SHAPE RETRIEVAL USING MULTISCALE ELLIPSE DESCRIPTOR

Xiaohong Zhang *Jianwen Xiang* *Shengwu Xiong**

Hubei Key Lab. of Transportation of IoT, School of Computer Science and Technology,
Wuhan University of Technology, Wuhan, China

ABSTRACT

In this paper, a novel multiscale ellipse descriptor (MED) method is proposed for shape description and matching. MED extracts the competitive features of shape contour by measuring the spatial location relationship between contour sample points and topology structure information of segmented multiscale zone. This method not only has the discriminative ability to describe the global and local information, but also is robustness to various linear (rotation, scale and translation transforms) and non-linear (irregular intra-class deformation) transforms. Experimental results on two public available databases consistently demonstrate that, our proposed method is effective and efficient when compared with other state-of-the-art shape retrieval benchmarks (such as 9.72% higher and 64 times faster than popular IDSC method on leaf100 dataset).

***Index Terms*—** shape retrieval, multiscale invariant descriptor, shape matching

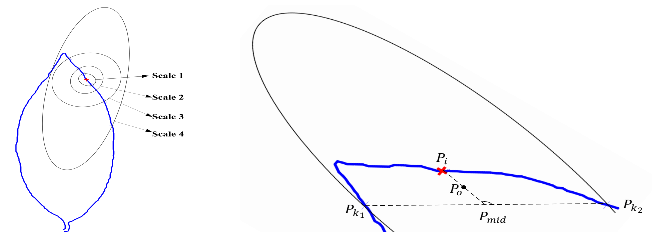
1. INTRODUCTION

Shape-based recognition and matching is a challenging task which plays an important role in many applications, such as plant leaf species retrieval [1][2][3][4], fish image retrieval [2] and animal species retrieval [3][5][6][7]. Reflected on the human intuition, a good shape descriptor should tolerate the larger intra-class variation (geometric differences of objects from the same class) and the smaller inter-class variation (discriminate objects from different classes)[8]. Theoretically, shape-based feature should be robustness to the linear transforms (translation, rotation and scale transforms) and other non-linear transforms (irregular intra-class deformation)[5], meanwhile have the discriminative ability of shape information. To construct a shape descriptor which overcomes these challenges is still an open problem that needs to be solved.

To address these challenges, many shape descriptors has been proposed. Belongie et al. [9] described a shape context (SC) by capturing 2-D histogram distributions, however it is sensitive to the articulation variance. Ling et al. [10] extended SC to the Inner-distance shape context (IDSC) by replacing the Euclidean distance with inner-distance. Daliri et al. [6] used aligning corresponding points to extract the symbolic representation (such as quantized angles, quantized distance from center of gravity). Kaothanthong et al. [11] proposed the statistical method distance interior ratio (DIR) to depict the maximum-alignment and mean-alignment histogram of bin interval information in shape contour. However these shape descriptors of global information are sensitive to non-linear transform for the larger changes of feature histogram.

In order to be invariant to the above transform, multiscale shape descriptors have been proposed to extract both global and local shape information. Shape tree (ST) [12], hierarchical procrustes matching (HPM) [13] and shape vocabulary learning [14] [7] decomposed the shape into both global and local contour fragments. Curvature scale space (CSS) [15] and multiscale convexity concavity (MCC) [16] used the Gaussian kernel smoothing to find location of curvature zero-crossing contour as the multiscale standard. Hu et al. [4] proposed a multiscale distance matrix (MDM) method to capture the geometric structure of shape with Euclidean distance matrix. Wang et al. [2] presented a hierarchical string cuts (HSC) method to segment a shape into multiple coarse-to-fine level curve, in each level the string cut features are given as the height deviation distributions [8] to string cut on both side, the imbalance number of string cut and the degree of bending on the string cut curve. However these shape descriptors of discriminative ability are limited with feature spatial location relationship. Cao et al. [17] proposed a novel multiscale R-angle descriptor to define the angle between intersection points and sample point. Mouine et al. [1] constructed four multiscale triangular representations to describe the relationship between neighboring sample points, named as TAR, TSL, TOA and TSLA. Yang et al. [3] constructed three invariant shape descriptors to capture the discriminative features in each multiple circle scales, which are defined as the ratio of major area, the ratio of segmented arc length and the offset distance between weighted center of major zone and sample point. However all these features described in triangle or circle segmented zone can not well describe the topology structure information of multiscale segmented zone,

