

# Invariant multi-scale descriptor for shape representation, matching and retrieval<sup>☆</sup>



Jianguo Yang<sup>a</sup>, Hongxing Wang<sup>b,\*</sup>, Junsong Yuan<sup>c</sup>, Youfu Li<sup>d</sup>, Jianyang Liu<sup>e</sup>

<sup>a</sup>School of Urban Rail Transportation, Soochow University, Suzhou, Jiangsu, China

<sup>b</sup>School of Software Engineering, Chongqing University, Chongqing, China

<sup>c</sup>School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore

<sup>d</sup>Department of Mechanical and Biomedical Engineering, City University of Hong Kong, Hong Kong

<sup>e</sup>Department of Mechanical Engineering, Southwest Jiaotong University, Mount Emei City, Sichuan Province, China

## ARTICLE INFO

### Article history:

Received 19 July 2015

Accepted 12 January 2016

Available online 21 January 2016

### Keywords:

Invariant descriptor

Shape representation

Shape matching

Contour

## ABSTRACT

Shape matching and retrieval have been some of the fundamental topics in computer vision. Object shape is a meaningful and informative cue in object recognition, where an effective shape descriptor plays an important role. To capture the invariant features of both local shape details and visual parts, we propose a novel invariant multi-scale descriptor for shape matching and retrieval. In this work, we define three types of invariants to capture the shape features from different aspects. Each type of the invariants is used in multiple scales from a local range to a semi-global part. An adaptive discrete contour evolution method is also proposed to extract the salient feature points of a shape contour for compact representation. Shape matching is performed using the dynamic programming algorithm. The proposed method is invariant to rotation, scale variation, intra-class variation, articulated deformation and partial occlusion. Our method is robust to noise as well. To validate the invariance and robustness of our proposed method, we perform experiments on multiple benchmark datasets, including MPEG-7, Kimia and articulated shape datasets. The competitive results indicate the effectiveness of our proposed method for shape matching and retrieval.

© 2016 Elsevier Inc. All rights reserved.

## 1. Introduction

The shape of an object contour is a compact clue for object recognition. Shape matching is an essential problem in computer vision as it has been widely studied for various applications, e.g., character recognition [1], biomedical image analysis [2], hand gesture recognition [3], robot navigation [4], human gait recognition [5], etc. To capture the salient shape feature for matching, an effective descriptor is necessary. In recent years, there has been a vast set of literatures in shape matching and retrieval algorithms, which report promising results [6–10]. The geometric transformations (translation, rotation, scaling, etc.) have been handled by different methods in the literature. However, it is still a challenging problem to match shapes with intra-class variations and nonlinear deformations (noise, articulation and occlusion). The intra-class variation here specifically indicates the varying shapes of the

instances from the same class as shown in Fig. 1, rather than the nonlinear deformations. The shapes with intra-class variations can be easily classified to the same class by humans, while they are very different in shape matching. The nonlinear deformations are more challenging especially when multiple deformations are combined simultaneously with intra-class variations and geometric transformations. Therefore, it is critical to extract a shape descriptor which is representative, discriminative and robust for shape matching and retrieval. The desired descriptor should not only tolerate geometric transformations, nonlinear deformations and intra-class variations, but also efficient to discriminate shapes from different classes.

Most of the traditional descriptors use local or global information of shapes, which cannot solve the problems of shape deformations or intra-class variations with sufficient expertise. The local descriptors [9,11–13] are capable of representing the local shape features well, but they do not consider the global shape structure. On the other hand, the global descriptors [8,14–16] are robust to local noise and deformations, but they fail to capture the detailed local shape features and cannot deal with occlusion. To satisfy all these contrasting requirements together, it is desirable to combine the advantages of both the global descriptors and local

<sup>☆</sup> This paper has been recommended for acceptance by P. Thomas Fletcher.

\* Corresponding author.

E-mail addresses: [jyyang@suda.edu.cn](mailto:jyyang@suda.edu.cn) (J. Yang), [ihxwang@cqu.edu.cn](mailto:ihxwang@cqu.edu.cn) (H. Wang), [jsyuan@ntu.edu.sg](mailto:jsyuan@ntu.edu.sg) (J. Yuan), [meyfli@cityu.edu.hk](mailto:meyfli@cityu.edu.hk) (Y. Li), [manchestsc@gmail.com](mailto:manchestsc@gmail.com) (J. Liu).

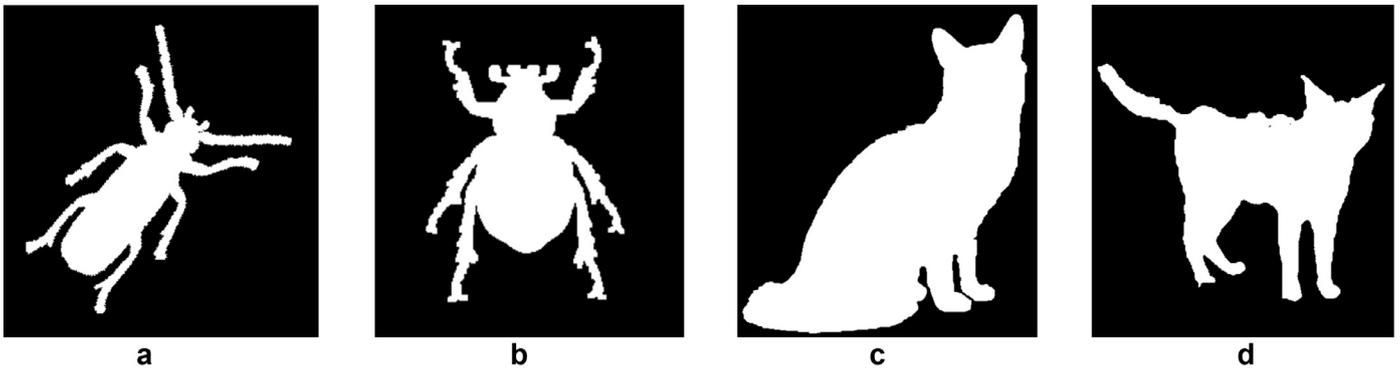


Fig. 1. Two pairs of same-class shapes with large intra-class variations. (a) and (b) are shapes from the “beetle” class in different types. (c) and (d) belong to the “cat” class with different poses.

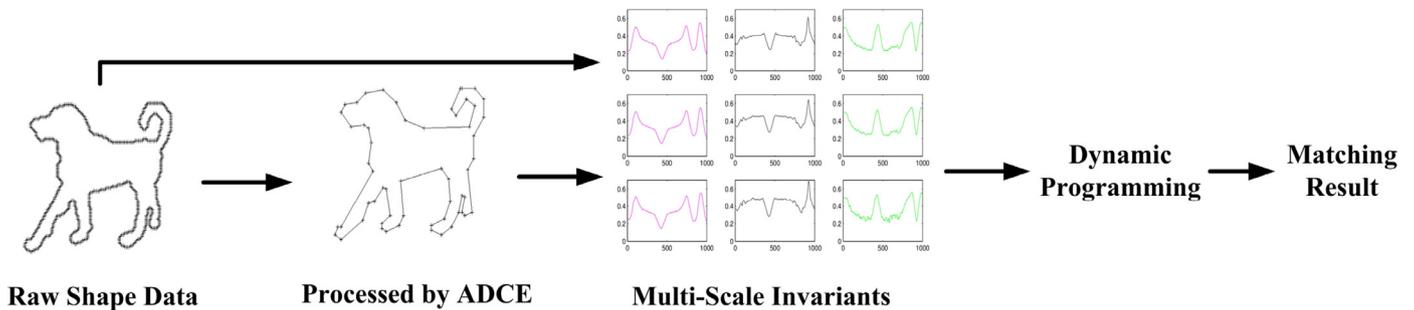


Fig. 2. The pipeline of our method. The raw shape contour is processed by ADCE first, and then the raw contour and the processed contour are used together to obtain the multi-scale description. Finally, pairwise similarity is calculated via the dynamic programming algorithm.

descriptors within a single framework for an exhaustive and discriminative representation scheme. Moreover, the descriptor should be invariant for geometric transformations. A number of recent works [10,17,18] combine the global and local shape features and achieve good shape matching performances on the popular benchmark dataset, MPEG-7 dataset [19]. Besides a series of distance learning based methods [20–22] report the state-of-the-art results. However, these methods still rely on the effective descriptor to achieve a good performance.

In this paper, we propose an invariant multi-scale descriptor for shape representation, matching and retrieval. The descriptor is defined with three types of invariants in multiple scales. Inspired by the characteristics of both global descriptors and local descriptors, we use smaller scales to capture shape details and larger scales to represent semi-global features, e.g., limbs, to obtain “rich” shape information. To make the descriptor discriminative, we use three types of invariants to capture salient shape features from different aspects, including: area, arc length and local central distance. These three shape features are combined to characterize shapes from different classes. The three types of invariants are normalized parameters between 0 and 1 to adaptively capture the inconsistent variations within one shape, e.g., only part of a shape is scaled. The dimension of the extracted feature descriptor for a contour point is between 9 and 15, which is much less than that of most of the previous descriptors. To extract the salient points of shape contour and improve the representativeness of shapes, we modify the discrete contour evolution (DCE) [23] algorithm in an adaptive way to preprocess a closed shape contour. We show that the contours processed by the adaptive discrete contour evolution (ADCE) is more compact. The similarity between shapes is calculated in term of the pairwise distance, and the dynamic programming (DP) [24] algorithm is employed to find the best correspondence between shapes. The pipeline of our method is shown in Fig. 2.

The invariance and robustness of the proposed descriptor is evaluated through extensive experiments. The proposed method

is invariant to geometric transformations and nonlinear deformations, especially to articulated variations and partial occlusion. The experimental results validate the invariance of our method to these variations. Our method is also verified to be capable of capturing the common shape features for shapes with large intra-class variations. We also validate the robustness of our method to noise. The effectiveness in shape matching is evaluated in the experiments of shape retrieval on several benchmark datasets, including: MPEG-7 dataset [19], Articulated dataset [8], Kimia’s 99 dataset and Kimia’s 216 dataset [25]. The comparable results with state-of-the-art demonstrate that the proposed method is effective for shape matching and retrieval. The distance learning method is also applied on our method to further improve the retrieval rate, and our method outperforms other distance learning based methods.

The remainder of this paper is organized as follows. We begin by reviewing the relevant works in the next section. The proposed invariant multi-scale descriptor is presented in Section 3. Section 4 describes the ADCE algorithm for shape preprocessing. The DP based shape matching is detailed in Section 5. Section 6 gives the experimental evaluation of the invariance and robustness of our method as well as the performances of shape retrieval on the benchmark datasets. This paper is concluded in Section 7.

## 2. Related work

There is a rich literature in shape matching that aims to improve shape retrieval rate and object recognition accuracy in two scenarios. One is to design representative and discriminative descriptors to capture the intra-class similarity as much as possible, while maintaining a larger inter-class distance. The other is to learn an optimized similarity measurement utilizing the pre-knowledge of the pairwise distances of a dataset based on the existing descriptors, called distance learning methods. In this

section, we mainly review the methods with different descriptions of shapes which are most related to our work.

There are two main categories of shape descriptors: global feature descriptors and local shape descriptors. Global descriptors capture the spatial distribution of shape and the relationship among contour points. One type of global descriptors represents a shape with all the contour points (contour-based descriptor). The most typical contour-based global descriptor is the shape context (SC) [14] descriptor proposed by Belongie et al. The SC descriptor uses the spatial relationship between one contour point to all other contour points to represent each contour point. For each point, the spatial distribution of all other points is summarized as a 60 dimensional histogram. After that, Ling and Jacobs [8] proposed a inner-distance shape context (IDSC) descriptor that modified the Euclidean distance in SC with inner distance inside the shape, which is insensitive to articulated variations. The inner distance is the shortest path between two related contour points within a shape silhouette. The triangle area representation (TAR) [7,15] is another global descriptor which represents each contour point with the signed areas of triangles formed by the boundary points at different scales. The curvature of corresponding contour points is measured by the triangle area, and the convexity/concavity of the contour is reflected by the sign of the area value. Grigorescu and Petkov [26] propose a shape descriptor based on distance sets from each contour point to all other contour points. The common idea of the contour-based global shape descriptors is exploring the one-to-all spatial relationship to represent each contour point. Another type of global shape descriptors use all the pixel information within a shape region to derive shape description (region-based descriptor). Souza and Marana [16] propose a shape descriptor via Hough transformation and its neighborhood based modification. Zhang and Lu [27] propose the Fourier descriptor to transform the polar raster sampled image into the spectral domain by 2D Fourier transformation. The Fourier descriptor captures shape features at multiple resolutions in both the two dimensions of the polar coordinates. Srestasathiern and Yilmaz [28] use the conic basis to take the conic-section coefficients for shape representation. The extracted projective parameters are invariant to projective transformation. Direkoglu and Nixon [6] design a multi-scale 2D Fourier descriptor based on the assumption that the shape parts far from the center of shape contribute more than the central part. Then, the exterior parts of objects are emphasized in this representation, while the central part is less weighted. In the work of Kim and Kim [29], a set of Zernike moments are used to represent shapes with regions, which is geometric invariant and efficient in computing. The global descriptors represent a shape as a whole, so that the distribution of the shape or contour points is captured, while the local shape features are lost.

