Multi-Scale Embedded Descriptor for Shape Classification

Chen Huang[a], Tony X. Han[a], Zhihai He[a,*]

[a]*Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO, USA, 65211*

Abstract

We present a new shape descriptor that are robust to deformation and capture part details. In our framework, the shape descriptor is generated by 1) using running angle to transforming a shape into a 2-D description image in the position and scale space. 2) performing circular wavelet-like sub-band decomposition of this 2-D description image based on its periodic convolution with orthogonal kernel functions. Each sub-band is described by the histogram of its decomposition coefficients. To capture unique and discriminative part, we compare the decomposition coefficients across sub-band to detect singularity in the position and scale space. The singularity information is encoded with a tree of binary bits. The coded feature vectors of all sub-bands and singularity trees are pooled together to form the descriptor of the shape. The shapes are classified with linear SVM. Our performance evaluations on several public datasets, demonstrating that the proposed method significantly outperforms state-of-the-art methods.

Keywords: shape descriptor, running angle, wavelet transform, object recognition

*Corresponding author

Email addresses: chenhuang@mail.missouri.edu (Chen Huang), hantx@missouri.edu (Tony X. Han), HeZhi@missouri.edu (Zhihai He)

1. Introduction

Shape description, matching and classification are fundamental problems in computer vision with important applications like image retrieval, object detection and recognition. Compared to other image features, shape is invariant to lighting conditions and changes in object colors and texture [1]. However, shape contours obtained from object segmentation often exhibits a large degree of intra-class of variations due to different view points, changing illumination, segmentation errors, shape parts articulation variations, non-rigid deformations, etc [2]. Furthermore, shape contours of objects often have strong inter-class ambiguities. For example, tree leaves or animals of similar species, are often very similar to each other, except some small or even tiny distinguishable features embedded in a large amount of intra-class shape contour variations.

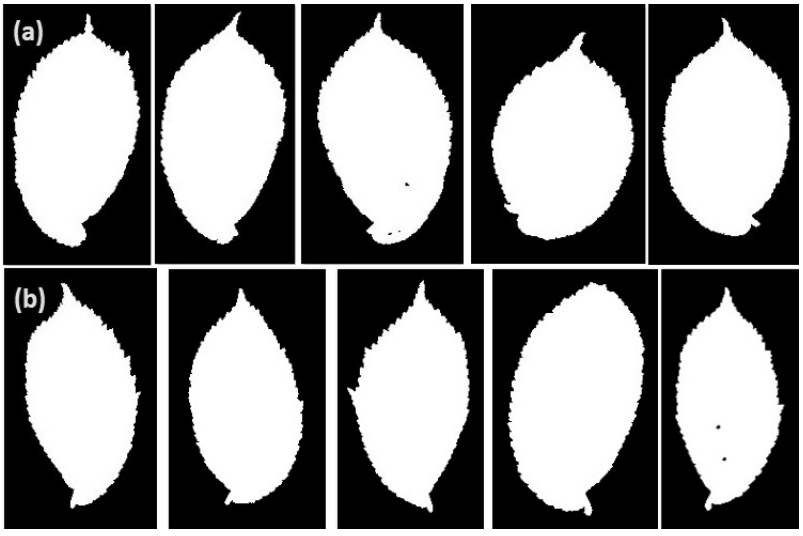

Figure 1: Two groups of leaves, Group-1 (the first row) and Group 3 (the second row) from the Swedish Leaf Database [3].

For example, Fig. 1 shows two types (Group-1 and Group-3) of leaves from the Swedish Leaf Database [3]. We can see that the inter-class ambiguity is very strong. Human eyes, even experts, cannot tell them apart.