This work was partially supported by the National Natural Science Foundation of China (Grant No. 61672398), the Key Natural Science Foundation of Hubei Province of China (Grant No. 2015CFA069), and the Applied Fundamental Research of Wuhan (Grant No. 20160101010004).



(a) multiscale ellipse (b) specific definition in each scale

Fig. 1. Definition of multiscale ellipse.

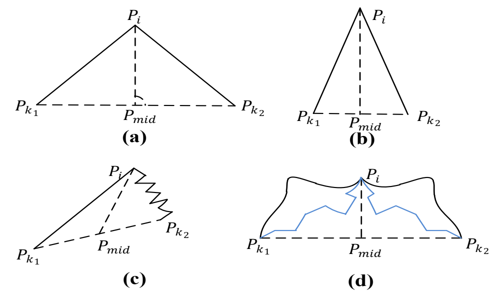


Fig. 2. The explanatory figures under different cases .

which is limited in discriminative ability of shape feature. In mulitscale zone, ellipse has its advantage on discriminative ability of shape feature, which can not only capture the spatial information but also topological structure of segmented zone.

In this paper, we propose a novel multiscale ellipse descriptor (MED) for shape retrieval. This descriptor firstly segment a shape into multiscale ellipse, then the spatial location relationship of contour sample points and topological structure information of segmented ellipse zone are expressed by four invariant ellipse features. This scheme provides a powerful discriminative ability with both global and local details, which gets a bulls-eye score of 98.07% on the MPEG-7 shape dataset. The contribution of our work can be summarized as follows: (1) the multiscale ellipse segmented pattern is used to extract both global and local shape information; (2) four invariant ellipse features are given to describe the discriminative information of shape contour curves in each segmented scale; (3) all the designed features are invariant to linear (rotation, scale and translation transforms) and non-linear (irregular intra-class deformation) transforms.

2. MULTISCALE ELLIPSE DESCRIPTOR

2.1. Definition of multiscale ellipse

A closed shape can be effectively expressed by a sequences of uniformed sample points [3][2][8]. The benefits of this expression way is that, the approximate shape contour information can be reserved without costing the extra time to compute the key points. Therefore, a shape contour $\mathcal{C}$ can be represented as $\mathcal{C} = \{P_i(x_i, y_i)|i = 1, 2, \cdots, N\}$ in an anticlockwise direction, where i is the index number of contour sample point, (x_i, y_i) is the corresponding coordinates location of P_i and N is the total number of shape contour sample points (for segmentation convenience, the number should be the time of 2).

Let the interval distance t between sample point and neighboring point denotes the standard scale size, where $t = 1, 2, \cdots, \log N - 1$. Physically, the scale of larger distance can capture the global shape information, while the smaller distance focus on the local details of shape contour (see in Fig. 1 (a)). Let P_{k_1}, P_{k_2} separately represent the neighboring points of sample point P_i, where $k_1 = i - 2^t, k_2 = i + 2^t$. P_{mid} is calcluated as the middle point of two neighboring points P_{k_1}, P_{k_2} ($P_{mid} = (P_{k_1} + P_{k_2})/2$), we define that both sample point P_i and middle point P_{mid} are regarded as the two focus points of ellipse. P_0 is the middle point of P_{mid} and P_i ($P_0 = (P_i + P_{mid})/2$), which can be regarded as the center axis point of ellipse. Therefore, in each scale the ellipse can be depicted through the trajectory of a moving point, meanwhile the ellipse should satisfy three properties: the center point is P_0, the two focus points are P_i, P_{mid} and the summation of distance values between moving point and two focus points equals to ($|\mathbf{P_{k_1}P_i}| + |\mathbf{P_{k_1}P_{mid}}|$) (see in Fig. 1 (b)).

