natural replacement to the well-known Euclidean distance. In order to apply the inner-distance for shape matching and retrieval, the authors extended the shape context method [5] using this distance and called it inner-

distance shape context (IDSC). Then, DP is used for matching shapes after calculating the IDSC distances. The retrieval performance on the MPEG-7 dataset is 85.4% which is the highest published performance so far. The main disadvantage of this method is the use of different parameter settings for different databases. This method has not been tested on geometric transformed shapes.

Adamek and O'Connor proposed a multi-scale representation for a single closed contour that makes use of both concavities and convexities of all contour points [2]. It is called multi-scale convexity concavity (MCC) representation where different scales are obtained by smoothing the boundary with Gaussian kernels of different widths. Then, a new measure for the curvature was proposed that is based on the relative displacement of a contour point with respect to its position in the preceding scale level. This idea is motivated by the observation that when smoothing a closed contour, convex and concave points are moved inside and outside the contour, respectively. Afterwards, the matching is done using a DP approach. The MCC was able to achieve 84.9% retrieval accuracy on the MPEG-7 dataset. However, the MCC suffers from being computationally expensive, $O(N^3)$ where N is the number of contour points. Also, the robustness of the MCC method to affine transformation was not demonstrated which limits its application to 3-D images.

Petrakis *et al* proposed an approach for matching open and closed shapes using DP [19]. In their approach, implicit multi-scale matching takes place through matching merged contour segments in order to avoid the cost of computing the scale space explicitly. The DP algorithm examines all possible merges of small segments of one shape to match with larger segments of the other and selects the best merge. The authors did not report superior performance over other existing methods. Other limitations of this method include the lack of robustness to the general affine transformations and the high computational complexity of the matching process.

Sebastian *et al* proposed a curve alignment approach, which is called curve edit distance (CED), for matching open and closed curves [22]. In their method, the correspondence between the points of the two curves is controlled by the relative difference in their spatial location and their curvature. Then, a matching function is defined as the minimum cost of such correspondence. The search for the optimal correspondence is made efficient by decomposing each curve into segments, which is ideally solved using DP. Moreover, merging, deletion and addition of curve segments are allowed in

order to account for shape deformations. The complexity of this method is $O(N^2 \log N)$ for closed curves and the authors reported 78.2% accuracy for MPEG-7 part B retrieval test.

Arica and Vural proposed another descriptor for closed contours based on the curvature information of all boundary points, which is called Beam Angle Statistics (BAS) [3]. In BAS, the curvature at each boundary point is viewed as a random variable that draws its values from the angles between each equally-distant neighboring points at that point. Then, few order moments are computed for the random variable at each point. For measuring the similarity, DP is used to find the best correspondence that minimizes the Euclidean distance between the signatures of two shapes. The authors reported 82.4% accuracy using the MPEG-7 part B retrieval test.

In another recent work, Latecki *et al* presented a shape matching approach that works directly on the the closed boundaries [11, 13]. It is based on visual parts (VP), where (part of) a database shape is simplified in the context of the query shape prior to their matching. The simplification process includes the elimination of particular points from the database shape such that the similarity to the query shape is maximized. The main disadvantage of this method is the high computational complexity of the matching algorithm, which is $O(N^3 \log N)$ where N is the number of the boundary points.

3 Triangle Area Representation of Closed Boundaries

Many researchers have used the area of the triangle, formed by the boundary points, as the basis for shape representations, for example [20, 7, 26, 25]. These methods use a global measure for the signature normalization. In this report, a shape representation that is based on the area of the triangle is introduced, namely triangle area representation (TAR). Unlike previous methods, the signature at each scale, i.e, triangle side length, is normalized locally based on that scale. Fig. 1 shows the difference between local and global normalization of the triangle areas (here, t_s is the triangle side length, n is the boundary point index, and N is the number of boundary points). Clearly, the local normalization provides higher discrimination between shape properties at different scales. Unlike some other shape description methods,

which use a finite number of discriminative boundary points (such as corners or inflection points), our representation equally employs all boundary points in a systematic way.

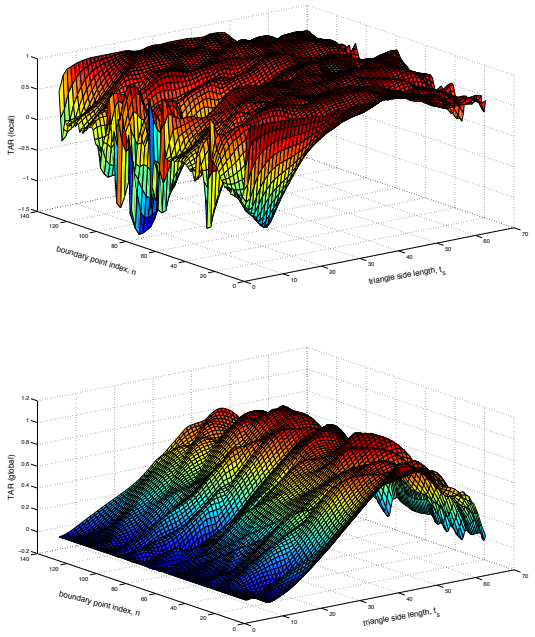

Figure 1: 3D plots of the TAR signatures ($t_s = 1$ to 63 and $N = 128$) for the shape in Fig. 3. The two plots differ only in the normalization method of the signatures. TAR signatures normalized locally per scale (top) and normalized globally according to [26] (bottom).