In the literature, a number of efficient shape descriptors and shape similarity measures have been developed for representing, matching, classifying, and recognizing shapes. One typical approach for shape similarity measure is to construct some physical shape models and then measure the amount of energy required to deform one shape contour into another [4]. Latecki et al. [5] developed an elastic partial shape matching algorithm to model a possible non-rigid shape deformation. A hierarchical matching approach has been developed in [6]. It has been observed that this type of approaches are often very sensitive to strong, local shape variations. Extracting invariant descriptors of the shape is another important approach. In order to capture the local features, Belongie and Malik [7] introduce the 2-dimensional non-linear histogram, Shape Context(SC), to describe the distance and angles between contour points. Since SC cannot solve the problem of matching articulated shape, Ling and Jacobs [8] modified SC by using the geodesic distance inner shape instead of Euclidean distance to represent shapes, which is called Inner Distance Shape Context(IDSC). Bending invariants [9] for 2D and 3D shapes can be achieved by using geodesic distances. For both 2D and 3D shape analysis, topology invariants have been developed in [10]. To capture the inherent part structure of the shape contour, skeleton based approaches, particularly the shock graph method in [2], have been developed. Given a shape and its boundary, shocks are defined as curve segments of the medial axis with monotonic flow. The shocks are then organized into a shock graph, which forms a hierarchical representation of the shape and naturally captures its part structure. These

invariant descriptors have demonstrated their effectiveness in handling intra-class variations. However, as observed in [11], invariants to larger groups of deformation often come at a price of reduced inter-class discriminative.

Apart from the existing efforts in the literature to construct shape descriptors and similarity measures and then use nearest neighbor approaches for shape classification [2, 6, 12, 8, 13], in this work, we propose a global shape descriptor and a learning-based approach for shape classification, aiming to develop trainable classifiers for shape classification and recognition. During the past several years, we have witnessed the success of data-driving and machine learning approaches in computer vision applications, e.g., object detection, classification, and recognition [14, 15]. These approaches are able to discover and construct important features to resolve inter-class ambiguities during classification. We believe that this learning and discovery capability is also important and beneficial in shape-based domain. However, we observe that shapes are much different from conventional image or video data for object detection and recognition. Shape contour is basically a 1-D data. An enclosed contour has no starting point. The shape feature should be invariant or insensitive under changes in scale, rotations, and non-rigid deformations. How to extract low-level from shapes for successful learning and training of accurate and robust shape classifiers remains a challenging and interesting research problem.

In this work, we propose to develop a global multi-scale embedded description scheme for shape classification. We are mainly motivated by the following observation: our human vision system often examines and compares shapes at different scales. Some classes of shapes can be well separated by shape features at coarse scales. However, for shapes of some closely related classes, we have to

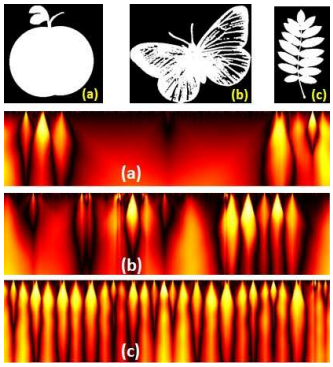

Figure 2: We first transform a shape contour into a 2-D shape descriptor image in the geodesic index and scale space. (a), (b), (c) is the original shape contour, and below are their corresponding 2-D shape descriptor image

examine or discover detailed features at fine scales. To develop the multi-scale embedded shape description, we construct a dense description of shapes or object contours using a low-level feature called running angle, transforming the whole shape into a 2-D shape description image in the geodesic index and scale space, as Fig. 2. We then perform circular wavelet-like sub-band decomposition of this shape descriptor image based on its periodic convolution with orthogonal kernel functions. We then extract invariant features from these decomposition coefficients and pool them across sub-band to form a dense description of the shape. The shape is classified with linear SVM. Our performance evaluations on public datasets demonstrate that the proposed method significantly outperforms state-of-the-art methods.

The rest of the paper is organized as follows. Section 2 presents the 2-D shape description image. Section 3 presents our global multi-scale shape descriptor and the SVM-based shape classification framework. The experimental results are pre-

sented in Section 4. Section 5 concludes the paper.