2.2. Invariant ellipse feature

A good shape feature descriptor should have the discriminative ability to distinguish the smaller inter-class variation and be robustness to the bigger intra-class variation.

In each segmented scale, to describe the feature of contour curve some state-of-the-art descriptors are presented, such as height distance to tangent line [8], procruste distance [13], symbol representation [6], and so on. However these descriptors have limited discriminative ability with less features (see in Part 1). In order to better describe the shape with more features, in this paper we propose a novel multiscale ellipse descriptor (MED) in terms of four invariant ellipse features, to describe the shape of spatial location relationship between sample points and topology structure information of segmented ellipse zone, namely: the distance ($f_1^t(i)$) between two ellipse focus points, the angle ($f_2^t(i)$) formed by the line connected with two neighboring points and central long axis of ellipse, the enclosed area ($f_3^t(i)$) of shape contour and the eccentricity ($f_4^t(i)$) of formed ellipse. Mathematically they can be expressed as

$$f_1^t(i) = |\mathbf{P_iP_{mid}}| \tag{1}$$

$$f_2^t(i) = \cos(\angle P_i P_{mid} P_{k_2}) \tag{2}$$

$$f_3^t(i) = S_t^*(i) \tag{3}$$

$$f_4^t(i) = \frac{|\mathbf{P_i P_{mid}}|}{|\mathbf{P_i P_{k_1}}| + |\mathbf{P_{k_1} P_{mid}}|} \quad (4)$$

In Eq. (1) the distance $f_1^t(i)$ between two ellipse focus points P_i, P_{mid} can reflect the concavity and convexity of segmented shape curve (see in Fig. 2), (a) and (b) shows that the longer the distance is, the curve fluctuation is relatively larger. In Eq. (2) the angle $\angle P_i P_{mid} P_{k_2}$ formed by the line $\mathbf{P_{k_1} P_{k_2}}$ connected with two neighboring points and central long axis of ellipse $\mathbf{P_i P_{mid}}$, can reflect the curve bending degree of each neighboring side (see in Fig. 2), (a) and (c) shows that the more the sample point is inclined to neighboring point, the stronger the bending degree of curve would be. In order to guarantee the monotonicity of changing angle in the interval of $[0, \pi]$, the cos function is adopted ($\cos(\angle P_i P_{mid} P_{k_2})$). In Eq. (3) the enclosed area $f_3^t(i)$ of shape contour can reflect the contour interior changing trend (see in Fig. 2 (d)), the enclosed area means that, if the pixel is located in the enclosed area surrounded by the shape contour curve and the line of two neighboring points, then the pixel would be calculated as one; otherwise it would be zero, the value of enclosed area is the summation of all the calculated pixels. In Eq. (4) the eccentricity $f_4^t(i)$ of formed ellipse can reflect the flat degree of segmented ellipse zone (see in Fig. 1 (a)), the larger the value of eccentricity is, the more flat the ellipse is, this description can verified that the size variation of coverage ellipse zone has the relationship with shape contour curve distribution.

In this part, the invariant properties of our proposed shape descriptor are analyzed. It should be mentioned that, four invariant features have been normalized by dividing the maximum of feature itself to make sure the invariance of scale transform. In order to be robustness of rotation transform, the magnitudes of their Fourier transform coefficients are computed as

$$F_g^t(m) = \frac{1}{N} \left| \sum_{i=0}^{N-1} f_g^t(i) exp(\frac{-j2\pi im}{N}) \right| \quad (5)$$

where $g = 1, 2, 3, 4$; m is the order of Fourier transform ($m = 0, 1, \cdots, N-1$)(specific description see in [2]). Finally the multiscale ellipse descriptor (MED) of shape contour information $\varphi_g^t = \{F_g^t(m), \delta_g^t\}$ is the combination of Fourier magnitude coefficients $F_g^t(m)$ and corresponding standard deviation δ_g^t of invariant ellipse features (designed for enhancing the discriminative ability).