3.1 TAR signatures

The TAR signature is computed from the area of the triangles formed by the points on the shape boundary. Each contour point is represented by its x and y coordinates and separated parameterized contour sequences x_n and y_n are obtained and re-sampled to N points. Then, the curvature of each point is measured using the triangle area representation (TAR) as follows. For each three consecutive points (x_{n-t_s}, y_{n-t_s}), (x_n, y_n), and (x_{n+t_s}, y_{n+t_s}),

where $n \in \langle 1, N \rangle$ and $t_s \in \langle 1, T_s \rangle$ is the triangle side length. The signed area of the triangle formed by these points is given by:

$$TAR(n, t_s) = \frac{1}{2} \begin{vmatrix} x_{n-t_s} & y_{n-t_s} & 1 \\ x_n & y_n & 1 \\ x_{n+t_s} & y_{n+t_s} & 1 \end{vmatrix} \quad (1)$$

When the contour is traversed in counter clock-wise direction, positive, negative and zero values of TAR mean convex, concave and straight-line points, respectively. Fig. 2 demonstrates these three types of the triangle areas. The triangles at the edge points are formed by considering the periodicity of the closed boundary. Fig. 2 also shows the complete TAR signature for the hammer shape. By Increasing the length of the triangle sides, i.e., considering farther points, the function of (1) will represent longer variations along the contour.

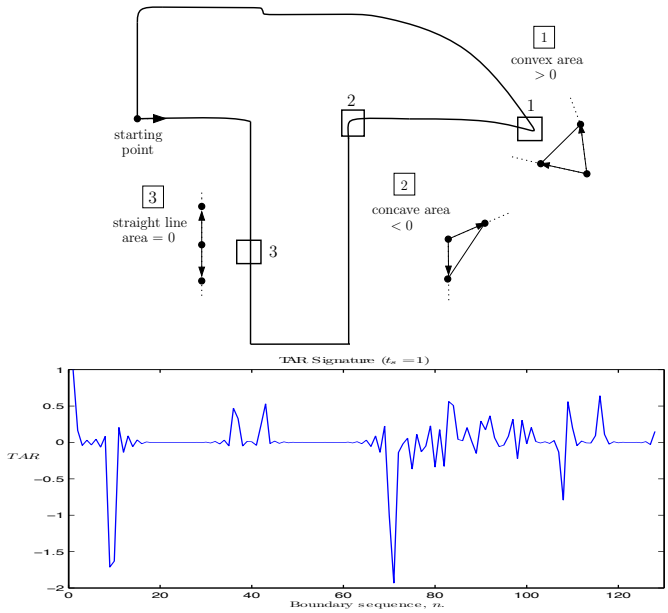

Figure 2: Three different types of the triangle-area values and the TAR signature for the hammer shape.

The choice of the number of scales, i.e., triangle side lengths (T_s), is

constrained by the implied periodicity of the closed boundaries. More specifically, for a closed contour of N points:

$$TAR(n,t_s) \;=\; \begin{cases} -TAR(n,N+1-t_s) & t_s = 1\ldots\lfloor\frac{N-1}{2}\rfloor \\ 0 & at\ t_s=\frac{N}{2}\ and\ N\ is\ even \\ does\ not\ exist & at\ t_s=\frac{N}{2}\ and\ N\ is\ odd \end{cases} \quad (2)$$

where $\lfloor\frac{N-1}{2}\rfloor$ is the floor value of $\frac{N-1}{2}$. The first line in (2) shows the odd symmetry property of the triangle area versus the triangle side length t_s. Also, at t_s equals the middle point of the boundary, the value of the triangle area depends on N, the total number of points on that boundary. If N is odd, then there will be no zero-crossing points on the area curve. Usually, researchers tend to use an even number of points on the shape boundary. In this case, the inflection (zero-crossing) point exists at $t_s = \frac{N}{2}$, where $TAR(n,\frac{N}{2}) = 0$.

Fig. 3 illustrates the odd symmetry property of the TAR signature that is computed for only one point on the shape boundary of the Misk shape. A complete 3D plot of the TAR signatures for the Misk shape is shown in Fig. 1. The two plots shows the difference between local (left) and global (right) normalization of the signatures. In this report, the local normalization is performed at each scale (or triangle side length) with respect to the maximum area at that scale. In [20, 7, 26, 25], the normalization was carried out by dividing on the total sum of the signatures, i.e., global.