The local shape descriptors capture the local shape details of the shape contour. The differential invariants based shape descriptors [11,12] are sensitive to local noise, which require necessary smoothing on shape data. However, it is unavoidable to accept the unfortunate side effect of smoothing that meaningful information will be lost as well. To solve this problem, Siddharth et al. [13] propose the distance and area integral invariants to represent local shape features, where the area integral invariant is an approximation of the curvature of shape contour. The integral invariants are robust to noise and invariant for geometric transformations, however, only the curvature is far from enough to represent complex shapes. The curvature scale space (CSS) [30] method uses the locations of curvature zero crossing points to represent shapes. But the scale is hard to determine and the convex shapes have no curvature zero crossing points. The Riemannian optimization has also been employed in Elastic Shape Analysis [31,32], which solved the shape matching and comparison jointly. The Biswas method [9] used a variety of simple invariants to represent shape features

for shape indexing and retrieval. Daliri and Torre [33] transformed shape contour into a string of symbols and matched shapes via kernel edit distance. As the symbols represent shape locally, this method is invariant to similarity transforms and handles partial occlusion. The local shape descriptors extract the local shape features well, but the spatial structure and distribution of shape are ignored, as well as the shape contour context.

To obtain a “rich” representation of shape, some recent methods make use of both local and global features to derive a hierarchical descriptor. McNeill and Vijayakumar [17] proposed a hierarchical procrustes matching (HPM) method for shape matching based on shape contour segmentation. The hierarchical representation avoids the problems associated with pure global or local descriptors. Xu et al. [10] propose the contour flexibility method to extract both global and local features that depicts the deformable potential at contour points. Felzenszwalb and Schwartz [18] proposed the shape tree method which segments a curve into two halves in the middle, and the two halves are further segmented into respective halves. Raftopoulos and Kollias [34] proposed a method based on global-local transformation to represent shape curvature with view area representation, which is robust to noise. Bai et al. [35] proposed the shape vocabulary representation with the idea of bag of words (BoW), where the shape contours are segmented into fragments and represented as words of shape contours in different scales.

In recent years, a series of distance learning based methods [20–22] have been reported promising results of shape retrieval on benchmark datasets, e.g., the MPEG-7 dataset [19]. These methods use distance learning algorithms on the previous descriptors to improve the retrieval rates, e.g. the label propagation (LP) [20,21] and the locally constrained diffusion process (LCDP) [22]. Here, the previous descriptors, e.g., SC [14] and IDSC [8], are used as the basic in these methods to calculate the shape similarity, and then the rank of retrieval results are optimized by distance learning. It is undeniable that the distance learning algorithms increase the retrieval rates significantly based on the descriptors. However, an effective shape descriptor still makes an important role in the shape retrieval tasks. In this work, we aim to propose a representative and discriminative shape representation to obtain better retrieval rates with and without distance learning.

In this paper, we use not only local shape features, but also semi-global shape structure to make a “rich” representation of shape. Different from the previous descriptors, the multi-scale invariants are designed to combine both the advantages of global and local descriptors while eliminate the deficiency of them. For local shape features, three types of invariants are proposed with the idea of integral, which are robust to local noise. For global information, the salient features of shape parts are included in the invariants, which is robust to intra-class variations and partial occlusion. Moreover, with the proposed idea of major zone and major segment, our method is insensitive to articulated variations. The proposed ADCE method extracts the compact shape contour with meaningful points, which intensify the capability of the proposed descriptor for geometric transformations, nonlinear deformations and intra-class variations.

### 3. Multi-scale shape descriptor

#### 3.1. Motivation: invariance, discrimination and robustness

For a shape matching and retrieval engine, a good performance is significantly relevant to three properties of shape representation: *invariance*, *discrimination* and *robustness*. In the shape matching, both small intra-class distances and large inter-class distances are expected, which rely on the invariance and the discrimination of descriptor, respectively. The robustness is also necessary

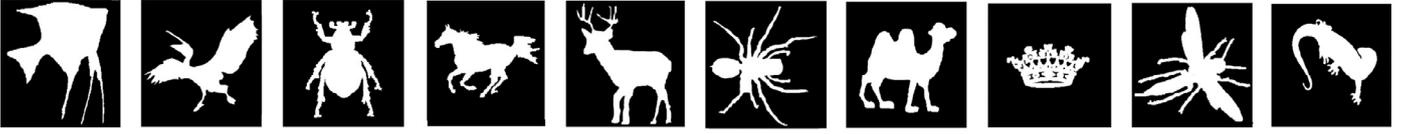


Fig. 3. Complex shapes.

for both large scale disturbance, e.g., partial occlusion, and small scale disturbance, e.g., high-frequency noise, which are two critical roadblocks for a good retrieval result. However, it is very hard to well satisfy these properties simultaneously, especially for complex shapes. As shown in Fig. 3, both the global structures and the local details of the shapes are complex.

To represent a shape concerning these three properties, we consider two aspects: (1) various forms of shape features representing the salient shape information in the identical scale and (2) different scales of shape features capturing both local and global shape information. It is intuitive that a descriptor with multiple shape features in various forms can capture more shape information than a single feature, therefore, the merged multiple features is more discriminative than a single feature. For complex shapes, besides the detailed local shape feature, the global structure, e.g., the body part information of a shape, should be considered as well. That is the reason for which the part based shape decomposition methods [3,35,36] are more popular. In this sense, the invariance of a descriptor depends on the invariants in different scales. The small scale invariants can represent the invariant features of detailed local shape, while the large scale invariants capture the invariant features of body parts. The robustness of a descriptor also benefits from the multi-scale features, e.g., the global feature is robust to local noise while the local features is robust to occlusion, and an appropriate utilization of both the advantages is preferred.

In view of the above analysis, building a multi-scale shape descriptor with various invariants is needed for a good performance in shape matching and retrieval. Basically, the capability of a descriptor depends largely on feature selection and descriptive structure. In the following, we propose a novel invariant multi-scale shape descriptor using three different types of invariants and each type is used in different scales to represent local and semi-global shape features, which can be used for shape matching and retrieval.

### 3.2. Definition of The Invariant Multi-Scale Descriptor (IMD)

Since the raw data of a shape is a closed contour which is composed of a series of sample points, the descriptor is used to represent the shape features of each sample point on the contour. Denote  $S = \{p(i) | i \in [1, n]\}$  the closed planar shape contour with a sequence of sample points  $p(i)$ , where  $n$  is the length of the contour. The sample point  $p(i)$  is parameterized as  $p(i) = \{u(i), v(i)\}$ , where  $u(i)$  and  $v(i)$  are coordinates in the image. The IMD descriptor  $M$  is defined as follows:

$$M = \{s_k(i), l_k(i), c_k(i) | k \in [1, m], i \in [1, n]\}, \quad (1)$$

where  $s_k$ ,  $l_k$  and  $c_k$  are three invariants: normalized area  $s$ , arc length  $l$  and central distance  $c$  in the scale  $k$  (as shown in Fig. 4).  $k$  is the scale label, and  $m$  is the total scale number. These invariants are defined separately in the following.

**Definition 1.** Consider a circle  $C_k(i)$  with radius  $r_k$  centered at  $p(i)$ , one or more zones may be occupied by the shape (see Fig. 5 (b), zone A and zone B). Zone A is so called the major zone to  $p(i)$  because  $p(i)$  is on the edge of this zone (direct connection), while the other zones have no connection to  $p(i)$  via the shape within the circle. For convenience, here we indicate with  $Z_k(i)$  the major

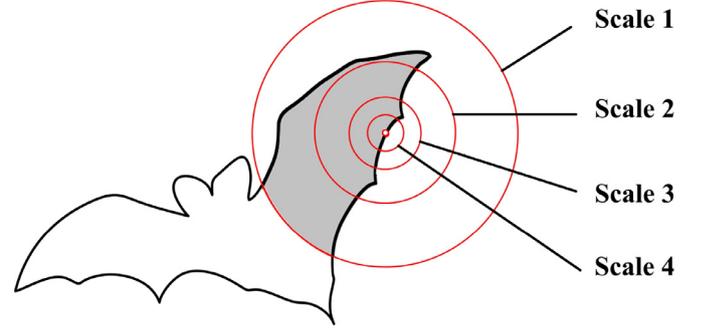


Fig. 4. The four circles indicate three scales 1–4. The grey area in each circle represents the area invariant in respective scale. The contour segment in each circle represents the arc length invariant. The central distance is from the weight-center of the grey area to the circle center.

zone of  $p(i)$  in scale  $k$ . Denote the area of  $Z_k(i)$  (Zone A in Fig. 5, (b)) by  $s_k^*(i)$ , the normalized area  $s_k(i)$  is defined as the ratio of  $s_k^*(i)$  to the area of  $C_k(i)$ , as follows:

$$s_k^*(i) = \int_{C_k(i)} B(Z_k(i), x) dx, \quad (2)$$

$$s_k(i) = \frac{s_k^*(i)}{(\pi r_k^2)}, \quad (3)$$

where  $B(Z_k(i), x) : \mathbb{R}^2 \times \mathbb{R}^2 \mapsto \{0, 1\}$  is define to be an indicator function on the interior of  $Z_k(i)$ ,

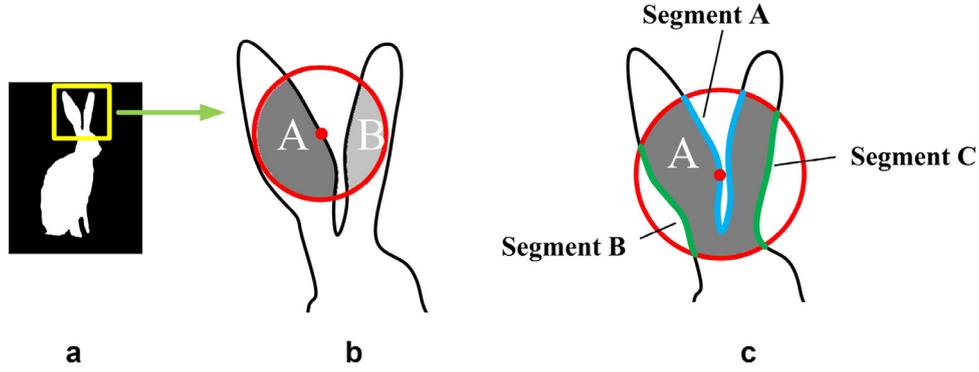
$$B(Z_k(i), x) = \begin{cases} 1, & \text{if } x \text{ is inside } Z_k(i), \\ 0, & \text{if } x \text{ is outside } Z_k(i). \end{cases} \quad (4)$$

Since  $s_k^*(i)$  is less than the area of  $C_k(i)$ , the ratio  $s_k(i)$  of them ranges from 0 to 1. In our method, only the major zone to the center  $p(i)$  is considered, while other zones without circle-inside connection to  $p(i)$  are not calculated, e.g., Zone B in Fig. 5. In the following definitions of  $l_k$  and  $c_k$ , only the major zone and major segment are considered as well. The reason and benefit of this selection is explained in Section 3.4.