2. Multi-Scale 2-D Shape Description Image and Circular Sub-band Decomposition

In this work, we focus on enclosed shape contours. We uniformly sample the contour and denote this sequence of point samples by

$$\mathbf{C} = \{\mathbf{C}(s) = (x_s, y_s) | 1 \leq s \leq N\}. \tag{1}$$

s is the geodesic distance along the curve from the starting position. Later on, we will see that our algorithm does not depend on this starting position.

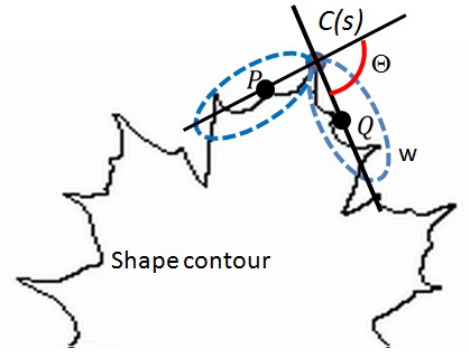

Figure 3: 2-D shape descriptor image.

Next, we will compute the so-called running angle at each contour sample. As illustrated in Fig. 3, at point $\mathbf{C}(s)$, we consider a looking-out window of size w, which consists of w samples before $\mathbf{C}(s)$ and w samples after $\mathbf{C}(s)$. We compute the centroids of contour samples at both sides as follows:

$$\mathbf{P}(s,w) = \sum_{[s-w]_N}^{[s-1]_N} \mathbf{C}(s), \quad \mathbf{Q}(s,w) = \sum_{[s+1]_N}^{[s+w]_N} \mathbf{C}(s). \tag{2}$$

The running angle of the contour at index s with a looking-out window size w is defined to be the angle between vectors $\mathbf{P}(s,w) - C(s)$ and $C(s) - \mathbf{Q}(s,w)$. We denote this angle by $\theta(s,w)$. When the window size w is large, the running angle θ will capture more global information around the point $\mathbf{C}(s)$. When w is small, θ will focus on local and more detailed structures of the shape. This represents a 2-D image containing multi-scale information, with each image point taking a value between $[0, 2\pi]$. We refer to this image as the running angle image. Fig. 2 shows the running angle images for three contour samples from the MPEG-7 shape dataset. We can see that, for different shapes, their running angle images are drastically different. This running angle image aims to capture local shape structures at different scales.

To extract salient features from this running angle image to characterize the shape contour, we propose to perform circular subband decomposition. Continuous subband decomposition or wavelet analysis on the enclosed curve have been studied by Holschneider [16]. We use two bi-orthogonal wavelet kernel functions, the low-frequency $G_0(s)$ and high-frequency $G_1(s)$. First, we perform horizontal circular subband decomposition. As illustrated in Fig. 4(b), for each w, we perform circular convolution of $\theta(s,w)$ in the s domain using these two kernel functions and then down sample the output by 2 as following low-frequency and high-frequency subbands:

$$L(w,s) = [\theta(s,w) \otimes_N G_0(s)] \downarrow 2, \quad (3)$$

$$H(w,s) = [\theta(s,w) \otimes_N G_1(s)] \downarrow 2. \quad (4)$$

Here, $\otimes_N$ represents the circular convolution with a period of N. This circular convolution attempts to remove the effect of starting point selection on the final shape description. We then repeat this subband decomposition procedure on the

low-frequency subband $L(w, s)$ for M times to create M high-frequency subbands plus one low-frequency subband. The subband structure is shown in Fig. 4(c). We denote this subband decomposition output by Θ_H. During our experiments, we observe that this horizontal multi-scale subband decomposition along the geodesic index s allows us to capture sudden changes on the contour, which often correspond to singularity structures and distinctive features for shape classification.