3.2 The relation between TAR and the curvature function

The curvature function can be rewritten as:

$$c(n) = \frac{\dot{x}_n\ddot{y}_n - \ddot{x}_n\dot{y}_n}{(\dot{x}_n^2+\dot{y}_n^2)^{3/2}} = \frac{TAR(n+1,1)}{(ds_n)^3} \quad (3)$$

where $TAR(n+1,1)$ is the triangle areas at $t_s = 1$ and the distance $ds_n = \sqrt{\dot{x}_n^2+\dot{y}_n^2}$ corresponds to the length of the first triangle side, i.e., the distance between the first and the second vertices of the triangle formed by the points $(x_n, y_n), (x_{n+1}, y_{n+1})$, and (x_{n+2}, y_{n+2}). Equation (3) clearly shows the relation between the triangle area and the discrete curvature function. The one-point shift between the triangle area and the curvature can be easily

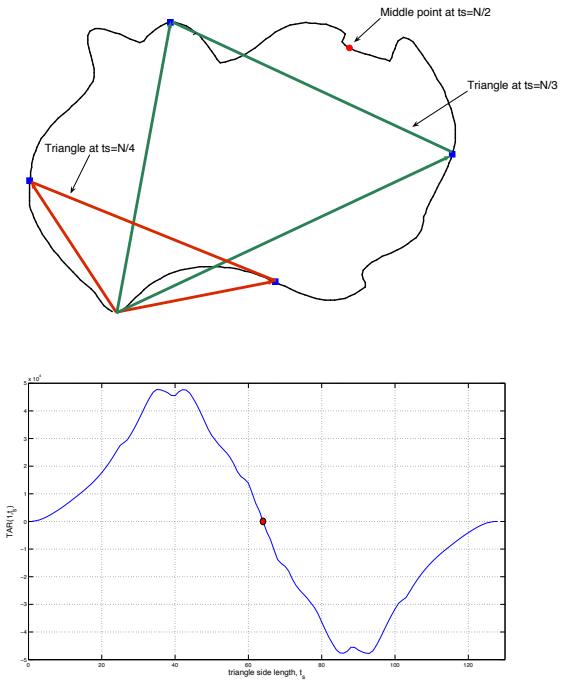

Figure 3: Illustration of the odd symmetry of the triangle-area signatures. Two virtual triangles are shown (top) for computing the TAR signature (bottom). The circled point on the boundary is the middle-point, which is also considered as the inflection point for the TAR signature.

eliminated by cyclic shifting one of them against the other. The distance ds only depends on two points of the triangle triplets. Furthermore, cubing this distance increases its domination over the triangle area for the extreme values (the very small and the very large values of ds).

In [1, 17], it is proved that the zero-crossings of the curvature function are invariant under the general affine transformation. However, non-zero curvature function points are not invariant to these transformations. Fig. 4 shows a comparison between the TAR and curvature function signatures. Clearly, the TAR is invariant to the skew transformation whereas the curvature function is not.

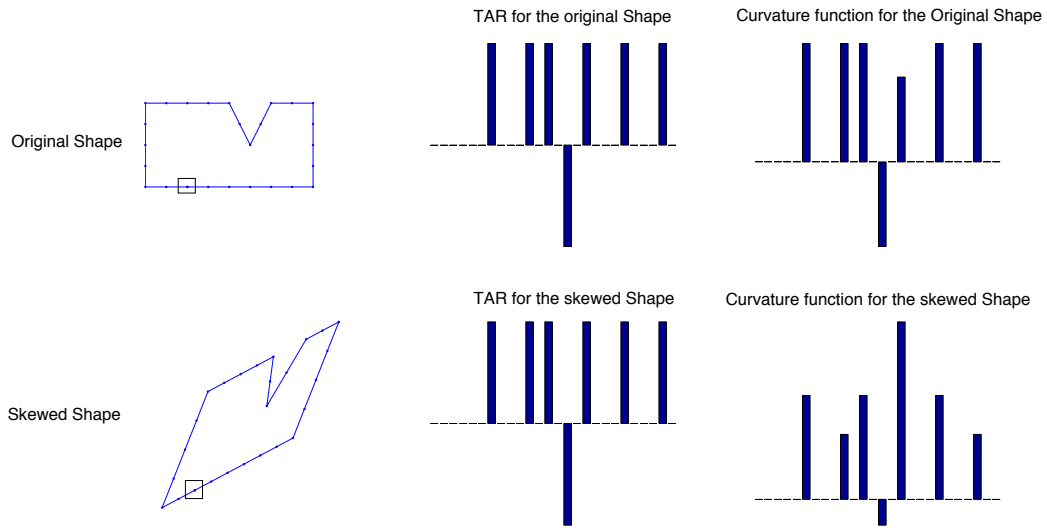

Figure 4: A comparison between the TAR and the curvature function under affine transformation.