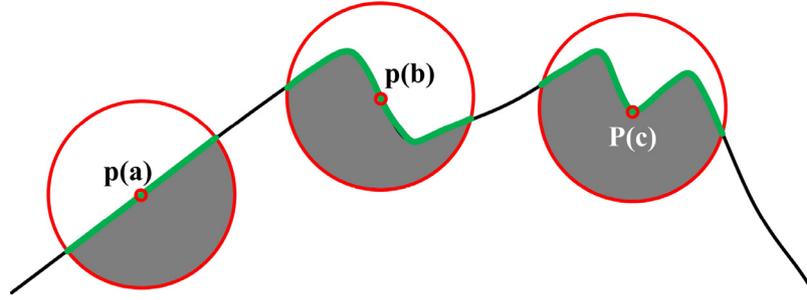
**Definition 2.** In the same circle  $C_k(i)$ , there are one or more contour segments of the shape contour that are segmented by  $C_k(i)$  (see Fig. 5, segment A, B and C in (c)). Similar to the major zone in Definition 1, we only consider the major segment A. Denote the arc length of the major segment (segment A in this case) by  $l_k^*(i)$ . The normalized arc length  $l_k(i)$  is defined as the ratio of  $l_k^*(i)$  to the circumference of  $C_k(i)$  and computed as follows:

$$l_k(i) = \frac{l_k^*(i)}{(2\pi r_k)}. \quad (5)$$

The length  $l_k^*(i)$  is usually shorter than the circumference of  $C_k(i)$  in most of the frequently seen object shapes and benchmark datasets, e.g., the MPEG-7 [19] dataset used in our experiments. Therefore, the normalized  $l_k(i)$  ranges from 0 to 1. If a bigger  $l_k^*(i)$  than the circumference of  $C_k(i)$  exists in a special dataset, the range of  $l_k(i)$  can be restricted to  $[0, 1]$  by dividing a bigger normalization parameter.



**Fig. 5.** Invariants in the articulated parts. The grey zones A and B on the rabbit ears in (b) are not directly connected within the circle, while the grey zone A in (c) is a major zone. However, the two segments B and C in (c) are still not directly connected to  $p(i)$  thus only the blue segment A passing the circle center is used to calculate the arc length invariant. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 6.** Three circles at different positions of a shape contour segment.  $p(a)$  and  $p(b)$  have the same area invariant, but the arc length invariants of them are different. The arc length of  $p(b)$  and  $p(c)$  are the same, while the areas and central distances of them are different.

**Definition 3.** The central distance  $c_k^*(i)$  is defined as the distance between  $p(i)$  and  $w_k(i)$ :

$$c_k^*(i) = ||p(i) - w_k(i)||, \quad (6)$$

where

$$w_k(i) = \frac{\int_{C_k(i)} B(Z_k(i), x) dx}{\int_{C_k(i)} B(Z_k(i), x) dx}, \quad (7)$$

is the weight-center of the major zone in  $C_k(i)$ , which is calculated by averaging the coordinates of all the pixels in the zone. Then,  $c_k(i)$  is calculated by normalizing with  $r_k$ :

$$c_k(i) = \frac{c_k^*(i)}{r_k} = \frac{||p(i) - w_k(i)||}{r_k}. \quad (8)$$

Since the weight-center must be inside the simply connected zone,  $c_k^*(i)$  is less than  $r_k$ , and  $c_k(i)$  ranges from 0 to 1. As the methods of calculating area  $s_k^*(i)$ , arc length  $l_k^*(i)$  and weight-center  $w_k(i)$  are straight forward, and there are efficient implementations in the tool box of Matlab and OpenCV, we do not specify the formula here for brevity. The functions of these invariants of example shapes are shown in Fig. 10.

For each point  $p(i)$ , there are  $3 \times m$  invariants in the descriptor to represent the shape feature in different aspects (area, arc length and central distance) and different scales (semi-globally and locally). The combination of these invariants includes an exhaustive shape information at  $p(i)$ . The area can only represent the concavity and convexity of a simple shape, while the arc length can only represent the complexity of the local shape that a long arc length indicates a complex shape. However, if these two invariants are combined, the representation is improved significantly in terms of its discriminability, as shown in Fig. 6. Similarly, merging the invariants in different scales increases the capability of the descriptor as well.

### 3.3. The setting of $r$ and $k$

In representing a shape with IMD, the radius  $r_k$  of the circle  $C_k$  needs to be set first. In our method, the radius  $r_k$  in different scales is set with respect to an initial radius  $R$ :

$$r_k = \frac{R}{2^k}, \quad (9)$$

where  $r_k$  is half of  $r_{k-1}$  in the prior scale. That is,  $r_1 = R/2$  is the radius of  $C_1$  in the highest scale ( $k = 1$ ), and the circles in the following scales ( $k = 2, 3, \dots, m$ ) have their radii decreased half by half.

The setting of initial  $R$  is important for representing both the local and semi-global information. We use the invariants in the first scale, i.e.,  $s_1$ ,  $l_1$  and  $c_1$ , to represent the semi-global information, while the invariants in the last scale, i.e.,  $s_m$ ,  $l_m$  and  $c_m$ , to represent the local information. Consider that the points far from the shape center need larger circles to cover and represent shape parts, e.g., in Fig. 5, the points at the top of the rabbit ears need a larger circle than the points at the middle of the ears to cover the part of ears. Instead of using a fixed radius, we set the initial  $R(i)$  as half of the distance from the contour point  $p(i)$  to the shape center  $p_{center}$ :

$$R = \frac{||p(i) - p_{center}||}{2}, \quad (10)$$

where

$$p_{center} = \frac{\int_{shape} x dx}{\int_{shape} dx}, \quad (11)$$

is the weight-center of the whole shape.  $p_{center}$  is calculated by averaging the coordinates of all the pixels in the shape. The circle  $C_1(i)$  in the first scale covers the meaningful body part near  $p(i)$ , and the invariants in this scale capture the semi-global invariant shape features in this part. In contrast, the circle in the last scale

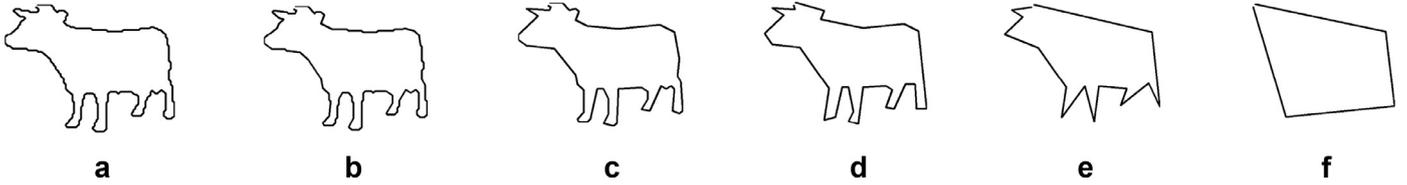


Fig. 7. The DCE revolution results under different evolution degrees. The evolution degree gets deeper from left to right.

is very small, which covers so small an area that the invariants can only represent the detailed yet simple shape features, e.g., the curvature of the local shape. Therefore, the invariants in different scales represent full information of the shape. In Fig. 4, an example shape with circles from Scale 1 to Scale 4 are shown. As seen that  $C_1$  covers a visual part of the cup, while the  $C_4$  covers only a local shape pattern.

The scale number  $m$  influences the capability of the descriptor as well, which needs to be carefully selected. A big scale number is not always better than a small one. For simple shapes, e.g., a regular square, the normalized area at the corner points are always 0.25 in different scales. Thus a bigger scale number does not increase the discriminative power of the descriptor, and may not benefit the shape matching performance. Moreover, the computational cost is also increased by a bigger scale number. Therefore, the scale number  $m$  should be set according to the complexity of shapes. In the proposed descriptor, the bigger scales are always used regardless of the complexity, and the scale number  $m$  determines only the smallest scale. Complex shapes include more salient local features which need more scales to represent. Simple shapes with few local features, however, need less scales to represent the shape. Therefore, we set the scale number  $m$  according to the following convergence condition: if the average difference of the invariants between two neighbor scales  $m$  and  $m + 1$  is less than a threshold, e.g.,  $1e-2$ , the invariants in scale  $m + 1$  is unnecessary.

### 3.4. Properties of IMD and differences from the integral invariants

Although the previous area integral invariant [13] and the proposed area invariant  $s_k$  are both normalized area inside the local circle centered at  $p(i)$ , our method is significantly different from the integral invariants in the following aspects:

- The definitions are different. Our method only uses the simply connected zone  $Z$ , while the integral invariants use all the zones within the circle. Our method is more robust for articulated variation: if a circle covers two articulated parts, additional area will be calculated in the integral invariants, which increase the intra-class distance. This problem is demonstrated by the rabbit ears in Fig. 5. As seen in the figure we can find that the circle centered upper on the ears (Fig. 5, (b)) covers two zones A and B, where only the A zone is calculated regardless the distance between ears. However, if the circle center is near the roots of the ears as in Fig. 5, (c), the whole grey area is calculated. Therefore, our method is more invariant to articulated deformation. This problem also exists in another example of the horse legs shown in Fig. 12. Using the simply connected curve, the calculation of the normalized arc length  $l_k$  benefits for this problem as well.
- Siddharth et al. [13] indicates that the integral invariants has an approximate relation with the curvature at  $p(i)$ , however, this property does not always hold for complex shapes in big scales with serious variation of curvature, e.g., the circle zone in Fig. 5, (c). From Fig. 5, (c), the shape within the circle cannot be represented by simply the curvature at  $p(i)$ . Hence, the approximate calculation of integral invariants is not reliable in the case in

(b). In contrast, the straight forward calculation of area in our method is more intuitive.

- The integral invariants also use a multi-scale representation with different radius, but the radius increases by arithmetic progression, i.e.,  $r/R = \{0.1, 0.2, \dots, 0.5\}$ . On the contrary, in our method, the radius decreases by geometric progression, which can converge fast. Hence, our method represents more shape information by less scales. Furthermore, our method is not sensitive to the initial  $R$ , since the invariants are normalized parameters, and the little variation of the radius in the same scale will not significantly change the descriptor.

### 4. Contour evolution for salient feature points

Most of the previous methods use all the sample points on the shape contour, which have three limitations: (1) there are many redundant points that do not capture salient shape features, but they have the same weight as the salient feature points, which makes the descriptor not representative for the shape features; (2) the redundant points are sensitive to noise so that the variations of them increase the alignment error and affect shape matching; (3) the redundant points increase the computational cost in both representation and matching.

To extract shape features and weed out redundant points, Latecki and Lakamper [23] proposed the discrete contour evolution (DCE) method. The DCE method deletes the redundant points while preserving the salient feature points. The main idea of DCE is that, in each time evolution, the point with minimum contribution to target identification is deleted. The contribution of a point  $p(i)$  is defined by a relevant measure function  $K$ :

$$K(i) = \frac{B(i) \cdot b(i, i-1) \cdot b(i, i+1)}{(b(i, i-1) + b(i, i+1))}, \quad (12)$$

where  $b(i, i-1)$  is the segment length between  $p(i)$  and  $p(i-1)$ ,  $b(i, i+1)$  is the segment length between  $p(i)$  and  $p(i+1)$ , and  $B(i)$  is the angle of these two segments. The length  $b$  is normalized with respect to the contour perimeter. A higher  $K$  indicates a bigger contribution to the shape. However, the deficiency of this method is that it cannot evolve adaptively due to the local calculation of  $K$ , which conducts insufficient evolution with border noises (Fig. 7, (b)), or over evolution into a convex polygon (Fig. 7, (f)).