After the horizontal circular subband decomposition in the geodesic index domain, we then apply vertical non-circular subband decomposition to Θ_H in the scale w-domain, which yields the subband structure as illustrated in Fig. 4(d). Fig. 4(e) shows one example of the final decomposition output for the running angle images in (a).

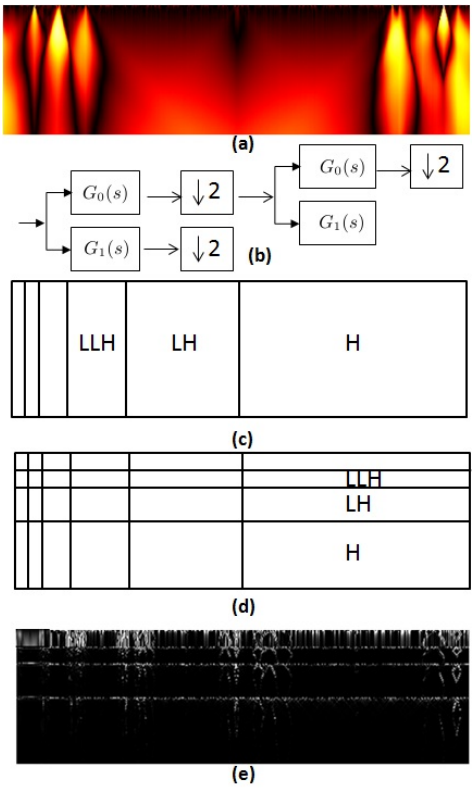

Figure 4: (a) the running angle image of one example shape contour; (b) the subband decomposition scheme; (c) horizontal circular subband decomposition structure; (d) vertical subband decomposition structure; (e) the final decomposition coefficients for (a).

3. Embedded Shape Description and SVM Classification

In this section, we explain how a dense description is extracted from the subband decomposition of the running angle image. We denote the subbands by $\mathcal{B}_{i,j}$, where $1 \leq i \leq I$ is the index for the vertical subband in the scale domain, $1 \leq j \leq J$ is the horizontal subband index in the geodesic distance domain. Note that, within each subband $\mathcal{B}_{i,j}$, each row, denoted by $\mathcal{B}_{i,j}(k, \cdot)$ has information about the whole shape contour. To ensure that the feature is shift-invariant and

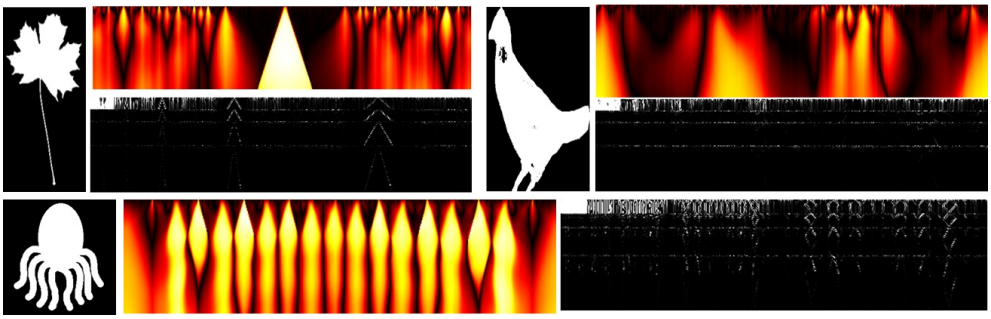

Figure 5: (a) the running angle image of one example shape contour; (b) the subband decomposition scheme; (c) horizontal circular subband decomposition structure; (d) vertical subband decomposition structure; (e) the final decomposition coefficients for (a).

rotation-invariant, we propose to use the histogram features. More specifically, we quantize the decomposition coefficients in $\mathcal{B}_{i,j}(k,\cdot)$ into 8 bins and compute its histogram $\mathcal{H}_{i,j}(k,\cdot)$. As illustrated in Fig. 6, the histogram vectors of all rows of all subbands are cascaded together to build a high-dimensional feature vector denoted by $\mathcal{H}$ to describe the shape contour.

