IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 29, 2019, accepted August 14, 2019, date of publication August 16, 2019, date of current version August 30, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2935879

Shape Matching Based on Multi-Scale Invariant Features

ZHANG KUN[1,2], MA XIAO[3,4], AND LI XINGUO[1,2]

[1]School of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China
[2]Shaanxi Aerospace Flight Vehicle Design Key Laboratory, Xi'an 710072, China
[3]National Key Laboratory of Shock Wave and Detonation Physics, Mianyang 621900, China
[4]Institute of Fluid Physics, China Academy of Engineering Physics, Mianyang 621900, China

Corresponding author: Zhang Kun (zhangkun02804@163.com)

This work was supported in part by the National Natural Science Fund of China under Grant 11672235, and in part by the National Defense Basic Research Project of China under Grant A0420132102.

ABSTRACT Shape matching has been extensively used in various fields. The local feature-based or global feature-based algorithms can hardly describe the shape comprehensively due to their inherent defects. Combining the local and global feature to describe the shape has become a trend. In this paper, an improved discrete curve evolution algorithm is proposed which combines the discrete curve evolution with the uniform sampling and achieves a better description of the shape contour. Three simple and intuitive multi-scale features which represent both the global and local features of shapes are designed from aspects of the spatial relationship of contour points, structural information of contour sequence, and shape geometry feature. A cyclic Smith-Waterman algorithm is introduced to solve local contour matching and starting point selection. Experimental results demonstrate that our proposed features are translation, rotation, and scaling invariant, and have good robustness to deformation. Retrieval accuracies of Kimia99, Kimai216, and MPEG-7 indicate that our method can bring out a better performance.

INDEX TERMS Shape matching, improved discrete curve evolution, multi-scale invariant features, cyclic Smith-Waterman algorithm.

I. INTRODUCTION

As an important issue in image processing, shape matching has achieved fruitful achievements in various fields, such as target recognition [1], image retrieval [2], [3], and biometric recognition [4], [5]. Shape matching is to measure the similarity between shapes by calculating the similarity between features according to certain criteria. In practice, image is often interfered by deformation, scaling, and occlusion resulting in the reduction of intra-class similarity, and even influencing the retrieval accuracy. Therefore, how to construct stable feature descriptors becomes a key part of shape matching.

Research on shape matching has been extensively studied by many scholars. The global feature-based algorithms usually analyze the spatial distribution of image pixels or feature points from a global perspective to construct feature descriptors, and achieve shape matching through the similarity between features, such as shape context (SC) [6], [7], Fourier transform [3], curvature scale space (CSS) [8], feature invariants [9], [10], geometric hash value [11], mapping by tone (MTM) [12], etc. The global features have the property of being insensitive to noise, but the matching performance on deformation and occlusion is not good due to the missing of local details.

The local feature-based algorithms usually focus on a local region of the shape and construct feature descriptors by analyzing characteristics of the local region, such as scale-invariant feature transform (SIFT) [13], histograms of oriented gradient (HOG) [14], integral invariants [15], multi-resolution polygon [16], multi-scale concave-convex representation [17], etc. The local feature-based algorithms can deal with occlusion and deformation, but they are susceptible to noise. Also, the matching accuracy is degraded because of the lack of global shape information.

Recently, combining global and local features has become a tendency [2], [18]–[21]. It can ensure that feature descriptors could retain the global information of the shape while preserving local details. Using the combined features can characterize the shape more comprehensive, improve the robustness to the noise, and deal with deformation and occlusion effectively.

The associate editor coordinating the review of this article and approving it for publication was Abdel-Hamid Soliman.

In this paper, we proposed a shape matching algorithm based on multi-scale invariant features of the contour sequence. Based on the analysis of discrete curve evolution (DCE), an improved DCE (IDCE) algorithm which effectively combines DCE with uniform sampling is introduced. By analyzing the contour sequence, it is characterized by three different features: the spatial relationship of contour points which is represented by the contour angle of adjacent contour points, structural information of contour sequence which can be described by the local binary pattern based on contour points (CP-LBP) and shape geometry feature by Zernike moment of neighborhood regions. To solve the problem of starting point selection of matching sequences, a cyclic Smith-Waterman (CSW) algorithm is proposed. Experimental results indicated that our proposed method can achieve better performance.

The organization of this paper is as follows: Section II introduces related work on shape matching. Section III gives the detailed IDCE algorithm. The proposed multi-scale features and CSW are explained in section IV and section V separately. Experimental results are discussed in section VI, and conclusions are given in section VII.

II. RELATED WORK

Global feature-based algorithms analyze the spatial distribution of image pixels or feature points in the shape region from a global perspective to construct feature descriptors. Hu moment [9] and Zernike moment [10] have been widely used in shape matching. Belongie *et al.* [6] proposed SC which contains abundant shape information and reflects the distribution of contour sampling points. Ling and Jacobs [7] used the inner-distance instead of Euclidean distance and achieved good performance with the image under non-rigid deformation. Mokhtarian *et al.* [8] applied CSS to the feature description. Yang *et al.* [3] analyzed the distribution of shape contour points by Fourier transform. For each row of the distance matrix, only the amplitudes of M low-frequency signals are retained as features which can effectively reduce the impact of different starting points on the matching result. To deal with the non-linear tone variation, Hel-Or *et al.* [12] introduced the matching by tone mapping (MTM). This method achieves the goal by finding a global optimal tone mapping between the reference image and the input image. However, it only adopts the image tone information. Jiao *et al.* [22] divided the target region into 24 circular areas in 8 directions. The feature of each circular area at the same scale is calculated by Census transform and rotated according to the main direction. Wei *et al.* [23] introduced a framework for shape-based object detection which could convert the object detection to a graph-based search problem.

Local feature-based algorithms usually focus on a local region of the image and construct feature descriptors by analyzing characteristics of the nearby region. As classical feature descriptors, SIFT [13] and HOG [14] are commonly used in shape matching. Yoo *et al.* [24] defined the principal gradient and built the histogram of the principal gradient to solve the matching problem under a great change of image scale. Adamek and O'Connor [17] adopted the concave and convex of the shape to build multi-scale feature descriptors. In this method, the concave and convex are determined by the Euclidean distance between the corresponding contour points in two adjacent scales whose position was smoothed by the Gaussian function. This method is able to resist the influence of local noise. Klassen *et al.* [25] employed the distance and area integral invariants to deal with the noise sensitivity problem of differential invariants. Zhao and Du [26] found the extreme points of each ring region in the resulting image of ring projection transformation. The feature sequence containing the pixel gray values and shape structure information is extracted by analyzing the distribution of extreme points. Yu *et al.* [19] segmented the contours and introduced local directional chamfer matching (DCM) and oriented chamfer matching (OCM). The matching problem is converted into the optimal combination of contour segments. Kaothanthong *et al.* [27] introduced a shape signature named distance interior ratio (DIR) as the feature of local shape. The DIR is defined as the ratio of the length of the line segment inner the shape and the length of the line segment. Deng *et al.* [28] proposed a local feature descriptor, point pair local topology (PPLT). The PPLT is a histogram which is constructed by the weighting of distance measures and angle measures based on local pairs.

Because of inherent defects of global and local features, combining global and local features has attracted more and more attention. Alajlan *et al.* [20] obtained triangles at different scales by controlling the distance between contour sampling points. The concave and convex of the shape are determined by the sign of the triangle area. This method can obtain local and global shape features through the feature descriptors at different scales and is invariant to rotation, translation, and scaling. Donoser *et al.* [21] introduced the angle of the contour chord. The feature matrix is generated by calculating the angle between the reference point and all other points. The diagonal nearby features represent the local information while features far away from the diagonal indicate the global shape information. Yang *et al.* [2] built feature descriptors with chord length ratio, area ratio, and distance between contour point and centroid at different scales. Local shape features can be acquired at a small scale while semi-global shape features can be obtained at a large scale. This method performs well in dealing with deformations, such as joint deformation. Tu and Yuille [29] defined a set of informative features including local shape features and global features. The local features are constructed by calculating the statistical characteristics of the angle between contour points in a local range, while the global features hold the statistical characteristics of the angle between all contour points. Yang and Yu [18] proposed a multiscale Fourier descriptor based on triangular features. The global and local features can be obtained by controlling the side length of the triangular. Convolution neural network (CNN) has also been adopted in shape matching and retrieval [30], [31]. Li and Yang [31]

designed an adaptive weighted CNN features-based Siamese network. In this method, the feature extraction network is designed to derive feature maps while the feature integration network is adopted to adaptively assign weights to each region of the feature maps and integrate feature maps.

