Pattern Recognition 45 (2012) 3222–3230

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

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

Perceptually motivated morphological strategies for shape retrieval

Rong-Xiang Hu [a,b], Wei Jia [a,*], Yang Zhao [a,b], Jie Gui [a]

[a] Hefei Institutes of Physical Science, CAS, P.O. Box 1130, Hefei 230031, China
[b] Department of Automation, University of Science and Technology of China, Hefei 230027, China

ARTICLE INFO

Article history:
Received 22 June 2011
Received in revised form
21 February 2012
Accepted 23 February 2012
Available online 5 March 2012

Keywords:
Shape retrieval
Shape contexts
Morphological operation
Perceptual custom

ABSTRACT

In this paper, two perceptually motivated morphological strategies (PMMS) are proposed to enhance the retrieval performance of common shape matching methods. Firstly, two human perception customs are introduced, which have important relations to shape retrieval. Secondly, these two customs are properly modeled by morphological operations. Finally, the proposed PMMS is applied to improve the retrieval performances of a popular shape matching method named Inner-Distance Shape Contexts (IDSC), and then the Locally Constrained Diffusion Process (LCDP) method is exploited to further enhance the retrieval performance. This combination achieves a retrieval rate of 98.56% on MPEG-7 dataset. We also conduct the experiments on Swedish Leaf dataset, the ETH-80 dataset and the Natural Silhouette dataset. The experimental results obtained from four datasets demonstrate clearly the effectiveness of the proposed method.

1. Introduction

Shape is one of the most important features of an object. It plays a key role in human perception, and has been widely exploited for many computer vision applications such as object classification, object recognition and object retrieval. Generally, in a shape-based object retrieval application, given a query object, the most similar objects will be retrieved from a dataset according to certain similarity or distance measures, which are generated by shape matching algorithms. In the past decade, many shape matching algorithms have been proposed [1–10]. Designing a suitable shape matching method for robust shape retrieval, however, is a very difficult task. Some researchers therefore start to explore new perspectives to enhance the retrieval performances. Recently, many novel shape retrieval methods providing new perspectives have been proposed [11–19], and have achieved promising retrieval performances. Here, for convenience's sake, we call them as enhancing methods. Further, they could be divided into three categories, i.e., context-based, knowledge-based and fusion-based.

Context-based methods are the most widely studied enhancing methods so far. Traditionally, in a common shape matching method, matching a pair of shapes would generate a similarity or distance measure, which has no relationship with other shapes. Recently, it is believed that all shapes should be considered as a group rather than pairs of shapes in shape retrieval. Precisely, when the target similarity or distance measure is computed between one pair of shapes, all similarity or distance measures between other shapes within the group could provide useful and complementary information to further correct the target measure. In this way, more satisfactory retrieval results could be obtained. Here, the similarity or distance measures between other shapes within the group are regarded as the context of the target measure. Thus, those methods based on this strategy are called as context-based methods. Yang et al. [11,15] proposed a method to improve the shape retrieval performance through graph transduction by propagating the model through all existing shapes. They took advantage of the manifold formed by the existing shapes, and used an unsupervised graph transduction approach to learn a better measure for retrieval. Kontschieder et al. [12] proposed a modified mutual K nearest neighbor graph to consider the underlying structure of the shape manifold. The manifold is estimated from the shape similarity measures between all the shapes within a dataset. Yang et al. [13] later proposed the method of Locally Constrained Diffusion Process (LCDP) to better enhance the retrieval performance, in which the influence of other shapes is propagated as a diffusion process on a graph formed by the given set of shapes. However, the classical diffusion process is unstable for sparse space, so they add some local constraints using K nearest neighbor graph to achieve a more robust diffusion process. Egozi et al. [16] proposed a meta-shape-similarity approach to characterize a given shape by its similarity to its K nearest neighbors. This approach does not propagate similarities, but aims to compare local graph structures as intrinsic similarity measures. Recently, Yang et al. [19] proposed a novel affinity learning method using Tensor Product

* Corresponding author. Tel./fax: +86 5515591108.
E-mail address: icg.jiawei@gmail.com (W. Jia).

0031-3203/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2012.02.020

Graph, in which not only local but also long range similarities among graph nodes are explicitly considered as higher order relations. They also proposed a novel way to construct the neighborhood structure called Dominant Neighborhood to automatically determine the optimal number of neighbors, which is also a free parameter in LCDP.

Recently, exploiting prior knowledge to improve the retrieval performance becomes a new trend of the enhancing methods, which inspires the knowledge-based methods. As we know, most existing shape matching methods are invariant to scaling and rotation, and are robust to deformation to some extent. In shape retrieval applications, however, the ultimate goal is to obtain human-like retrieval results, where prior knowledge and high level understanding of shapes are very necessary. Thus, in recent years, how to utilize prior knowledge of human visual perception customs to improve the retrieval performance has received more attentions. Temlyakov et al. [17] proposed two perceptually motivated strategies for shape retrieval. The first strategy is to find the strand structure of the shape, a very thin and elongated shape part, and then decomposed the whole shape into a base structure and a set of inward or outward pointing strands. To match two different shapes, the base structures and strand structures of these two shapes are compared, respectively. The second strategy is to identify the bilateral symmetry of the shapes and unify the aspect ratio of shapes according to the symmetric axis before comparison. These two strategies could be applied simultaneously to improve the overall retrieval performance. Gopalan et al. [18] also proposed a knowledge-based approach operating on the contour points rather than the shape image, which considers the changes in 2D shape due to 3D articulations assuming a weak perspective camera model. In this approach, the parts of the shape through approximate convex decomposition is firstly estimated, and then part-wise affine normalization is performed to all contour points.

Fusion-based methods have been widely exploited in many areas in computer vision, such as multimodal biometrics, but they are seldom reported in shape retrieval. It is well known that different shape descriptors could capture different properties of the shapes, which might be complementary to each other. So fusing different measures might lead to a better retrieval performance. Recently, Bai et al. [14] proposed the co-transduction algorithm to fuse different similarity measures for robust shape retrieval. Given two similarity measures and a query shape, this algorithm iteratively retrieves the most similar shapes using one measure and assigns them to a pool for other measures to do a re-ranking and vice-versa. They also evaluated a classical fusion method, i.e., Sum rule, and proved that the proposed fusion algorithm is very effective.