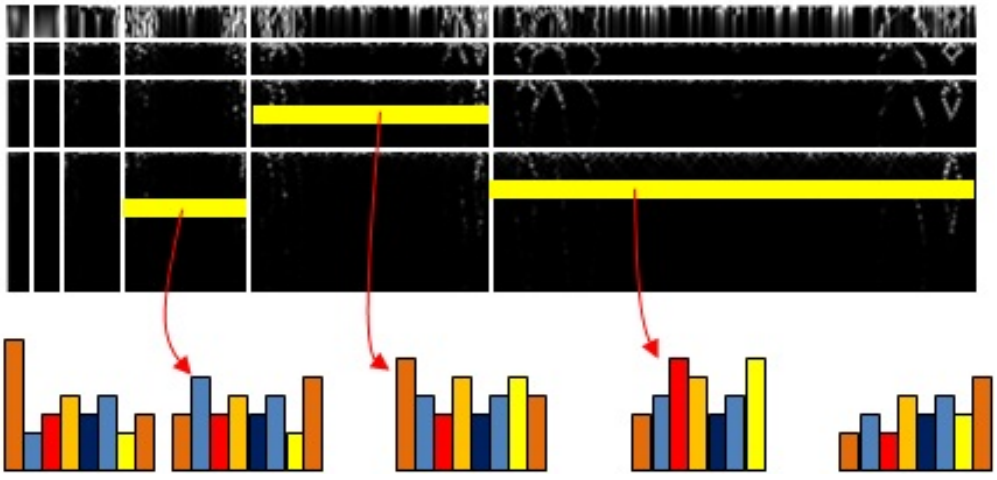

Figure 6: the histogram vectors of all rows of all subbands are cascaded together to build a high-dimensional feature vector.

To further capture how the shape structure change over different scales, we introduce a subband feature called *singularity tree*. As illustrated in Fig. 7, the subband decomposition has an inherent parent-children tree structure.

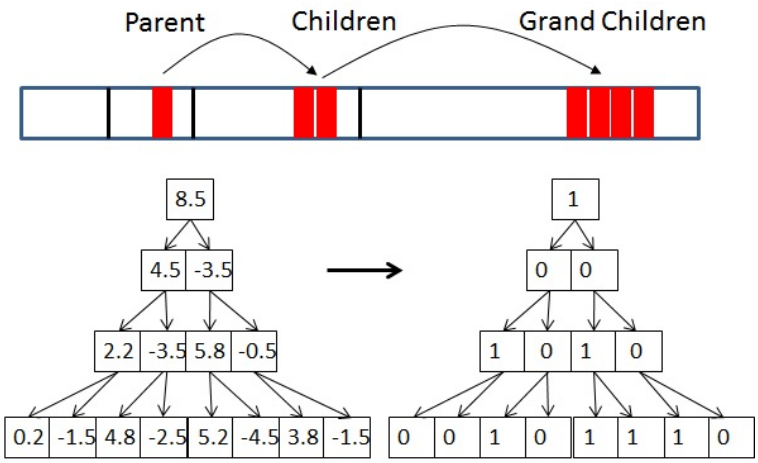

Figure 7: 2-D shape descriptor image.

For example, as illustrated in Fig. 7, in subbands $\mathcal{B}_{i,j}$ and $\mathcal{B}_{i,j+1}$, in the k-th row, the coefficient $\mathcal{B}_{i,j}(k,m)$ is the parent node of the two children nodes, $\mathcal{B}_{i,j+1}(k,2m)$ and $\mathcal{B}_{i,j+1}(k,2m+1)$, which in turn have four grandchildren nodes in subband $\mathcal{B}_{i,j+2}$: $\mathcal{B}_{i,j+2}(k,4m)$, $\mathcal{B}_{i,j+2}(k,4m+1)$, $\mathcal{B}_{i,j+2}(k,4m+2)$, and $\mathcal{B}_{i,j+2}(k,4m+3)$. This results in a binary tree, as shown in Fig. 7. Let $\Omega[\mathcal{B}_{i,j+1}(k,m)]$ be the subtree of $\mathcal{B}_{i,j+1}(k,m)$, the set of children nodes with $\mathcal{B}_{i,j+1}(k,m)$ as the node. We assign a binary label to each tree node as follows.

