Contour-Based Similarity Retrieval of Trademarks Using Ellipse-Ring Layers

Nai-Chung Yang, Chung-Ming Kuo, Wei-Han Chang, Chih-Wei Chen
Dept. of Information Engineering, I-Shou University
E-Mail: ncyang@isu.edu.tw

ABSTRACT

This paper focuses on solving asymmetrical deformation problem of retrieving, and introduces a contour-based shape descriptor for retrieving both simple and complex shaped-images. In this method, we use the morphological operation to extract the boundary points from binary shape images. Then, the extracted points are partitioned into three groups according to Principal Component Analysis. We compute the statistics properties of the distribution information for each group. This method is rotation, scaling and translation invariant. Experimental results show that the proposed method is robust and efficient. In addition, this method can effectively overcome the limitation of contour-based shape descriptors.

1: INTRODUCTION

In recent years, there is increasing interest in multimedia content description for content-based image retrieval [1-12]. The multimedia standard MPEG-7 is developed to address this problem. In MPEG-7, six criteria have been specified for an ideal shape descriptor, which includes good retrieval accuracy, small memory, compact features, general applications, small computational cost and hierarchical coarse-to-fine structure. Various shape descriptors have been developed in literatures; we can broadly categorize them into two types: contour-based [12-16] and region-based shape descriptors [1, 3, 7-10].

Contour-based shape descriptors extract the boundary information, but they don't preserve the information of interior contents of objects. For example, Fourier descriptor, wavelet descriptor, curvature scale space (CSS) [12] are typical contour-based shape descriptors. Contour-based shape descriptors cannot deal with disjoint shapes; therefore, its applications are usually limited to describe single closed-contours. On the contrary, region-based shape descriptors collect all the pixels within a shape region instead of boundary information only. They can capture the interior contents of shapes. This type of descriptors can describe non-connected and complex shapes. As a result, the region-based shape descriptors are more popular than the contour-based shape descriptors at the present time.

There are several desirable advantages on contour-based method such as small memory, compact features, low computational cost and compatible with human perception. In this paper, we aim to overcome the limitation of conventional contour-based method.

In this work, we use the morphological operations to extract the boundaries from shape images. In principle, the extracted boundaries are very similar to human perception because it is incorporated important features by one or several contours. When the boundary points have been extracted, the shape features could be determined. To preserve scaling invariant, we compute the distance between each boundary point to centroid of image region, and then normalize the distances. After this procedure, we propose a layer-partition concept which divides the boundary points into several groups based on their distances. The details are presented in the following section. It will be shown that this method not only overcomes the limitation of conventional contour-based method but also achieves higher retrieval accuracy than that of conventional region- or contour-based methods.

In this paper we first present a brief overview of the proposed method. In Section 2, we explain the rationale for using visual features and details how they are computed. Results are presented in Section 3, followed by discussion. Conclusions are presented in Section 4.

2: THE PROPOSED METHOD

The proposed method has three main steps.

1. Pre-processing:
   Erosion operation is applied as basic operators for extracting boundary shapes.
2. Coordinate system transformation and layer division:
   The extracted boundaries are transformed into polar coordinate system, where the original point is shifted to the center of mass of the region. The image plane is partitioned into several ring typed layers (regions) which with the same center of ellipse. Then, the shape is partitioned into several groups according to which layer the points located.
3. Feature extraction:
   The distances, angles and number of shape pixels are extracted as features from each layer for retrieval.

The details of the proposed method are explained in the following subsections.

2.1: Boundary extraction

Humans are sensitive to shape boundaries. To obtain the boundaries, the operations of erosion and dilation are employed. An example for a binary image is shown in Fig. 1, where intensity of a pixel is 0 (white block) or 1 (dark block). Initial extraction is achieved by a 3×3 mask, and erosion is defined as

$$p(x,y)=\begin{cases}0, \ for \quad AND\{I(x\pm i, y\pm j)\}=0\\ 1, \ for \quad AND\{I(x\pm i, y\pm j)\}=1\end{cases}, \quad (1)$$

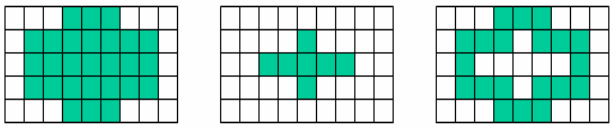
$i, j \in \{-1,0,1\}$,

where $p(x,y)$ is the eroded image, and $AND\{\}$ implies the logical AND. We assume $A \supseteq B$, where A is the data set of the original image pixels, and B is the eroded result of A. See Fig. 1. The image shape can be obtained by the operation of relative complement of B in A, i.e., $\Omega = A - B = \{p \,|\, p \in A, \ p \notin B\}$, where Ω is the set of image boundary. The result of an eroded image is shown in Fig. 1(c). For the operation of dilation, it is defined as

$$p(x,y)=\begin{cases}0, \ for \quad OR\{I(x\pm i, y\pm j)\}=0\\ 1, \ for \quad OR\{I(x\pm i, y\pm j)\}=1\end{cases}, \quad (2)$$