Generally speaking, the context-based methods make full use of the information of the dataset; the knowledge-based methods take the prior information of human visual perception customs into account; and the fusion-based methods integrate useful information from diverse properties of different shape matching methods. So these three categories methods are intrinsically independent, and could be applied simultaneously to enhance the retrieval performance. In fact, Temlyakov et al. [17] have combined one context-based method with their knowledge-based method. And Bai's method [14] integrates one context-based method and one fusion-based method, which achieves impressive retrieval results.

Notice that, all methods mentioned above used the method of Shape Contexts (SC) [1] or Inner-Distance Shape Contexts (IDSC) [1,2,5] to compute the basic distance measures. The method of SC with Weighted Bipartite Matching was proposed by Belongie et al. [1], which describes the relative spatial distribution (distance and orientation) of the feature points by the information of other points. Given n sample points x_1, x_2, …, x_n on a shape, the shape context at point x_i is defined as a 2-D histogram h_i of the distance and angle joint distribution of the remaining $n-1$ points. Later, SC was extended by Ling et al. [2,5] to IDSC with Dynamic Programming (DP), where Euclidian distance in SC is simply replaced by inner distance and the natural constraint of points order is introduced to apply DP. It is proved that the inner distance is more suitable for non-rigid shape matching, and DP is more efficient than Weighted Bipartite Matching. Up to now, SC and IDSC with DP are regarded as the most robust and efficient shape matching methods and have been used in a wide range of computer vision applications [10,14,18]. Especially, the method of IDSC is regarded as the only effort addressing the planar articulation problem [18].

As mentioned above, in knowledge-based methods, perceptually motivated strategies have been successfully exploited to improve the retrieval performance [17,18]. However, research on this topic is still very preliminary. Some issues should be further studied in depth, e.g., "is there any new perceptually motivated strategy that could be used?". In [17], the knowledge-based method proposed by Temlyakov et al. explicitly removes the inward and outward strand structures of the shapes and adjusts the original shape to some standard form according to the bilateral symmetry, both of which are motivated by human perception customs. However, in their method the strands structures are only restricted to thin and elongated type, which is not enough to significantly improve the retrieval performance. In this paper, we propose a novel knowledge-based method, which is also motivated by human perception properties and is based on morphological operations. Following the methods mentioned above, we also select the IDSC algorithm to compute the basic distance measure. The proposed method could improve both the accuracy and robustness of the retrieval. We then combine our method with the LCDP method to further enhance the retrieval performance as previous knowledge-based enhancing method did [17]. The main contribution of this paper is twofold. First, we confirm that two human perception customs introduced in Section 2 should be taken into consideration for shape retrieval. Second, it is demonstrated that morphological operations could simulate these two human perception customs properly and the main structure of the shape could be obtained after the morphological processing. Meanwhile, two morphological strategies have been explored, and our experiments demonstrate the better one in terms of performance and usability.

The rest of this paper is organized as follows. Section 2 introduces two human perceptual customs. Section 3 introduces the basic morphological concepts and explains two morphological strategies. In Section 4, experimental results on four datasets are reported which demonstrate the effectiveness of the proposed method. In Section 5, we conclude the whole paper.

2. Two human perceptual customs

When people retrieve objects based on shape features, the retrieval results would be easily influenced by their perceptual customs, such as dividing the shape into base structure and strand structures, and identifying the bilateral symmetry of shapes. These two customs have been well handled by Temlyakov et al. [17], and can help get more human-like shape retrieval results. Inspired by their work, we also introduce two customs that could strongly affect the retrieval results. First, people would tend to neglect small deformation of the shape and only use the main structure of the shape for retrieval. Second, if the shape consists of main structure and inward parts, people would tend to neglect the inward parts and only regard the shape as the main

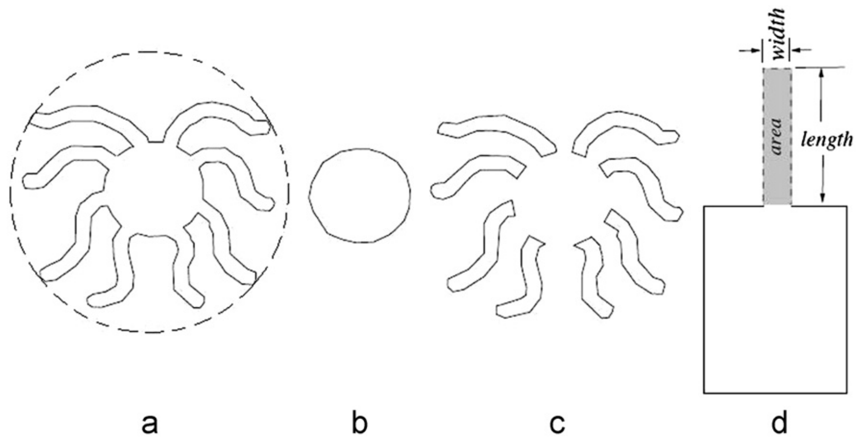

Fig. 1. Difference between the base and strand structures in Temlyakov et al.'s method and the main structure in our approach: (a) the original shape in real line and one possible main structure in dashed line; (b) the base structure; (c) the strand structures; (d) thin and elongated strand that could be processed in Temlyakov et al.'s method defined by small area and small ratio of width to length.

structure. These two customs are active whenever people retrieve shapes. Therefore, we can design a shape retrieval approach taking these two customs into account. Here, it should be noted that the main structure of a shape is different from the base structure defined in Temlyakov et al.'s method. The main structure is not a single shape that depicts the object, but a possible region that is consistent with the shape perceived in our minds. To illustrate the difference between the main structure defined by us and the base structure defined in Temlyakov et al.'s method, an example is shown in Fig. 1. Fig. 1(a) shows an original shape represented by real line. Its base and strand structures are depicted in Fig. 1(b) and (c), respectively. However, its main structure may be a circle represented by dash line as shown in Fig. 1(a) since one person may think it is an approximately circular object according to his/her basic perception. In fact, the given main structure is only one choice among many possible candidates, on the contrary, the base structure is fixed.

2.1. Modeling custom one

Generally, small deformations between shapes are always inevitable due to inter-class variation or image noise. In practice, people are inclined to neglect these deformations and regard the shapes with similar main structure as of the same class. To reduce the effect of these small deformations, some processings can be applied to the original shapes, such as contour smoothing [20] and morphological operation, which would be carefully explained in the next section. After using one of these two processings, small deformations may be removed, and main structures of the shapes can be obtained. In other words, these two processings can be used to model custom one, and some examples are shown in Fig. 2(a).