Contour-based matching methods usually alter the shape matching to the similarity measure of two feature sequences. The similarity measure refers to the calculation of the similarity between two shape features by certain criteria. Generally, the greater the similarity, the higher the similarity between the two shapes. Euclidean distances, such as the sum of squared differences (SSD), the sum of absolute differences (SAD), and Manhattan distance, have advantages of small computational complexity and easy implementation, however, they are susceptible to noise. Yoo and Han [32] utilized the normalized cross-correlation (NCC) as the similarity measurement and achieved a good anti-noise performance, but NCC is computationally expensive. Euclidean distance and NCC can only handle features with the same length. When there exists deformation or occlusion, Euclidean distance and NCC appear powerless. Edit distance (ED) is first applied to the string matching by finding the minimum operations that equalize two sequences through operations of substitution, insertion, and deletion. With a threshold to determine whether the two features are equal or not, ED can cope with the matching of sequences in different lengths. However, ED only counts the minimum operations, the actual difference between the two features is not considered once it is within the threshold. Zhao and Itti [33] employed dynamic time warping (DTW) to turn the sequence matching into the shortest path problem, but the shortest path would contain the correspondence of all points. The longest common subsequence (LCSS) can effectively solve the matching of sequences in different lengths and reduce the influence of singular points. However, similar to ED, LCSS only counts the maximum length of the common sub-sequence. Chen *et al.* [34] proposed a local contour matching method based on the Smith-Waterman (SW) algorithm. The proposed method adopts the weighted distance between two feature points as the similarity function which measure the similarity of two feature points intuitively. SW allows to skip singular points, but unlike LCSS, it can avoid the problem of excessively insertion or deletion by the gap function.

III. IMPROVED DISCRETE CURVE EVOLUTION

A. DISCRETE CURVE EVOLUTION

Generally, the shape contour obtained by the contour extraction method has many redundancy points and is susceptible to noise. Latecki *et al.* [35] utilized DCE as the pre-processing which can remove noise points while retaining salient feature points, so as to simplify the contour and reduce the sensitivity of the contour to noise.

Given a shape contour C, $S_1, \cdots, S_n$ are contour segments of C, $P_1, \cdots, P_n$ are contour sample points. The correlation

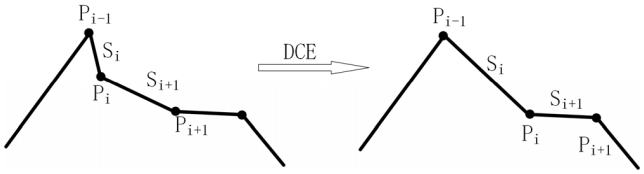

FIGURE 1. The schematic diagram of discrete curve evolution.

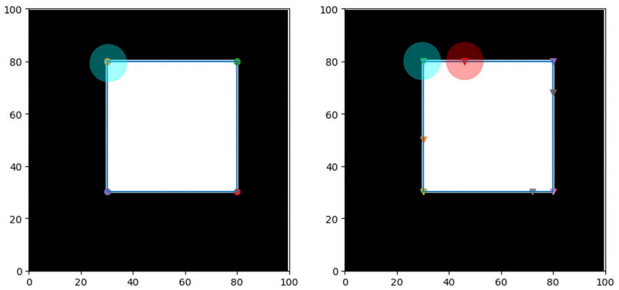

FIGURE 2. Comparison results of DCE and IDCE of the square. (a) Result of DCE. (b) Result of IDCE.

K at point P_i is defined as (1):

$$K(S_i, S_{i+1}) = \frac{\beta(S_i, S_{i+1})\, l(S_i)\, l(S_{i+1})}{l(S_i) + l(S_{i+1})} \tag{1}$$

where S_i and S_{i+1} are line segments formed by point P_i and P_{i-1}, P_{i+1}, respectively. $l(S_i)$ and $l(S_{i+1})$ are lengths of S_i and S_{i+1}, $\beta(S_i, S_{i+1})$ is the turn angle of S_i and S_{i+1}. A larger K means that the point contributes a lot to the contour and should be preserved, while a smaller K point can be eliminated. In each iteration, point P_i corresponding to the minimum correlation K will be substituted by P_{i+1}, as shown in Fig. 1. In each evolution, at least one point will be removed resulting in a simplified contour with salient feature points preserved.

B. IMPROVED DISCRETE CURVE EVOLUTION

DCE can effectively simplify the contour and retain salient feature points, but it is easy to over-simplify in practice. As indicated in Fig. 2(a), DCE can describe the square well only through four corner points, but there is obviously a lack of expression of the four sides. Suppose the feature descriptor is defined as the ratio of the inner contour area in the circle to the circle area. The feature value of the corner point whose neighborhood region marked by green is constantly 0.25 as shown in Fig. 2(a) and Fig. 2(b), while the side point marked by red is 0.5 as shown in Fig. 2(b). It can be seen that there is a certain lack of expression when the square contour consists of only four corner points without any side points. Therefore, an improved DCE (IDCE) algorithm is proposed in this section. By the IDCE, salient feature points and certain redundant points will be appropriately retained, and the descriptive ability of the contour will be highly improved.

Mark $K_1, \cdots, K_n$ as the correlation sequence of contour sequence $P_1, \cdots, P_n$, σ_K is the variance of correlation sequence. For each contour point P_i, a subsequence

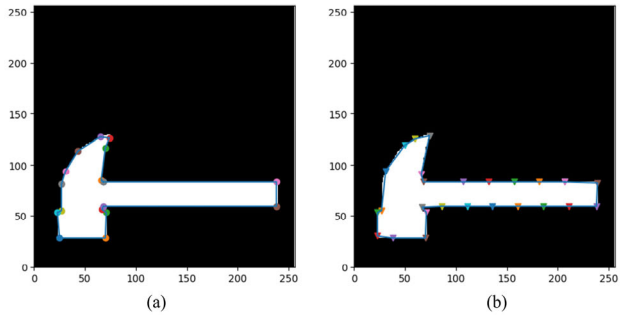

FIGURE 3. Results of DCE and IDCE of the hammer (a) Result of DCE (b) Result of IDCE.

$K_{Pi} = K_i, K_{i+1}, \cdots, K_{i+N}$ of length N is extracted along the contour direction and its variance is marked as σ_{Pi}. The variance, as a statistical parameter of the sequence distribution, can represent the dispersion characteristic of the sequence. When $\sigma_{Pi} \geq \sigma_K$, the dispersion of the current subsequence is greater than the average dispersion, that is, there is a salient feature point in the subsequence with high probability. The point with the largest correlation in the subsequence is preserved as the salient contour point. $\sigma_{Pi} < \sigma_K$ implies that the dispersion of current subsequence is relatively small and the shape curvature changes slowly. The point K_{i+N+1} is added to update the subsequence and the calculation is repeated with the new subsequence. When the length of subsequence N reaches the length threshold $M(M \geq N)$, the point corresponding to the maximum correlation in the subsequence is taken as the salient feature point and preserved.

Comparison results of IDCE and DCE are shown in Fig. 3. It is indicated that IDCE cannot only effectively describe the hammerhead, but also emphasize the region with slow curvature change such as the handle. IDCE is equivalent to the combination of DCE and uniform sampling. It can retain more feature points in the region with large curvature change to enhance the representation of the simplified contour while keeping the descriptive capability of the region with small curvature change through uniform sampling. In the worst case, IDCE is equivalent to the uniform sampling of length N, however, DCE can only remove a single point in each evolution. The proposed features are calculated with the original contour sequence on contour points retained by IDCE.

IV. MULTI-SCALE FEATURE DESCRIPTORS

By analyzing the contour sequence, in this section, the multi-scale invariant feature descriptors are introduced through the spatial relationship of contour points, structural information of contour sequence, and shape geometry feature.