3.3 TAR under general affine transformations

For a 2D shape, represented by its contour sequences, x_n and y_n, and subjected to an affine transformation, the relation between the original and the distorted sequences is given by:

$$\begin{bmatrix} \hat{x}_n \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} \tag{4}$$

where $\hat{x}$ and $\hat{y}$ are the affine distorted sequences, e and f represent translation, and a, b, c and d reflect scale, rotation and shear. The effect of the translation parameters is easily eliminated by normalizing the shape boundary with respect to its centroid. This normalization is achieved by subtracting from each boundary sequence its mean value. By substituting (4) into (1), we obtain:

$$\hat{TAR}(n, t_s) = (ad - bc)\, TAR(n, t_s) \tag{5}$$

where $\hat{TAR}$ is the affine transformed version of TAR. It is clear that $\hat{TAR}$ is relatively invariant to the affine transformations. Absolute invariance can

be achieved by dividing $T\hat{A}R$ by its maximum value. A complete affine test is given in the experimental results section.

4 Shape Matching Using Dynamic Space Warping

In this section, a brief review about the origin of Dynamic Time Warping[1] (DTW) and how some researchers applied it to the shape matching problem is presented. Then, the description of our DSW algorithm to measure the similarity between two shapes based on their TARs is given followed by the definition of the dissimilarity function. Finally, an indexing scheme using geometric features is shown.

4.1 DTW in Shape Matching

The idea of using dynamic programming for matching 1D sequences originally came from the speech recognition community [8, 21, 6] where optimal alignment between two 1D sequences is searched via dynamic programming, which was called Dynamic Time Warping (DTW). In the past few years, several researchers adopted DTW for 1D sequences alignment and matching [4, 27]. Unlike the Euclidean distance that provides one-to-one alignment, nonlinear alignment can be achieved by the DTW, where one point on the sequence can be aligned to one or more points on another sequence.

Recently, many researchers have applied DTW in the 2D shape matching problem. In [19], a DP table is used to find the least cost match between segments of two curves. Merging of segments is allowed during the matching to facilitate a more meaningful correspondence between segments. However, this increases the complexity of matching. In the MCC method [2], their DP algorithm searches for the optimal correspondence between the N-points boundaries. A window, which limits the optimal path to be around the diagonal, is used to make the search more efficient. Another constraint that limits a single point of one contour to correspond to a maximum of two points on the other contour is enforced, which limits the generality of the method and demands more computations. On an attempt to reduce the size of the

[1]Here, the analogous terminology Dynamic Space Warping (DSW) is used instead of DTW since still images are space-variant as opposed to speech which is time-variant.

DP search space, WARP method [4] applies DTW on normalized points after applying the inverse discrete Fourier transform. In [22], the optimal path in the DP table is used to define an edit distance metric that transforms one shape into the other.

4.2 Finding the Minimum Cost Distance Using DSW

Now, we describe our DSW algorithm that is used to compute the dissimilarity distance between two closed contours based on their TARs. At first, it is necessary to define the distance between two individual contour points. Let $TAR_A(n, t_s)$ and $TAR_B(n, t_s)$ be the TARs for shapes A and B, respectively, where $n \in \langle 1, N \rangle$ is the index of the boundary points and $t_s \in \langle 1, T_s \rangle$ is the triangle side length. Then, the distance between the two contour points $n \in A$ and $m \in B$ is defined as:

$$D(n, m) = \frac{1}{T_s} \sum_{t_s=1}^{T_s} |TAR_A(n, t_s) - TAR_B(m, t_s)| \tag{6}$$

Then, an $N \times N$ distance table, DT, is constructed to find the optimal correspondence between the points of the two contours. The columns of DT represent the points of one contour and the rows represent the points of the other. Initially, the elements of DT are set as:

$$DT_{initial}(n, m) = \begin{cases} 0, \ max(1, n - w + 1) \leq m \leq min(N, n + w - 1) \\ \infty, \ otherwise \end{cases} \tag{7}$$

Where $n, m \in \langle 1, N \rangle$, w is a predefined diagonal width for DT as illustrated in Fig. 8, and $max(a, b)$ and $min(a, b)$ are the maximum and minimum values of a and b, respectively. Only the elements of DT that fall within w are updated during the DSW search. This initialization of DT avoids computing the distances between all the points of two contours and restricts the distance computation to only those points which are more likely correspond to each other. Therefore, the computational complexity is largely reduced while more meaningful correspondences are obtained.