However, the contour smoothing would cause serious problems when it is applied to certain class of the object. Obviously, without knowing the original shapes, it can be seen from the second column of Fig. 2(b) that the Gaussian smoothed shapes might be recognized as something else (like the silhouette of some animals, e.g., snake and rabbit) rather than the triangle and square. So the contour smoothing cannot be used alone to simulate human perception. Specifically speaking, the contour smoothing is consistent with human perception when facing the small deformations, but it could not handle severe deformation properly. On the contrary, the morphologically processed images can well reflect the main structures of the shapes, and are consistent with human perception.

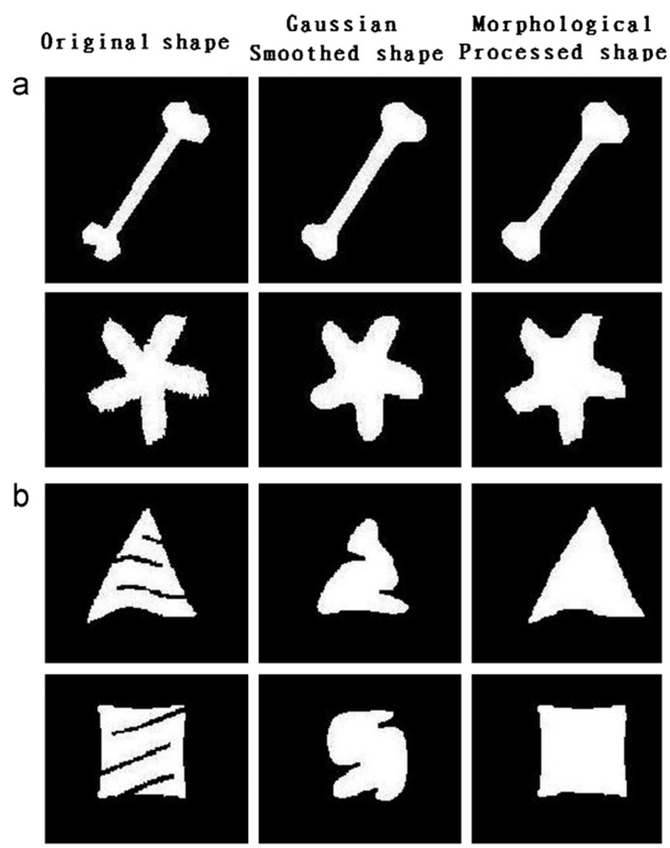

Fig. 2. Modeling custom one by contour smoothing and morphological operation.

2.2. Modeling custom two

When shapes consist of main structure and inward parts, people often neglect the inward parts and regard the shape as the main structure, which is the custom two we introduced above. In Temlyakov et al.'s method, thin and elongate inward parts have been detected and filled [17]. Their method could only handle thin and elongated strands, as shown in Fig. 1(d). However, the inward parts sometimes contain very complex structures and may be even larger than the base structure. In these cases, Temlyakov et al.'s method is invalid, such as original shapes depicted in Fig. 3(a). In order to conduct the retrieval more effectively, we try to extract possible main structure of the shape.

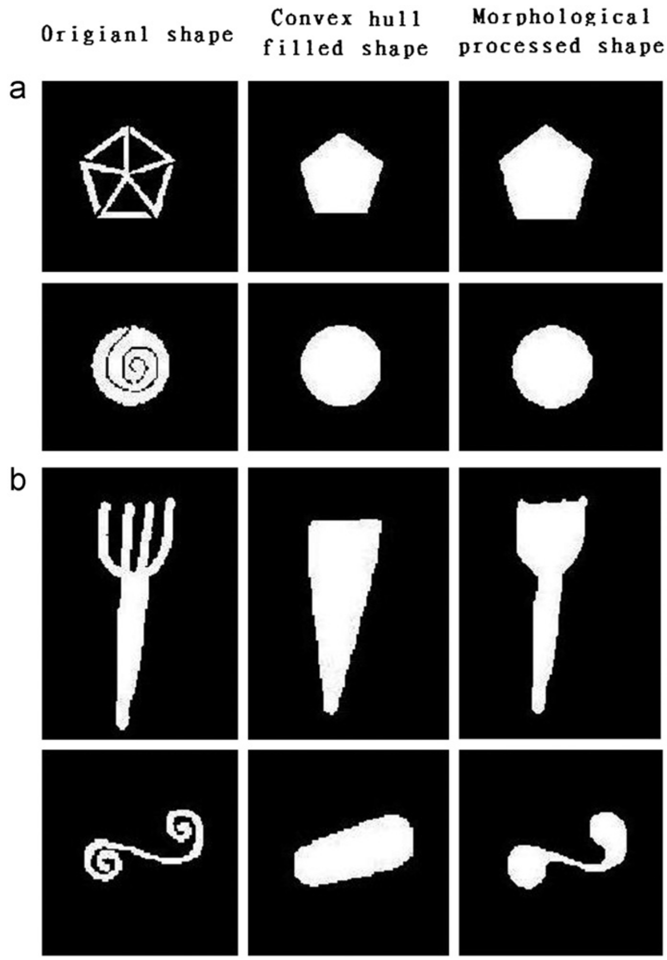

Fig. 3. Modeling custom two by convex hull and morphological operation.

An intuitive method is to represent the main structure by convex hull, and some examples are shown in the second column of Fig. 3(a). Though the convex hulls of some shapes are consistent with human perception, they are not the best choices since they might overprocess the shapes as shown in the second column of Fig. 3(b). In the third column of Fig. 3(a) and (b), the corresponding morphologically processed images are depicted, which are more consistent with the human perception. From Fig. 3, it can be preliminarily concluded that the morphological operations may be a good way to model custom two.

2.3. Other discussions about two human perceptual customs

In fact, people would tend to consider the shape and space gaps around it as a whole object, and ignore certain variations within this object area, which leads to these two customs introduced above. Exactly speaking, in the small scale filling the space gaps around the shape would equal to smoothing the contour, while in the middle scale it would equal to filling the inward parts. Surprisingly, when the space gaps are filled in the large scale, it is still consistent with our perception to some extent, and the original shape would be expanded to occupy meaningful space, as shown in Fig. 4. To define different scales, the ratio of the area of the filled gaps to the area of the original shape could be used; however, it is unnecessary to be computed explicitly. In most cases, filling the space gaps in the small and middle scales is enough, and the most suitable scale to simulate human perception would be determined through experiments.