A. CONTOUR ANGLE

The relative spatial position of different contour points can be described by the contour angle between contour segments as shown in Fig. 4. For each contour point P_i, its neighbor points with distance K are P_{i-k} and P_{i+k}. P_i and P_{i-k}, P_{i+k} form vectors $\overrightarrow{P_iP_{i-k}}$ and $\overrightarrow{P_iP_{i+k}}$ respectively. The contour angle is

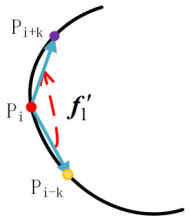

FIGURE 4. The contour angle.

defined as the turn angle from $\overrightarrow{P_iP_{i-k}}$ to $\overrightarrow{P_iP_{i+k}}$. The contour angle at contour point P_i of scale K can be calculated as (2).

$$f'_{1,i,k} = Angle\left(\overrightarrow{P_iP_{i-k}}, \overrightarrow{P_iP_{i+k}}\right) \tag{2}$$

The contour angle with small K represents a local feature indicating the local detail of shape contour, while a large K interprets the semi-global to global feature. The sampling distance K is adaptively adjusted according to the contour length, which can cope with contours with different lengths. To improve the noise sensitivity and the robustness to the shape deformation, here the contour angle is quantized into N intervals as (3), where $(x]$ represents the smallest integer greater than x. The quantized angle is more robust to deformation than the original angle when the shape deforms slightly.

$$f_{1,i,k} = \left(f'_{1,i,k} \cdot \frac{N}{360}\right] \tag{3}$$

B. LOCAL BINARY PATTERN BASED ON CONTOUR POINTS

Local binary pattern (LBP) as a local feature descriptor can reflect the texture information of the region around the center point. LBP is invariant to rotation and translation and insensitive to illumination change. Inspired by LBP, in this section, we utilize LBP based on contour points (CP-LBP) as features to describe structural characteristics of the shape. For each contour point P_i, extract the subsequence $Seq_{i,s} = \{P_{i-4s}, P_{i-3s}, P_{i-2s}, P_{i-s}, P_{i+s}, P_{i+2s}, P_{i+3s}, P_{i+4s}\}$ along the contour direction, s is the sampling step. Different sampling steps represent CP-LBP at different scales. Gradients of each sampling point are calculated and marked as $G_{i,s} = \{g_{i-4s}, g_{i-3s}, g_{i-2s}, g_{i-s}, g_{i+s}, g_{i+2s}, g_{i+3s}, g_{i+4s}\}$ and g_i. In order to reduce the noise sensitivity of the gradient, the average gradient in the neighborhood region of the point is employed instead of the original gradient. The CP-LBP can be calculated as (4).

$$f_{2,i,s} = \sum_{j=1}^{8} 2^{j-1} sgn\left(G_{i,s}(j) - g_i\right) \tag{4}$$

where $f_{2,i,s}$ is the CP-LBP value at point P_i with sampling step s. $sgn(\cdot)$ is the sign function defined as (5).

$$sgn(x) = \begin{cases} 1; & if\ x \geq 0 \\ 0; & else \end{cases} \tag{5}$$

The schematic diagram of CP-LBP is shown in Fig. 5. The subsequence $Seq_{i,s}$ is shown as the left-up of Fig. 5 in which

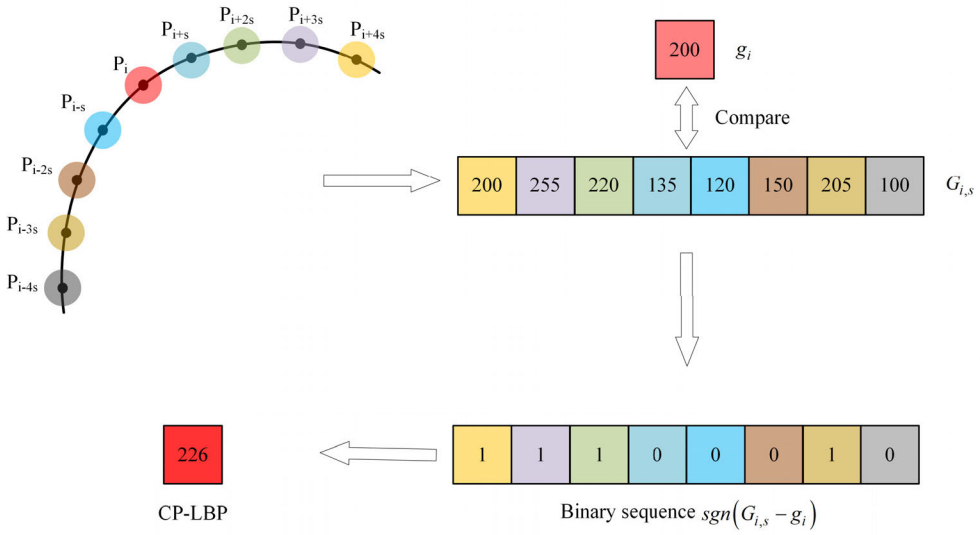

FIGURE 5. The LBP based on contour points.

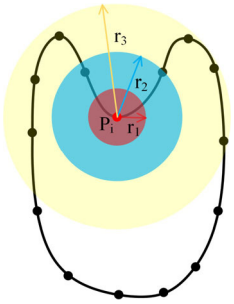

FIGURE 6. Zernike moments at multi-scales.

different contour points are marked with different colors while gradients of each point are depicted in the right-up of Fig. 5. The binary sequence obtained by comparing $G_{i,s}$ and g_i is shown in the lower-right. The final CP-LBP value can be achieved by converting the binary sequence to a decimal number which is marked with the red box.

C. ZERNIKE MOMENT

Zernike moment is a geometric feature of the shape with translation, rotation, and scale invariance. Therefore, Zernike moments of circular regions centered on the contour point with different radii are employed as features describing the geometric features at different scales as shown in Fig. 6. The Zernike moment of the red region represents the local feature of the shape at the small scale, while the yellow region indicates the semi-global to global feature depending on the radius. If the entire image is taken as the input, background pixels may affect the accuracy of Zernike moment to describe the shape. Therefore, the image region of the minimum enclosing rectangle of the shape is taken as the input instead of the original image, so as to avoid excessive background interference pixels and reduce the sensitivity to the background. Meanwhile, to decrease the influence of illumination variation, gray values of each region to be calculated are linearly changed with the maximum value setting to 255 and the minimum setting to 1. This adjustment could improve the robustness to both local and global illumination changes. To deal with the size of the calculation region at different scales, the region radius is set to be proportional to the original contour area. Zernike moment with small radius represents the local geometric feature while the large radius indicates the global geometric feature which could achieve the combination of local and global features with different scales. The Zernike moment at different scales can be calculated as (6).

$$f_{3,i,r} = Zernike_r\,(i) \tag{6}$$

where $f_{3,i,r}$ is the Zernike moment at contour point P_i with radius r.

After obtaining three multi-scale invariant features, the final feature at contour point P_i can be built by concatenating all three features as (7).

$$f_i = \left[f_{1,i} \mid f_{2,i} \mid f_{3,i}\right]^T \tag{7}$$

where $f_{1,i} = \{f_{1,i,k}, k = 1, 2, \cdots, \mathrm{N}\}$ is the feature sequence of contour angles at point P_i of different scales, $f_{2,i} = \{f_{2,i,s}, s = 1, 2, \cdots, \mathrm{N}\}$ is the CP-LBP sequence and $f_{3,i} = \{f_{3,i,r}, r = 1, 2, \cdots, \mathrm{N}\}$ represents the multi-scale Zernike moment sequence, N is the largest scale.

V. MATCHING ALGORITHM

A. SMITH-WATERMAN ALGORITHM

Contour-based matching algorithms generally alter the shape matching to the sequence matching which could be solved by dynamic programming (DP). The Smith-Waterman algorithm was proposed by Smith and Waterman [36] in 1981 to find similarity fragments in biological sequences. Different from DTW, SW is to search for the similar local segment in two sequences and has no constraint on the starting and ending point of two sequences.