Starting at an arbitrary TAR point for both contours A and B, the distance table DT is searched, through the diagonal window of width w, left-to-right and up-to-bottom starting from the upper-left element, as shown in Fig. 8. The first row and first column elements are initialized as the distance

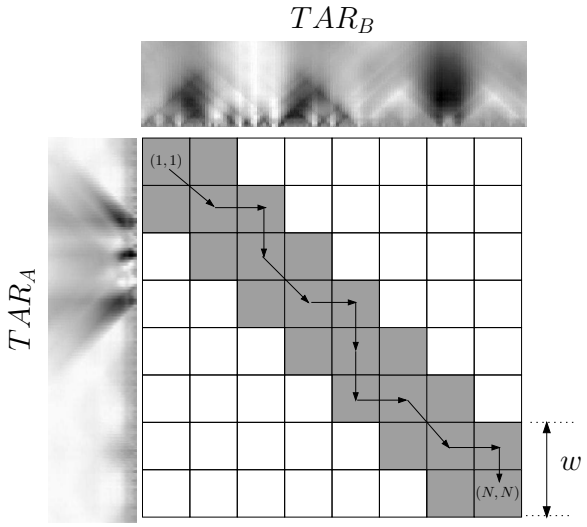

Figure 5: DSW table.

between the corresponding points using (6). Then, the rest of the zero-valued elements of DT are updated as:

$$DT(n,m) = D(n,m) + \min \begin{cases} DT(n-1,m) \\ DT(n-1,m-1) \\ DT(n,m-1) \end{cases} \tag{8}$$

The least cost path through the distance table is the value of element $DT(N,N)$, which corresponds to the best matching between the two TAR points according to the selected starting points. However, it is clear that the established correspondence is sensitive to the starting point of each TAR. In order to achieve starting point (or rotation) invariance, it is sufficient to fix the starting point of one TAR and try all N starting points of the other TAR. Moreover, invariance to the mirror transformation can be obtained by flipping the points of one TAR and repeat the search for the N starting points again. The final least cost correspondence is taken as the minimum value of $DT(N,N)$ among all $2N$ runs of the DSW table search, denoted by DT_{min}.

4.3 The Dissimilarity Distance Measure

Following the approach presented in [2], the dissimilarity distance (D_{dis}) is chosen as the minimum cost distance DT_{min}, normalized by the *shape*

complexity (SC) of each contour. The motivation behind this normalization is based on the observation that the sensitivity of the human perception to the boundary variations reduces as the shape complexity increases. Here, the shape complexity is considered as the average, over all boundary points, of the absolute differences between the maximum and minimum TAR values at all scale levels (or triangle side lengths):

$$SC = \frac{1}{N} \sum_{n=1}^{N} \left| \max_{t_s} \{TAR\,(n, t_s)\} - \min_{t_s} \{TAR\,(n, t_s)\} \right| \tag{9}$$

Then, the dissimilarity distance between two shapes, A and B, is given by:

$$D_{dis}\,(A, B) = \frac{DT_{min}\,(A, B)}{K + SC_A + SC_B} \tag{10}$$

where $DT_{min}(A, B)$ is the minimum cost distance between shapes A and B, computed using the DSW table search, and SC_A and SC_B are the complexities of shapes A and B, respectively. A constant K is added to prevent the domination of the denominator when the complexities are very small. In our experiments, K is set to 1.

4.4 Indexing Using Geometric Features

In many practical applications, it is highly desired that the shape descriptor provides means for indexing in order to organize the database efficiently. Abbasi *et al* [1] used a set of global features, i.e., circularity, eccentricity, and aspect ratio, at an initial stage to discard very dissimilar shapes and increase the discrimination power of the descriptor. Jain and Vailaya [9] used invariant moments and histograms of edge directions for fast pruning of the database.

In our method, a set of simple geometric features are used to further increase the discrimination ability of the dissimilarity distance, which includes aspect ratio (AR), eccentricity (E), and solidity (S). These features include considerable information about the global properties of a shape. However, since some dissimilar shapes have comparable global features, the indexing using the global features comes at the price of the accuracy. Therefore, the final dissimilarity distance between shapes A and B is given as:

$$D_f(A,B) = \alpha_{ar}\,|AR_A - AR_B| + \alpha_e\,|E_A - E_B| + \alpha_s\,|S_A - S_B| + D_{dis}(A,B) \tag{11}$$

where AR_A, E_A, and S_A are the aspect ratio, eccentricity, and solidity of shape A (same for shape B), and α_{ar}, α_e and α_s are the associated weights. Our experiments show that (11) performs effectively under a wide range of the weight values, which supports the generality of our approach.

5 Computational Complexity

The complexity of each of the TAR and the matching stage is evaluated separately. It should be noted that the complexity of the matching stage is more critical since the TARs of the database images can be computed prior to the time of the matching, whereas the matching usually takes place between the query image and most (if not all) the database images.

The TAR computation involves calculating the triangle area, according to (1), at each of the N points of the boundary. In addition, at each boundary point, the triangle area is calculated at different scales (or triangle side lengths). Typically, there are $\frac{N-1}{2}$ scales (see subsection 3.1). Therefore, the computational complexity of the TAR stage is $O\left(N.\frac{N-1}{2}\right)$ or $O\left(N^2\right)$.