$$\begin{aligned}\mathcal{L} \quad & [\mathcal{B}_{i,j+1}(k,m)] \\ = \quad & \begin{cases} 1, & if\ \mathcal{B}_{i,j+1}(k,m) > max\{\Omega[\mathcal{B}_{i,j+1}(k,m)]\} \\ 0, & elsewhere. \end{cases}\end{aligned}$$

The leaf node without any children are assigned lable 1 if it is above a given threshold Δ, otherwise, it is assigned label 0. Thus, we have a binary tree of

labels for each horizontal low-frequency subband coefficients at every row of the subband image. We scan this tree in the breadth first order and convert it into a binary vector, denoted by $\mathcal{L}_i(m)$, $1 \leq m \leq M$, where M is the number of coefficients is the horizontal low-frequency subband. We pool these binary vectors across all low-frequency subband coefficients to form an embedded singularity description:

$$\mathcal{L}_i = \uplus_{m=1}^{M} \mathcal{L}_i(m), \tag{5}$$

where $\uplus$ represents the OR operation on each dimension of the binary vector. The $\mathcal{L}_i$ vector of all rows are cascaded together to form the embedded singularity feature denoted by $\mathcal{S}$. The overall shape description is given by $[\mathcal{H}, \mathcal{S}]$ by cascading the subband histogram feature vector and the embedded singurlarity feature. With this shape descriptor, we use the linear SVM to train the shape classifier.

4. Experimental Results

4.1. implementation detail

In our experiment, the number of sampling points in the shape contour is 512, which means $1 <= s <= 512$. Maximum window size is 128, so $1 <= w <= 128$. the 2-D description image's resolution is $128 * 512$.

Given the 2-D description image $\theta(s, w)$, we do horizontal sub-band decomposition for 5 times, and vertical sub-band decomposition for 3 times. The final dimension for histogram is 6144, because we have $6*128$ $\mathcal{B}_{i,j+1}(k, .)$ in total, each sub-band, each row $\mathcal{B}_{i,j+1}(k, .)$ will generate a 8-dimensional small histogram, the total histogram will have $6 * 128 * 8 = 6144$ bins

Table 1: Classification rate on the Swedish Leaf dataset.

Method	Recognition rate
This Work	**98.40**%
Shape-tree [6]	96.28%
IDSC + DP [8]	94.13%
SC + DP [8]	88.12%
Fourier descriptors [8]	89.60%
Soderkvist [3]	82.40%

4.2. Datasets

We evaluate our method using the following two public shape benchmark datasets. The Swedish Leaf dataset [3] has shape images of 15 species of leaves, with 75 images per species for a total of 1125 images. Fig. 8 shows some example shapes from this dataset. Note that some species are close to each other with significant inter-class ambiguity. Similar to the methods in [8] and [6], we randomly select 25 training images from each species and classify the remaining images using the proposed SVM-based approach. We compare the proposed method with the following methods developed in the literature: (a) **Shape-tree**, the hierarchical tree matching method by [6]; (b) **IDSC+DP**, the inner-distance shape context plus dynamic programming approach [8]; (c) **SC+DP**, the shape context plus dynamic programming approach [8]; and (d) Soderkvist's method [3]. Table 1 shows the shape classification accuracy. The proposed method achieves a classification rate of 98.40%, outperforming all other methods. Table 2 shows the confusion matrix. We can see except few cases in classes 1, 3, 9, all other shapes are correctly classified.