In this work, we propose an adaptive discrete contour evolution (ADCE) method to overcome the above limitations by an adaptive ending of evolution. The ideal result is that all the visual parts correspond exactly to all significant visual parts of the original shape contour after evolution. Therefore, the evolved contour should maintain a certain similarity with the original contour. We define an area-based adaptive ending function  $F(t)$ , called evolution area difference function (EADF), as follows:

$$F(t) = \frac{n_0 \sum_{i=1}^t |S_i - S_{i-1}|}{S_0}, \quad (13)$$

where  $S_0$  is the area of the original contour,  $S_t$  is the area after  $t$  times of evolution and  $n_0$  is the total point number of the original contour. Note that we use the absolute difference between  $S_t$

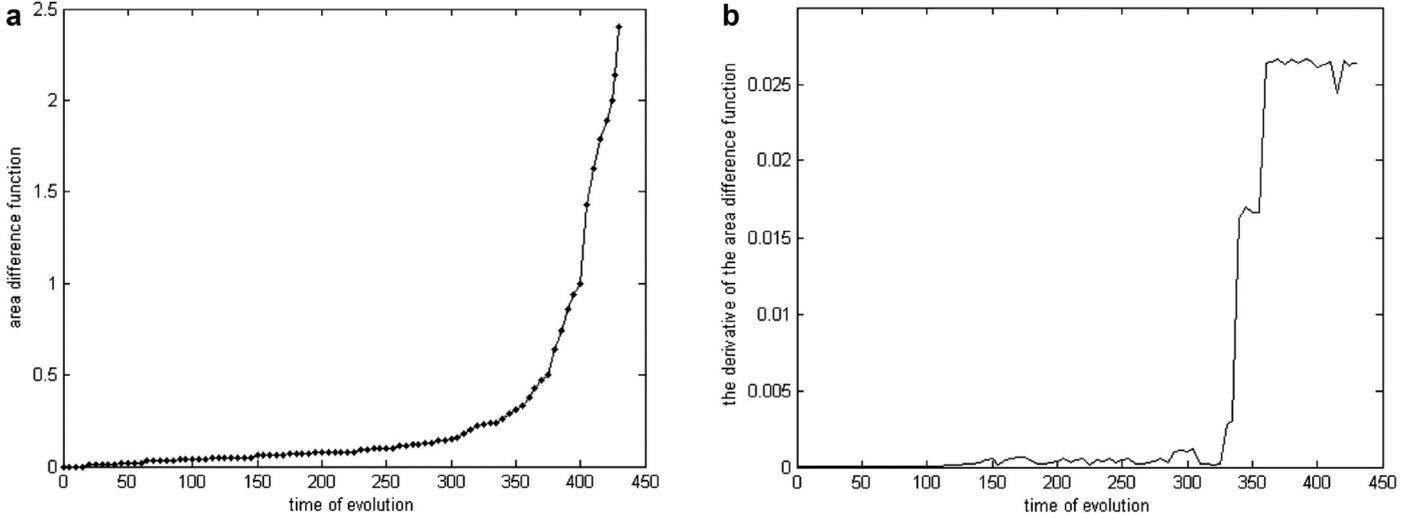


Fig. 8. (a) is the EADF function  $F(t)$  of a sample shape and (b) is the derivative function of  $F(t)$ . The threshold of  $F(t)$  is chosen at the point where the derivative function significantly increases (the inflection point at 325 for this shape).

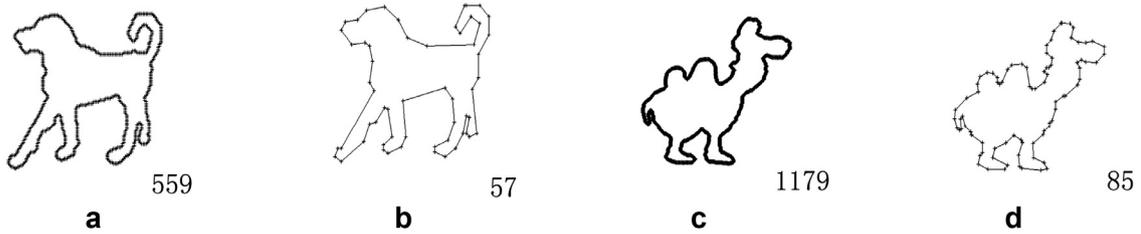


Fig. 9. (b) and (d) are the ADCE evolution results of the original shapes (a) and (c) under the threshold  $F = 0.5$ . The numbers aside the shapes are the point numbers of respective contours.

and  $S_{t-1}$  considering that both the convex and concave evolutions contribute to the total area difference. The contour evolution ends when  $F(t)$  is bigger than a given threshold. The function  $F(t)$  of a sample shape is plotted in Fig. 8, (a). The derivative function of  $F(t)$  is shown in Fig. 8, (b). The threshold for function  $F(t)$  is chosen at the point where its derivative significantly increases. This point is the inflection point at 325 for this shape in Fig. 8, (b). If the evolution time is less than 325, there would be still redundant points. Otherwise, the salient feature points are removed and the shape information is lost. The evolution results of two sample shapes are shown in Fig. 9 with the number of points of both the original and evolved contours. We can see that the evolved shapes preserve the salient shape features of the original shapes with less redundant points.

We should note that, the proposed ADCE step is only used to find the representative feature points, while the IMD of the evolved contour is still calculated from the original image at these feature points. That is, the length of the contour sequence is decreased by the ADCE, while the IMD values of the remained contour points are the same as that in the original shape. This is to preserve the original shape features of the salient feature points.

### 5. Shape matching by dynamic programming

As the IMD of the evolved shape contour is obtained, we match two shapes by calculating their dissimilarity. An intuitive idea is to compare their invariant functions (the functions of the IMD invariants) directly. However, the contour points of two shapes should be aligned beforehand. In this work, the dynamic programming (DP) [24] is employed to find the best correspondence between two shapes. Given two evolved shapes A and B, describe them with point sequences on their evolved contours:  $A = \{p_1, p_2, \dots, p_{n_A}\}$

and  $B = \{q_1, q_2, \dots, q_{n_B}\}$ . Without loss of generality, assume  $n_A \geq n_B$ . The matching cost of two points  $p_i$  and  $q_j$  is defined as the Euclidean distance of their IMD invariants:

$$d(p_i, q_j) = \sqrt{\sum_{k=1}^m ((s_k^p(i) - s_k^q(j))^2 + (l_k^p(i) - l_k^q(j))^2 + (c_k^p(i) - c_k^q(j))^2)}, \quad (14)$$

and the cost matrix  $D$  represents the set of costs between all pairs of points:

$$D(A, B) = \begin{bmatrix} d(p_1, q_1) & d(p_1, q_2) & \dots & d(p_1, q_{n_B}) \\ d(p_2, q_1) & d(p_2, q_2) & \dots & d(p_2, q_{n_B}) \\ \dots & \dots & d(p_i, q_j) & \dots \\ d(p_{n_A}, q_1) & d(p_{n_A}, q_2) & \dots & d(p_{n_A}, q_{n_B}) \end{bmatrix}. \quad (15)$$

Then, the DP algorithm is used to find the best correspondence  $\pi$  of the two point sequences, where  $p_j$  is mapped to  $q_{\pi(i)}$  if  $\pi(i) \neq 0$ , and  $p_i$  is unmapped otherwise.  $\pi$  should minimize the matching cost function  $f_{A, B}(\pi)$  of shapes:

$$f_{A, B}(\pi) = \sum_{i=1}^{n_A} d(p_i, q_{\pi(i)}). \quad (16)$$

As there is an over fitting problem in the directly optimization, a penalty employed to  $\pi$  is defined as  $d(p_i, q_{\pi(i)}) = \tau$  if  $\pi(i) = 0$ , where  $\tau$  is the penalty factor for the points with no suitable matches. The similarity  $sim(A, B)$  between shape A and B is the minimum value of  $f_{A, B}(\pi)$ :

$$sim(A, B) = \min f_{A, B}(\pi). \quad (17)$$

## 6. Experimental results

In this section, we evaluate the capability of the proposed method in three aspects: (1) demonstrate the invariant properties of the proposed IMD descriptor for articulated deformation, partial occlusion and intra-class variation including rotation and scale variation; (2) evaluate the representative and discriminative power by shape matching and retrieval experiments on benchmark datasets, including the MPEG-7 [19] dataset, Articulated [8] dataset, Kimia's 99 and Kimia's 216 [25] datasets by an extensive comparative study; (3) test the robustness to noise in shape description and retrieval.

### 6.1. Invariant properties of IMD

In this experiment we evaluate the invariant properties of IMD to rotation, scale variation, articulated deformation, intra-class variation, and partial occlusion. Since the rotation and scale variation often happens in most of object detection and recognition applications, we combine these two variations with the intra-class variation together to test our method first. In the test, we select two images with salient intra-class variations from the same class in the MPEG-7 dataset, and calculate their invariant functions as shown in Fig. 10. In the first column of Fig. 10, row1 and row2 show the original images with salient intra-class variations, row3 and row4 are their corresponding rotated images, row5 and row6 are scaled images, and finally row7 and row8 have the rotated and scaled images. Columns 2, 3, and 4 are the IMD functions corresponding to the shapes. From the figures we can find significant correspondences and similarities among the functions of different shapes. We should note that, the plots in the figures are the functions of the invariants of the original shapes with many redundant points. After the revolution by ADCE, the invariants may extract the shape information of the salient feature points more clearly, e.g., the peak values in the functions. The strong correspondence of the function peaks among shapes in the plots indicates the invariance of our method to intra-class variation. Fig. 11 visualizes the non-linear matching maps between shapes with integrated variations, from which we can get an intuitive perception about the properties of our method. The correspondences between the ADCE contour points are optimized by the DP algorithm. In spite of the salient variations, we can still find the corresponding parts of the shapes are well matched, which illustrates intra-class invariance, rotation invariance and scale invariance. This result also verifies that the DP algorithm is reasonable, and the calculated similarities among shapes show a reliable quantitative evaluation for shape matching and retrieval.

The articulated deformation is another challenging variation in shape description. We use shape samples of two shape classes, rabbit and horses, in the MPEG-7 dataset to demonstrate the invariance of our method. As mentioned in Section 3, the articulated deformation invariance of the rabbit ears (in Fig. 5) is indicated in the definition of the invariants using the major zones and segments. Here, we use another example, the horse legs, to evaluate the articulation invariance as shown in Fig. 12. As seen that the standing horse (the top image of the first column in Fig. 12) have two front legs parallel together, and the circle centered on the left leg covers both the two front legs. However, the circle centered at the corresponding point on the running horse (the bottom image of the first column in Fig. 12) covers only one leg. The functions of their invariants in the first scale are shown in Columns 2, 3, and 4 with the labels (black circles on the functions) corresponding to the circle centers on the horse legs. We can find the strong similarities between the corresponding functions at the labels which indicate the articulated deformation invariance of our method. The

right front leg of the standing horse does not affect the calculation of the invariants of the left front leg, although it is inside the circle. The articulated dataset [8] includes shapes with various articulated deformations, which is used to test the robustness of our method in shape matching and retrieval described in Section 6.2. We only show the functions in the first scale, because the first scale is affected the most by the articulated deformation. The higher scales have smaller circles to cover a more local range which cannot be affected by the articulated parts seriously. Thus, if the method is invariant of articulated deformation in the first scale, it is also invariant in the higher scales. The matching result of the two shapes is shown in Fig. 13. To emphasize the matching result of the articulated parts, the mapping of the common parts are not shown in this figure. From the result we can see that the legs of the horses are well matched, which demonstrates the articulated invariance of our method.

Occlusion is unavoidable in the real applications of object recognition by shape matching due to the obstacles in tracking. This experiment is to demonstrate the invariant property of the IMD descriptor on the partially occluded shapes. Sample shape of the camel class from the MPEG-7 dataset is partially occluded in our test, where the IMD invariants of both the occluded and original shapes are shown in Fig. 14. As seen, the camel head is occluded, and the squares in the function plots (Column 2–4) corresponds to the parts blocked by squares in the original and occluded images (Column 1). The blocked part of the occluded camel is a segment of line, and the functions inside the corresponding squares are much simpler than that of the original image. Besides the squares, the rest of the corresponding functions are similar, which indicates that our method is invariant for occlusion. The reason for using only the first scale of the functions in Fig. 14 is the same as that in the experiment of articulated deformation invariance. Fig. 15 shows the matching result of the two shapes. In this figure, the contour points of shapes are well mapped except the occluded part. The matching of the occluded part does not affect the matching of other parts, which verifies the occlusion invariance of the proposed descriptor. Note that, a too heavy occlusion, e.g., the main part of shape is occluded, may result in ambiguities in shape matching and retrieval due to losing too much shape information.