Suppose $A = a_1, a_2, \cdots, a_m$ and $B = b_1, b_2, \cdots, b_n$ are two feature sequences, m and n are lengths of A and B separately. $f(a_i, b_j)$ is the similarity function of two feature points a_i and b_j, $gap(a_i, b_j)$ is the gap function to limit the occurrence of continuous gaps. A gap means that sequence A and B has not been matched correctly at the current position and the current position should be skipped. A gap is equivalent to the deletion or insertion in Editing distance. The score matrix H of sequence A and B can be calculated as (8).

$$H(i,0) = H(0,j) = 0$$
$$H(i,j) = max \begin{cases} H(i-1,j-1) + f(a_i, b_j) \\ H(i-1,j) - gap(a_i) \\ H(i,j-1) - gap(b_j) \\ 0 \end{cases} \quad (8)$$

In the score matrix H, any position with the score less than 0 means that the previous subsequence of the two sequences has no similarity, and the score is set to 0 to ensure the similarity score calculated from this position will not be affected by the previous sequence. After obtaining the scoring matrix, the fragment with the highest local similarity of the two sequences can be found by backtracking. Backtracking starts with the element with the highest score according to the source of the current element until encountering an element with score 0.

B. SIMILARITY FUNCTION

In the SW algorithm, the Euclidean distance between feature points is used as the similarity function in this paper.

$$f(a_i, b_j) = \begin{cases} L - SAD(a_i, b_j); & if\ SAD(a_i, b_j) < 0.1L \\ -SAD(a_i, b_j); & else \end{cases} \quad (9)$$

where L is the length of multi-scale features at the current feature point, and also indicates the maximum similarity between two feature points. $SAD(\cdot)$ is the sum of the absolute difference between two feature points.

Before calculating the similarity using the similarity function above, the two sequences need to be pre-processed to cope with different amplitudes of features. When calculating the similarity between two sequences, generally we only pay attention to the relative quantity of the two. Therefore, the maximum of each feature of two sequences can be used as normalization coefficients as shown in (10).

$$feature_{max} = max_{col}\begin{pmatrix} A \\ B \end{pmatrix}$$
$$A_{normalized} = A/feature_{max}$$
$$B_{normalized} = B/feature_{max} \quad (10)$$

where $max_{col}(\cdot)$ represents the maximum value of each column of the sequence consisting of A and B by column. The normalized features are between 0 and 1.

In order to limit the occurrence of continuous gaps, the gap function is defined as (11):

$$gap(a_i) = d + a(L-1) \quad (11)$$

(a)

A:		1	2	3	4	5	6	7	8	9	10
B:	10	1	2	3	4						

(b)

A:	5	6	7	8	9	10	1	2	3	4
B:						10	1	2	3	4

FIGURE 7. Matching results of SW and CSW. (a). Matching result of SW. (b). Matching result of CSW.

d and a are constants representing the initial penalty and continuous penalty, L is the length of gaps. It can be seen that the gap function mainly serves as a punishment for the similarity when a gap appears. As the gap appears continuously the penalty accumulates, thereby limiting the occurrence of continuous gaps which is usually considered to barely occur.

C. CONTOUR STARTING POINT SELECTION

How to choose the sequence starting point is one of the key problems in contour sequence matching. Different starting points may lead to very different similarity results, especially in local sequence matching. Zhao and Du [26] adopted the cyclic shift (CS) method to obtain the maximum similarity by continuously shifting the sequence. Although the cyclic shift method can acquire the maximum similarity, the computation is expensive. Chen *et al.* [34] used a method of partial copying of the sequence, but the length is not easy to determine. Based on the matching result in each step, we propose a cyclic SW (CSW) matching algorithm.

Suppose the correspondence obtained by SW between sequence A and B is shown in Fig. 7(a). Obviously, the resulting correspondence is not the optimal result because of different starting points although sequence B is a subsequence of sequence A. The similarity and correspondence of two sequences can be acquired when the matching is completed at each step. It is considered that no similar subsequence has been found for each subsequence after the last matching position, such as the subsequence [5]–[10] in A. This may be caused by different starting points or local matching. Therefore, sequence A is truncated at the last matching position, and the sequence segment after the truncation position is copied to the front end to form a new sequence. The new sequence will be re-matched with B until the similarity is no longer increased. As shown in Fig. 7(b), the proposed CSW can deal with matching similarity degradation caused by different starting points.

VI. EXPERIMENTAL RESULTS

A. PARAMETER SETTING

The ratio of the area of simplified contour to the original contour is used as the stopping criterion of IDCE. The evolution ends when the area ratio is less than 0.95. The number of contour sampling points of different shapes can hardly be guaranteed consistent, so the initial sequence length N is set to $L_{init} = max(0.08L, 3)$ in the evolutionary process experimentally, L is the length of the current contour, the length threshold M is set to $L_{init} + 3$.

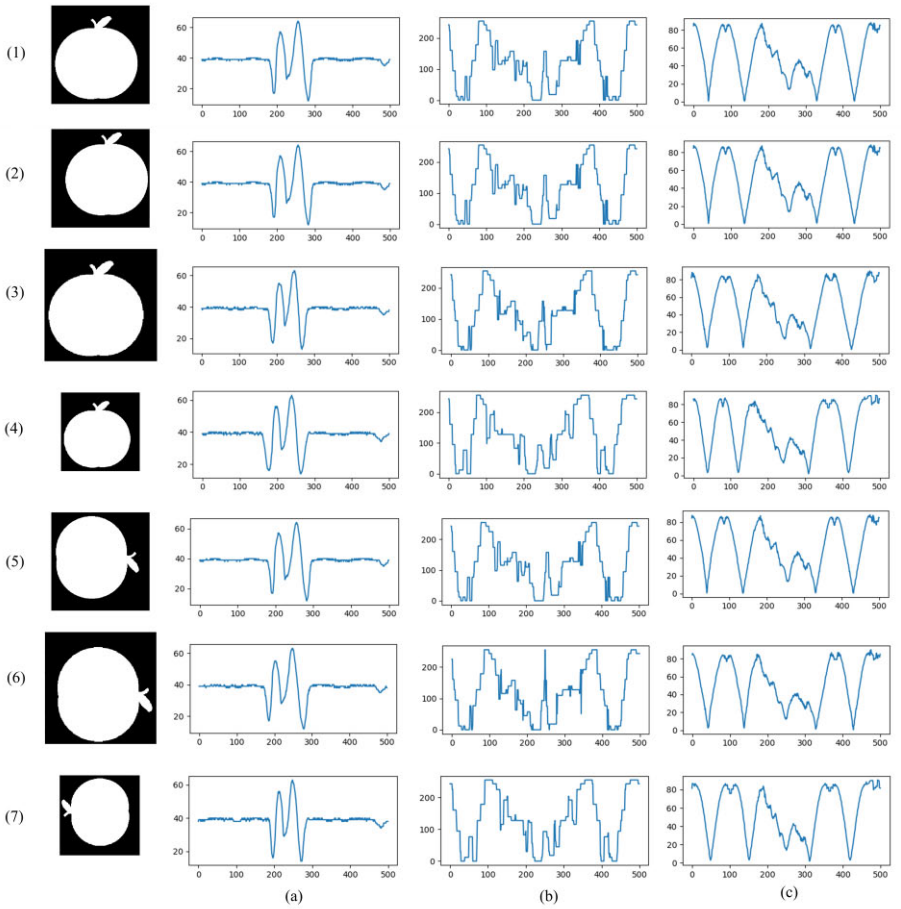

FIGURE 8. Invariance validation of the proposed features. (1) Features of the original image. (2) Features of the translated image. (3) Features of the enlarged image. (4) Features of the reduced image. (5) Features of the rotated image (90 degrees). (6) Features of the enlarged image with rotated 90 degrees and translation (7) Features of the reduced image with rotate −90 degrees and translation (a) Contour angle. (b) CP-LBP. (c) Zernike moment.

When calculating multi-scale features, the optimal sample steps or region radii are chosen experimentally as $k = 0.04L, 0.08L, 0.12L$, $s = 0.04L, 0.08L, 0.12L$, $r = 0.2A_0, 0.4A_0, 0.8A_0$, L is the length of the original contour, A_0 is the original contour area. The contour angles are quantized into 72 intervals.

