Fourier Transform to Group Feature on Generated Coarser Contours for Fast 2D Shape Matching

YAN ZHENG[1,2], FANJIE MENG[1], JIE LIU[2], BAOLONG GUO[1], YANG SONG[3], XUEBING ZHANG[4], AND LING WANG[2]

[1]Institute of Intelligent Control and Image Engineering, Xidian University, Xi'an 710071, China
[2]Department of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China
[3]School of Computer Science and Engineering, University of New South Wales, Sydney, NSW 2052, Australia
[4]College of Geology and Mining Engineering, Xinjiang University, Urumqi 830049, China

Corresponding author: Baolong Guo (1017708350@qq.com)

ABSTRACT Fourier descriptors are classical global shape descriptors with high matching speed but low accuracy. To obtain higher accuracy, a novel framework for forming Fourier descriptors is proposed and named as MSFDGF (multiscale Fourier descriptor using group feature). MSFDGF achieves multiscale description by generating coarser contours. Then, a group of complementary features are extracted on the generated coarser contours. Finally, Fourier transform is performed on the features. MSFDGF-SH is a new global descriptor using the MSFDGF framework and shape histograms. Experiments are conducted on four databases, which are MPEG-7 CE-1 Part B, Swedish Plant Leaf, Kimia 99 and Expanded Articulated Database, to evaluate the performance of MSFDGF-SH. The experimental results show that MSFDGF-SH is an effective and efficient global shape descriptor. This new descriptor has a high accuracy of 87.76%, which exceeds the Shape Tree on the MPEG-7 CE-1 Part B dataset. This is the first Fourier descriptor that surpasses the Shape Tree method in terms of both accuracy and speed on this dataset.

INDEX TERMS Fast 2D shape retrieval, fast shape matching, MPEG-7 CE-1 Part B, shape descriptor.

I. INTRODUCTION

Shape is an important feature in plant leaf retrieval [1], trademark retrieval [2] and object recognition in blurred images. Shape descriptor is an important tool for extracting shape features of objects in 2D images.

Although the researches on post-processing methods [3]–[13] in the field of shape retrieval have been extensive for years, many scholars are still working on designing better shape descriptors because they can provide the original dissimilarity/similarity between shapes, which is the basis of shape matching. A ineffective shape descriptor cannot obtain high accuracy in shape retrieval, no matter how advanced a post-processing method is combined with. Therefore, the study of shape descriptors has never stopped

and a large number of excellent descriptors [14]–[34] have been proposed.

These local descriptors, such as SC (shape context) [35], IDSC (inner-distance shape context) [36], TAR (triangle-area representation) [37] and Shape Tree [38], have achieved highly accurate experimental results on the MPEG-7 CE-1 Part B shape database, but they all perform poorly in terms of matching efficiency. The matching efficiency of global descriptors MDM (multiscale distance matrix) [39], FD (Fourier descriptor) [40] and WD (wavelet descriptor) [41], is very high, but their performance in terms of accuracy is poor. The shape descriptor AP&BAP (angular pattern and binary angular pattern) [42] has thus been proposed to achieve both matching accuracy and efficiency with multiscale description and efficient distance metrics.

Inspired by AP&BAP, researchers then focus more on designing gloabal descriptors that are efficient in the matching process. However, the design of this type of descriptors

is extremely difficult. Among new global descriptors only HSC (hierarchical string cuts) [43] is at the same level of discriminability as AP&BAP.

To this end, a novel framework for forming global Fourier descriptors and a new descriptor based on this framework are proposed in our study. Using the Fourier transform, we proposed various approaches to improve the discriminability of the global descriptor as much as possible. These approaches include constructing multiple scales and improving structure of the spatial signature. The performance of this descriptor on accuracy and speed is as good as that of the excellent global descriptors HSC [43] and AP&BAP [42]. It performs even better than Shape Tree [38] on MPEG-7 CE-1 Part B in terms of both accuracy and speed.

The rest of this paper is organized as follows: Section 2 discusses related work. In Section 3, the new framework and descriptor are introduced in detail. In Section 4, the computational complexity of the proposed method is estimated. In Section 5, a number of well-known databases are used to evaluate the performance of the new method, in terms of effectiveness and efficiency. Finally, Section 6 concludes the paper.

II. RELATED WORK

In the last fifteen years, shape representation methods based on contour sampling points have developed much faster than the area-based ones. Usually, a contour is a set of uniformly sampled points on the outline of a shape. In this section, these contour-based methods are discussed in detail.

A. LOCAL DESCRIPTORS

The describing process of most local descriptors is to calculate a feature for each contour point or segment. This feature is typically a vector or matrix. Ignoring the relative order between features, a shape is described as a feature set. The matching process of local descriptors is to find the best correspondence between two sets of elements (features). The optimization algorithm is used to find the optimal correspondence between the two sets of elements, and the matching cost under the optimal correspondence is the dissimilarity (distance) between two shapes.

The SC [35] has been one of the most important descriptors in the field of shape matching. It sets each point in the contour as a reference point in turn, then calculates the distance of other points relative to the reference point, and builds a shape histogram (distance histogram) to describe the corresponding reference point. Finally, N (the number of contour sampling points) shape histograms are obtained. These shape histograms are put together to form a set (SC feature), which describes the shape. The matching process of two shapes is to compute the distance between their SC features. Therefore, shape matching becomes matching of two sets of shape histograms. The χ^2 distance is used to measure the difference between two histograms. Matching two sets of shape histograms is to calculate the minimum sum of χ^2 distances between two sets of shape histograms.

Finally, the minimum sum (matching cost) is the dissimilarity between two shapes. Dynamic programming [3] can be used in the process to find the optimal correspondence between two sets of shape histograms.

SC has an enhanced version IDSC [36], which performs better than the original SC in representing articulated shapes. The IDSC uses inner-distance instead of Euclidean distance used in the original SC when describing the relationship between two contour points. Inner-distance refers to the shortest path connecting two points inside the shape. Another major advantage of IDSC is its strong compatibility. Many post-processing algorithms based on learning [3]–[9] use IDSC to obtain the matching results between shapes as basis of learning.

The SC and IDSC are two important local descriptors, as they all achieve high retrieval rates on MPEG-7 CE-1 Part B and they are complementary to obtain higher retrieval rates. The complementarity between IDSC and SC is described in detail in [4]. However, SC and IDSC still have shortcomings. In terms of practice, they run too slowly to meet the practical requirements as they all use DP (dynamic programming).

In [38], Felzenszwalb *et al.* describe a hierarchical representation for shapes that captures shape information at multiple levels of resolution. Usually this method is called Shape Tree, and it achieves very high retrieval rate (87.70%) on MPEG-7 CE1 Part B. This high retrieval rate had not been surpassed by Fourier descriptors before our method was proposed. Overall most local descriptors with DP perform effectively in terms of accuracy. However, all of them have high computational costs.

B. GLOBAL DESCRIPTORS

Generally, in a global descriptor, a shape is represented by a feature vector (or matrix) extracted from the whole contour, and matching is conducted by comparing such representation vectors (or matrices) [42]. In the matching process, a global descriptor is suitable for using efficient distance metrics such as Euclidean distance and city block distance.

FDs (Fourier descriptors) are classical global descriptors. FD-CCD (Fourier descriptor based on centroid contour distance) [40] is taken as an example to introduce the characteristics of such descriptors. The Euclidean distance from each contour point to the centroid point is put into a sequence in order. Then, Fourier transform is used on the Euclidean distance sequence and the transformed result is the FD-CCD feature. The dissimilarity is the city block distance between two FD-CCD features belonging to two shapes respectively. WDs (wavelet descriptors) [41], [44] are also global descriptors, and they have similar effectiveness and efficiency as FD-CCD.

MDM (multiscale distance matrix) [39], which captures the shape geometry while being invariant to translation, rotation, scaling, and bilateral symmetry, is an important shape descriptor in global descriptors. It combined multiscale

description and distance metrics to achieve high efficiency and effectiveness in plant leaf retrieval. However, unfortunately, MDM's discriminability is limited, and its performance on some important databases, such as the MPEG-7 CE-1 Part B, does not reach the level of local descriptors in terms of accuracy.