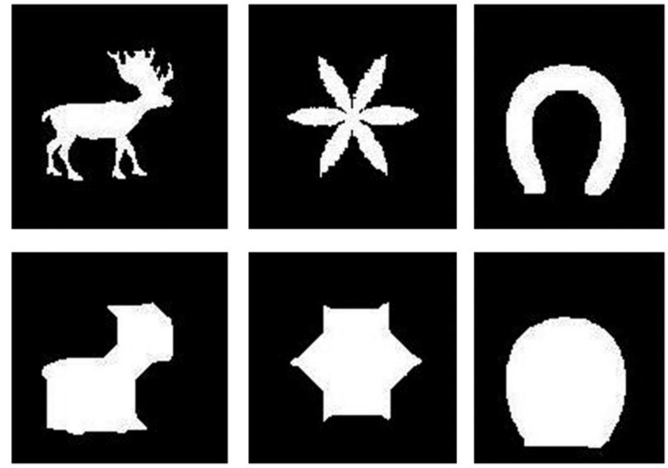

Fig. 4. Three original shapes (depicted in the first row), and their corresponding shapes (depicted in the second row) after filling the space gaps in the large scale.

The two customs introduced above could be explained to some extent by Gestalt Theory [21,22] for human perception. The fundamental principle of Gestalt Theory is the law of pragnanz (German for pithiness), which suggests that humans often order their experience in a manner that is regular, orderly, symmetric and simple. The refinements of the law of pragnanz include many gestalt laws, such as the law of closure and the law of convexity. The law of closure suggests that the mind may perceive illusory elements in order to complete a regular figure, which increases the regularity of the original figure. And the law of convexity means that if some elements suggest themselves as the boundary of a convex figure, the mind may perceive illusory convex contours to complete the convex figure. Obviously, the proposed customs could be easily explained in these two laws, since filling the space gaps would definitely increase the regularity and convexity of the shape. In general, the operation of filling the space gaps makes the shape more regular and simple as the law of pragnanz expects. Notice that, the original shape contains all the information and what is the best for retrieval is unknown, so we not only use the main structure alone for retrieval, but also propose a fusion procedure to incorporate the retrieval results using original shape and that using the main structure, which is detailed in Section 3.3.

3. Morphological strategies for simulating human perception

3.1. Mathematical morphology

Mathematical morphology is a geometric approach for nonlinear image processing, and could be applied for shape analysis in binary and grayscale images. Morphological operations are based on morphological operators, which are defined as combinations of basic numerical operations taking over an image A and a small object B. This small object B is called a Structuring Element (SE), which is used as a probe to scan image A and modify image A according to the specified rule. Different SEs and different modifying rules constitute all kinds of morphological operators, which are powerful tools for image processing and analysis. In other words, diverse morphological operators could be constructed to meet the user's demands.

In shape analysis problems, only binary image is needed, where object is usually depicted with value ones and background is depicted with value zeros. In such cases, binary-image mathematical morphology is enough. Two basic operators extensively presented in the literatures are known as dilation and erosion, which are defined respectively as

$$A \oplus B = \bigcup_{b \in B} A_b \tag{1}$$

$$A \Theta B = \bigcap_{b \in B} A_{-b} \tag{2}$$

where A_b means translation of image A with a displacement vector b contained in B. So the dilation of A by B is regarded as the set union of image A translated by all vectors contained in B and erosion is regarded as the intersection of similar translated images of image A. In general, the dilate operation would expand the image A and the erosion operation would shrink the image A.

3.2. Morphological strategy

Based on two basic operators, many morphology operations could be derived such as opening operation and closing operation, which are defined, respectively, as

$$A \circ B = (A \Theta B) \oplus B \tag{3}$$

$$A \bullet B = (A \oplus B) \Theta B \tag{4}$$

It could be seen that opening operation is defined as an erosion operation followed by a dilation operation using the same SE and closing operation is defined as a dilate operation followed by an erosion operation using the same SE. Both opening operation and erosion operation have the smoothing effect on the shape contour, however, the opening operation would remove thin and elongated strands and the closing operation would fill the gaps and holes. To simulate the human perception properties, we prefer using closing operation as our basic strategy. Exactly speaking, we prefer to process the image using dilate operation first and then using erosion operation.

As mentioned above, those two human perception customs are in different scales, so morphology operations should also be applied in different scales to simulate filling the gaps. Here, two optional strategies could be applied to conduct the simulation. First, the size of SE B could be changed to represent different scales, which means that just the closing operation is applied with some suitable SE. Second, the times of dilate operations and corresponding times of erosion operations could be increased to capture shape variations of different scales, which means that the size of SE is fixed and multiple times of dilate followed by the same times of erosion are applied. Notice that, in shape retrieval application, the rotation of shapes should not affect the final results, so the only suitable form of SE is 'disk' rather than other commonly used elements such as 'diamond' or 'square'. In Fig. 5, the processed images obtained by these two strategies are shown. It could be seen that both strategies could properly simulate human perception in different scales and result in similar results. In the small scale, the gaps of feet are filled, which means the small deformation of the feet is neglected. In the middle scale, the gaps between legs are filled, which means the orientation of the legs is neglected. In the large scale, the gaps between body and legs are filled, and how the body is supported is neglected. In all cases, the main structure of the bird is maintained and can be easily perceived by human. Notice that the processed images in a larger scale include the filling effect in a smaller scale and at the limit (as the SE expanded) the closing operation will approximate the convex hull, which is not the optimal simulation of human perception, so our goal is to find the most suitable scale to simulate human perception.