B. INVARIANCE VALIDATION

For shape matching, we often hope that the designed feature descriptors can effectively express the object shape and earn a large intra-class similarity with a big inter-class difference. However, objects in real images usually suffer translation, rotation, scaling, and deformation which may cause a great impact on the matching results, especially when the object deforms. Therefore, the designed features need to be able to deal with the problem above with translation, rotation, and scaling invariance and have good robustness to the slight shape deformation.

The invariance of the proposed multi-scale features is illustrated in Fig. 8 and Fig. 9 in the case of translation, rotation, and scaling. Three features of the original image, translated image, rotated image and scaled image are shown from row 1 to row 5, while row 6 and row 7 are the features of image suffering all changes above. Comparison results show that the three multi-scale features proposed in this paper have high similarity under different conditions indicating that the proposed features are translation, rotation and scaling invariant.

The discriminative ability of the proposed features is presented in Fig. 10. The first row and second row belong to the apple class, the third row is a sample of watch which is similar to the apple, the fourth and fifth row are samples of classes different from the apple. It can be seen that feature curves of the first two rows are highly similar except for the difference of local feature caused by apple leaves. Although, the feature curves of row 3 show some local similarities with row 1 and row 2, they are still quite different on the whole and can be easily distinguished. Row 4 and row 5 are quite different from any one of the first two rows and do not show any similarity. Experimental results indicate that our proposed features can ensure a high similarity of the same class while distinguishing different classes, that is, small intra-class differences and large inter-class differences can be guaranteed.

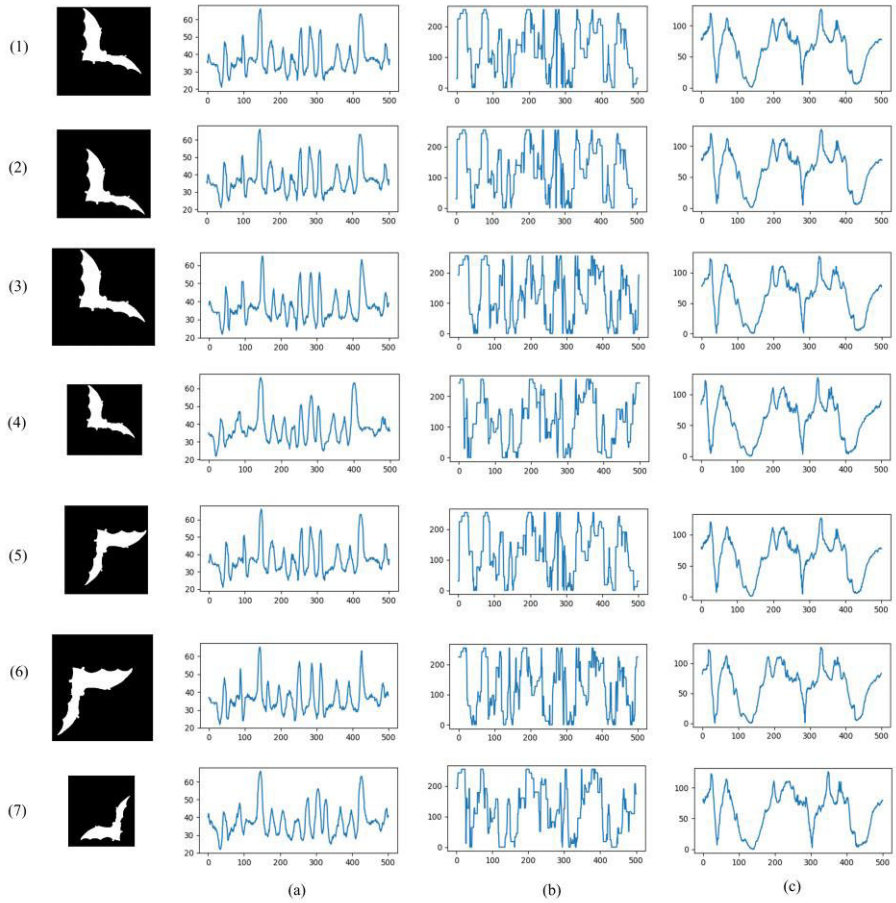

FIGURE 9. Invariance validation of the proposed features. (1) Features of the original image. (2) Features of the translated image. (3) Features of the enlarged image. (4) Features of the reduced image. (5) Features of the rotated image (90 degrees). (6) Features of the enlarged image with rotated 90 degrees and translation (7) Features of the reduced image with rotate −90 degrees and translation (a) Contour angle. (b) CP-LBP. (c) Zernike moment.

The matching results under shape deformation are given in Fig. 11. Shapes in Fig. 11(a) mainly suffer deformation and translation while shapes in Fig. 11(b) undergo deformation, rotation, translation and scaling change. The matching results show that the multi-scale features combined with the CSW algorithm proposed in this paper can obtain a well-matching performance even there exists a certain degree of deformation.

C. LOCAL CONTOUR MATCHING

In practice, objects often suffer occlusion making it impossible to obtain a complete contour. So, the proposed method should be able to deal with the matching of local contours. Another key problem of the matching algorithm of contour sequence is the selection of contour starting point. Different starting points tend to affect the matching results to a large extent resulting in completely different results.

Fig.12 demonstrates the matching results of complete contours with different starting points by the proposed features and CSW. Starting points of different contours are marked by blue diamond and red square. It can be seen from the result, in the first matching step, due to the different start points partial contour has not been matched, while it has been successfully matched after the contour starting point being adjusted according to the matching correspondence of the first step. Compared with the cyclic shift method, the proposed method only needs 2 steps to complete the shape matching with different starting points.

Fig. 13 gives the results of local contour matching with different starting points under three situations by our proposed method. The complete contour is depicted in blue, and the starting point is identified by a blue diamond, while the local contour is drawn in red, and the starting point is marked by a red square. Results show that the proposed algorithm can effectively solve the partial shape matching. At the same time, comparison of (b) and (c) shows that the CSW can adjust the starting point according to the correspondence of the first step, so as to effectively solve the problem of starting point selection.

The matching results under occlusion are shown in Fig. 14. In the case of occlusion, the shape contour is only partially

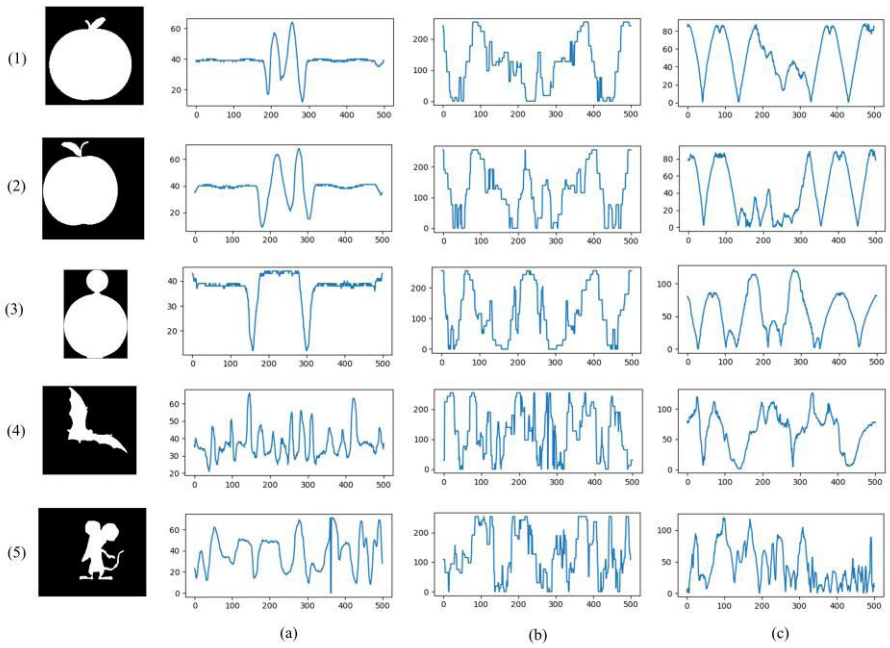

FIGURE 10. Verification results of the discriminative ability of the proposed features. (1) Features of apple1. (2) Features of apple2. (3) Features of the watch. (4) Features of the bat. (5) Features of the mouse. (a) Contour angle. (b) CP-LBP. (c) Zernike moment.