In [42], Hu *et al.* propose two novel shape features, AP (angular pattern) and BAP (binary angular pattern), and a multiscale integration of them (AP&BAP) for shape matching. AP&BAP is a much significant descriptor, which allows many scholars to see the hope of the global descriptor in terms of accuracy. The previous FD [40] and MDM [39] are still inaccurate, relative to the local descriptors. Yet, AP&BAP is both fast and accurate. The retrieval rate of AP&BAP with χ^2 distance (87.04%) even surpassed SC+DP (86.80%) and IDSC+DP (85.40%) on MPEG-7 CE-1 Part B shape database. In terms of speed, AP&BAP continues the advantages of the global descriptor.

The HSC (hierarchical string cuts) [43] method is proposed to partition a shape into multiple level curve segments of different lengths from a point moving around the contour to describe the shape gradually and completely from the global information to the finest details. HSC continues the great breakthrough of global descriptors. In the experiments, it gets a higher retrieval rate (87.31%) than AP&BAP (87.04%), on the MPEG-7 CE-1 Part B shape database with a faster speed.

Kaothanthong *et al.* [45] propose a shape signature named DIR (distance interior ratio) that utilizes intersection pattern of the distribution of line segments with the shape, and a histogram alignment method for adjusting the interval of the histogram according to the distance distribution. DIR is a recent attempt at global descriptors. Its speed is as fast as FD, which is faster than HSC. Its retrieval rate is 10% higher than FD on MPEG-7 CE-1 Part B shape database. However, it still does not reach the level of AP&BAP and HSC in terms of discriminability. This result shows how difficult it is to design an effective global descriptor.

III. METHOD

A. GROUP FEATURE

Fourier transform is a commonly used technique in fast shape matching. It usually transforms a spatial feature vector of a shape into a sequence of coefficients in the frequency domain. An element of the spatial feature vector is determined by its corresponding contour point. These elements are arranged in the same order as the contour points in the closed contour. An obvious problem is that different starting point positions in the closed contour will result in different spatial feature vectors. In the Fourier transform, the operation of abandoning phase information solves this problem (see Eq. (1)). This special Fourier transform can get the starting point position invariance, thus it avoids the computation process of finding the best starting point. This is the reason why the Fourier

transform technique is widely used for fast shape matching.

$$f(k) = \frac{1}{N} \left| \sum_{i=0}^{N-1} v_s(i_d) exp(\frac{-j 2\pi i_d k}{N}) \right|,$$
$$k = 0, 1, \ldots, N-1 \quad (1)$$

where v_s is the spatial feature vector, N is the length of the vector and f is the output frequency sequence.

With the uniqueness of transforming result of Eq. (1), distance metric is used to measure the difference between frequency coefficients sequences of two shapes. This matching process has very low computational consumption. But the discriminability of most previous Fourier descriptors is not so effective. This is mainly because the previous Fourier descriptors often rely on a single spatial feature, such as FD-CCD (Fourier descriptor based on the centroid contour distance) [40] and FD-FPD (Fourier descriptor based on the furthest point distance) [46]. Therefore, it is a feasible method to improve the discriminability of the Fourier descriptor that more and better spatial features are used together in transforming.

Such a combination of spatial features is called Group Feature (GF), and the GF-based Fourier descriptor is named as FDGF. For example, CCD and FPD can form a GF, which is named GF-CCD&FPD. The experiments can prove that the discriminability of FDGF-CCD&FPD is better than FD-CCD and FD-FPD, but still can not reach the level of a local descriptor (such as TAR [37]). In terms of discriminability, in order to reach the level of local descriptors, it is necessary to design an effective GF. The combination of CCD and FPD is a simple GF containing two feature units. An ideal GF should contain some feature units, which are highly complementary and lowly correlated. These feature units preferably have the ability to describe any contour points in various shapes. This is like the orthogonal basis in an Euclidean space.

B. BIN VECTOR

Shape histogram can be used as a GF, although it always appears in local descriptors [35], [36]. A histogram is rarely used in FD, probably because scholars are accustomed to use a single spatial feature vector. The histogram $h_s^{i_d}$ describes the distribution of the remaining points in the contour c_d relative to $c_d(i_d)$.

$$h_s^{i_d}(i_b) = \#\{c_d(j_d) : j_d \neq i_d, c_d(j_d) - c_d(i_d) \in bin(i_b)\},$$
$$i_d, j_d = 1, 2, \ldots, N, i_b = 1, 2, \ldots, B \quad (2)$$

where $h_s^{i_d}$ is the shape histogram of the $i_d th$ contour point $c_d(i_d)$. $h_s^{i_d}(i_b)$ is the value of the bth bin in $h_s^{i_d}$. These B bins uniformly divide the log-polar plane centered on $c_d(i_d)$. c_d is a contour represented by a sampling points sequence. $c_d(i_d), i_d \in Z$, is the $i_d th$ point in the contour c_d that has N sampling points. Since the contour is closed, $c_d(i_d) = c_d(i_d + N)$. $c_d(i_d - 1)$ and $c_d(i_d + 1)$ are two adjacent points of $c_d(i_d)$ on the contour. Therefore, shape histogram is a set of feature units that describe a contour point.

The feature units in this set are highly complementary and lowly correlated. In other words, the shape histogram is an excellent GF.

The Fourier descriptor based on the GF of shape histogram is called FDGF-SH. In the process of extracting the FDGF-SH feature, a new feature Bin Vector (BV) is required as Fourier transform cannot deal with N shape histograms directly. $v_b^{i_b}$ (a BV) is a column vector generated by Eq. (3). BV is the key that is used to transform a local feature to a global feature.

$$v_b^{i_b}(i_d) = h_s^{i_d}(i_b), \quad i_d = 1, 2, \ldots, N, \ i_b = 1, 2, \ldots, B \quad (3)$$

Setting $v_b^{i_b}$ as v_s, f^{i_b} is generated by Eq. (1). This Fourier coefficient sequence is still a column vector, just like the previous Fourier descriptor. Subsequently, $f^1, f^2, \ldots, f^B$ are used to form a feature matrix F.

$$F = \left[f^1, f^2, \ldots, f^B \right] \quad (4)$$

Since the lower frequency components are more stable than higher frequency ones in FDGF-SH, only a few low frequency coefficients are used in the matching process with weighted city block distance. The weighted city block distance between two contours C_1 and C_2 in the FDGF-SH feature space are represented as $D(C_1, C_2)$ in

$$D(C_1, C_2) = \sum_{i_b=1}^{B} \sum_{k=0}^{K} w_k \left| f_{C_1}^{i_b}(k) - f_{C_2}^{i_b}(k) \right| \quad (5)$$

where $f_{C_1}^{i_b}$ and $f_{C_2}^{i_b}$ are the f^{i_b} sequence of C_1 and C_2. In Eq. (5) K is far smaller than N. Usually, $K < 3log_2N$.

FDGF-SH can surpass local descriptors easily in terms of efficiency. However, in terms of accuracy, FDGF-SH may be still not at the level of local descriptors. Therefore, FDGF-SH should continue to be improved.

C. FEATURE ON A GENERATED COARSER CONTOUR

Human eyes sometimes automatically filter some local details and preserve coarser features to reduce interference caused by noises, when recognizing shapes. This approach can be used in the design of shape descriptors. In our study, three approaches for obtaining variable-level coarser contours are proposed.