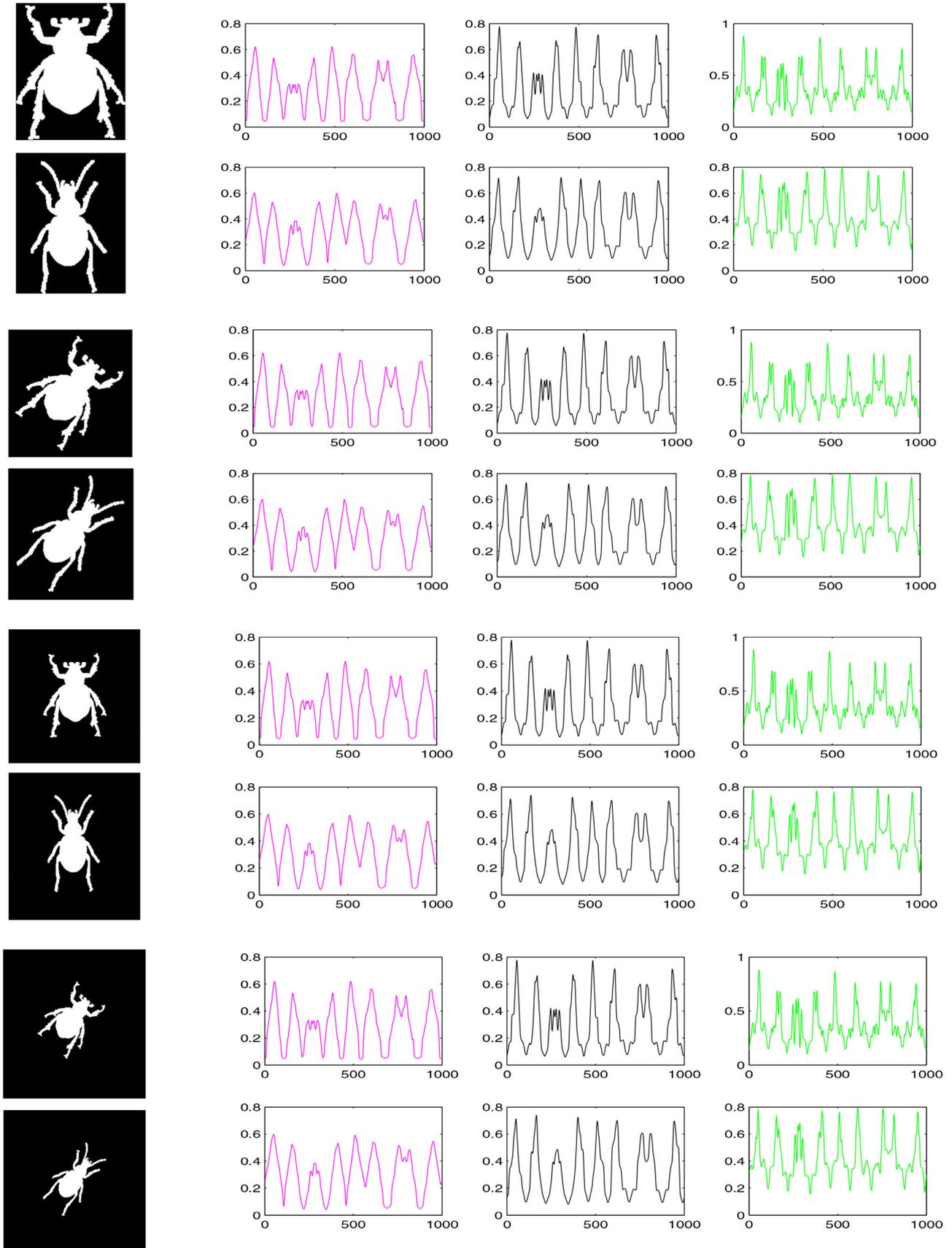
### 6.2. Shape retrieval

#### 6.2.1. Implementation setting

For shape representation, the scale number of the IMD descriptor is set to 4 for MPEG-7 dataset, while 3 for the Articulated dataset and Kimia's 99 dataset, according to the convergence condition in our method. We use the proposed ADCE method to extract the salient feature points, where the threshold  $F$  is 0.4 and the number of the extracted points ranges from 4 to 100. In the following we give the retrieval results and analysis on the benchmark datasets.

#### 6.2.2. MPEG-7 dataset

The MPEG-7 [19] is a standard dataset which is widely used to test the capability of shape matching and retrieval methods. It consists of 1400 binary images divided into 70 shape classes that each class contains 20 shapes. Fig. 16 shows some typical shapes of MPEG-7 with two shapes from each class. From the figure we can see that the shapes have considerable intra-class variations. Generally, the retrieval rate of a method is measured by the so-called bull's eye score. Every shape of the dataset is used as a query to compare with all other shapes retrieved by shape matching. Among the 40 most similar shapes, the number of shapes from the same class of the query shape is reported. The



**Fig. 10.** In the first column, row1 and row2 are the original images with salient variations, row3 and row4 are the corresponding rotated images, row5 and row6 are scaled images, row7 and row8 are rotated and scaled images. Column 2–4 are the invariant functions of  $s$ ,  $l$  and  $c$  respectively, corresponding to the images in the left.

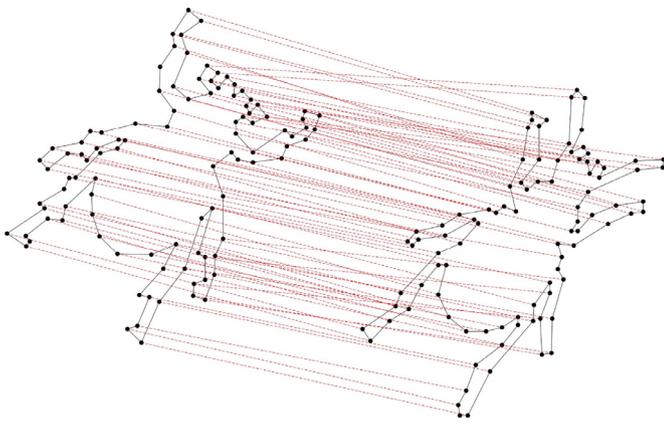


Fig. 11. Nonlinear mapping of shape contour points via DP matching of IMD description.

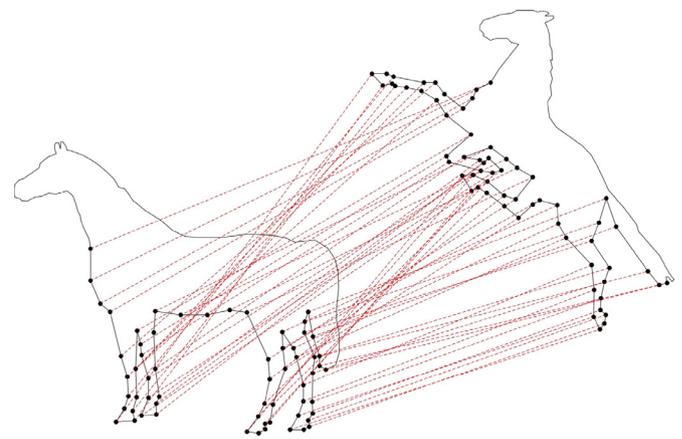


Fig. 13. Nonlinear mapping of articulated shapes.

bull's eye retrieval rate is the ratio of the total number of shapes from the same class to the highest possible number and the best possible rate is 100%. We calculate the retrieval rate of all the shapes in this dataset and divide the sum by 28000 ( $1400 \times 20$ ).

The bull's eye scores of our method compared with other methods are listed in Table 1, where our method has the best score comparing with other methods (without distance learning) in this table. The accuracy of our method is significantly higher than that of the popular IDSC [8] method which is a classic algorithm in the literature. The methods below the IDSC [8] in the table are proposed to improve the shape matching performance and our method shows better result than them. This validates the superior representative and discriminative characteristics of the proposed method for shape retrieval.

The retrieval rates of the 70 shape classes by our method are displayed by a histogram in Fig. 17, which shows that the rates of two classes, Class 29 and Class 32, are much lower than that of other classes. The shapes of these two classes are shown in Fig. 18, which are hard to retrieve for two reasons. (1) there are very few feature points to represent the shapes, e.g., only the points on the circle can represent the meaningful shape feature of Class 32. (2) there are too many interference points, e.g., the points of the chaotic curve inside the circle. This affects the performance of shape matching and increase the intra-class distances.

Table 1

Retrieval rates (Bulls-Eye) of different methods on the MPEG-7 dataset without distance learning.

Method	Retrieval rate (%)
Visual part [23]	76.45
SC [14]	76.51
Distance sets [26]	78.38
Gen. model [37]	80.03
SSC [38]	82.39
AISD [39]	84.26
IDSC [8]	85.40
HPM [17]	86.35
CDTW+SC [40]	86.73
TAR [15]	87.23
Shape tree [18]	87.70
ASC [41]	88.30
Contour flexibility [10]	89.31
Locally affine invariant [42]	89.62
Height functions [43]	89.66
Shape vocabulary [35]	90.41
<b>Our method</b>	<b>91.25</b>

We should note that, it is impossible for any descriptor to extract all the shape features to achieve a 100% retrieval rate. Furthermore, considering two shapes in matching, some differences between them are more relevant for shape similarity, while some differences are irrelevant for shape characterization. The pairwise

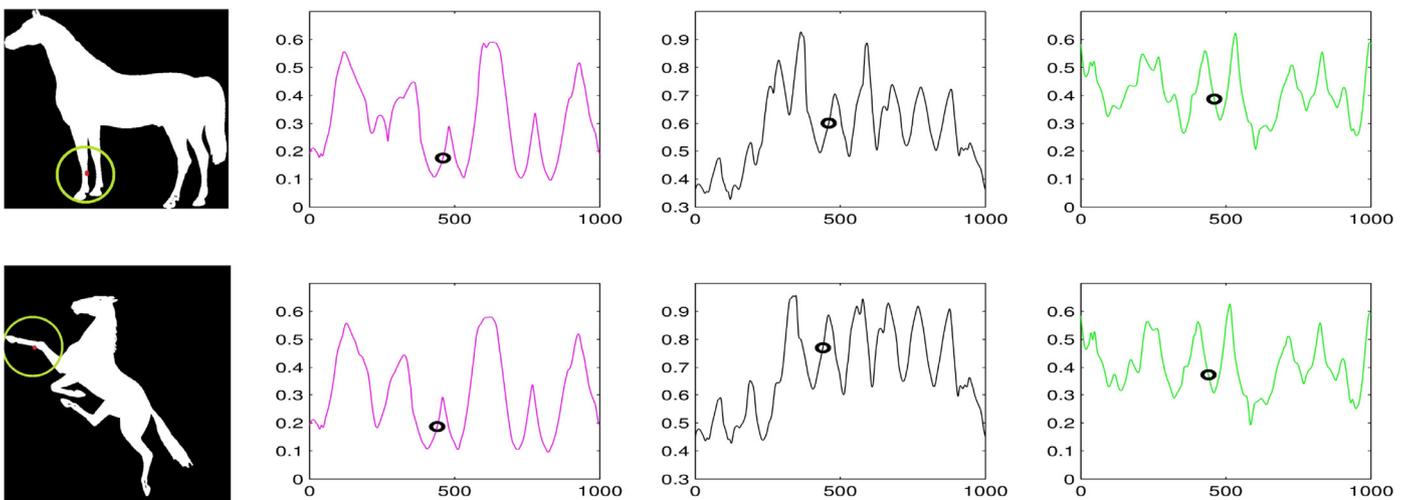


Fig. 12. The invariant functions of articulated shapes. The first column is two shapes with articulated deformation at the horse legs. The 2–4 columns are functions of the invariants  $s$ ,  $l$  and  $c$  in the first scale. The small circles in the invariant functions correspond to the circle center positions on the original shape contours.