3.3. Incorporating perceptual retrieval results

The proposed morphological strategy could be incorporated into any available shape matching method. For a certain shape matching method, it can generate a matching cost $C(S_1,S_2)$ between shapes S_1 and S_2. Usually, this matching cost measures the shape distance and is invariant to rotation, translation and scaling. Assuming that the shapes S_1 and S_2 are processed with the proposed morphological strategy, which result in two new shapes, i.e., M_1 and M_2, respectively, then another matching cost $C(M_1,M_2)$ using the same shape matching method could be computed. We believe that if S_1 and S_2 are similar under human perception, then $C(M_1,M_2)$ should be a smaller matching cost. So the following formula is used to determine the final distance measure $D(S_1,S_2)$ between S_1 and S_2:

$$D(S_1,S_2) = \min(C(S_1,S_2), \alpha \cdot C(M_1,M_2)) \tag{5}$$

where α is a scalar larger than one which depicts the penalty of processing S_1 and S_2 to M_1 and M_2, and it is determined by experiments. In fact, this Min rule is just one of many information fusion methods, and other rules could be used, such as Sum rule, or Product rule. However, in this case the Min rule has the clearest explanation, and the experiments demonstrate its superiority. In general, we believe that if the shape should be retrieved by its local variations, then $C(S_1,S_2)$ is more suitable, because this cost reflects global and local variations more equally than $C(M_1,M_2)$. Since the distance between any two processed shapes is computed, the computational complexity of using this fusion procedure is twice as much as that of using the method of IDSC alone.

4. Experiments

In this paper, the IDSC method is used to compute the matching cost $C(S_1,S_2)$ between shapes S_1 and S_2 for its popularity. And, the LCDP method is exploited to further improve the retrieval results. For IDSC, 128 points on the contour are uniformly sampled and the shape context is constructed with 8 bins for distances and 12 bins for angles. For LCDP, the best values of K and t are determined experimentally. The value of K is the number of robust neighbors and the value of t is the optimal times of graph propagation. The codes of IDSC are implemented in Matlab and C, while the codes of LCDP are in Matlab, both of which are obtained from their authors [5,13].

The proposed method is firstly evaluated on the most widely used MPEG-7 dataset (specifically the MPEG-7 CE-Shape-1 Part B) [23] that contains 70 shape classes and 20 different shapes per class. In total, the MPEG-7 dataset contains 1400 samples. Meanwhile, it is a very challenging dataset for shape retrieval application. The examples of shape depicted above are all from this dataset. In experiments, the most commonly accepted performance measure, Bull's eye retrieval rate, is computed to compare the retrieval performances of different methods, which is the ratio of total number of correct matches to maximum number of correct matches. Each image is used as a query, and the number of images that belong to the same class is counted in the top 40 similar images to the query image. Since there are 20 shapes in one class, the maximum number of correct matches for a single query image is 20, and the total number of correct matches is $1400 \times 20 = 28{,}000$. Besides retrieval rates, the recognition rate of the proposed method using the leave-one-out method is also reported, where each shape in turn is left out and used as a query.

To demonstrate the generalization ability of the proposed strategies, experiments are also conducted on the Swedish Leaf dataset [24], the ETH-80 dataset [25] and the Natural Silhouette dataset [26]. Here, it should be noted that only recognition

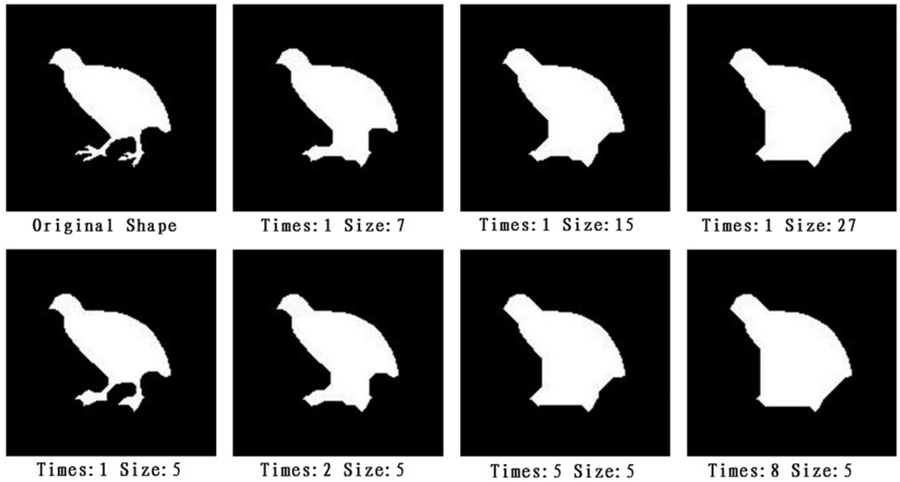

Fig. 5. Results of using different morphological strategies: changing the SE size (depicted in the first row) or changing the operation times (depicted in the second row).

Fig. 6. Eight samples from Swedish Leaf dataset.

Fig. 7. Eight samples from ETH-80 dataset.

experiments are conducted on Swedish Leaf dataset and ETH-80 dataset, and both recognition and retrieval experiments are conducted on Natural Silhouette dataset. The Swedish Leaf dataset contains isolated leaves from 15 different Swedish tree species, with 75 leaves per species. Fig. 6 shows some representative examples from Swedish Leaf dataset. Following the protocol of previous works [4,5], the first 25 images from each class are used as prototypes and the rest 50 images are exploited for test. The ETH-80 dataset contains 80 objects from eight categories. For each object, there are 41 images from different viewpoints. So, the dataset contains 3280 images in total. Fig. 7 shows some representative examples from ETH-80 dataset. To test the recognition rate, the leave-one-object-out method is used as previous works, which means that each shape is compared to all the shapes from other 79 objects and the recognition rate is averaged over all objects. The Natural Silhouette dataset consists of 490 shapes of 12 different classes. Twelve samples from this dataset are shown in Fig. 8. In the recognition experiments, a random selection of 396 shapes were used for training and the remaining 94 shapes were used for test, where the average error rates over 100 random splits are reported. In retrieval experiments, for each shape, the 15 closest shapes are retrieved and the average recall and precision rates are computed. Precision is defined as the ratio of the number of correct retrieved shapes over the total number of retrieved shapes. Recall is defined as the ratio of retrieved shapes over the total number of relevant shapes in the dataset.

Notice that the shapes in the images are not of the same scale, so normalization procedure is applied to make them suitable for the morphological operations of certain scale. We believe that the convex hull could properly represent the scale of the shape, and normalize all shapes in the images to have a convex hull's area near 5000.

Fig. 8. Twelve samples form Natural Silhouette dataset.