1) MEDIAN FILTERING TO A CONTOUR

Median filtering is used to generate the level t coarser contour. It is characterized by a linear increase of filtering core scale as t increases. The FDGF-SH feature on level t coarser contour is represented as FDGF-SH-MFtC in this median filtering approach. The level t coarser contour is generated with

$$c_{mf}^t(i) = \frac{1}{t+1}(\sum_{i_0=i}^{i+t} c_0(i_0)), \quad i = 1, 2, \ldots, N,$$
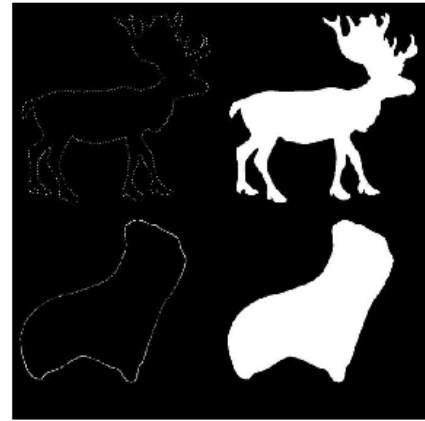$$t = 0, 1, 2, \ldots, N-1 \quad (6)$$

FIGURE 1. c_{mf}^t and its corresponding shape when $t = 0$ or $N/4$. c_0/c_{mf}^0 are in the first row. $c_{mf}^{N/4}$ and its corresponding shape are in the second row.

where the level 0 coarser contour c_0 is the original contour containing N sampling points. It can be seen that the number of sampling points for the level t coarser contour c_{mf}^t is still N. FDGF-SH-MFtC feature is extracted from contour c_{mf}^t using FDGF-SH method. The weighted city block distance is still used in the matching process of FDGF-SH-MFtC. $D_{MFtC}(C_1, C_2)$ is the distance (dissimilarity) in FDGF-SH-MFtC feature space between two shapes C_1 and C_2.

In practice, when t is large, for example $t = N/4$, the adjacent points in c_{mf}^t may overlap each other. Even if they don't overlap, sampling points are excessive as shown in Fig. 1. A large number of sampling points are used for describing a very simple shape, and they are nonuniformly distributed as shown in Fig. 1. It is not conducive to reduce the computional cost of the large scale features in extracting and matching processes. The advantage is that the variation of coarser contours between adjacent levels is small, so there are many levels to use. Enough levels makes it is easy to find a tacit combination of levels for more effective description. Theoretically, $0 \leq t \leq N - 1$ and $t \in Z$. When $t = 0$, c_{mf}^t is still c_0. When $t = N - 1$, c_{mf}^t is only one point. The shapes at different scales are shown in Fig. 2.

2) DOWNSAMPLING TO A CONTOUR

In this downsampling version N is required to be in the power of 2. The number of points in the downsampled contour decreases exponentially, as t increases. In this approach, there are fewer ($log_2 N + 1$) levels of coarser contours. The level t coarser contour using downsampling approach is generated by

$$c_{ds}^{t,n_0}(i) = \frac{1}{2^t} \sum_{i_0=(i-1)\cdot 2^t+n_0}^{i\cdot 2^t+n_0-1} c_0(i_0), \quad t = 0, 1, 2, \ldots, log_2N,$$
$$n_0 = 1, 2, 3, \ldots, 2^t, \ i = 1, 2, \ldots, N/2^t \quad (7)$$

FIGURE 2. The shapes of c_{mf}^t when $N = 512$ and $t = 0, 4, 8, \ldots, 116$ respectively. The shapes in the first row are corresponding to c_0, c_{mf}^4, c_{mf}^8, c_{mf}^{12} and c_{mf}^{16} from left to right respectively. The shapes in the second row are corresponding to c_{mf}^{20}, c_{mf}^{24}, c_{mf}^{28}, c_{mf}^{32} and c_{mf}^{36} in the same order as the first row. Other rows are in the same rules as the first and second rows.

FIGURE 3. The shapes of $c_{ds}^{1,1}$ and $c_{ds}^{1,2}$.

The physical meaning of Eq. (7) is to downsample 2^t points, which are consecutive from the $((i-1) \cdot 2^t + n_0)th$ point in c_0, to form a new point $c_{ds}^{t,n_0}(i)$. However, a new problem has arisen. Though when t is small shapes of c_{ds}^{t,n_0} with different n_0 are so similar (see Fig. 3), when t is larger, the geometric features of the contour c_{ds}^{t,n_0} are influenced by what the variable n_0 is.

When $t = 6$ and $n_0 = 4, 8, 12, \ldots, 2^t$, the shapes of c_{ds}^{t,n_0} are shown in Fig. 4. It is clear that the difference between them is large. The difference makes the FDGF feature of c_{ds}^{t,n_0} unstable. There is a solution for this problem. At first, each c_{ds}^{t,n_0} is computed out for $n_0 = 1, 2, 3, \ldots, 2^t$ respectively. Then, the shape histograms of these contours are used to calculate the average shape histogram $\bar{h}_s^{i_d}$ with

$$\bar{h}_s^{i_d}(b) = \frac{1}{2^t} \sum_{n_0=1}^{2^t} h_s^{i_d,n_0}(b) \tag{8}$$

where $h_s^{i_d,n_0}$ is the shape histogram of $c_{ds}^{t,n_0}(i_d)$ in the contour c_{ds}^{t,n_0}. Finally, the FDGF-SH-DStC feature is calculated base on the set of these average shape histograms. The weighted city block distance is still used in the matching process. The distance between two shapes C_1 and C_2 in the FDGF-SH-DStC feature space is expressed as $D_{DStC}(C_1, C_2)$.

The number of contour sampling points at each level is different from each other, the larger the t, the fewer the contour sampling points. In addition, there are fewer levels
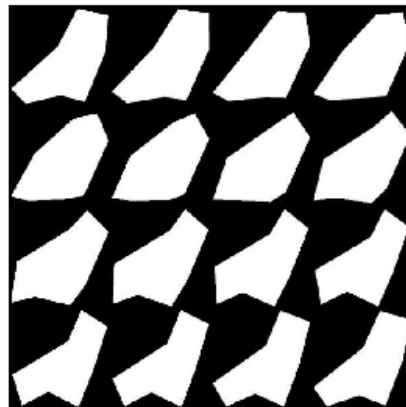

FIGURE 4. The shapes of c_{ds}^{6,n_0} when $n_0 = 4, 8, 12, \ldots, 2^6$ respectively. These shapes are arranged in the same rule as in Fig. 2.

FIGURE 5. The shapes of c_{ds}^{t,n_0} when $n_0 = 1$, $N = 512$ and $t = 0, 1, 2, \ldots, 8$ respectively. These shapes are arranged in the same rule as in Fig. 2.

in this downsampling approach than in FSCtS-mf. The above characteristics make the FSCtS-ds more favorable for acceleration.

Theoretically, in downsampling approach, $0 \le t \le \log_2 N$ and $t \in Z$. When $t = 0$, c_{ds}^{t,n_0} is still c_0. When $t = \log_2 N - 1$, c_{ds}^{t,n_0} is a line segment. When $t = \log_2 N$, c_{ds}^{t,n_0} is a point. c_{ds}^{t,n_0} at each level is shown in Fig. 5.

3) SPATIAL FILTERING TO A SHAPE

Inspired by [5], spatial filtering is also incorporated. Closing operation, which is defined as a dilate operation followed by an erosion operation using the same SE (structuring element), is used to generate coarser contours, as it can reduce some finer features and reserve coarser features of a shape. In this spatial filtering approach, the FDGF-SH feature of the level t coarser contour is represented as FDGF-SH-SFtC. To generate the level t coarser contour the original image is processed by a SE of level t size. The 'disk' SE is used in this approach.

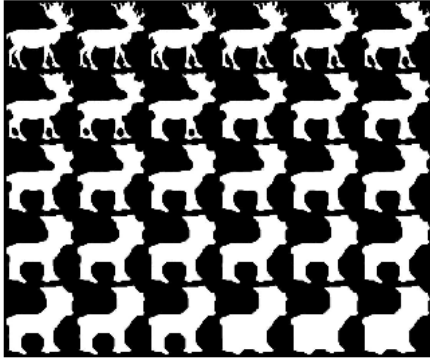$$im^t = im \cdot se^t = (im \oplus se^t) \ominus se^t \tag{9}$$

FIGURE 6. The shapes of c_{sf}^t when $t = 1, 2, 3, \ldots, 30$ respectively. The size of SE at level t equals $4t$. These shapes are arranged in the same rule as in Fig. 2.

The level t coarser shape is extracted from im^t. Then, N sampling points are uniformly extracted from the contour of the level t shape to form c_{sf}^t. The FDGF-SH-SFtC feature is extracted from c_{sf}^t. The weighted city block distance is still used in the matching process. The distance between two shapes C_1 and C_2 in the FDGF-SH-SFtC feature space is expressed as $D_{SFtC}(C_1, C_2)$. Different levels coarser shapes are shown in Fig. 6.