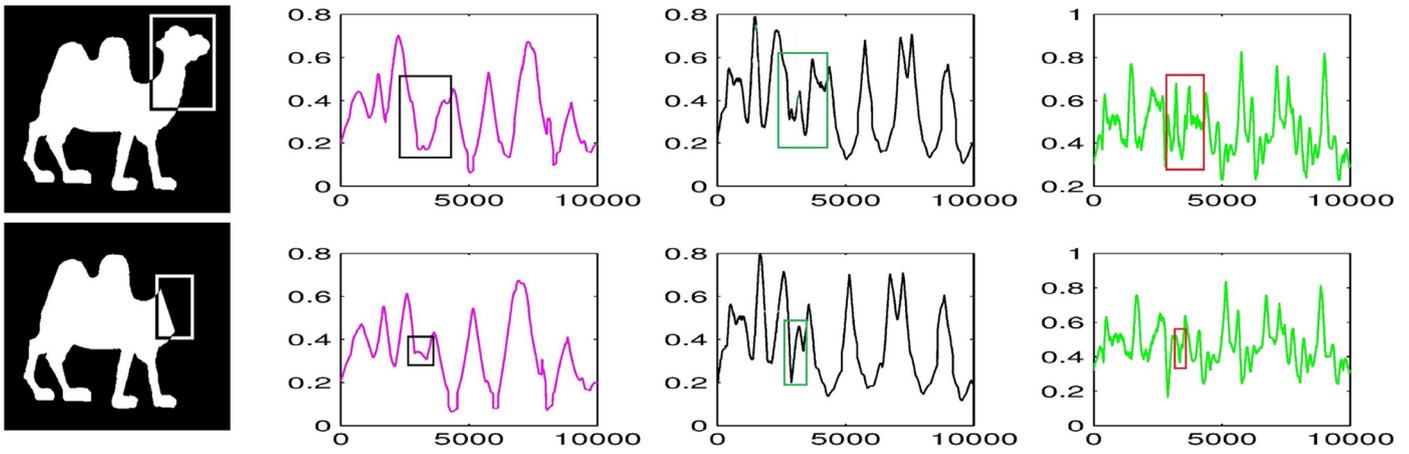


Fig. 14. The invariant functions of occluded shape compared with original shape. The first column are the original shape and an occluded shape. The occluded part is blocked by squares in both images. The 2–4 columns are functions of the invariants  $s$ ,  $l$  and  $c$  in the first scale. The squares in the functions correspond to the occluded part of the image.

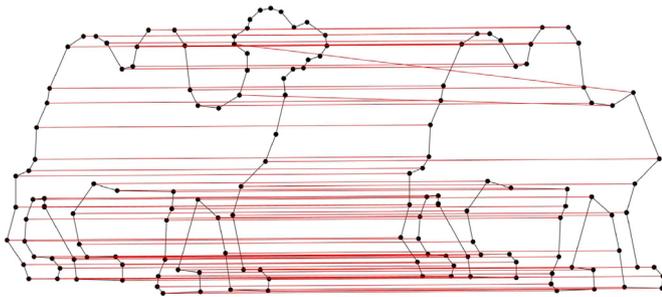


Fig. 15. Nonlinear mapping of occluded shapes.

matching with shape description cannot measure the weight of shape feature because the pre-knowledge of the dataset is not utilized. In this sense, it is necessary to make use of the excellent distance learning method to further improve the retrieval rate.

To evaluate the capability of our method with distance learning method, the proposed MID invariants of the ADCE shape contour are used as the basic descriptor and the LP [21] distance learning method is applied to improve the retrieval rate. The distance learning methods [21,22] are also applied on the IDSC[8] descriptor for comparison. The IDSC descriptor is a popular shape descriptor

Table 2  
Retrieval rates (Bulls-Eye) on MPEG-7 dataset with distance learning.

Method	Retrieval rate (%)
IDSC [8]	85.40
Our Method	91.25
IDSC+LP [20]	91.61
IDSC+LCDP [22]	92.36
<b>Our method+LP</b>	<b>94.51</b>

which is widely used as the basic descriptor to design the distance learning methods for shape matching [20–22]. Hence, this experiment is fair to evaluate the accuracy of our method with distance learning.

The results are listed in Table 2, where the methods above the boundary line are the original methods without distance learning while the methods below the boundary line are methods applied with distance learning. From the results we can see that the retrieval rates of both the IDSC and our method are improved by the distance learning methods, i.e., the LP and the LCDP. With distance learning, our method shows an impressive performance with the dominating retrieval rate of 94.51%, compared to others. The results indicate that improving the capability of descriptor is

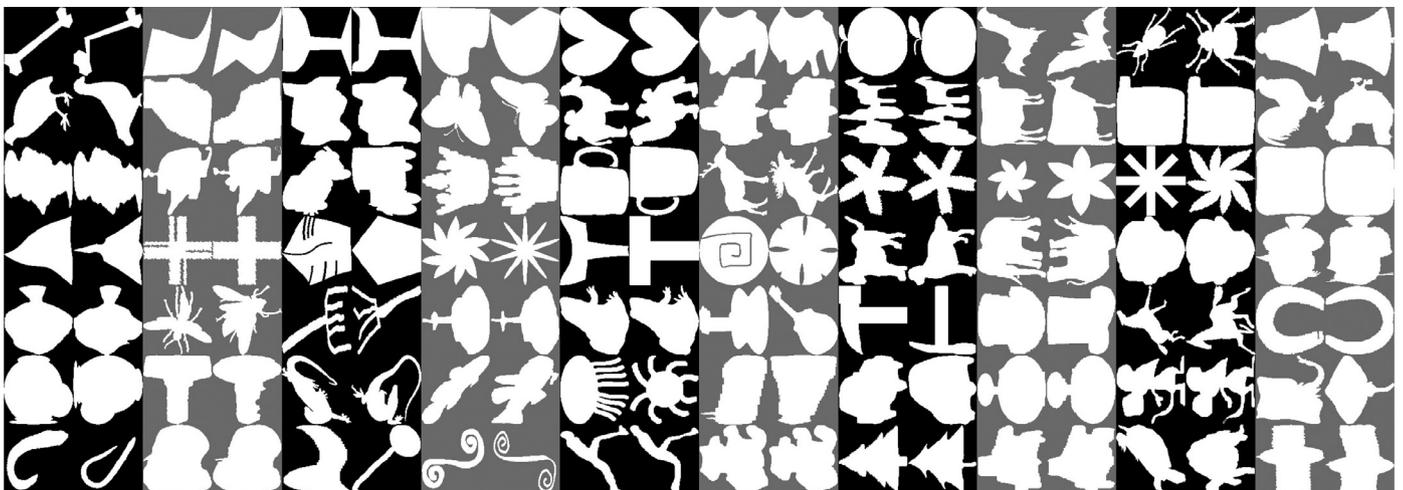


Fig. 16. Typical shapes of MPEG-7 with two shapes from each class.

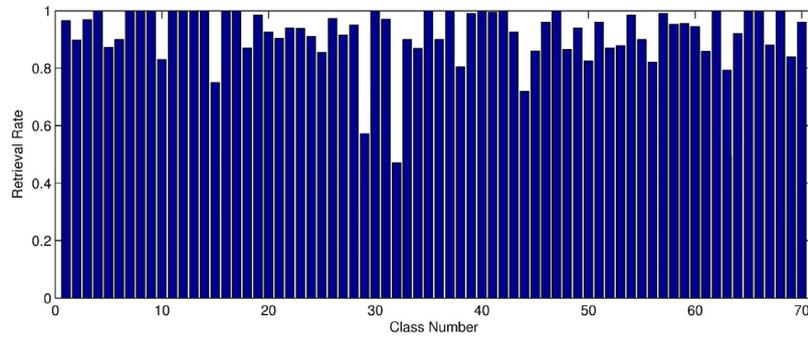


Fig. 17. Retrieval rates of the 70 classes in the MPEG-7 dataset.

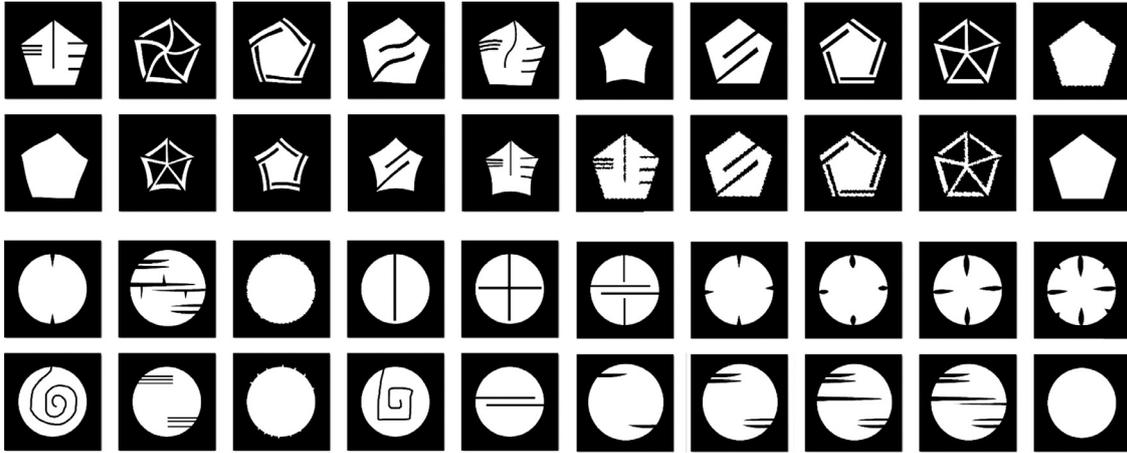


Fig. 18. Shapes of the two classes which have the lowest retrieval rate of the 70 classes in the MPEG-7 dataset.

Table 3

Retrieval results on Kimia's 99 dataset.

Method	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
SC [14]	97	91	88	85	84	77	75	66	56	37
CDPH+EMD [44]	96	94	94	87	88	82	80	70	62	55
Gen. model [37]	99	97	99	98	96	96	94	83	75	48
Efficient indexing [9]	99	97	98	96	97	97	96	91	83	75
Path similarity [45]	99	99	99	99	96	97	95	93	89	73
<b>Our method</b>	<b>99</b>	<b>99</b>	<b>99</b>	<b>99</b>	<b>98</b>	<b>97</b>	<b>95</b>	<b>94</b>	<b>90</b>	<b>83</b>

Table 4

Retrieval results on Kimia's 216 dataset.

Method	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th
SC [14]	214	209	205	197	191	178	161	144	131	101	78
CDPH+EMD [44]	215	215	213	205	203	204	190	180	168	154	123
<b>Our method</b>	<b>216</b>	<b>216</b>	<b>214</b>	<b>210</b>	<b>207</b>	<b>207</b>	<b>201</b>	<b>194</b>	<b>188</b>	<b>182</b>	<b>163</b>

important for shape matching, even if a distance learning method can optimize the accuracy.

### 6.2.3. Kimia's dataset

The Kimia's dataset [25] is another widely used benchmark dataset for evaluating shape matching and retrieval methods, including Kimia's 25, Kimia's 99 and Kimia's 216. As the Kimia's 25 dataset is too small (contains only 25 shape samples in total), we choose the Kimia's 99 and Kimia's 216 datasets to test our method for shape retrieval. All the shape samples of these two datasets are shown in Fig. 19 and Fig. 20. The Kimia's 99 dataset contains 99 shapes grouped into 9 classes. In our test, each shape is used as a query to match all other shapes and their similarities are computed. The retrieval rates are summarized as the number of shapes

from the same class among the top1 to 10 most similar shapes, and the best possible result for each of them is 99. The Kimia's 216 dataset consists of 18 shape classes with 12 shapes in each class. We test our method on this dataset in the same manner as the test on the Kimia's 99 dataset and the best possible score of each result is 216. We only use our method to calculate the pairwise similarity of the shapes, where the distance learning methods are not employed. The results on these two datasets comparing with other methods are shown in Tables 3 and 4, which show that our method has comparable results.