For the matching function given by (11), each of the first three terms involves a single computation of the absolute difference operation; therefore, the forth term D_{dis} given by (10) governs the complexity of the matching stage. Furthermore, each of the shape complexity terms SC_A and SC_B in (10) requires $O\left(N\right)$ complexity as given by (9). For the minimum cost distance term DT_{min}, the DSW table search is restricted within the diagonal w-width window; thus, the DSW table search complexity is $O\left(wN\right)$ (usually $w << N$). Since the DSW search is repeated for N starting points, the complexity becomes $O\left(wN^2\right)$. Finally, by considering the flipping operation, the total complexity of the matching stage turns out to be $O\left(2wN^2\right)$ or $O\left(N^2\right)$ (for $N = 128$, our experiments show that $w = 3$ is good enough and larger w doesn't achieve better results).

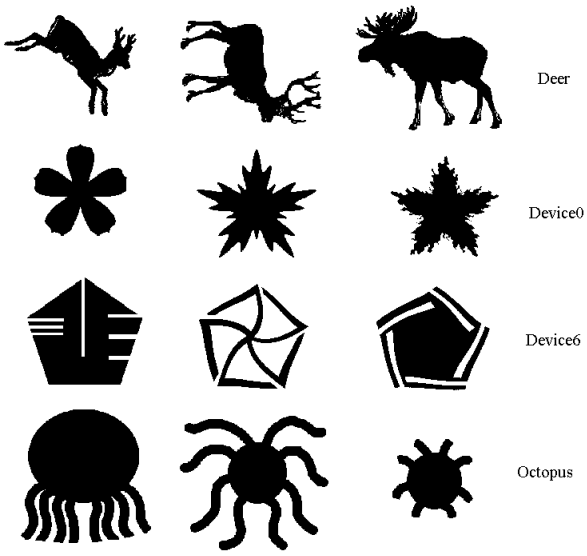

Figure 6: Samples of the MPEG-7 CE-shape-1 database.

6 Experimental Results

In this section, we demonstrate the performance of our method using four standard experiments on two shape databases. The first database is the well-known MPEG-7 CE-Shape-1 database [12] which consists of 1400 images semantically classified into 70 classes. This database contains a mixture of natural and man-made objects under various rigid and non-rigid deformations (a sample of the database is shown in Fig. 6). The other database is the Kimia's database [23] which contains 99 images for 9 categories as shown in Fig. 7. There are 11 images for each category and most of the images are partially occluded.

6.1 Robustness to Scaling and Rotation

Here, the results of our method, according to the MPEG-7 Core Experiment CE-Shape-1 part A1 (for scaling) and part A2 (for rotation) tests, are pre-

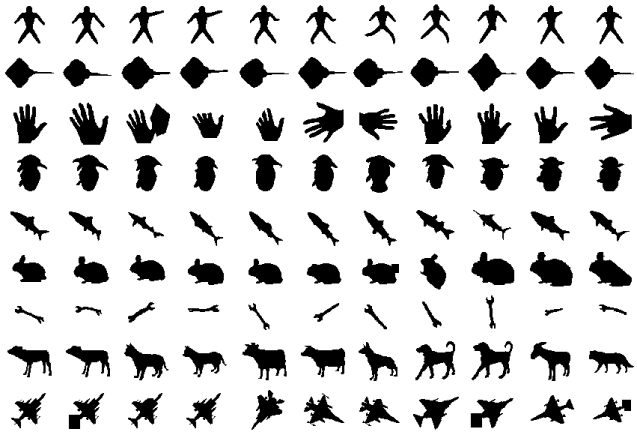

Figure 7: Kimia's database.

sented. The database used in part A1 includes 420 shapes, 70 basic shapes from the MPEG-7 CE-Shape-1 database (one shape per class) and 5 derived shapes from each basic shape by scaling the images with factors 2, 0.3, 0.25, 0.2, and 0.1. Each of the 420 shapes was used as a query and the number of correct matches were founded among the first 6 retrieved shapes.

Similarly, the database used for part A2 test consists of 420 shapes as in part A1, but the derived images are obtained by rotating each basic image with angles 9, 36, 45, 90, and 150 degrees. The correct matches were evaluated as in part A1.

Table 1 summarizes the results of both tests for our method along with three other methods from the literature. The CSS method is due to Mokhtarian *et al* [17, 1] and has been selected as the MPEG-7 standard for the boundary-based shape descriptor after comprehensive experiments. The other two methods are the BAS [3] and the VP [11] (see section 2 for details). The results in the table clearly shows that the DSW matching given by (10) performs better than the others even without using the global parameters of the shapes. Using DSW+global given by (11) further improves the performance.

Table 1: Comparison of the results of different methods on the MPEG-7 CE-shape-1 part A test.

Test	BAS [3]	CSS [17]	VP [11]	DSW	DSW+Global
Part A1	90.87%	92.86%	88.65%	95.1%	98.02%
Part A2	100%	100%	100%	100%	100%

Table 2: Comparison of the results of different methods on the MPEG-7 CE-shape-1 part B (bulls-eye) test.