D. MULTISCALE FOURIER DESCRIPTOR
FDGF-SH features on multiple generated coarser contours can be used to generate the multiscale Fourier descriptor. It is obvious that the higher the level, the coarser the feature, and the lower the level, the finer the feature.

To achieve the goal of multiscale description, multiple FDGF-SH features are used together. FDGF-SH-MFtC, FDGF-SH-DStC and FDGF-SH-SFtC features are used to form multiscale features MSFDGF-SH-MF, MSFDGF-SH-DS and MSFDGF-SH-SF respectively. MSFDGF-SH-MF is used as an example to illustrate how to integrate. The minimum distance (MD) and the sum distance (SD) are two common approaches. In Eq. (10), D_m^{mf} is the MD between the MSFDGF-SH-MF features of two shapes.

$$D_m^{mf} = \min_{t \in S_v} \alpha_m^t D_{MFtC} \quad (10)$$

Since the range of the distance at each level of the coarser contour is different from each other, α_m^t (α_m^t increases when t increases) is used to normalize the distances. S_v is a set of the used values of t, which denotes the level of the coarser contour. Only a few levels are used in the matching process.

D_s^{mf} is the SD as shown in Eq. (11). Since the importance of the distance at each level is different from each other, α_s^t (α_s^t decreases when t increases) is used to normalize the distances.

$$D_s^{mf} = \sum_{t \in S_v} \alpha_s^t D_{MFtC} \quad (11)$$

How to generate S_v is a problem. For consideration of training speed, the Sequential Forward Selection method in [42] is used in selecting the combination of scales S_v. First, 30% images in dataset are randomly selected as a training subset. The single scale with the highest accuracy is set as the starting point of the scale combination, and the remaining scales are set as candidates. Nextly, each single candidate is put into the combination by turn in order to find the best candidate, which makes the new combination obtain the highest accuracy. Then, this best candidate is put into the combination and removed from the candidates. This process of finding best candidate is performed iteratively until no new scale is put into the combination S_v, which means that integrating a larger scale combination will damage the performance of the descriptor in terms of accuracy.

When the MD metric approach is used with three descriptors (MSFDGF-SH-MF+MD, MSFDGF-SH-DS+MD and MSFDGF-SH-SF+MD), the α_m^t is computed by the Eq. (12).

$$\alpha_m^t = (1 + dif_m \cdot t)^{pwr_m}, \quad t \le T_m, dif_m \ge 0 \quad (12)$$

where the T_m is the max level of the generated coarser contour. In SD Eq. (13) is used. In MD $dif_m \ge 0$, but in SD $dif_s \le 0$ always.

$$\alpha_s^t = (1 + dif_s \cdot t)^{pwr_s}, \quad t \le T_m, dif_s \ge -1/T_m \quad (13)$$

In weighted city block distance w_k (computed out by Eq. (14)) decreases linearly as k increases in both MD and SD.

$$w_k = 0.008((K + 1)/2 - k) + 1, \quad k \le K, \quad (14)$$

MSFDGF is a framework to form a multiscale Fourier descriptor. In this framework, many GFs can be used and SH is just one of them. Three approaches in GCC (generated coarser contour) are used to implement multiscale description. MSFDGF-SH-MF, MSFDGF-SH-DS and MSFDGF-SH-SF are three descriptors using SH (shape histogram) base on the MSFDGF framework.

IV. COMPUTATIONAL COMPLEXITY
Eqs (15)-(16) are two commonly used formulas in this section.

$$S_n = \sum_{s_i=1}^{n} a_1 q^{s_i-1} = a_1 \cdot \frac{1 - q^n}{1 - q} \quad (15)$$

$$S_n = \sum_{s_i=1}^{n} s_i q^{s_i} = q(1-q^n)/(1-q)^2 - nq^{n+1}/(1-q) \quad (16)$$

The computational complexities of MSFDGF-SH (MSFDGF-SH-MF, MSFDGF-SH-DS and MSFDGF-SH-SF) are divided into several parts. To generated all levels coarser contours, MSFDGF-SH-MF spends $O(N^2)$, MSFDGF-SH-DS spends $O(N \log N)$ (see Eq. (17)) and MSFDGF-SH-SF spends $O(n_s N^2)$ where n_s means how many

levels there are in MSFDGF-SH-SF.

$$\sum_{s_i=0}^{\log_2 N - 2} (N/2^{s_i} \cdot 2^{s_i}) = \sum_{s_i=0}^{\log_2 N - 2} N = N(\log_2 N - 1) \quad (17)$$

To extract shape histograms of all coarser contours, MSFDGF-SH-MF spends $O(N^3)$, MSFDGF-SH-DS spends $O(N^2)$ (see Eq. (18)) and MSFDGF-SH-SF spends $O(n_s N^2)$ respectively, as the shape histograms of one contour cost $O(N^2)$ [36].

$$\sum_{s_i=0}^{\log_2 N - 2} ((N/2^{s_i})^2 \cdot 2^{s_i}) = \sum_{s_i=0}^{\log_2 N - 2} N^2/2^{s_i}$$
$$= N^2 \sum_{s_i=0}^{\log_2 N - 2} (1/2)^{s_i}$$
$$= N^2 \cdot \frac{1 - (1/2)^{\log_2 N - 1}}{1 - 1/2}$$
$$= 2N^2 - 4N \quad (18)$$

On Fourier transform, MSFDGF-SH-MF spends $O(BN^2 \log N)$, MSFDGF-SH-DS spends $O(BN \log N)$ computed by Eq. (19), and MSFDGF-SH-SF spends $O(n_s BN \log N)$.

$$\sum_{s_i=0}^{\log_2 N - 2} B(N/2^{s_i}) \log_2(N/2^{s_i})$$
$$= BN \sum_{s_i=0}^{\log_2 N - 2} 2^{-s_i} (\log_2 N - s_i)$$
$$= BN \log_2 N \sum_{s_i=0}^{\log_2 N - 2} 2^{-s_i} - BN \sum_{s_i=0}^{\log_2 N - 2} 2^{-s_i} s_i$$
$$= (2BN \log_2 N - 4B \log_2 N) - (2BN - 4B \log_2 N)$$
$$= 2BN(\log_2 N - 1) \quad (19)$$

Then, in matching stage, MSFDGF-SH-MF spends $O(BN \log N)$, MSFDGF-SH-DS spends $O(B(\log N)^2)$ (see Eq. (20)), MSFDGF-SH-SF spends $O(n_s B \log N)$ when all the scales are used. In fact, only a few scales are used in this stage.

$$\sum_{s_i=0}^{\log_2 N - 2} B \log_2(N/2^{s_i}) = B \sum_{s_i=0}^{\log_2 N - 2} (\log_2 N - s_i)$$
$$= B \sum_{s_i=0}^{\log_2 N - 2} \log_2 N - B \sum_{s_i=0}^{\log_2 N - 2} s_i$$
$$= (B \log_2^2 N - B \log_2 N)$$
$$- (B \log_2^2 N - 3B \log_2 N + 2B)/2$$
$$= B(\log_2^2 N + \log_2 N - 2)/2 \quad (20)$$

The computational complexity in matching process plays a decisive role in online large database retrieval [43], therefore the computational complexities of some state-of-the-art descriptors in matching stage is used to compare with MSFDGF-SH, as shown in Table 1.