### 6.2.4. Articulated dataset

In this experiment, we implement our method on the articulated dataset [8] to evaluate the robustness against articulated

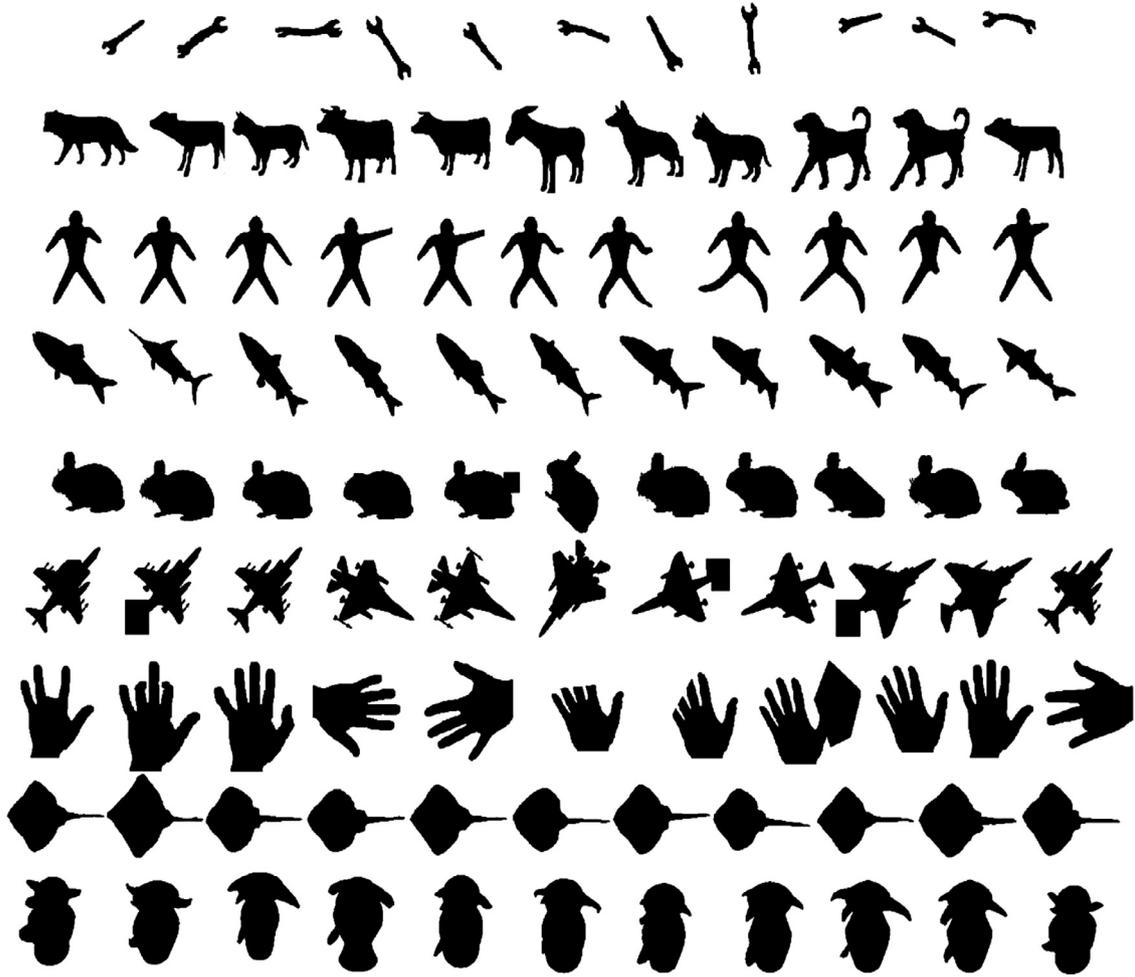


Fig. 19. All the shapes of the Kimia's 99 dataset.

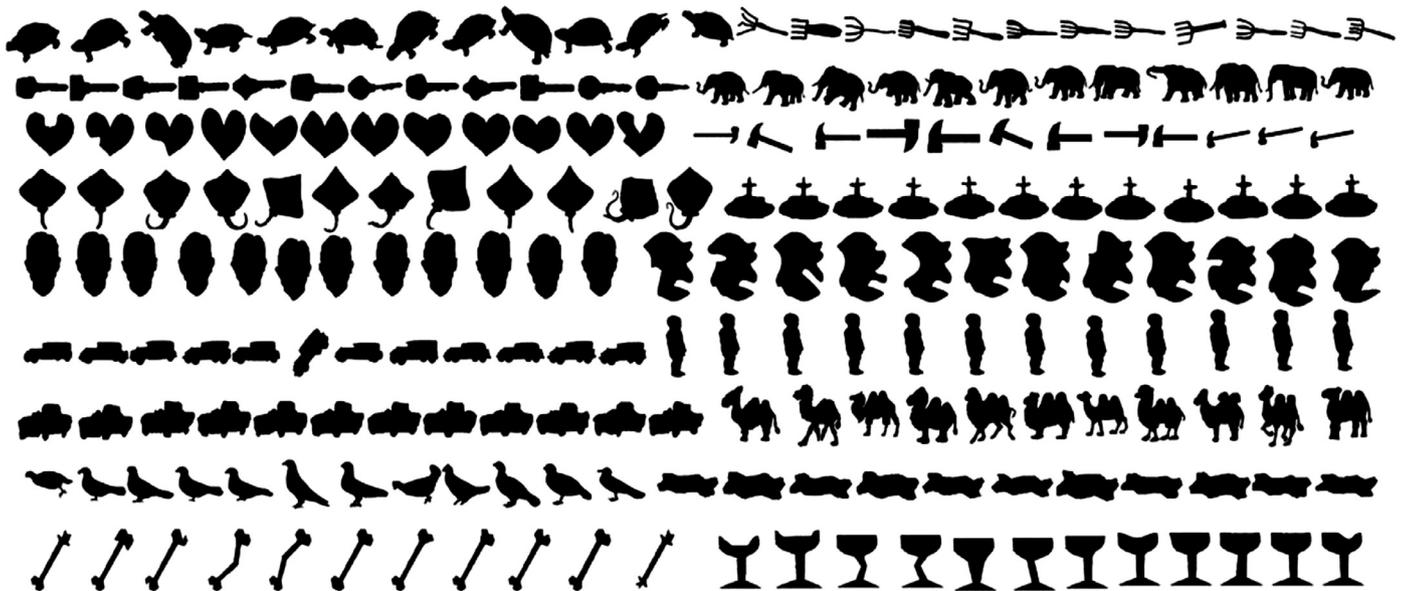


Fig. 20. All the shapes of the Kimia's 216 dataset.

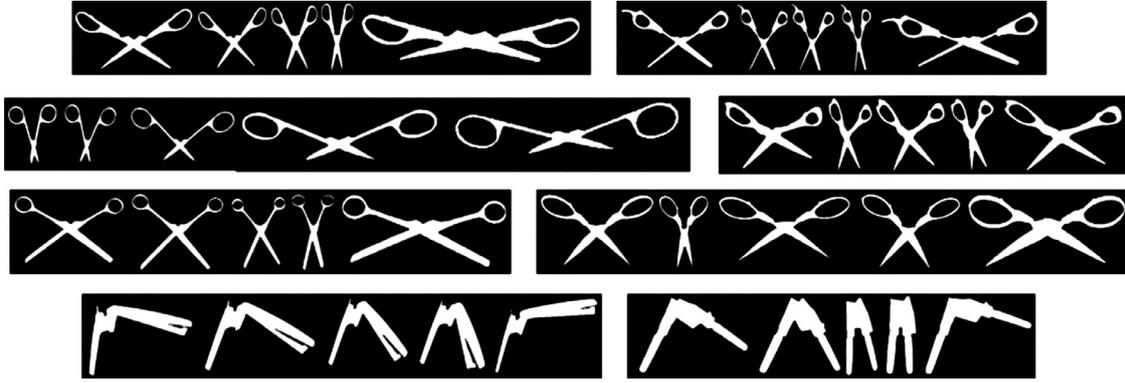


Fig. 21. All the shapes of the articulated shape dataset.

Table 5  
Retrieval results on the articulated shape dataset.

Method	Top 1	Top 2	Top 3	Top 4
L2 (base line) [8]	25	15	12	10
SC [8]	20	10	23	5
MDS+SC [8]	36	26	17	15
IDSC [8]	40	34	35	27
<b>Our method</b>	<b>40</b>	<b>37</b>	<b>35</b>	<b>30</b>

Table 6  
Retrieval rates with different initial  $R$ .

Ratio of $R$	0.3	0.4	0.5	0.6	0.7
MPEG-7	88.16	90.42	<b>91.25</b>	88.74	85.32
Kimia's 99	90.22	93.17	<b>94.46</b>	91.78	86.65

deformation. The articulated dataset contains 40 shapes from 8 classes of objects (six classes of different scissors and two classes of staplers) as shown in Fig. 21. In each class, the five shapes have serious articulated deformation contrast to each other. Besides the articulated deformation, another challenge to implement shape retrieval on this dataset is the significant inter-class similarity. The test is carried out in the same manner as the experiments on the Kimia's datasets, and the results are shown in Table 5. Our method has the highest scores in this table, especially outperforms the IDSC [8] method which is specially designed for articulated shapes. Although the curves of the holes inside the scissors are not used as contours in our method, the holes have effects on the values of the IMD invariants, which increase the discrimination of our method on this dataset.

#### 6.2.5. Validation of initial radius $R$ and total scale number $m$

This experiment demonstrates how the parameters  $R$  and  $m$  affect the performance of our method. Firstly, the accuracy of our method is tested when the initial radius  $R$  is set to different values. The  $R$  is set to different ratios of the distance from the contour point to the shape center (defined in Section 3.3, Eqs. 10 and 11). In this experiment, we set the ratio from 0.3 to 0.7, and the performances of shape retrieval on the MPEG-7 dataset and Kimia's 99 dataset are shown in Table 6 with the bull's eye scores. From the results we can find that the ratio of 0.5 renders the best score in both datasets. The score drops when the ratio is higher or lower. That is because a too small initial  $R$  covers a very local area and the semi-global information is lost, while a too large  $R$  covers nearly the whole shape, thus the description of different contour points are similar. Thus the idea of our descriptor to represent both

Table 7  
Retrieval rates with different scale number  $m$ .

Scale number $m$	1	2	3	4	5
MPEG-7	87.37	88.79	90.10	<b>91.25</b>	89.41
Kimia's 99	91.85	93.21	<b>94.46</b>	92.62	89.56

local and semi-global shape features is reasonable and the setting of initial  $R$  to 0.5 is validated.

In validation of the total scale number  $m$ , we set  $m$  from 1 to 5, and calculate the retrieval accuracy of our method on the two datasets.  $m = 1$  means that only the first scale with initial  $R$  is used to represent shape features, while  $m = 5$  is big enough to include detailed local shape features in these datasets. An extreme  $m$  is to make the smallest circle cover only one pixel, which is meaningless in real application. The experimental results are shown in Table 7, where  $m = 4$  makes the best score in MPEG-7 dataset and  $m = 3$  renders the best result in the Kimia's 99 dataset. The scores with other scale numbers are lower in respective datasets. Less scales cannot represent shape feature details well, while too many scales cannot increase the representativeness of the descriptor but introduce extra noises and errors.

#### 6.3. Robustness for noise

This experiment is carried out to evaluate the robustness of our method against noise. Gaussian noises are added to the original shape contours in the Kimia's 99 dataset. The shape contours are perturbed by a Gaussian random function with zero mean and deviation  $\sigma$  in both  $x$  and  $y$  directions. The noisy shapes with different deviations  $\sigma$  are shown in Fig. 22. The IMD functions of the original shape (Fig. 22, (a)) and the noisy shapes with  $\sigma = 0.4$  and  $\sigma = 0.8$  (Fig. 22, (c), (e)) are shown in Fig. 23. From the figure we can find that the IMD invariants preserves invariance under noise, and the increasing of  $\sigma$  has very little effect on our method. The invariants  $s$  and  $c$  with different  $\sigma$  are almost the same. The absolute values of  $l$  increases when  $\sigma$  gets bigger, because the noise increases the total arc length of the contour as well as the normalized  $l$ .

We also test our method on the perturbed Kimia's 99 dataset with different  $\sigma$ . This test is implemented in the same manner as before to calculate the retrieval rate. The result is shown in Table 8. From the results we can see that the retrieval rates of our method is rarely affected under noise level from  $\sigma = 0.2$  to  $\sigma = 0.6$ , while the rates of  $\sigma = 0.8$  still maintain a relatively high score. This result verifies the robustness of our method against noise in shape retrieval.

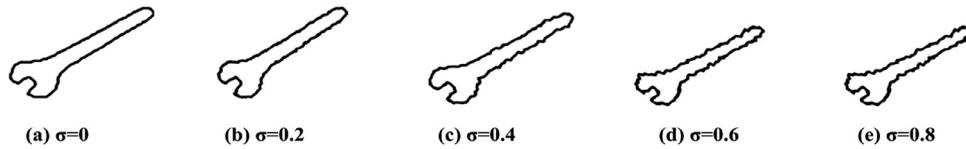


Fig. 22. The same shape with Gaussian noises in different degrees.