4.1. Results on MPEG-7 dataset

In the retrieval experiments on MPEG-7 dataset, using the closing operator to process the images with different sizes of the SE is firstly tested. The retrieval results are listed in Table 1 and corresponding α in Eq. (5) is shown in the brackets below.

It can be seen that the best performance is 90.24% when the size of SE is 33 and the value of α is 1.6. Notice that in Temlyakov et al.'s methods, a strategy based on similar motivation could only enhance the performance of IDSC to 87.68% [17], which is obviously lower than our results. Then the LCDP method is applied to further improve the retrieval results, which are also shown in Table 1. Here the best performances of the method

Table 1
Retrieval results of different SE's size.

Size of SE	5	9	15	21	27	33
Rate (%) (α)	89.54 (1.3)	89.68 (1.4)	89.78 (1.5)	90.04 (1.5)	90.14 (1.5)	**90.24** (1.6)
Rate (%)+LCDP (K,t)	97.94 (15,14)	98.14 (15,23)	98.46 (15,26)	98.49 (15,23)	**98.53** (15,20)	98.48 (15,23)

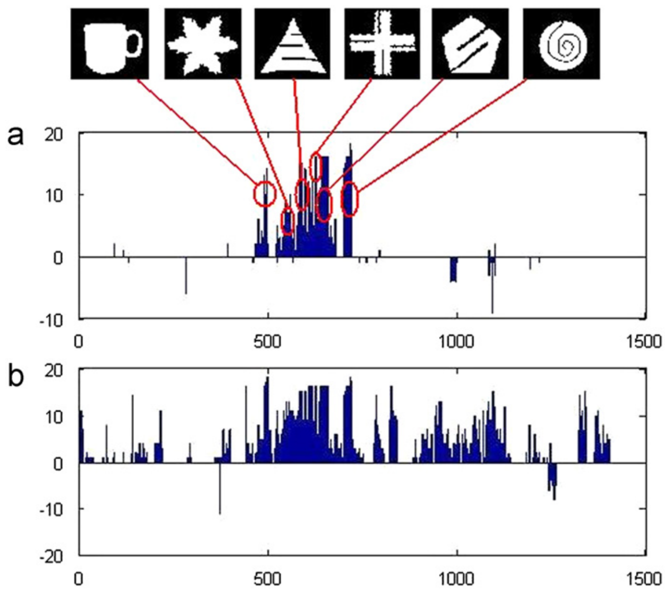

Fig. 9. Improved correct retrieval numbers, where the horizontal axis is the index of the samples. (a) The results of using morphological strategy alone. (b) The results of using morphological strategy and LCDP.

'IDSC+Morphological strategy+LCDP' are listed, and the corresponding parameters of LCDP are in the brackets with the form of (K, t). The best performance is achieved when the size of SE is 27, which indicates that a suitable size would produce more robust distance measures and the retrieval performance could benefit from the LCDP algorithm. In fact a higher retrieval rate could be achieved without LCDP with size larger than 33, but the best result with LCDP has been achieved in a suitable size.

The improved correct retrieval numbers of each query are shown in Fig. 9, where (a) depicts the results of using morphological strategy alone and (b) depicts results of using morphological strategy and LCDP. It can be seen that about one quarter of all the samples have gained positive effect from the proposed morphological strategy, and most of them are shapes with complex inwards parts. Meanwhile, most of the samples have gained positive effect from further enhancement by LCDP.

The second morphological strategy is tested, which uses the dilate operations multiple times followed by erosion operations for the same number of times with the fixed SE. Here, the SE is a disk with the size of 5. The retrieval results are listed in Table 2.

It can be seen that the best performance of IDSC+Morphological strategy is 90.18% when the operation time is 6 or 8. Then the LCDP method is applied to further improve the retrieval results, as shown in the third row of Table 2. Again, the best performance of 'IDSC+Morphological strategy+LCDP' is achieved in a different time, which indicates that a suitable time would produce more robust distance measures and could benefit the LCDP algorithm. In Table 3, the LCDP performances with different K and t for operation time of 4 are shown, and the retrieval is robust in a wide range of parameters. The best K is 15 in all cases, and the best t varies around 26.

Using the first morphological strategy would achieve a slightly lower total performance, and determining the suitable size of SE is troublesome. Obviously, tuning the operation time is quite natural and easy, so the second morphological strategy is a better choice both for accuracy and usability.

The missing retrieval numbers of each query are shown in Fig. 10(c), and it can be seen that most of the shapes could acquire correct retrieval results. The missing retrieval numbers focus on class 'spoon', which is very difficult to retrieval due to great interclass variation, as shown in Fig. 10(d). Some other missing samples are due to the larger stretch of the shapes, which our strategy could not deal with, but they could be handled by bilateral symmetry identification method proposed by Temlyakov, and samples are shown in Fig. 10(a) and (b).

The MPEG-7 dataset is widely used; therefore, many shape matching methods have reported performances on it. To the best of our knowledge, all the retrievals results reported better than IDSC and their corresponding recognition rates are listed in Table 4. Obviously, the proposed method achieves higher retrieval rate than Temlyakov et al.'s method, which uses similar motivations. It is noticed that the performances of combining IDSC, Affine Normalization and LCDP (IDSC+Affine Normalization+LCDP) have not been reported, so this combination is tested here, and it achieves a 100% Bull's eye retrieval rate. In this case, larger and more complex shape datasets are in demands for further research in this area. Surprisingly, though Affine Normalization could dramatically increase the retrieval rate, it would degrade the recognition performance to 91.07%, which is much lower than other methods. On the contrary, our proposed method could achieve very robust retrieval and recognition performances simultaneously.

4.2. Results on Swedish Leaf dataset and ETH-80 dataset

The recognition rates obtained on these two datasets are listed in Table 5. In these experiments, only the strategy of changing the operation times is applied. Notice that the reported performances of IDSC and our reproduced ones are slightly different. A possible reason about this phenomenon is that the segmentation method for converting the original color images to binary object shape might be different. Here, using reproduced ones for performance comparison may be a fair choice. On both datasets using morphological strategy could improve the recognition results. Though the improvement is not significant, considering that the recognition task is usually more dependent on details, which is not the main focus of Gestalt Theory, the effectiveness of the proposed method is still positive.

4.3. Results on Natural Silhouette dataset

Both recognition and retrieval experiments have been conducted on this dataset, where only the results of IDSC and that of IDSC with our strategy are reported. For results of other methods on this dataset, one could refer to [8,26].