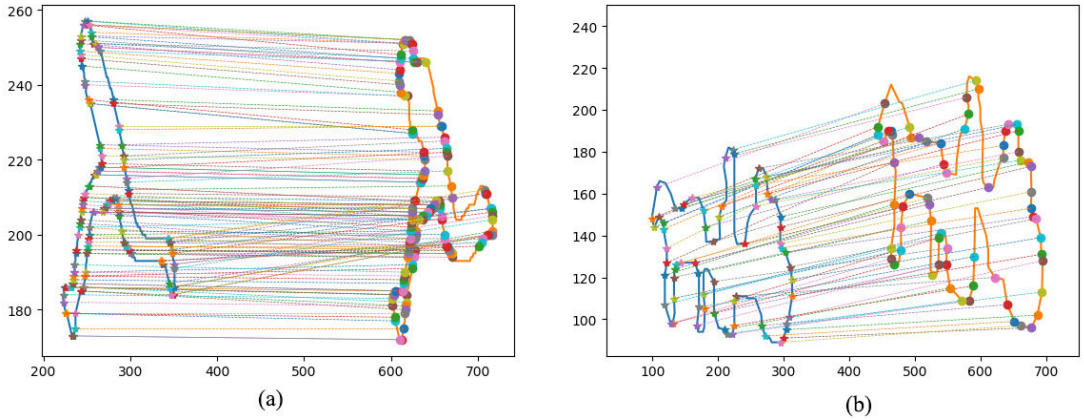

FIGURE 11. The matching result of deformed shapes (a) Matching result of shapes with deformation and translation. (b) Matching result of shapes with deformation, translation, rotation and scale change.

achieved causing the shape matching under occlusion to be turned into local contour matching. Experimental results show that the proposed method can solve the shape matching under occlusion. Meanwhile, it should be noted that since the features are described at multi-scales, features at large scale at the edge of occlusion part tend to be different from features of the original contour, so that the edge of occlusion part has not been properly matched.

D. TESTING ON BENCHMARK DATASETS

To verify the performance of our proposed method, two small datasets (Kimia99 and Kimia216) and a large dataset (MPEG-7) are used as validation sets in this section. Kimia [38] and MPEG-7 [39] are widely used benchmark datasets for evaluating shape matching and retrieval

The Kimia99 dataset consists of nine categories with 11 shapes in each for a total of 99 shapes. In this section, we calculate the similarity between all shapes in the dataset and the reference shape sequentially and rank the similarity of each reference shape. The evaluation method is to count the number of shapes in the 10 closest matches to each reference image which are in the same class. The retrieval results of different methods on Kimia99 are shown in TABLE 1. Results indicate that our proposed method can achieve a better retrieval accuracy on Kimia99.

The Kimia216 dataset contains 216 shapes consisting of 18 classes with 12 shapes in each. Similar to the evaluation

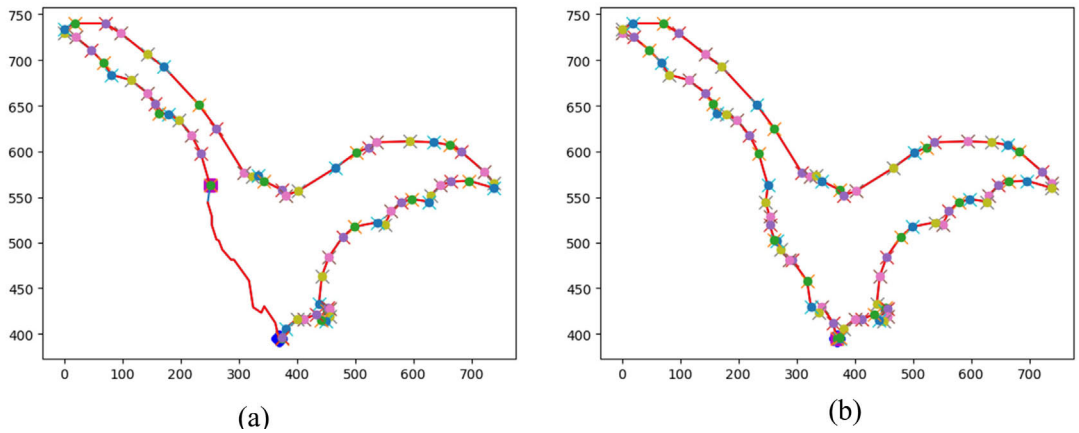

FIGURE 12. Complete contours matching results with different starting points. (a) The first evolution step. (b) The second evolution step.

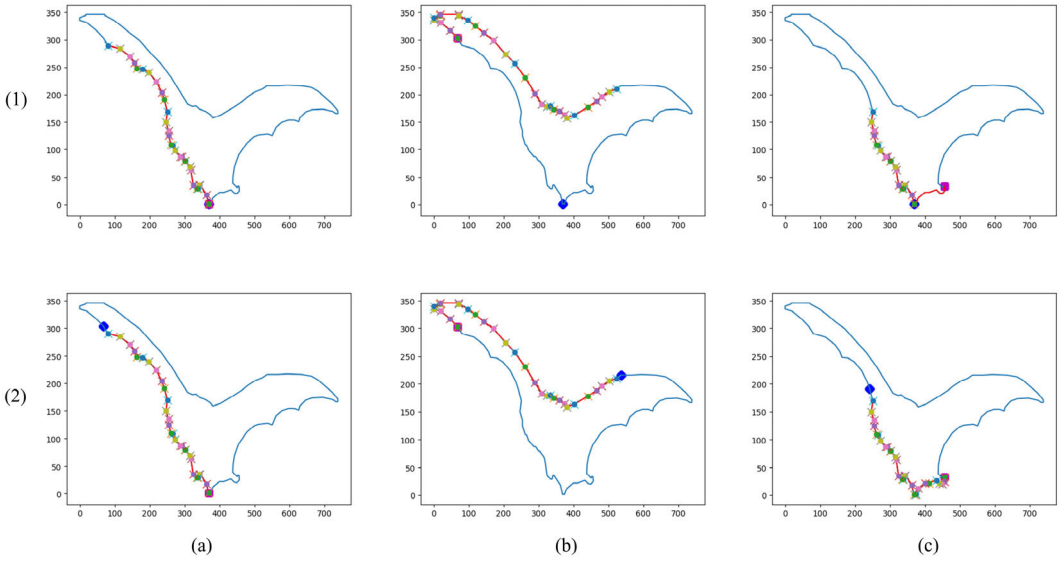

FIGURE 13. Local contour matching results with different starting points. (1) The first evolution step. (2) The second evolution step. (a) Situation1: The starting point of the local contour is the same as the starting point of the complete contour. (b) Situation2: The starting point of the local contour is after the starting point of the complete contour. (c) Situation3: The starting point of the local contour is before the starting point of the complete contour.

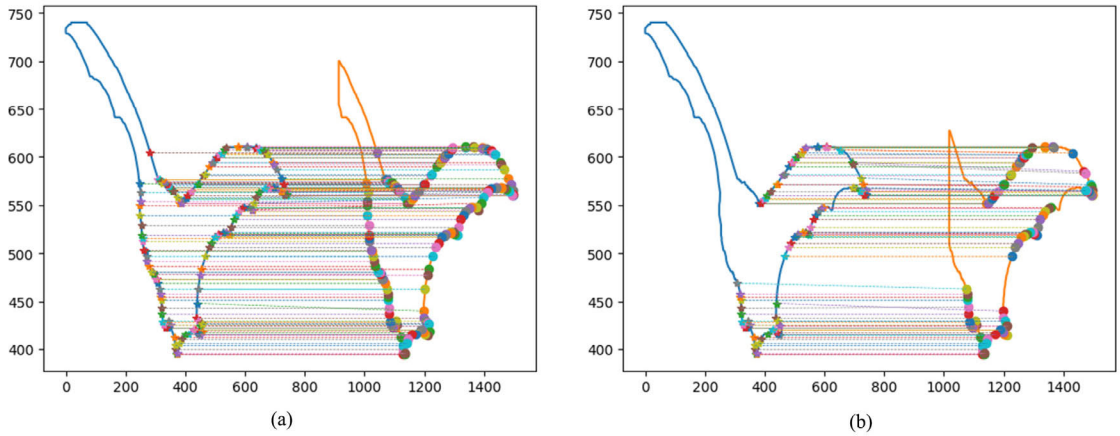

FIGURE 14. Matching results under different occlusion conditions.