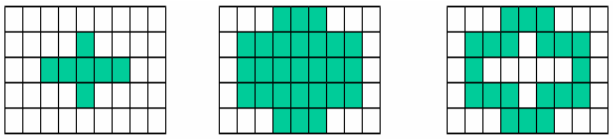
$i, j \in \{-1,0,1\}$,

where $p(x,y)$ is the dilated image, and $OR\{\}$ implies the logical OR operation. Assume that $A \subseteq B$, where A is the data set of the original image pixels and B is the dilated result of A. Similarly, the shape is obtained by the operation of relative complement of A in B, i.e., $\Omega = B - A = \{p \,|\, p \in B, p \notin A\}$, where Ω is the set of image boundaries. A dilated example is shown in Fig. 2.



(a) (b) (c)

Fig. 1. (a) Original binary image (0 : white block, 1 : dark block), (b) The eroded image, (c) The boundary image.



(a) (b) (c)

Fig. 2. (a) Original binary image (0 : white block, 1 : dark block). (b) The dilated image, (c) The boundary image.

2.2: Coordinate system transformation and layer partition

2.2.1: Coordinate system transformation. After shapes have been extracted, we transform the data set Ω to polar coordinate system. The center of mass of each shape is given by

$$(x_c, y_c) = (\frac{1}{n}\sum_{i=1}^{n} x_i, \frac{1}{n}\sum_{i=1}^{n} y_i), \quad (3)$$

where n is the number of total region points of an image.

We shift the origin to (x_c, y_c) as new coordinate system. Then, each boundary point is transformed form Cartesian coordinate system to polar coordinate system. For each point, we have

$$r_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}, \quad (4)$$

and

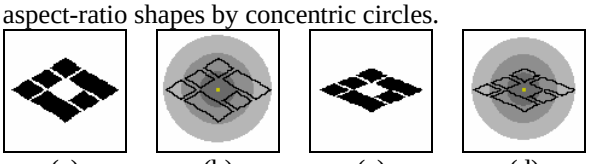
$$\theta_i = \tan^{-1}\frac{y_i - y_c}{x_i - x_c}, \quad (5)$$

where r_i and θ_i represent the norm and angle, respectively. Then, the data set Ω is consisted of boundary points represented using polar coordinate system, i.e.,

$$\Omega = \{p_1(r_1,\theta_1), \ p_2(r_2,\theta_2), ..., \ p_n(r_n,\theta_n)\}. \quad (6)$$

It is expected that the image boundaries are represented as function of normalized distance and angle so that preserve rotation and translation invariance.

2.2.2: Layer partition. It has been mentioned that the shape is partitioned into several groups according to which layer the points located for extracting boundary features. However, the boundary points of asymmetrical deformation images may not uniformly distributed in each layer, which will lead to some layers not having enough samples to gather feature information. As an example, an original image and its deformed image are shown in Fig. 3(a) and Fig. 3(c), respectively. Note that the deformed image has higher aspect ratio. Fig. 4 illustrates the portioned layers by using concentric rings. It can be observed that the two distributions of the extracted boundary points for each layer are quite different. For example, the middle ring of the original image and its deformed image are shown in Fig. 4(b) and Fig. 4(e), respectively. Obviously, they are not similar at all because their numbers of boundary points are not different. This leads to violate the human perceptual judgment, and then causes large distance value between the two shapes. Therefore, we can conclude that it is not appropriate to partition high aspect-ratio shapes by concentric circles.



(a) (b) (c) (d)

Fig. 3. (a) Original image. (b) Extracted the contour of (a) and layer partition by concentric circles. (c) Asymmetrical deformation image of (a). (d) Extracted the contour of (c) and layer partition by concentric circles.

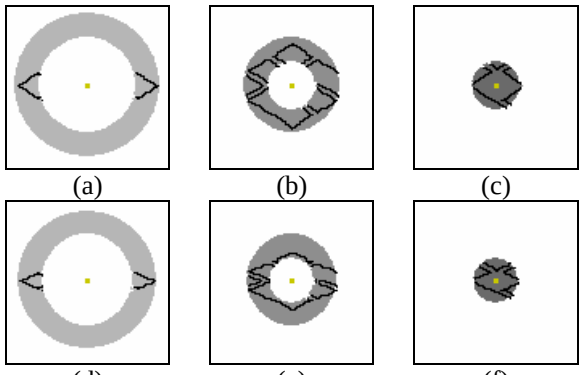


Fig. 4. (a)(b)(c) are the partitioned layers of Fig. 3(b). (d)(e)(f) are the partitioned layers of Fig. 3(d).