Test	BAS [3]	CSS [17]	VP [11]	MCC [2]	WARP [4]	IDSC [14]	CED [22]	DSW	**DSW+Global**
Part B	82.37%	81.12%	76.45%	84.93%	58.50%	85.40%	78.17%	85.03%	**87.23%**

6.2 Similarity Retrieval Test

The retrieval effectiveness of our method is evaluated using the MPEG-7 Core Experiment CE-Shape-1 part B test (also called *bulls-eye* test), which is the main part of CE-Shape-1. All the 1400 images of the MPEG-7 database were used here. Each image was used as a query; then, the number of correct matches was counted in the first 40 retrieved shapes. As stated in [12], a 100% retrieval rate in this case is not possible using only the shape information since many classes contain very different objects. In our opinion, this shape retrieval test is the most challenging in the literature so far.

Table 2 shows the results of our DSW method and many recent methods; namely, BAS [3], CSS [17], VP [11], MCC [2], WARP [4], IDSC [14], and CED [22] (see section 2). The best performance based on the bulls-eye test was reported as 85.4% [14]. Our DSW method achieves high retrieval accuracy (85.03%) even without using the global parameters. When the global parameters are incorporated in the matching function, our method outperforms all existing methods in the retrieval accuracy. Fig. 8 shows the breakdown of the total retrieval rate into the retrieval rates for each class for both DSW+Global and the MCC method [2].

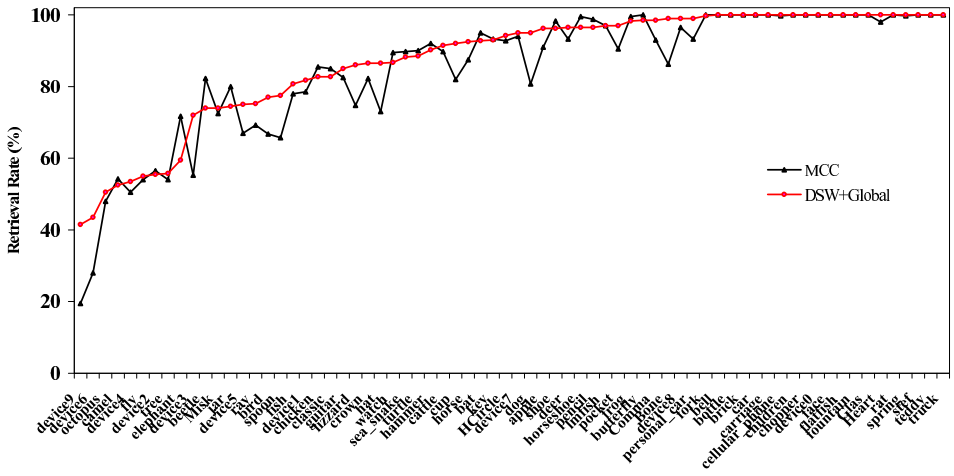

Figure 8: Results of the MPEG-7 CE-shape-1 part B test for each class for both DSW+Global and MCC methods.

6.3 Retrieval Using the Kimia's Database

In this test, the Kimia's database [23] is used. As shown in Fig. 7, partial occlusion is the main factor of variation among shapes of the same category. Each shape in the database is considered as a query and the first 10 retrieved shapes, excluding the query, are determined. Then, the correct retrievals for each ranking, over all 99 shapes, are counted. Table 3 summarizes these results for the DSW and three other methods. Note that the maximum number of correct retrievals in each case is 99. The shock graph edit (SGE) method [24] breaks down a shape's skeleton into parts (or shocks) and represents them as graph nodes and their relations as graph edges; thus, it handles partial occlusions explicitly. The SGE outperforms the shape context method [5] in this test. However, the performance of the SGE method on the MPEG-7 part B test was not reported. In contrast, the IDSC method [14] slightly outperforms the SGE method. Table 3 also shows the results of DSW given by (10), DSW+Global given by (11) with same parameter set of the previous tests, and DSW+Global* given by (11) with parameters tuned up specifically for this test (as in the IDSC method [14]). Our method achieves satisfactory performance without adjusting the parameters and achieves comparable

Table 3: results on Kimia database of 99 shapes. The table shows the number of correct retrievals, over all 99 shapes, at different rankings. See text for details.

Method	Ranking of the retrieved shape									
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
Shape Context [5]	97	91	88	85	84	77	75	66	56	37
SGE [24]	99	99	99	98	98	97	96	95	93	82
IDSC [14]	99	99	99	98	98	97	97	98	94	79
DSW	99	99	96	97	96	94	91	84	70	45
DSW+Global	99	99	99	96	95	97	95	84	71	50
DSW+Global*	99	99	99	98	98	97	98	95	93	80

performance to the IDSC method when the parameters are tuned up.