TABLE 1. Retrieval results of different methods on Kimia99.

Methods	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
SC[6]	97	91	88	85	84	77	75	66	56	37
Path similarity[37]	99	99	99	99	96	97	95	93	89	73
TAR[20]	99	99	99	98	98	97	98	95	93	80
Multi-scale representation[2]	99	99	99	99	98	97	95	94	90	83
Our method	99	99	99	99	99	98	95	93	93	88

TABLE 2. Retrieval results of different methods on Kimia216.

Methods	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th
SC[6]	214	209	205	197	191	178	161	144	131	101	78
Path similarity[37]	216	216	215	216	213	210	210	207	205	191	177
Multi-scale representation[2]	216	216	214	210	207	207	201	194	188	182	163
Our method	216	216	215	215	210	213	209	207	205	201	190

method of Kimia99, it is to count the number of shapes in the 11 closest matches which are in the same class with the reference image. The retrieval results of different methods are given in TABLE 2. It can be seen that our proposed method can obtain a better performance on Kimia216 compared with other methods.

MPEG-7 dataset is a widely used benchmark dataset to evaluate the retrieval accuracy of similarity-based methods. MPEG-7 contains a total of 1400 shapes divided into 70 categories with 20 shapes each. The retrieval accuracy is determined by the bulls-eye rating in which each image is used as the reference and compared to all of the other shapes. The mean percentage of correct shapes in the top 40 matches is taken as bullseye rating. The bulls-eye rates of different methods are depicted in TABLE 3. Though the multi-scale features designed in this paper are all simple and intuitive shape descriptors, our proposed method could achieve a better performance, especially when compared with the method in [2] which also employs simple and intuitive multi-scale features.

TABLE 3. Retrieval accuracy of different methods on MPEG-7.

Methods	Accuracy (Bullseye)(%)
SC[6]	76.51
CSS[8]	75.44
MCC[17]	84.93
TAR[20]	87.23
Shape tree[40]	87.70
Contour flexibility[41]	89.31
Shape vocabulary[42]	90.41
Multi-scale representation[2]	91.25
Our method	92.27

TABLE 4. Comparison of time consumption of methods for calculating features.

method	SC	Multi-scale representation	our method
Time (s)	0.062	0.671	0.752

E. EFFICIENCY COMPARISON

To verify the efficiency of our proposed method, SC [6] and Multi-scale representation [2] (MSR) are used as comparisons. In order to ensure the validity of the verification, all three methods are implemented strictly according to the algorithm under python 3.6 without any optimization to avoid the influence of different optimization methods on the results. Two simple situations are discussed here. Situation 1 represents the matching of contour sequences with the same starting point in which sequence A is matched with itself, where sequence A is a sequence with 100 contour points. Situation 2 stands for the matching of sequences with different starting points in which sequence A is matched with sequence B, where B is the sequence with 10 points shift of A.

The comparison of time consumption of three methods for calculating features is shown in TABLE 4. Results indicate that multi-scale features can significantly increase the processing time to calculate features. TABLE 5 is the comparison results of situation 1. Mtime is the processing time of feature sequence matching only, while Ttime is the processing time from feature sequence calculation to matching completion. Accuracy is the ratio of the correctly matched points to all points. Our method takes a little bit longer because that CSW needs at least two matching cycles. TABLE 6 depicts the comparison results of situation 2. CS represents that cyclic shift is used in the dynamic programming (DP) to find the global optimal solution which stops when the accuracy reaches 100. Results indicate that SC and MSR can not match the contour sequence completely due to the different

TABLE 5. Comparison of time consumption of methods for feature matching (A, A).

Methods	Mtime(s)	Ttime(s)	Accuracy(%)
SC+DP	0.156	0.220	100
Multi-scale representation	0.171	0.843	100
our method	0.296	1.109	100

TABLE 6. Comparison of time consumption of methods for feature matching (A, B).

Methods	Mtime(s)	Ttime(s)	Accuracy (%)
SC+DP	0.156	0.220	90
Multi-scale representation	0.172	0.849	90
SC+DP+CS	1.234	1.312	100
Multi-scale representation + CS	1.234	1.921	100
our method	0.468	1.296	100

starting points. When CS is used, the accuracy can be guaranteed, however, the processing time is significantly increased. Compared with the first two methods with CS, the time consumption of our method is significantly reduced with only three matching cycles to achieve the global optimal results. When compared with results in TABLE 5, there is no significant increase in the processing time of our method. That to say, our method can complete the more complex cyclic matching problem with a lower time cost, thus effectively solving the problem of sequence matching with different starting points which is ubiquitous in shape matching based on contour points. It should be pointed out that the processing time of sequence matching in situation 1 can be reduced to about 10 milliseconds with proper optimization which means that the computational efficiency of our proposed method will be greatly improved by proper optimization.

VII. CONCLUSION

In this paper, a multi-scale contour-based shape matching method is proposed. Aiming at the deficiency of DCE, an improved DCE which combines DCE with the uniform sampling and achieves a better description of the shape is introduced. Three simple and intuitive multi-scale features are designed from aspects of the spatial relationship of contour points, structural information of contour sequence, and shape geometry feature. Also, for the problem of local contour matching and starting point selection, a cyclic SW algorithm is presented. Experimental results show that our proposed features are able to cope with situations of translation, rotation, scaling, and deformation. The proposed algorithm can solve the local contour matching and contour starting point selection without using cyclic shift algorithm. Retrieval accuracies of Kimia99, Kimai216, and MPEG-7 indicate that our method achieves a better performance.

ACKNOWLEDGMENT

The authors would like to acknowledge their help with information and financial support.