2.3. Shape dissimilarity measurement

Given two shapes A and B with their corresponding MED descriptors $\varphi(A) = \{F_g^{(A)t}(m), \delta_g^{(A)t}\}$ and $\varphi(B) = \{F_g^{(B)t}(m), \delta_g^{(B)t}\}$, the dissimilarity matching measurement is computed based on the Euclidean distance of their MED descriptors as:

$$D(A, B) = \sum_{t=1}^{\log N-1} \sum_{m=0}^{M-1} \sum_{g=1}^{4} |w_g(\varphi_g^{(A)t}(m) - \varphi_g^{(B)t}(m))| \quad (6)$$

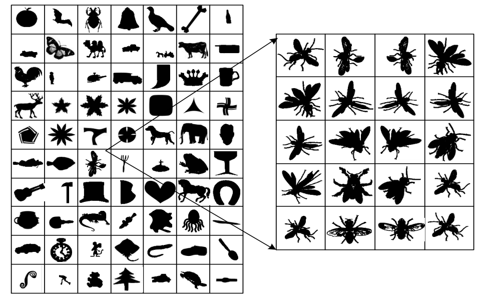


Fig. 3. Shape examples of MPEG-7 shape dataset

where m is the order of Fourier transform coefficient, $w_g(g = 1, 2, 3, 4)$ are the corresponding weight values to judge the contribution of each invariant ellipse features. To further improve the dissimilarity measurement,inspired by [18][19], we use model shapes as each other's contexts in propagation to lower the distance caused by intra-class variance.

3. EXPERIMENTAL RESULTS

To evaluate the effectiveness and efficiency of proposed MED descriptor, the experiments are conducted on two widely used MPEG-7 shape dataset [3] [2] [8] [11][20] and leaf100 dataset [2]. In the following experiments, the total number of shape contour sample points is $N = 256$ and each corresponding weight values of invariant ellipse features are $w_1 = 0.28, w_2 = 0.15, w_3 = 0.25, w_4 = 0.15$ (the best parameters computed from the experiments).

3.1. MPEG-7 shape dataset

The MPEG-7 CE-1 Part B shape dataset is widely used to evaluate the performance of proposed method in shape retrieval. This dataset has 1400 shape images, belonging to 70 classes of various shapes with 20 images in each class (see in Fig. 3). The retrieval accuracy is measured by the well-known "bulls-eye" score, in this measurement, each shape is used as a query image to match with all other shapes. The number of shape which belongs to the same class of query image is counted among the 40 most similar shapes, the maximum number for one query image is 20. Therefore, the bulls-eye score is the ratio of total matched shapes to the possible maximum numbers (1400 $\times$ 20).

In Table 1, our proposed method is compared with other state-of-the-art methods. It can be seen that the proposed method achieves highest accuracy of 98.07%, which is separately 1.64% higher than HG+Mutual graph [20], 3.56% higher than IMD +LP [3], 8.41% higher than HF [8] and 10.76% higher than HSC [2]. These results indicate that, the proposed MED descriptor not only has the better discriminative ability to distinguish the different classes, but also is robustness to the linear and non-linear transforms in the same class.

shape descriptors	bulls-eye scores(%)
MDM [4]	69.19
CSS [15]	75.44
MCC [16]	84.93
IDSC+DP [10]	85.40
SC+DP [9]	86.80
TAR+DP [21]	87.13
HSC [2]	87.31
Shape Tree [12]	87.70
Height Function [8]	89.66
IMD+LP [3]	94.51
HG+Mutual graph [20]	96.43
Our MED	**98.07**