For the recognition experiments, the average recognition rates and the standard deviations from 100 random split are shown in Table 6. It could be seen that, no significant improvement is achieved. With the results in Table 5, two conclusions could be

Table 2
Retrieval results of different operation times.

Operation times	1	2	4	5	6	8
Rate (%) (α)	89.54 (1.3)	89.64 (1.3)	89.89 (1.5)	90.04 (1.5)	**90.18** (1.5)	**90.18** (1.6)
Rate (%)+LCDP (K,t)	97.94 (15,14)	98.00 (15,26)	**98.56** (15,26)	98.52 (15,23)	98.52 (15,20)	98.43 (15,23)

Table 3
Retrieval results of different LCDP parameters.

K	t						
	14	17	20	23	26	29	32
20	98.23	98.14	97.98	97.72	97.42	97.26	97.05
15	98.45	98.52	98.49	98.55	**98.56**	98.51	98.43
10	98.09	98.16	98.21	98.26	98.29	98.38	98.40

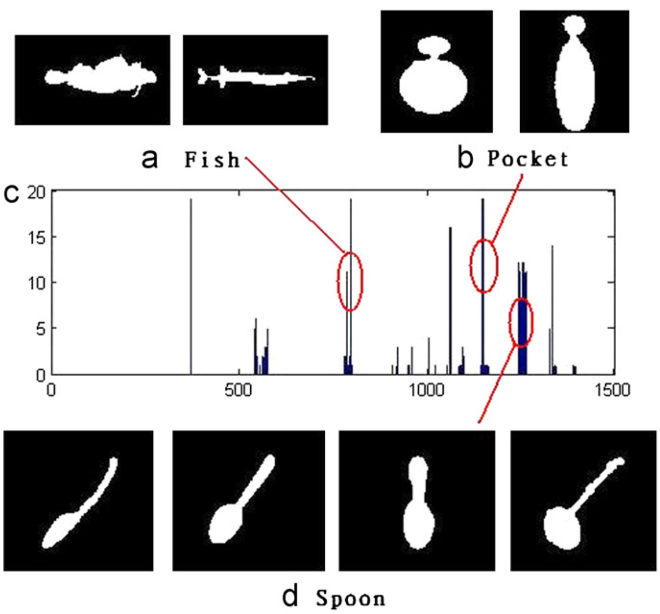

Fig. 10. Missing retrieval results and samples, where the horizontal axis is the index of the samples.

drawn. First, the effect of morphological strategy on recognition experiments is positive but not significant. Second, the operation time of 2 also results in the best improvement.

For the retrieval experiments, the precision–recall curve is shown in Fig. 11. It could be seen that with the morphological strategy the retrieval results of IDSC are obviously enhanced, and the best improvement is achieved when operation time is 3 with α equal to 1.3. It could be concluded that the morphological strategy would have larger positive effect on retrieval than on the recognition.

5. Conclusions

In this paper, we introduced two human perception customs that should be simulated by a knowledge-based method to enhance the shape retrieval performance of a common shape matching method. The first is that people would tend to neglect small deformations of the shape in retrieval and the second is that people would tend to neglect the inward parts of shape. To take into account both customs, we proposed two binary-image morphological strategies to process the shapes, which are based on dilate operation followed by erosion operation. The first strategy changes the size of the structuring element while the second strategy changes the times of the morphological operations. The experiments prove that the second strategy is slightly better than the first strategy.

Table 4
Retrieval performance of different methods.

Algorithm	Retrieval rate (%)	Recognition rate (%)
IDSC [2,5]	85.40	98.21
Symbolic representation [8]	85.92	98.57
HPM [3]	86.35	95.71
IDSC (using EMD) [6]	86.56	NA
SC (using DP) [1]	86.80	**99.14**
Triangle area [7]	87.13	NA
Shape tree [4]	87.70	NA
ASC [10]	88.30	NA
Layered graph [27]	88.75	NA
Variational shape matching [28]	89.05	98.86
Contour flexibility [9]	89.31	NA
IDSC+LP [11,15]	91.00	NA
IDSC+LCDP [13]	92.36	98.71
SC+GM+Meta [16]	92.51	NA
SC+LP [11,15]	92.91	NA
IDSC+Mutual graph [12]	93.40	NA
IDSC+Affine normalization [18]	93.67	91.07
IDSC+Two perceptual strategies+LCDP [17]	95.60	NA
ASC+LCDP [10]	95.96	NA
ASC+TPG [19]	96.47	NA
SC+IDSC+Co-transduction [14]	97.72	NA
IDSC+Morphological strategy+LCDP (**The proposed method**)	98.56	98.86
IDSC+Affine normalization+TPG [19]	99.99	NA
IDSC+Affine normalization+LCDP	**100.00**	96.79

Table 5
Recognition rates on Swedish Leaf dataset and ETH-80 dataset.

Dataset		IDSC only		IDSC+Morphological strategy		
		Reported [5]	Reproduced	Times 1	Times 2	Times 3
Swedish Leaf	Rate (%)	94.13	93.73	93.87	**94.80**	94.67
	α	NA	NA	1.03	1.06	1.08
ETH-80	Rate (%)	88.11	87.47	87.71	**88.04**	87.59
	α	NA	NA	1.1	1.3	1.6

Table 6
Recognition rates on Natural Silhouette dataset.

	IDSC only	IDSC+Morphological strategy			
		Times 1	Times 2	Times 3	Times 4
Rate (%)	96.79	97.05	**97.34**	97.17	96.98
Std (%)	1.81	1.74	**1.71**	1.72	1.77
α	NA	1.1	1.2	1.3	1.4

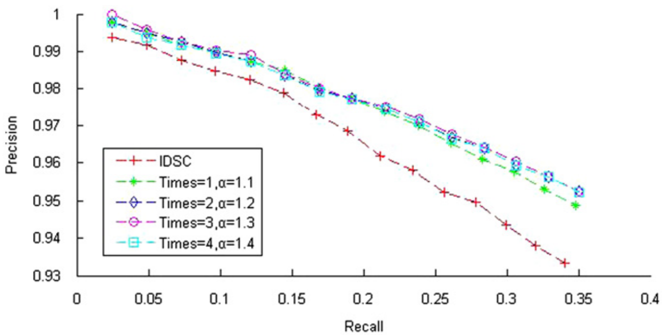

Fig. 11. Precision–recall curve on Natural Silhouette dataset.