REFERENCES

[1] S. Mattoccia, F. Tombari, and L. Di Stefano, "Efficient template matching for multi-channel images," *Pattern Recognit. Lett.*, vol. 32, no. 5, pp. 694–700, Apr. 2011. doi: 10.1016/j.patrec.2010.12.004.
[2] J. Yang, H. Wang, J. Yuan, Y. Li, and J. Liu, "Invariant multi-scale descriptor for shape representation, matching and retrieval," *Comput. Vis. Image Understand.*, vol. 145, pp. 43–58, Apr. 2016. doi: 10.1016/j.cviu.2016.01.005.
[3] C. Yang, H. Wei, and Q. Yu, "A novel method for 2D nonrigid partial shape matching," *Neurocomputing*, vol. 275, pp. 1160–1176, Jan. 2018. doi: 10.1016/j.neucom.2017.09.067.
[4] T. Zhang, J. Li, W. Jia, J. Sun, and H. Yang, "Fast and robust occluded face detection in ATM surveillance," *Pattern Recognit. Lett.*, vol. 107, pp. 33–40, May 2018. doi: 10.1016/j.patrec.2017.09.011.
[5] G. Ghinea, R. Kannan, and S. Kannaiyan, "Gradient-orientation-based PCA subspace for novel face recognition," *IEEE Access*, vol. 2, pp. 914–920, 2014. doi: 10.1109/ACCESS.2014.2348018.
[6] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002. doi: 10.1109/34.993558.
[7] H. Ling and D. W. Jacobs, "Shape classification using the inner-distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 286–299, Feb. 2007. doi: 10.1109/TPAMI.2007.41.
[8] F. Mokhtarian, S. Abbasi, and J. Kittler, *Efficient and Robust Retrieval by Shape Content through Curvature Scale Space* (Series on Software Engineering and Knowledge Engineering), vol. 8. Singapore: World Scientific, 1998, pp. 51–58.
[9] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inf. Theory*, vol. 8, no. 2, pp. 179–187, Feb. 1962. doi: 10.1109/TIT.1962.1057692.
[10] A. Khotanzad and Y. H. Hong, "Invariant image recognition by Zernike moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 5, pp. 489–497, May 1990. doi: 10.1109/34.55109.
[11] F. Fraundorfer and H. Bischof, "Affine invariant region matching using geometric hashing of line structures," in *Proc. 27th Workshop Austral. Assoc Pattern Recognit.*, Laxenburg, Austria, 2003, pp. 57–64.
[12] Y. Hel-Or, H. Hel-Or, and E. David, "Matching by tone mapping: Photometric invariant template matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 317–330, Feb. 2014. doi: 10.1109/TPAMI.2013.138.
[13] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Sep. 1999, pp. 1150–1157.
[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.
[15] S. Manay, D. Cremers, B.-W. Hong, A. J. Yezzi, and S. Soatto, "Integral invariants for shape matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1602–1618, Oct. 2006. doi: 10.1109/TPAMI.2006.208.
[16] E. Attalla and P. Siy, "Robust shape similarity retrieval based on contour segmentation polygonal multiresolution and elastic matching," *Pattern Recognit.*, vol. 38, no. 12, pp. 2229–2241, 2005. doi: 10.1016/j.patcog.2005.02.009.
[17] T. Adamek and N. E. O'Connor, "A multiscale representation method for nonrigid shapes with a single closed contour," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 5, pp. 742–753, May 2004. doi: 10.1109/TCSVT.2004.826776.
[18] C. Yang and Q. Yu, "Multiscale Fourier descriptor based on triangular features for shape retrieval," *Signal Process., Image Commun.*, vol. 71, pp. 110–119, Feb. 2019. doi: 10.1016/j.image.2018.11.004.
[19] Q. Yu, H. Wei, and C. Yang, "Local part chamfer matching for shape-based object detection," *Pattern Recognit.*, vol. 65, pp. 82–96, May 2017. doi: 10.1016/j.patcog.2016.11.020.
[20] N. Alajlan, I. El Rube, M. S. Kamel, and G. Freeman, "Shape retrieval using triangle-area representation and dynamic space warping," *Pattern Recognit.*, vol. 40, no. 7, pp. 1911–1920, 2007. doi: 10.1016/j.patcog.2006.12.005.
[21] M. Donoser, H. Riemenschneider, and H. Bischof, "Efficient partial shape matching of outer contours," in *Computer Vision—ACCV*. Berlin, Germany: Springer, 2010, pp. 281–292.
[22] J. Jiao, X. Wang, Z. Deng, J. Cao, and W. Tang, "A fast template matching algorithm based on principal orientation difference," *Int. J. Adv. Robotic Syst.*, vol. 15, no. 3, 2018, Art. no. 1729881418778223. 10.1177/1729881418778223.

[23] H. Wei, C. Yang, and Q. Yu, "Efficient graph-based search for object detection," *Inf. Sci.*, vols. 385–386, pp. 395–414, Apr. 2017. doi: 10.1016/j.ins.2016.12.039.
[24] J. Yoo, S. S. Hwang, S. D. Kim, M. S. Ki, and J. Cha, "Scale-invariant template matching using histogram of dominant gradients," *Pattern Recognit.*, vol. 47, no. 9, pp. 3006–3018, 2014. doi: 10.1016/j.patcog.2014.02.016.
[25] E. Klassen, A. Srivastava, W. Mio, and S. H. Joshi, "Analysis of planar shapes using geodesic paths on shape spaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 3, pp. 372–383, Mar. 2004. doi: 10.1109/TPAMI.2004.1262333.
[26] D. Zhao and F. Du, "A novel approach for scale and rotation adaptive estimation based on time series alignment," *Vis. Comput.*, vol. 2018, pp. 1–5, Sep. 2018. doi: 10.1007/s00371-018-1598-3.
[27] N. Kaothanthong, J. Chun, and T. Tokuyama, "Distance interior ratio: A new shape signature for 2D shape retrieval," *Pattern Recognit. Lett.*, vol. 78, pp. 14–21, Jul. 2016. doi: 10.1016/j.patrec.2016.03.029.
[28] W. Deng, H. Zou, F. Guo, L. Lei, and S. Zhou, "Point-pattern matching based on point pair local topology and probabilistic relaxation labeling," *Vis. Comput.*, vol. 34, no. 1, pp. 55–65, 2018. doi: 10.1007/s00371-016-1311-3.
[29] Z. Tu and A. Yuille, "Shape matching and recognition—Using generative models and informative features," in *Proc. Eur. Conf. Comput. Vis.*, in Lecture Notes in Computer Science, vol. 3, 2004, pp. 195–209.
[30] H. A. Atabay, "Binary shape classification using convolutional neural networks," *IIOAB J.*, vol. 7, pp. 332–336, Oct. 2016.
[31] C. Li and B. Yang, "Adaptive weighted CNN features integration for correlation filter tracking," *IEEE Access*, vol. 7, pp. 76416–76427, 2019. doi: 10.1109/ACCESS.2019.2922494.
[32] J.-C. Yoo and T. H. Han, "Fast normalized cross-correlation," *Circuits, Syst. Signal Process.*, vol. 28, pp. 819–843, Dec. 2009. doi: 10.1007/s00034-009-9130-7.
[33] J. Zhao and L. Itti, "shapeDTW: Shape dynamic time warping," *Pattern Recognit.*, vol. 74, pp. 171–184, Feb. 2018. doi: 10.1016/j.patcog.2017.09.020.
[34] L. Chen, R. Feris and M. Turk, "Efficient partial shape matching using Smith-Waterman algorithm," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Anchorage, AK, USA, Jun. 2008, pp. 1–6.
[35] L. J. Latecki, V. Megalooikonomou, Q. Wang, and D. Yu, "An elastic partial shape matching technique," *Pattern Recognit.*, vol. 40, no. 11, pp. 3069–3080, 2007. doi: 10.1016/j.patcog.2007.03.004.
[36] T. F. Smith and M. S. Waterman, "Identification of common molecular subsequences," *J. Molecular Biol.*, vol. 147, no. 1, pp. 195–197, 1981. doi: 10.1016/0022-2836(81)90087-5.
[37] X. Bai and L. J. Latecki, "Path similarity skeleton graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1282–1292, Jul. 2008. doi: 10.1109/TPAMI.2007.70769.
[38] T. B. Sebastian, P. N. Klein, and B. B. Kimia, "Recognition of shapes by editing their shock graphs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 550–571, May 2004. doi: 10.1109/TPAMI.2004.1273924.
[39] L. J. Latecki, R. Lakamper, and T. Eckhardt, "Shape descriptors for non-rigid shapes with a single closed contour," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2000, pp. 424–429.
[40] P. F. Felzenszwalb and J. D. Schwartz, "Hierarchical matching of deformable shapes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[41] C. Xu, J. Liu, and X. Tang, "2D shape matching by contour flexibility," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 180–186, Jan. 2009. doi: 10.1109/TPAMI.2008.199.
[42] X. Bai, C. Rao, and X. Wang, "Shape vocabulary: A robust and efficient shape representation for shape matching," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3935–3949, Sep. 2014. doi: 10.1109/TIP.2014.2336542.

ZHANG KUN received the B.S. degree from Northwestern Polytechnical University, Xi'an, China, in 2012, where he is currently pursuing the Ph.D. degree in flight vehicle design with the School of Astronautics. His current research interests include image guidance, flight dynamics, image process, and target recognition.

MA XIAO received the B.S., M.S., and Ph.D. degrees from Northwestern Polytechnical University, Xi'an, China, in 2012, 2015, and 2019, respectively. He is currently a Researcher with the China Academy of Engineering Physics. His current research interests include flight dynamics, image guidance, and impact dynamics.

LI XINGUO received the B.S., M.S., and Ph.D. degrees from the School of Astronautics, Northwestern Polytechnical University, Xi'an, China, in 1987, 1990, and 1997, respectively, where he is currently a Professor with the School of Astronautics. His current research interests include flight dynamics, reentry guidance and control, image guidance, and image process.