Table 1. Bulls-eye scores conducted on the MPEG-7 dataset

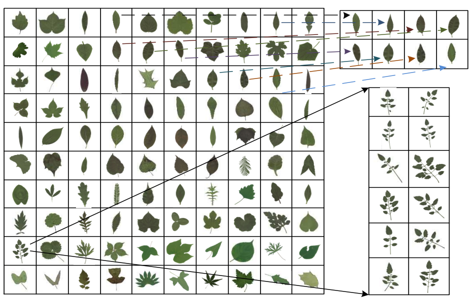


Fig. 4. Shape examples of the leaf100 dataset

3.2. Leaf100 dataset

The real and challenging leaf100 dataset is widely used to validate the efficiency of shape descriptor, it contains 1200 leaf images separated into 100 plant species (12 different images in each class). Different from MPEG-7 shape dataset, the challenge of this dataset lies in the high inter-class similarity between some leaf shapes and large intra-class variations of some classes (see in Fig. 4). The same performance measurement as used in MPEG-7 shape dataset is adopted in this leaf100 dataset.

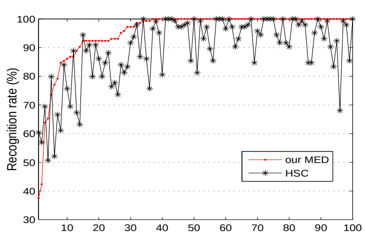


Fig. 5. The recognition rates of each class on leaf100 dataset

Table 2 shows the comparative results of the proposed MED descriptor and other benchmarks. It can be seen that the proposed descriptor achieves the highest accuracy of 95.36%,

shape descriptors	bulls-eye scores(%)	computational complexity	matching time (s)
MDM [4]	64.84	-	-
MCC [16]	77.10	$\mathcal{O}(N^3)$	-
TAR+DP [21]	77.66	$\mathcal{O}(N^3)$	-
IDSC+DP [10]	85.64	$\mathcal{O}(KN^2)$	21.0
SC+DP [9]	86.82	$\mathcal{O}(KN^2)$	24.2
Height Function [8]	87.81	$\mathcal{O}(N^2)$	32.8
HSC [2]	89.40	$\mathcal{O}(M \log N)$	0.226
our MED	**95.36**	$\mathcal{O}(M \log N)$	**0.326**

Table 2. Computational results conducted on the leaf100 dataset

which is separately 5.96% higher than HSC [2], 7.55% higher than HF [8] and 8.54% higher than SC+DP [9]. These results consistently verify the effectiveness of our proposed MED descriptor (in [3][20] don't have directly results on leaf100 dataset). To further analysis the performance of proposed descriptor, Fig. 5 shows the specific recognition rates of each class on the leaf100 dataset, most of the result values in our descriptor are better than HSC. This indicates that the ability of the overall description in this method has been strengthened, which can be used in more challenging dateset.

Here we only discuss the computational complexity of matching measurement which plays more important part in big shape data retrieval. The complexity of calculating Eq. (6) is $\mathcal{O}(4(M+1)(\log N - 1))$, where 4 represents the four invariant ellipse descriptors, M is the order of Fourier coefficient and N is total number of shape sample points (closed to $\mathcal{O}(M \log N)$). In Table 2, we compare our MED descriptor with other benchmarks about computational complexity and corresponding CPU time. It can be seen that, our proposed MED method still have the advantage of lower computational complexity and matching time (both slightly higher than HSC but still acceptable). Note that the experimental platform is CPU Core(TM) i3-3110M CPU 2.40GHz and matlab R2012b.

4. CONCLUSIONS

In this paper, we propose a novel multiscale ellipse descriptor (MED) method for shape retrieval. In each scale both spatial location relationship between sample points and topology structure information of segmented multiscale zone, are used to extract the coarse-to-fine shape details. Meanwhile this descriptor is robustness to various linear and non-linear transforms. Experimental results conducted on the MPEG-7 shape dataset and leaf100 dataset demonstrate that, this method is effective and efficient in shape retrieval compared with most state-of-art methods.