TABLE 1. The computational complexities of some state-of-the-art methods and MSFDGF-SH at the stage of dissimilarity measurement.

method	time complexity
MSFDGF-SH-MF+MD/SD	$O(BN \log N)$
MSFDGF-SH-DS+MD/SD	$O(B(\log N)^2)$
MSFDGF-SH-SF+MD/SD	$O(n_s B \log N)$
AP&BAP [42]	$O(N B_a + N 4^{M_b})$
HSC [43]	$O(M_h \log N)$
IDSC+DP [36]	$O(BN^2)$
SC+DP [3], [35]	$O(BN^2)$
TAR+DP [37]	$O(N^3)$
Shape Tree [38]	$O(N^4)$

TABLE 2. The parameters of MSFDGF-SH-MF, MSFDGF-SH-DS and MSFDGF-SH-SF when combined with MD and SD.

method	dif_m/dif_s
MSFDGF-SH-MF+MD	0.007
MSFDGF-SH-DS+MD	0.036
MSFDGF-SH-SF+MD	0.14
MSFDGF-SH-MF+SD	-0.001
MSFDGF-SH-DS+SD	-0.1
MSFDGF-SH-SF+SD	-0.07

In Table 1, n_s means how many levels are used in MSFDGF-SH-SF and it is smaller than 7 in experiments. B means how many bins in the shape histogram, and it is smaller than N usually. B_a means how many bins in AP, and it is 24 in [42]. M_b means how many bits the BAP festure has at the largest scale. In the experiments in [42], $M_b = 12$. In HSC $M_h \ll N$, and $M_h = 7$ in the experiments in [43].

V. EXPERIMENTAL RESULTS

The MSFDGF-SH-MF, MSFDGF-SH-DS and MSFDGF-SH-SF are evaluated in terms of both effectiveness and efficiency. These evaluating databases include MPEG-7 CE1 Part B, Kimia 99 [47], Swedish Plant Leaf [48] and Expanded Articulation Database [36]. All the algorithms are written using Matlab and run on a PC with Intel(R) Core(TM) i7-7700K 4.20 GHz CPU and 16 GB DDR4 RAM under Windows 10. As the DP part in SC+DP and IDSC+DP consumes large computation, it is implemented in C in order to be comparable to global descriptors like AP&BAP [42], HSC [43] and MSFDGF-SH.

N is always 512 in all the experiments. In MSFDGF-SH-SF, to make one SE is suitable to all shapes, shapes are normalized to have a convex hull's area near 5000 [5]. The size of SE is $5t$ at level t coarser contour.

In the experiment on each dataset, the SFS technique is used to find a good combination of scales S_v. $K = 23$, $pwr_m = 1$, $pwr_s = 2$, and the values of dif_m and dif_s (as shown in the Table 2) are set emperically.

A. RESULTS ON MPEG-7 CE-1 PART B SHAPE DATABASE

MPEG-7 CE-1 Part B shape database [35], [36], [43] is widely used in shape matching research. This database contains 70 categories, each containing 20 different shapes,
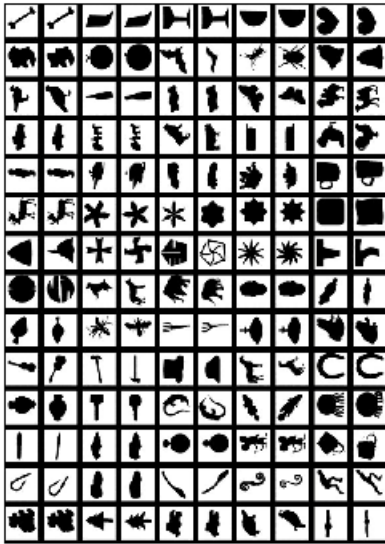
FIGURE 7. Some examples of MPEG-7 CE-1 Part B shape database.

method	retrieval rate	matching time (ms)
MSFDGF-SH-SF+MD	**87.76%**	**21**
MSFDGF-SH-SF+SD	**87.76%**	**41**
MSFDGF-SH-MF+MD	85.10%	52
MSFDGF-SH-MF+SD	85.65%	61
MSFDGF-SH-DS+MD	85.07%	49
MSFDGF-SH-DS+SD	85.46%	49
Shape Tree [38]	87.70%	8926
TAR+DP [37]	87.13%	5810
SC+DP [3], [35]	86.80%	4837
IDSC+DP [36]	85.40%	3829
HSC [43]	87.31%	23
AP&BAP [42]	87.03%	85
IDSC−wFW [49]	82.34%	9
SITVRT [50]	80.20%	NA
DIR [45]	77.69%	5
ASD&CCD [27]	76.20%	NA
MDM [39]	70.46%	30
FD-CCD [40]	67.94%	3
FD-FPD [46]	64.29%	3

TABLE 4. The area enclosed by the Precision-Recall curve and the coordinate axis.

methods	area
MSFDGF-SH-SF+SD	**0.847558**
MSFDGF-SH-SF+MD	0.844857
HSC [43]	0.833407
AP&BAP [42]	0.830903
SC+DP [3], [35]	0.834662
IDSC+DP [36]	0.817020
MDM [39]	0.658289
FD-CCD [40]	0.608322

TABLE 5. The recognition rates of some state-of-the-art methods and MSFDGF-SH on Swedish Plant Leaf database.

method	recognition rate	matching time (ms)
MSFDGF-SH-MF+MD	95.20%	13.9
MSFDGF-SH-MF+SD	95.07%	16.3
MSFDGF-SH-SF+MD	95.20%	5.7
MSFDGF-SH-SF+SD	94.80%	11.3
MSFDGF-SH-DS+MD	95.47%	13.1
MSFDGF-SH-DS+SD	95.47%	13.1
HSC [43]	96.91%	6.2
SC+DP [3], [35]	88.12%	1171.3
IDSC+DP [36]	94.13%	1015.6
AP&BAP [42]	94.80%	5.7
IDSC−wFW [49]	93.71%	6.1
MDM [39]	93.60%	8.1
ASD&CCD [27]	93.08%	NA

so this database contains 1400 silhouette images. Two examples in each category are shown in Fig. 7.

The test method is called "Bull-eye test" [35], [36], [43]. In Bull-eye test, a shape in the database is set as a query in a retrieval and matched with all the shapes in the database. The correct (that is the query shape and the retrieved one belong to the same category) matches of the top 40 most similar (smallest dissimilarity) shapes are counted. The number of correct matches divided by 20 is the score of a retrieval. The retrieval rate of Bull-eye test is the average score of all retrievals where each shape is set as the query in a retrieval.

Matching time is used to test the performance of each algorithm in terms of efficiency. Matching time refers to the time it takes to match the feature of the query to features of all shapes in the database.

In Table 3, it can be seen that MSFDGF-SH-SF+MD and MSFDGF-SH-SF+SD get the highest retrieval rate (87.76%) among all the descriptors including Shape Tree (87.70%) [38], HSC (87.31%) [43], AP&BAP (87.03%) [42], SC+DP (86.80%) [3], [36], IDSC+DP (85.40%) [36] and so on. It is the first time that a Fourier descriptor performs better than Shape Tree [38] in terms of accuracy on MPEG-7 CE-1 Part B shape database.

In terms of matching time, MSFDGF-SH-SF+MD (21ms), MSFDGF-SH-SF+SD (41ms), HSC (23ms) and AP&BAP (85ms) consume less time as global descriptors than IDSC+DP (3829ms) and SC+DP (4837ms) as local descriptors.

The Precision-Recall curves of some descriptors are shown in Fig. 8. For quantitative analysis, the area enclosed by the curve and the coordinate axis is used to determine which the best is. The Table 4 shows that MSFDGF-SH-SF+SD has the biggest area 0.847558.