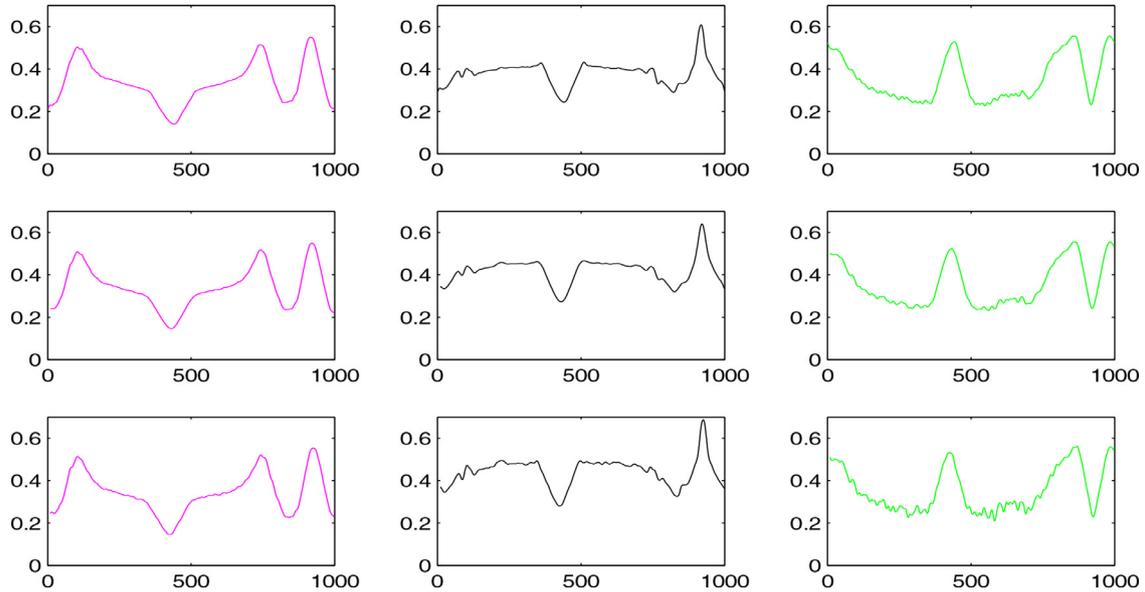


Fig. 23. The invariant functions with Gaussian noises, where  $\sigma = 0, 0.4$  and  $0.8$  in the 1–3 rows respectively. The columns are different types of invariants in the first scale.

Table 8

Retrieval results on Kimia’s 99 dataset with noise.

$\sigma$	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
0	99	99	99	99	98	97	95	94	90	83
0.2	99	99	99	98	97	96	94	93	88	82
0.4	99	99	98	97	97	95	93	92	87	79
0.6	99	99	98	97	96	96	93	91	86	77
0.8	99	99	97	96	96	95	92	88	83	74

Table 9

Efficiency of different methods on the MPEG-7 dataset.

Method	Time (ms)
SC [14]	200
IDSC [8]	310
HPM [17]	25
TAR [15]	70
Shape Tree [18]	500
<b>Our method without ADCE</b>	<b>65</b>
<b>Our method</b>	<b>12</b>

6.4. Efficiency comparison

In this experiment, we test the efficiency of our method comparing with other important methods. The average calculation time of each query shape in the MPEG-7 dataset is tested for different methods, and the results are listed in Table 9. From the table we can see that our method needs 12 ms for each query shape, which is the best score in this table. Furthermore, this result validates that our method is capable of real time implementation under at least 80 fps. We also test the efficiency of our method without using the ADCE algorithm, and the result is 65 ms. This result indicates that the ADCE processing improves the efficiency of shape matching significantly.

7. Conclusion

In this paper, a novel invariant multi-scale descriptor is proposed for shape representation, matching and retrieval. Different types of invariants in the proposed descriptor capture shape features from different aspects. Invariants are used in different scales, which are capable of representing both local and global information simultaneously. The experimental results verify that this descriptor is invariant to rotation, scaling, partial occlusion and intra-class variations, insensitive to articulation and robust for noises. The invariant functions show that the local and semi-global features are both captured by the invariants in different scales. The proposed ADCE algorithm is capable of extract the salient feature points. The retrieval experiments on the benchmark datasets verify that the proposed method has an excellent advantage on retrieval accuracy. The experiment of our method with distance learning algorithms validates that the capability of descriptor is essentially important for a good retrieval performance.

Acknowledgments

This work was funded by research grants from the National Natural Science Foundation of China (Grant No.61305020), the Natural Science Foundation of Jiangsu province, China (Grant No.BK20130316) and the MoE Tier-1 project M4011272.040.

References

- [1] X. You, Y.Y. Tang, Wavelet-based approach to character skeleton, in: Proceedings of the IEEE Transaction on Image Processing, vol. 16, 2007, pp. 1220–1231.
- [2] J. Wang, Y. Li, X. Bai, Learning context-sensitive similarity by shortest path propagation, Pattern Recognit. 41 (10–11) (2011) 2367–2374.
- [3] R. Zhou, J. Yuan, J. Meng, Z. Zhang, Robust part-based hand gesture recognition using kinect sensor, in: Proceedings of the IEEE Transaction on Multimedia, vol. 15, 2013, pp. 1110–1120.
- [4] D. Wolter, L.J. Latecki, Shape matching for robot mapping, in: Proceedings of the Pacific Rim International Conference on Artificial Intelligence, 2004.

- [5] S.D. Choudhury, T. Tjahjadi, Gait recognition based on shape and motion analysis of silhouette contours, *Comput. Vis. Image Underst.* 117 (12) (2013) 1770–1785.
- [6] C. Direkoglu, M.S. Nixon, Shape classification using multiscale fourier-based description in 2-d space, in: *Proceedings of the IEEE International Conference on Signal Processing*, 2008, pp. 820–823.
- [7] N. Alajlan, M. Kamel, G. Freeman, Geometry-based image retrieval in binary image databases, *IEEE Trans. Pattern Anal. Machine Intell* 30 (6) (2008) 1003–1013.
- [8] H. Ling, D. Jacobs, Shape classification using the inner-distance, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (2) (2007) 286–299.
- [9] S. Biswas, G. Aggarwal, R. Chellappa, Efficient indexing for articulation invariant shape matching and retrieval, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, 2007, pp. 1–8.
- [10] C. Xu, J. Liu, X. Tang, 2d shape matching by contour flexibility, *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (1) (2009) 180–186.
- [11] T.B. Sebastian, P.N. Klein, B.B. Kimia, On aligning curves, *IEEE Trans. Pattern Anal. Mach. Intell.* 25 (1) (2003) 116–125.
- [12] E. Klassen, A. Srivastava, W. Mio, Analysis of planar shapes using geodesic paths on shape spaces, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (3) (2004) 372–383.
- [13] M. Siddharth, C. Daniel, H. Byung-Woo, Integral invariants for shape matching, *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (10) (2006) 1602–1618.
- [14] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (4) (2002) 509–522.
- [15] N. Alajlan, I.E. Rube, M.S. Kamel, G. Freeman, Shape retrieval using triangle area representation and dynamic space warping, *Pattern Recognit.* 40 (2007) 1911–1920.
- [16] G.B.D. Souza, A.N. Marana, HTS and HTSn: New shape descriptors based on hough transform statistics, *Comput. Vis. Image Underst.* 127 (2014) 43–56.
- [17] G. McNeill, S. Vijayakumar, Hierarchical procrustes matching for shape retrieval, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1, 2006, pp. 885–894.
- [18] P.F. Felzenszwalb, J. Schwartz, Hierarchical matching of deformable shapes, in: *Proceedings of the CVPR: IEEE International Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.
- [19] L. Latecki, R. Lakamper, T. Eckhardt, Shape descriptors for non-rigid shapes with a single closed contour, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Hilton Head Island, SC, USA, vol. 1, 2000, pp. 424–429.
- [20] X. Bai, X. Yang, L. Latecki, Learning context-sensitive shape similarity by graph transduction, *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (5) (2010) 861–874.
- [21] X. Yang, X. Bai, L.J. Latecki, Improving shape retrieval by learning graph transduction, *Proceedings of the European Conference on Computer Vision (ECCV)*, Springer Berlin Heidelberg, 2008, pp. 788–801.
- [22] X. Yang, S. Koknar-Tezel, L.J. Latecki, Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval, in: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 357–364.
- [23] L. Latecki, R. Lakamper, Shape similarity measure based on correspondence of visual parts, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (10) (2000) 1185–1190.
- [24] T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, *Introduction to Algorithms*, second, MIT Press, 2001.
- [25] T. Sebastian, P. Klein, B.B. Kimia, Recognition of shape by editing their shock graphs, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (5) (2004) 550–571.
- [26] C. Grigorescu, N. Petkov, Distance sets for shape filters and shape recognition, *IEEE Trans. Image Process.* 12 (10) (2003) 1274–1286.
- [27] D.S. Zhang, G. Lu, Generic fourier descriptor for shape-based image retrieval, in: *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, 2002, pp. 425–428.
- [28] P. Srestasathiern, A. Yilmaz, Planar shape representation and matching under projective transformation, *Comput. Vis. Image Underst.* 115 (11) (2011) 1525–1535.
- [29] W.Y. Kim, Y.S. Kim, A region-based shape descriptor using zernike moments, *Signal Process. Image Commun.* 16 (1) (2000) 95–102.
- [30] F. Mokhtarian, S. Abbasi, J. Kittler, Efficient and robust retrieval by shape content through curvature scale space, in: A. Smeulders, R. Jain (Eds.), *Image Databases and Multi-Media Search*, 1997, pp. 51–58.
- [31] A. Srivastava, E. Klassen, S.H. Joshi, H.L. Jermyn, Shape analysis of elastic curves in euclidean spaces, *IEEE Trans. on PAMI* 33 (7) (2011) 1415–1428.
- [32] W. Huang, K.A. Gallivan, A. Srivastava, P.A. Absil, Riemannian Optimization for Registration of Curves in Elastic Shape Analysis, *Journal of Mathematical Imaging & Vision*, in press.
- [33] R. Daliri M., V. Torre, Shape recognition based on kernel-edit distance, *Comput. Vis. Image Underst.* 114 (10) (2010) 1097–1103.
- [34] A. Raftopoulos K., D. Kollias S., The global–local transformation for noise resistant shape representation, *Comput. Vis. Image Underst.* 115 (8) (2011) 1170–1186.
- [35] X. Bai, C. Rao, X. Wang, Shape vocabulary: A robust and efficient shape representation for shape matching, *IEEE Trans. Image Process.* 23 (2014) 3935–3949.
- [36] Z. Ren, J.S. Yuan, W. Liu, Minimum near-convex shape decomposition, in: *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 35, 2013, pp. 2546–2552.
- [37] Z. Tu, A. Yuille, Shape matching and recognition using generative models and informative features, in: *Proceedings of the European Conference on Computer Vision (ECCV)*, Prague, Czech Republic, 2004, pp. 195–209.
- [38] V. Premachandran, R. Kakarala, Perceptually motivated shape context which uses shape interiors, *Pattern Recognit.* 46 (8) (2013) 2092–2102.
- [39] H. Fu, Z. Tian, M. Ran, Novel affine-invariant curve descriptor for curve matching and occluded object recognition, *IET Comput. Vis.* 7 (4) (2013) 279–292.
- [40] V. Palazn-Gonzlez, A. Marzal, On the dynamic time warping of cyclic sequences for shape retrieval, *Image Vis. Comput.* 30 (12) (2012) 978–990.
- [41] H. Ling, X. Yang, L. Latecki, Balancing deformability and discriminability for shape matching, in: *Proceedings of the European Conference on Computer Vision (ECCV)*, Heraklion, Crete, Greece, 2010, pp. 411–424.
- [42] Z. Wang, M. Liang, Locally affine invariant descriptors for shape matching and retrieval, *IEEE Signal Process. Lett.* 17 (9) (2010) 803–806.
- [43] J. Wang, X. Bai, X. You, Shape matching and classification using height functions, *Pattern Recognit. Lett.* 33 (2) (2012) 134–143.
- [44] X. Shu, X. Wu, A novel contour descriptor for 2d shape matching and its application to image retrieval, *Image Vis. Comput.* 29 (4) (2011) 286–294.
- [45] X. Bai, L. Latecki, Path similarity skeleton graph matching, *IEEE Trans. Pattern Anal. Mach. Intell.* 30 (7) (2008) 1282–1292.