5. REFERENCES

[1] Sofiene Mouine, Itheri Yahiaoui, and Anne Verroust-Blondet, "A shape-based approach for leaf classification using multiscale triangular representation.," in *ACM Conference on International Conference on Multimedia Retrieval*, 2013, pp. 127–134.

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

[3] Jianyu Yang, Hongxing Wang, Junsong Yuan, Youfu Li, and Jianyang Liu, "Invariant multi-scale descriptor for shape representation, matching and retrieval," *Computer Vision and Image Understanding*, vol. 145, no. C, pp. 43–58, 2016.

[4] Rongxiang Hu, Wei Jia, Haibin Ling, and Deshuang Huang, "Multiscale distance matrix for fast plant leaf recognition," *IEEE Transactions on Image Processing*, vol. 21, no. 11, pp. 4667–72, 2012.

[5] Mingqiang Yang, Kidiyo Kpalma, and Joseph Ronsin, "A survey of shape feature extraction techniques," vol. PP, no. 43-90, 2008.

[6] Mohammad Reza Daliri and Vincent Torre, "Robust symbolic representation for shape recognition and retrieval," *Pattern Recognition*, vol. 41, no. 5, pp. 1782–1798, 2008.

[7] Xinggang Wang, Bin Feng, Bai Xiang, Wenyu Liu, and Longin Jan Latecki, "Bag of contour fragments for robust shape classification," *Pattern Recognition*, vol. 47, no. 6, pp. 2116–2125, 2014.

[8] Junwei Wang, Xiang Bai, Xinge You, Wenyu Liu, and Longin Jan Latecki, "Shape matching and classification using height functions," *Pattern Recognition Letters*, vol. 33, no. 2, pp. 134–143, 2012.

[9] Serge Belongie, Jitendra Malik, and Jan Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 24, pp. 509–522, 2002.

[10] Haibin Ling and David W. Jacobs, "Shape classification using the inner-distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 286–299, 2007.

[11] Natsuda Kaothanthong, Jinhee Chun, and Takeshi Tokuyama, "Distance interior ratio: a new shape signature for 2d shape retrieval," *Pattern Recognition Letters*, vol. 78, pp. 14–21, 2016.

[12] P. F Felzenszwalb and J. D Schwartz, "Hierarchical matching of deformable shapes," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[13] G Mcneill and S Vijayakumar, "Hierarchical procrustes matching for shape retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 885–894.

[14] Xiang Bai, Cong Rao, and Xinggang Wang, "Shape vocabulary: A robust and efficient shape representation for shape matching," *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 3935–49, 2014.

[15] Sadegh Abbasi, Farzin Mokhtarian, and Josef Kittler, "Curvature scale space image in shape similarity retrieval," *Multimedia Systems*, vol. 7, no. 6, pp. 467–476, 1999.

[16] T Adamek and N. E O'Connor, "A multiscale representation method for nonrigid shapes with a single closed contour," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 5, pp. 742–753, 2004.

[17] Jie Cao, Bin Wang, and Douglas Brown, "Similarity based leaf image retrieval using multiscale r-angle description," *Information Sciences*, vol. 374, pp. 51–64, 2016.

[18] Herve Jegou, Cordelia Schmid, Hedi Harzallah, and Jakob Verbeek, "Accurate image search using the contextual dissimilarity measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 2–11, 2010.

[19] Xiang Bai, Bo Wang, Cong Yao, Wenyu Liu, and Zhuowen Tu, "Co-transduction for shape retrieval," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2747–2757, 2012.

[20] Cong Yang, Christian Feinen, Oliver Tiebe, Kimiaki Shirahama, and Marcin Grzegorzek, "Shape-based object matching using interesting points and high-order graphs," *Pattern Recognition Letters*, vol. 83, pp. 251–260, 2016.

[21] N. Alajlan, I El Rube, M. S. Kamel, and G. Freeman, "Shape retrieval using triangle-area representation and dynamic space warping," *Pattern Recognition*, vol. 40, no. 7, pp. 1911–1920, 2007.