Both strategies could be incorporated to any common shape matching method to improve the retrieval performance, and the experiments using IDSC have demonstrated the effectiveness of the proposed method. Since the knowledge-based method and the context-based method are independent, we also combined our proposed methods with LCDP, and this combination achieved the state-of-the-art performance of 98.56% on MPEG-7 dataset. In the future, based on the proposed method, we will exploit fusion-based methods to further enhance the retrieval performance.

Acknowledgments

This work is supported by the grants of the National Science Foundation of China, nos. 61175022, 61100161, 61005010, 60705007, 60975005 and 60905023; and the grants of the Knowledge Innovation Program of the Chinese Academy of Sciences (Y023A11292 and Y023A61121).

References

[1] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (4) (2002) 509–522.
[2] H. Ling, D.W. Jacobs, Using the inner-distance for classification of articulated shapes, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 719–726.
[3] G. McNeill, S. Vijayakumar, Hierarchical procrustes matching for shape retrieval, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2006, pp. 885–894.
[4] P.F. Felzenszwalb, J.D. Schwartz, Hierarchical matching of deformable shapes, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[5] H. Ling, D.W. Jacobs, Shape classification using the inner-distance, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (2) (2007) 286–299.
[6] H. Ling, K. Okada, An efficient Earth mover's distance algorithm for robust histogram comparison, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (5) (2007) 840–853.
[7] N. Alajlan, M.S. Kamel, G.H. Freeman, Geometry-based image retrieval in binary image databases, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (6) (2008) 1003–1013.
[8] M.R. Daliri, V. Torre, Robust symbolic representation for shape recognition and retrieval, Pattern Recognition 41 (5) (2008) 208–220.
[9] C. Xu, J. Liu, X. Tang, 2D shape matching by contour flexibility, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (1) (2009) 180–186.
[10] H. Ling, X. Yang, L.J. Latecki, Balancing deformability and discriminability for shape matching, in: Proceedings of the European Conference on Computer Vision, 2010, pp. 411–424.
[11] X. Yang, X. Bai, L.J. Latecki, Improving shape retrieval by learning graph transduction, in: Proceedings of the European Conference on Computer Vision, 2008, pp. 788–801.
[12] P. Kontschieder, M. Donoser, H. Bischof, Beyond pairwise shape similarity analysis, in: Proceedings of the Asian Conference on Computer Vision, 2009, pp. 655–666.
[13] X. Yang, Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 357–364.
[14] X. Bai, B. Wang, X. Wang, Co-transduction for shape retrieval, in: Proceedings of the European Conference on Computer Vision, 2010, pp. 328–341.
[15] X. Bai, X. Yang, L.J. Latecki, Learning context-sensitive shape similarity by graph transduction, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (5) (2010) 861–874.
[16] A. Egozi, Y. Keller, H. Guterman, Improving shape retrieval by spectral matching and meta similarity, IEEE Transactions on Image Processing 19 (5) (2010) 1319–1327.
[17] A. Temlyakov, B.C. Munsell, J.W. Waggoner, Two perceptually motivated strategies for shape classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 2289–2296.
[18] R. Gopalan, P. Turaga, R. Chellappa, Articulation-invariant representation of non-planar shapes, in: Proceedings of the European Conference on Computer Vision, 2010.
[19] X. Yang, L.J. Latecki, Affinity learning on a tensor product graph with applications to shape and image retrieval, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2011.
[20] F. Mokhtarian, S. Abbasi, J. Kittler, Efficient and robust retrieval by shape content through curvature scale space, Image Databases and Multi-Media Search (1997) 51–58.
[21] K. Koffka, Principles of Gestalt Psychology, Harcourt, Brace and Company, 1935.
[22] A. Desolneux, L. Moisan, J.-M. Morel, From Gestalt Theory to Image Analysis: A Probabilistic Approach, Springer, 2008.
[23] L.J. Latecki, R. Lakamper, U. Eckhardt, Shape descriptors for non-rigid shapes with a single closed contour, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2000, pp. 424–429.
[24] O.J.O. Soderkvist, Computer vision classification of leaves from Swedish trees, Master Thesis, Department of Electronic Engineering, Linkoping University, Linkoping, Sweden, 2001.
[25] B. Leibe, B. Schiele, Analyzing appearance and contour based methods for object categorization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2003.
[26] L. Gorelick, M. Galun, E. Sharon, Shape representation and classification using the poisson equation, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (12) (2006) 1991–2005.
[27] L. Lin, K. Zeng, X. Liu, Layered graph matching by composite cluster sampling with collaborative and competitive interactions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1351–1358.
[28] K. Nasreddine, A. Benzinou, R. Fablet, Variational shape matching for shape classification and retrieval, Pattern Recognition Letters 31 (2010) 1650–1657.

Rong-Xiang Hu received the B.E. degree in Electronic Information Engineering from Hefei University of Technology, Hefei, China, in 2006. From September 2006, he is a Master-Doctoral Program student in the Department of Automation, University of Science and Technology of China, Hefei, China. His research interests include pattern recognition, machine learning and image processing.

Wei Jia received the B.Sc. degree in informatics from Central China Normal University, Wuhan, China, in 1998, the M.Sc. degree in computer science from Hefei University of Technology, Hefei, China, in 2004, and the Ph.D. degree in pattern recognition and intelligence system from University of Science and Technology of China, Hefei, China, in 2008. He is currently an associate professor in Hefei Institutes of Physical Science, Chinese Academy of Science. His research interests include biometrics, pattern recognition, and image processing.

Yang Zhao received the B.E. degree in department of automation, University of Science and Technology of China, Hefei, China, in 2008. From September 2008, he is a Master-Doctoral Program student in the Department of Automation, University of Science and Technology of China, Hefei, China. His research interests include pattern recognition, machine learning and image processing.

Jie Gui received the B.E degree in Computer Science in Hohai University in 2004, the M.Sc. degree in computer science from Chinese Academy of Science, China, in 2007, and the Ph.D. degree in pattern recognition and intelligence system from University of Science and Technology of China, Hefei, China, in 2010. He is currently an assistant professor in Hefei Institutes of Physical Science, Chinese Academy of Science. His research interests are machine learning, pattern recognition, and image processing.