7 Concluding Remarks

In this report, we presented a closed-boundary shape representation that employs the triangle, formed by the boundary points at different scales, to measure the convexity/concavity of each boundary point. For the matching, the optimal correspondence between the points of two shapes is searched efficiently using a dynamic space warping algorithm. Based on the established correspondence, a dissimilarity distance is derived. Global features (aspect ratio, eccentricity, and solidity) of the shapes are incorporated in the dissimilarity distance to further increase the discrimination ability and to facilitate the indexing in large shape databases. The proposed technique is invariant to translation, rotation, and scaling and robust against noise and moderate amounts of noise and occlusion. For the MPEG-7 CE-shape-1 part B test, which is considered the most comprehensive shape retrieval test yet, our method outperforms all existing methods by a good margin.

References

[1] S. Abbasi, F. Mokhtarian, and J. Kittler. Curvature scale space image in shape similarity retrieval. *MultiMedia Systems*, 7(6):467–476, 1999.

[2] T. Adamek and N. E. O'Connor. A multiscale representation method for nonrigid shapes with a single closed contour. *IEEE Trans. on Circuits and Systems for Video Tech*, 14(5):742–753, May 2004.

[3] N. Arica and F. Vural. Bas: a perceptual shape descriptor based on the beam angle statistics. *Pattern Recognition Letters*, 24(9-10):1627–1639, 2003.

[4] I. Bartolini, P. Ciaccia, and M. Patella. Warp: accurate retrieval of shapes using phase of fourier descriptors and time warping distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):142–147, 2005.

[5] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(24):509–522, 2002.

[6] J. Deller, J. Hansen, and J. Proakis. *Discrete-Time Processing of Speech Signals*. Wiley-IEEE Press; Reprint edition, 1999.

[7] H. H. S. Ip and D. G. Shen. An affine-invariant active contour model (ai-snake) for model-based segmentation. *Image and Vision Computing*, 16(2):135–146, 1998.

[8] F. Itakura. Minimum prediction residual principle applied to speech recognition. *IEEE Trans Acoustics Speech Signal Process ASSP*, 23:5272, 1975.

[9] A. K. Jain and Aditya Vailaya. Shape-based retrieval: A case study with trademark image databases. *Pattern Recognition*, 31(9):1369–1390, 1998.

[10] S. Lambert, E. de Leau, and L. Vuurpijl. Using pen-based outlines for object-based annotation and image-based queries. In *Visual Information and Information Systems, 3rd International Conference*, pages 585–592, Amsterdam, The Netherlands, June 1999.

[11] L. J. Latecki and R. Lakamper. Shape similarity measure based on correspondence of visual parts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1185–1190, 2000.

[12] L. J. Latecki, R. Lakamper, and U. Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 424–429, 2000.

[13] L. J. Latecki, R. Lakamper, and D. Wolter. Optimal partial shape similarity. *Image and Vision Computing*, 23:227–236, 2005.

[14] H. Ling and D. Jacobs. Using the inner distance for classification of articulated shapes. In *IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 719–726, San Diego, CA, USA, 20-26 June 2005.

[15] S. Loncaric. A survey of shape analysis techniques. *Pattern Recognition*, 31(8):983–1001, 1998.

[16] J. M. Martinez. Mpeg-7 overview (version 9). Technical Report ISO/IEC JTC1/SC29/WG11N5525, ISO/IEC JTC1/SC29/WG11, International Organisation for Standardisation, Coding of Moving Pictures and Audio, March 2003.

[17] F. Mokhtarian and M. Bober. *Curvature Scale Space Representation: Theory, Applications, and MPEG-7 Standardization*. Kluwer Academic Publishers, 2003.

[18] The MPEG Home Page. http://www.chiariglione.org/mpeg.

[19] E.G.M. Petrakis, A. Diplaros, and E. Milios. Matching and retrieval of distorted and occluded shapes using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(11):1501–1516, November 2002.

[20] K. Roh and I. Kweon. 2-d object recognition using invariant contour descriptor and projective refinement. *Pattern Recognition*, 31(4):441–445, 1998.

[21] H. Sakoe and S.Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Signal Process.*, 26:43–49, 1978.

[22] T. Sebastian, P. Klein, and B. Kimia. On aligning curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(1):116–124, 2003.

[23] T. Sebastian, P. Klein, and B. Kimia. Computationally efficient wavelet affine invariant functions for shape recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):550–, 2004.

[24] T. Sebastian, P. Klein, and B. Kimia. Recognition of shapes by editing their shock graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):550–571, 2004.

[25] D. G. Shen, H. H. S. Ip, and E. K. Teoh. Affine invariant detection of perceptually parallel 3d planar curves. *Pattern Recognition*, 33(11):1909–1918, 2000.

[26] D. G. Shen, W. Wong, and H. H. S. Ip. Affine invariant image retrieval by correspondence matching of shapes. *Image and Vision Computing*, 17(7):489–499, 1999.

[27] K. Wang and T. Gasser. Alignment of curves by dynamic time warping. *Annals of Statistics*, 25(3):1251–1276, 1997.

[28] D. Zhang and G. Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19, 2004.