B. RESULTS ON SWEDISH PLANT LEAF DATABASE

The Swedish Plant Leaf database is a plant leaf database that contains 15 categories, each of which containing 75 shapes. Some examples of this database are shown in Fig. 9. This database is often used to test the classification ability of a shape descriptor. The test method in [36] is used to test the classification ability of the MSFDGF-SH descriptor. 25 images randomly selected from each species are used as models and the remaining images are used as testing images [43].
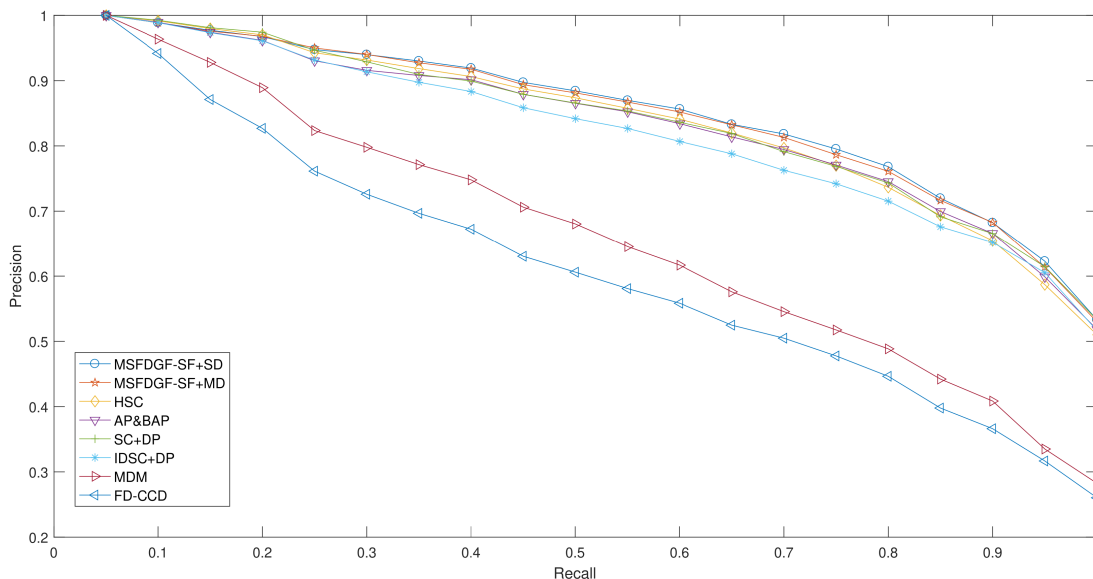
FIGURE 8. Precision-Recall curves of some state-of-the-art methods and MSFDGF-SH-SF on MPEG-7 CE-1 Part B shape database.
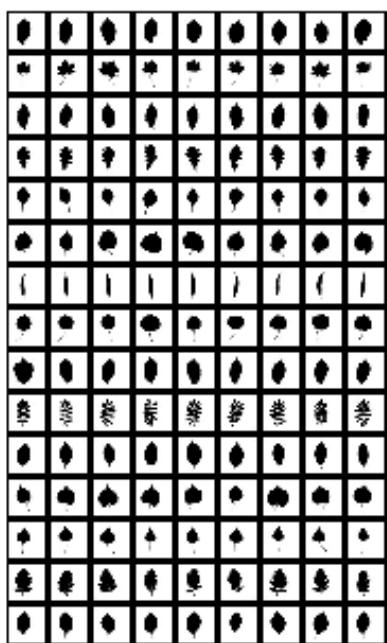

FIGURE 9. Some examples of Swedish Plant Leaf database.

TABLE 6. The retrieval results of some state-of-the-art methods and MSFDGF-SH on Kimia 99 database.

method	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	total	matching time
MSFDGF-SH-SF+MD+LP	99/99	99/99	99/99	99/99	99/99	99/99	99/99	99/99	99/99	99/99	**990/990**	0.75 ms
MSFDGF-SH-SF+SD+LP	99/99	99/99	99/99	99/99	99/99	99/99	99/99	99/99	99/99	99/99	**990/990**	0.51 ms
MSFDGF-SH-MF+MD+LP	99/99	99/99	99/99	99/99	99/99	99/99	99/99	99/99	99/99	99/99	**990/990**	0.51 ms
MSFDGF-SH-MF+SD+LP	99/99	99/99	99/99	99/99	99/99	99/99	99/99	99/99	99/99	99/99	**990/990**	0.51 ms
MSFDGF-SH-DS+MD+LP	99/99	99/99	99/99	99/99	99/99	99/99	99/99	99/99	99/99	99/99	**990/990**	0.96 ms
MSFDGF-SH-DS+SD+LP	99/99	99/99	99/99	99/99	99/99	99/99	99/99	99/99	98/99	99/99	989/990	0.89 ms
IDSC+DP+LP [3], [36]	99/99	99/99	99/99	99/99	99/99	99/99	99/99	99/99	97/99	99/99	988/990	392.05 ms
SC+DP+LP [3], [35]	98/99	96/99	98/99	98/99	98/99	99/99	99/99	98/99	96/99	99/99	979/990	592.73 ms
DIR+LP [3], [45]	99/99	99/99	99/99	99/99	99/99	99/99	96/99	98/99	90/99	98/99	976/990	0.85 ms
AP&BAP+LP [3], [42]	97/99	97/99	97/99	97/99	97/99	98/99	98/99	98/99	95/99	92/99	966/990	9.57 ms
HSC+LP [3], [43]	98/99	93/99	95/99	96/99	96/99	96/99	96/99	96/99	84/99	77/99	927/990	2.89 ms
MDM+LP [3], [39]	96/99	95/99	93/99	88/99	90/99	88/99	86/99	84/99	77/99	79/99	876/990	4.78 ms
FD-CCD+LP [3], [40]	93/99	86/99	86/99	83/99	79/99	67/99	71/99	80/99	49/99	37/99	731/990	0.61 ms
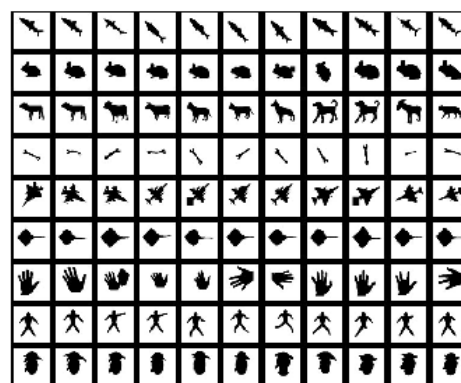
FIGURE 10. All the shapes of Kimia 99 database.

Matching time is also used to evaluate the performance of each algorithm in terms of efficiency. In the experiment of recognition, matching time refers to the time it takes to match the feature of a testing image to features of all model images.

In Table 5, it can be seen that MSFDGF-SH-DS+SD (95.47%), MSFDGF-SH-DS+MD (95.47%), MSFDGF-SH-MF+MD (95.20%), MSFDGF-SH-MF+SD (95.07%), MSFDGF-SH-SF+MD (95.20%) and MSFDGF-SH-SF+SD (94.80%) perform better than IDSC+DP (94.13%) [36],

IDSC−wFW (93.71%), MDM (93.60%) [39], ASD&CCD (93.08%) [27] and SC+DP (88.12%) [35], [36]. HSC (96.91%) [43] obtains the highest recognition rate. The matching time of each algorithm goes on with the same trend as in MPEG-7 CE-1 Part B, and global descriptors have absolute advantages.

TABLE 7. The retrieval results of some state-of-the-art methods and MSFDGF-SH on expanded articulated database.

method	1st	2nd	3rd	4th	total	matching time (ms)
MSFDGF-SH-SF+ID+MD	71/75	70/75	56/75	44/75	241/300	2.12
MSFDGF-SH-SF+ID+SD	71/75	69/75	56/75	51/75	247/300	2.41
MSFDGF-SH-DS+ID+MD	73/75	69/75	69/75	56/75	267/300	1.23
MSFDGF-SH-DS+ID+SD	75/75	70/75	70/75	56/75	**271/300**	1.23
MSFDGF-SH-MF+ID+MD	75/75	72/75	60/75	53/75	260/300	2.12
MSFDGF-SH-MF+ID+SD	74/75	71/75	67/75	51/75	263/300	2.12
IDSC+DP [36]	73/75	66/75	52/75	55/75	246/300	308.56
IDSC−wFW [49]	70/75	63/75	53/75	51/75	237/300	1.96
DIR [45]	62/75	41/75	37/75	22/75	162/300	0.69
FD-CCD [40]	49/75	28/75	19/75	21/75	117/300	0.53
HSC [43]	44/75	30/75	19/75	18/75	111/300	2.69
SC+DP [3], [35]	40/75	20/75	21/75	11/75	92/300	467.21
MDM [39]	42/75	24/75	15/75	9/75	90/300	3.25
AP&BAP [42]	15/75	7/75	5/75	8/75	35/300	7.52
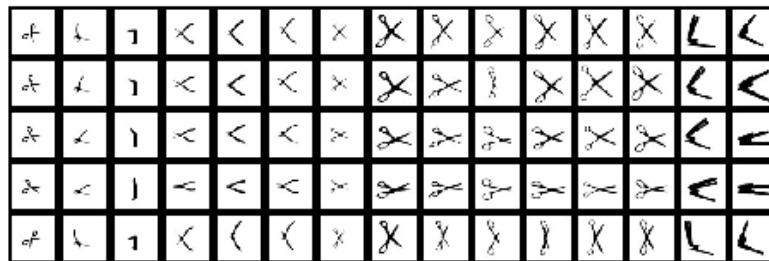
FIGURE 11. All the shapes of expanded articulation database.