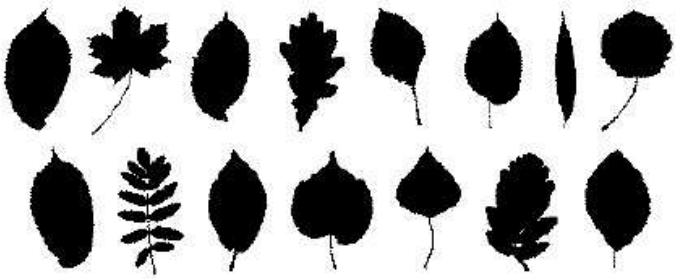

Figure 8: Samples from the Swedish Leaf Dataset, one leaf per species. Note the similarity among some species

Figure 9: Samples from the MPEG-7 shape dataset. One image per class for the first 40 classes (the database has 70 classes)

The MPEG-7 shape database [12] is widely used for evaluating shape recognition methods. The database has 1400 silhouette images, with 20 images per object class from a total of 70 different classes. Fig. 9 shows some sample shape images from this database. This is a challenging dataset due to the large intra-class variations. The standard method (often referred to as bull's eye approach) for evaluting the recognition rate of an algorithm in this dataset is as follows. For every image in the database, it finds the 40 most similar images and count how many of those are in the same class as the query image. The final score of the test is the number of correct hits divided by the class size minus one, which is 19. In other words, if we have 19 matches in the top 40, the accuracy will be 100%. Note that our

Table 2: Confusion matrix for shape classification on the Classification rate on the Swedish Leaf dataset.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	25	0	1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	22	0	0	0	0	0	1	0	0	0	0	0	0
4	0	0	0	25	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	25	0	0	0	0	0	0	0
9	0	0	2	0	0	2	0	0	24	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	25	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	25	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	25	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	25	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	25	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25

approach is training-based. One major issue with the MPEG-7 dataset is that the number of available samples in each class is small, 20. To provide reasonable comparison, we use the SVM classification score as the shape distance. In our experiments, we use 10 samples for training, and 10 for testing. We try to follow the standard test protocol: the classification rate reported in Table 3 is the number of correct hits in the top 20 matches divided by 10. Our approach has achieved the classification rate of 92.88% which is very close the best performance reported in the literature. We believe that this classification rate will be significantly improved when the training data size becomes larger. Currently, it is just 10. This performance comparison is just to demonstrate the effectiveness of our shape descriptor and the shape classification performance on the MPEG-7 dataset, although we do recognize that this performance is not totally fair, given that the criteria in the standard method is relatively loose. Another factor is that the proposed approach is a trained classifier, while other methods are based on distance measurements and using k-nearest neighbors as classification.

5. Concluding Remarks

In this paper, we have developed a dense description of shape contours using a low-level feature called running angle, transforming a shape into a 2-D shape description image in the geodesic distance and scale space. We performed circular wavelet-like subband decomposition of this shape descriptor image based on its periodic convolution with orthogonal kernel functions. Each subband is described by the histogram of its decomposition coefficients. To further capture unqiue and distinguished features of the shape, we compared the decomposition coefficients across subband to detect singularity in the position and scale space. The coded

Table 3: Classification rate on the MPEG7 shape dataset.

Method	Classification rate
IDSC+LCDP [17]	95.60%
This Work	**92.88**%
IDSC+LCDP [18]	92.36%
Contour Flexibility [19]	89.31%
IDSC [17]	88.39%
Shape-tree [6]	87.70%
Triangle Area [20]	87.23%
Multiscale Representation [21]	84.93%

feature vectors of all subbands and singularity trees are pooled together to form a dense representation of the shape. The shape is classified with linear SVM. Our performance evaluations on public datasets demonstrate that the proposed method significantly outperforms state-of-the-art methods.