C. RESULTS ON KIMIA 99 DATABASE

The Kimia 99 [47] database is a common database. This database contains 9 categories, each containing 11 shapes, as shown in Fig. 10. In this experiment, each shape is set as the query and matched to the remaining shapes. Then the correct matches of the top 10 most similar shapes of each query are counted. The post-processing algorithm LP (label propagation) [3] on shape retrieval performs well on this database when used in combination with IDSC+DP. In order to be fair, all algorithms are combined with LP.

In Table 6, it can be seen that MSFDGF-SH-DS+MD+LP (990), MSFDGF-SH-MF+MD+LP (990), MSFDGF-SH-MF+SD+LP (990), MSFDGF-SH-SF+MD+LP (990) and MSFDGF-SH-SF+SD+LP (990) all get the perfect sum of the correct hits (990). MSFDGF-SH-DS+SD (989) performs a little lower. MSFDGF-SH performs best among all methods.

In terms of matching time, the approaches using MSFDGF-SH (less than 1 ms), HSC (2.89 ms) and AP&BAP (9.57 ms) consume less time as global descriptors than IDSC+DP (392.05 ms) and SC+DP (592.73 ms) as local descriptors.

D. RESULTS ON EXPANDED ARTICULATED DATABASE

The Articulated database [36] is a database to test the articulation insensitivity of shape descriptors. It contains 8 categories, each containing 5 shapes. The Tools database [51], which has the same function, contains 7 categories, each containing 5 shapes. In our study, these two databases are merged into a new database Expanded Articulated Database. Obviously, Expanded Articulated Database contains 15 categories, each containing 5 shapes (see Fig. 11).

The test method is as the same as that in the experiment on Articulated Database in [36]. In this test method, each shape is set as the query and matched with other shapes in the database. Then the correct matches of the top 4 most similar shapes of each query are counted. The combination (IDSC) of SC and ID (inner-distance) [36] performs well on Articulated Database. ID can also be used in combination with MSFDGF-SH, so this database is also used to test the compatibility of MSFDGF-SH with ID.

In Table 7, it can be seen that MSFDGF-SH-DS+SD gets the largest sum of the correct hits (271) among all the methods including MSFDGF-SH-DS+MD (267), MSFDGF-SH-MF+MD (260), MSFDGF-SH-MF+SD (263), MSFDGF-SH-SF+MD (241), MSFDGF-SH-SF+SD (247), IDSC+DP (246) [36], IDSC−wFW (237) [49], SC+DP (92) [3], [35], HSC (111) [43], AP&BAP (35) [42] and so on.

In terms of matching time, the approaches using MSFDGF-SH (less than 2.5 ms), HSC (2.69 ms) and AP&BAP (7.52 ms) consume less time as global descriptors than IDSC+DP (308.56 ms) and SC+DP (467.21 ms) as local descriptors.

E. DISCUSSION

Three versions (MSFDGF-SH-MF, MSFDGF-SH-DS and MSFDGF-SH-SF) of MSFDGF-SH all exceed the classical IDSC+DP in terms of both effectiveness and efficiency on MPEG-7 CE-1 Part B shape database. Surprisingly, one version MSFDGF-SH-SF exceeds HSC, AP&BAP and even Shape Tree. This is the first time that a Fourier descriptor exceeds Shape Tree on this dataset in terms of both accuracy and speed. MSFDGF-SH performs better than IDSC+DP, SC+DP and AP&BAP on Swedish Plant Leaf. The performance show that MSFDGF-SH have strong robustness for different application scenarios. On other datasets MSFDGF-SH also performs effectively, especially On Kimia 99, on which all three versions of MSFDGF-SH achieve the perfect performance. On Expanded Articulated Database, all the 3 versions of MSFDGF-SH exceed IDSC+DP, which is so good at dealing with articulated shapes.

VI. CONCLUSION

AP&BAP [42] is a milestone for global shape descriptors. Many researchers have attempted to design effective global descriptors, however it is difficult to achieve both effectiveness and efficiency. The MSFDGF frame and MSFDGF-SH descriptor proposed in this article is a new attempt.

Our experiments show that MSFDGF is a flexible and effective framework to form a global descriptor. The descriptors using MSFDGF, such as MSFDGF-SH, are both efficient and effective.

REFERENCES

[1] S. Mouine, I. Yahiaoui, and A. Verroust-Blondet, "A shape-based approach for leaf classification using multiscale triangular representation," in *Proc. 3rd ACM Conf. Int. Conf. Multimedia Retr.*, 2013, pp. 127–134.

[2] C.-H. Wei, Y. Li, W.-Y. Chau, and C.-T. Li, "Trademark image retrieval using synthetic features for describing global shape and interior structure," *Pattern Recognit.*, vol. 42, no. 3, pp. 386–394, Mar. 2009.

[3] X. Bai, X. Yang, L. J. Latecki, W. Liu, and Z. Tu, "Learning context-sensitive shape similarity by graph transduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 861–874, May 2010.

[4] X. Bai, B. Wang, C. Yao, W. Liu, and Z. Tu, "Co-transduction for shape retrieval," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2747–2757, May 2012.

[5] R.-X. Hu, W. Jia, Y. Zhao, and J. Gui, "Perceptually motivated morphological strategies for shape retrieval," *Pattern Recognit.*, vol. 45, no. 9, pp. 3222–3230, Sep. 2012.

[6] V. Premachandran and R. Kakarala, "Perceptually motivated shape context which uses shape interiors," *Pattern Recognit.*, vol. 46, no. 8, pp. 2092–2102, Aug. 2013.

[7] P. Kontschieder, M. Donoser, and H. Bischof, "Beyond pairwise shape similarity analysis," in *Proc. Asian Conf. Comput. Vis.* Berlin, Germany: Springer, 2009, pp. 655–666.

[8] X. Yang, X. Bai, L. J. Latecki, and Z. Tu, "Improving shape retrieval by learning graph transduction," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2008, pp. 788–801.

[9] X. Yang, S. Koknar-Tezel, and L. J. Latecki, "Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 357–364.

[10] S. Bai and X. Bai, "Sparse contextual activation for efficient visual re-ranking," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1056–1069, Mar. 2016.

[11] B. Mokhtari, K. E. Melkemi, D. Michelucci, and S. Foufou, "Unsupervised geodesic convex combination of shape dissimilarity measures," *Pattern Recognit. Lett.*, vol. 98, pp. 46–52, Oct. 2017.

[12] D. C. G. Pedronette, J. Almeida, and R. D. S. Torres, "A graph-based ranked-list model for unsupervised distance learning on shape retrieval," *Pattern Recognit. Lett.*, vol. 83, pp. 357–367, Nov. 2016.

[13] Y. Zheng, Y. Yan, B. Guo, and W. He, "O2O method for fast 2D shape retrieval," *IEEE Trans. Image Process.* vol. 28, no. 11, pp. 5366–5378, Nov. 2019.

[14] E. Milios and E. G. M. Petrakis, "Shape retrieval based on dynamic programming," *IEEE Trans. Image Process.*, vol. 9, no. 1, pp. 141–147, Jan. 2000.

[15] S. Loncaric, "A survey of shape analysis techniques," *Pattern Recognit.*, vol. 31, no. 8, pp. 983–1001, Aug. 1998.

[16] 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.

[17] C. Grigorescu and N. Petkov, "Distance sets for shape filters and shape recognition," *IEEE Trans. Image Process.*, vol. 12, no. 10, pp. 1274–1286, Oct. 2003.

[18] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognit.*, vol. 37, no. 1, pp. 1–19, Jan. 2004.

[19] M.-K. Hu, "Visual pattern recognition by moment invariants," *IEEE Trans. Inf. Theory*, vol. IT-8, no. 2, pp. 179–187, Feb. 1962.