6. Reference

[1] J. Shotton, A. Blake, R. Cipolla, Multiscale categorical object recognition using contour fragments, Pattern Analysis and Machine Intelligence, IEEE Transactions on 30 (2008) 1270–1281.

[2] T. B. Sebastian, P. N. Klein, B. B. Kimia, Recognition of shapes by editing their shock graphs, Pattern Analysis and Machine Intelligence, IEEE Transactions on 26 (2004) 550–571.

[3] O. Söderkvist, Computer vision classification of leaves from swedish trees (2001).

[4] J. Coughlan, A. Yuille, C. English, D. Snow, Efficient deformable template detection and localization without user initialization, Computer Vision and Image Understanding 78 (2000) 303–319.

[5] L. J. Latecki, V. Megalooikonomou, Q. Wang, D. Yu, An elastic partial shape matching technique, Pattern Recognition 40 (2007) 3069–3080.

[6] P. F. Felzenszwalb, J. D. Schwartz, Hierarchical matching of deformable shapes, in: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, IEEE, 2007, pp. 1–8.

[7] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, Pattern Analysis and Machine Intelligence, IEEE Transactions on 24 (2002) 509–522.

[8] H. Ling, D. W. Jacobs, Shape classification using the inner-distance, Pattern Analysis and Machine Intelligence, IEEE Transactions on 29 (2007) 286–299.

[9] A. Elad, R. Kimmel, On bending invariant signatures for surfaces, Pattern Analysis and Machine Intelligence, IEEE Transactions on 25 (2003) 1285–1295.

[10] A. M. Bronstein, M. M. Bronstein, R. Kimmel, M. Mahmoudi, G. Sapiro, A gromov-hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching, International Journal of Computer Vision 89 (2010) 266–286.

[11] H. Ling, X. Yang, L. J. Latecki, Balancing deformability and discriminability for shape matching, in: Computer Vision–ECCV 2010, Springer, 2010, pp. 411–424.

[12] L. J. Latecki, R. Lakamper, T. Eckhardt, Shape descriptors for non-rigid shapes with a single closed contour, in: Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on, volume 1, IEEE, 2000, pp. 424–429.

[13] Y. Cao, Z. Zhang, I. Czogiel, I. Dryden, S. Wang, 2d nonrigid partial shape matching using mcmc and contour subdivision, in: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE, 2011, pp. 2345–2352.

[14] Y. Jia, C. Huang, T. Darrell, Beyond spatial pyramids: Receptive field learning for pooled image features, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 3370–3377.

[15] Q. V. Le, Building high-level features using large scale unsupervised learning, in: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, IEEE, 2013, pp. 8595–8598.

[16] M. Holschneider, Wavelet analysis on the circle, Journal of Mathematical Physics 31 (1990) 39–44.

[17] A. Temlyakov, B. C. Munsell, J. W. Waggoner, S. Wang, Two perceptually motivated strategies for shape classification, in: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE, 2010, pp. 2289–2296.

[18] X. Yang, S. Koknar-Tezel, L. J. Latecki, Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval, in: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, 2009, pp. 357–364.

[19] C. Xu, J. Liu, X. Tang, 2d shape matching by contour flexibility, Pattern Analysis and Machine Intelligence, IEEE Transactions on 31 (2009) 180–186.

[20] N. Alajlan, M. S. Kamel, G. H. Freeman, Geometry-based image retrieval in binary image databases, Pattern Analysis and Machine Intelligence, IEEE Transactions on 30 (2008) 1003–1013.

[21] T. Adamek, N. E. O'Connor, A multiscale representation method for non-rigid shapes with a single closed contour, Circuits and Systems for Video Technology, IEEE Transactions on 14 (2004) 742–753.

[22] E. Begelfor, M. Werman, Affine invariance revisited., in: CVPR (2), 2006, pp. 2087–2094.

[23] D. Bryner, A. Srivastava, E. Klassen, Affine-invariant, elastic shape analysis of planar contours, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 390–397.