[20] C. C. Chang, S. M. Hwang, and D. J. Buehrer, "A shape recognition scheme based on relative distances of feature points from the centroid," *Pattern Recognit.*, vol. 24, no. 11, pp. 1053–1063, Jan. 1991.

[21] B. Wang, J. Wu, H. Shu, and L. Luo, "Shape description using sequency-ordered complex Hadamard transform," *Opt. Commun.*, vol. 284, no. 12, pp. 2726–2729, Jun. 2011.

[22] X. Fan, C. Qi, D. Liang, and H. Huang, "Probabilistic contour extraction using hierarchical shape representation," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1, Oct. 2005, pp. 302–308.

[23] G. McNeill and S. Vijayakumar, "Hierarchical procrustes matching for shape retrieval," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2006, pp. 885–894.

[24] J. Xie, P.-A. Heng, and M. Shah, "Shape matching and modeling using skeletal context," *Pattern Recognit.*, vol. 41, no. 5, pp. 1756–1767, May 2008.

[25] T. B. Sebastian, P. N. Klein, and B. B. Kimia, "On aligning curves," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 1, pp. 116–125, Jan. 2003.

[26] 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, Dec. 2005.

[27] F. Fotopoulou and G. Economou, "Multivariate angle scale descriptor of shape retrieval," in *Proc. Signal Process. Appl. Math. Electron. Commun.*, 2011, pp. 105–108.

[28] Y. Zheng, B. Guo, C. Li, and Y. Yan, "A Fourier descriptor of 2D shapes based on multiscale centroid contour distances used in object recognition in remote sensing images," *Sensors*, vol. 19, no. 3, p. 486, 2019.

[29] R. Gopalan, P. Turaga, and R. Chellappa, "Articulation-invariant representation of non-planar shapes," in *Computer Vision—ECCV* (Lecture Notes in Computer Science), vol. 6313. Berlin, Germany: Springer, 2010.

[30] B. Wang, Y. Gao, C. Sun, M. Blumenstein, and J. La Salle, "Can walking and measuring along chord bunches better describe leaf shapes?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6119–6128.

[31] Q. Jia, X. Fan, Y. Liu, H. Li, Z. Luo, and H. Guo, "Hierarchical projective invariant contexts for shape recognition," *Pattern Recognit.*, vol. 52, pp. 358–374, Apr. 2016.

[32] C. Yang, H. Wei, and Q. Yu, "A novel method for 2D nonrigid partial shape matching," *Neurocomputing*, vol. 275, pp. 1160–1176, Jan. 2018.

[33] P. Govindaraj and M. S. Sudhakar, "A new 2D shape retrieval scheme based on phase congruency and histogram of oriented gradients," *Signal, Image Video Process.*, vol. 13, no. 4, pp. 771–778, Jun. 2019.

[34] J. Guo, Y. Rong, Y. Gao, Y. Liu, and S. Xiong, "A novel multi-scale invariant descriptor based on contour and texture for shape recognition," in *Proc. Asian Conf. Comput. Vis.* Cham, Switzerland: Springer, 2018, pp. 551–564.

[35] 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.

[36] 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.

[37] 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, Jul. 2007.

[38] P. F. Felzenszwalb and J. D. Schwartz, "Hierarchical matching of deformable shapes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.

[39] R. Hu, W. Jia, H. Ling, and D. Huang, "Multiscale distance matrix for fast plant leaf recognition," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4667–4672, Nov. 2012.

[40] D. Zhang and G. Lu, "Study and evaluation of different Fourier methods for image retrieval," *Image Vis. Comput.*, vol. 23, no. 1, pp. 33–49, Jan. 2005.

[41] H. S. Yang, S. U. Lee, and K. M. Lee, "Recognition of 2D object contours using starting-point-independent wavelet coefficient matching," *J. Vis. Commun. Image Represent.*, vol. 9, no. 2, pp. 171–181, Jun. 1998.

[42] R.-X. Hu, W. Jia, H. Ling, Y. Zhao, and J. Gui, "Angular pattern and binary angular pattern for shape retrieval," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1118–1127, Mar. 2014.

[43] B. Wang and Y. Gao, "Hierarchical string cuts: A translation, rotation, scale, and mirror invariant descriptor for fast shape retrieval," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4101–4111, Sep. 2014.

[44] G. C.-H. Chuang and C.-C.-J. Kuo, "Wavelet descriptor of planar curves: Theory and applications," *IEEE Trans. Image Process.*, vol. 5, no. 1, pp. 56–70, Jan. 1996.

[45] 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.

[46] A. El-Ghazal, O. Basir, and S. Belkasim, "Farthest point distance: A new shape signature for Fourier descriptors," *Signal Process., Image Commun.*, vol. 24, no. 7, pp. 572–586, Aug. 2009.

[47] 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.

[48] O. J. O. Söderkvist, "Computer vision classification of leaves from Swedish trees," M.S. thesis, Linköping Univ., Linköping, Sweden, 2001.

[49] Y. Zheng, B. Guo, C. Li, and Y. Yan, "A weighted Fourier and wavelet-like shape descriptor based on IDSC for object recognition," *Symmetry*, vol. 11, no. 5, p. 693, 2019.

[50] B. Wang and Y. Gao, "Structure integral transform versus radon transform: A 2D mathematical tool for invariant shape recognition," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5635–5648, Dec. 2016.

[51] A. M. Bronstein, M. M. Bronstein, A. M. Bruckstein, and R. Kimmel, "Analysis of two-dimensional non-rigid shapes," *Int. J. Comput. Vis.*, vol. 78, no. 1, pp. 67–88, Jun. 2008.

JIE LIU received the bachelor's and master's degrees from Tsinghua University, China, and the Ph.D. degree from UC Berkeley, in 2001. He is the Dean of AI research with the Harbin Institute of Technology, a Professor with the School of Computer Science and Technology, the Director of the IoT and Ambient Intelligence Lab, and a Ph.D. advisor of both computer science and control. Before joining the HIT, he managed the research and development of ambient intelligence at Microsoft cloud and AI platforms, and was a partner of a company. He has more than 15 years of industrial research and research leadership experience at Microsoft Research and Xerox PARC, working in areas such as the IoT, cyber-physical systems, embedded AI, and energy-efficient systems. He is an ACM Distinguished Scientist and ACM Distinguished Speaker.
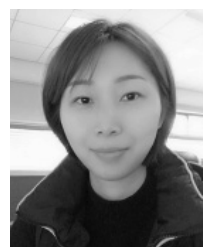
BAOLONG GUO received the B.S., M.S., and Ph.D. degrees from Xidian University, Xi'an, China, in 1984 and 1995, respectively. He is currently a Professor with the Department of Electronics Engineering, Xidian University. His research interests include computer graphics, multimedia information processing, computer vision, pattern recognition, and machine learning. He is a Senior Member of the IEEE System Council.

YANG SONG received the Ph.D. degree in computer science from The University of Sydney, in 2013. She was an Australian Research Council (ARC) Discovery Early Career Researcher Award (DECRA) Fellow of the School of Information Technologies, The University of Sydney, from 2015 to 2018. She is currently a Lecturer with the School of Computer Science and Engineering, University of New South Wales, Australia. Her research interests include biomedical imaging informatics, computer vision, and machine learning.

YAN ZHENG received the B.S. degree from Northeast Petroleum University, Daqing, China, in 2012. He is currently pursuing the Ph.D. degree with the Department of Electronics Engineering, Xidian University, Xi'an, China. His research interests include computer vision and machine learning.

XUEBING ZHANG received the Ph.D. degree from Jilin University, Changchun, China, in 2017. His research interests include deep learning and ore genesis.

FANJIE MENG received the B.S. degree in control technology and instruments, the M.S. degree in pattern recognition and intelligent systems, and the Ph.D. degree in circuits and systems from Xidian University, Xi'an, China, in 2001, 2004, and 2015, respectively. Since 2012, she has been an Associate Professor with the ICIE Institute, School of Aerospace Science and Technology, Xidian University. Her research interests include signal processing and image analysis.

LING WANG received the Ph.D. degree in computer and electronics engineering from the University of Nevada, in 2003. She is a Professor with the School of Computer Science and Technology, Harbin Institute of Technology.

• • •