To overcome this problem, we propose principal component analysis [9] to modify the layer-partition process [17]. First, let $l_i = [x_i \ y_i]^T$ be the location of a pixel i and N be the total number of pixels belonging to the image region. The mean position vector, m_L, is defined as

$$m_L = \frac{1}{N}\sum_{i=1}^{N} l_i , \tag{7}$$

and the covariance matrix, C_L, is defined as

$$C_L = \frac{1}{N}\sum_{i=1}^{N} l_i l_i^T - m_L m_L^T . \tag{8}$$

Then, eigenvectors e_j and its corresponding eigenvalues λ_j for $j = 1,2$ can be obtained by solving $C_L e_j = \lambda_j e_j$. The two solved eigenvectors imply the directions of boundary-point distribution, That is, boundary points mainly distribute along the larger eigenvector direction, and secondarily along the smaller eigenvector direction. Based on this fact, we partition the image plane into L ellipse-ring layers. Then the members of the data set Ω are partitioned to L groups. We define the partitioned groups as

$$\Omega_p = \sum_{l=1}^{L} R_l , \tag{9}$$

where

$$R_l = \{(r_i,\theta_i) \mid R_{l-1}(\theta_i) \le r_i < R_l(\theta_i), 0 \le \theta_i < 2\pi\}_{i=1,\ldots,n}, \tag{10}$$

$$R_l(\theta_i) = \sqrt{(\frac{l}{L}a_L)^2 \cos^2\theta_i + (\frac{l}{L}b_L)^2 \sin^2\theta_i} , \tag{11}$$

$$a_L = \max\{|r_1 \cos(\frac{\pi(\theta_1-\theta_b)}{180})|,\ldots,|r_n \cos(\frac{\pi(\theta_n-\theta_b)}{180})|\}, \tag{12}$$

$$b_L = \max\{|r_1 \sin(\frac{\pi(\theta_1-\theta_b)}{180})|,\ldots,|r_n \sin(\frac{\pi(\theta_n-\theta_b)}{180})|\}. \tag{13}$$

Notation θ_b is the angle of the bigger eigenvector. In Eq.(10), R_l is the pixel data set of layer l, and n is the total number in layer l. Fig. 5 is an example for an image boundaries divided into three ellipse layers. The advantage of this structural design is that the weighting of each region is able to be tuned for different applications; furthermore, it captures interior information of target images. An example is shown in Fig. 6. By using ellipse layers, the second ring of the original image and its deformed image, shown in Fig. 6(b) and Fig. 6(e), are similar. It will be shown that this improves retrieval performance.

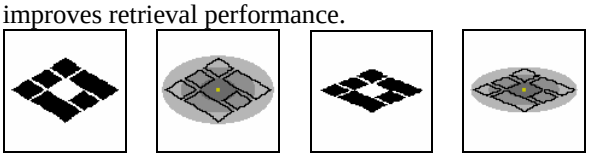


Fig. 5. (a) Original image. (b) Extracted the contour of (a) and layer partition by principal component analysis. (c) Asymmetrical deformation image of (a). (d) Extracted the contour of (c) and layer partition by principal component analysis.

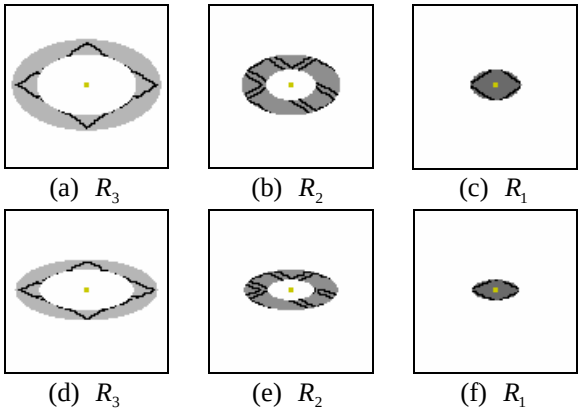


Fig. 6. (a)(b)(c) are the partitioned layers of Fig. 5(b). (d)(e)(f) are the partitioned layers of Fig. 5(d).

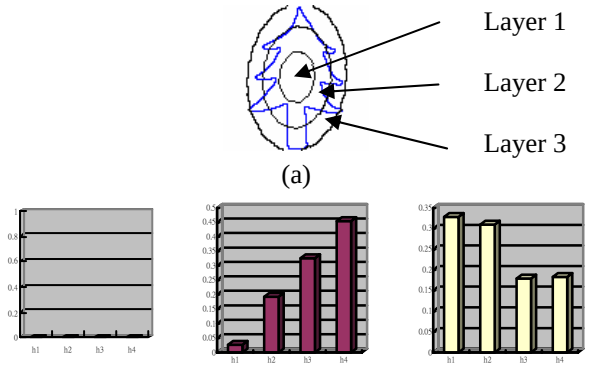
2.3: Shape descriptor

In order to describe the extracted shape, we select distances, angles and pixel number from each layer as the features for shape descriptor. The details of feature extraction are described as follows.

2.3.1: Distance. For extracting the distance, the distances of partitioned set Ω_p in Eq.(9) have to be normalized. The purpose of this normalization step is to preserve scaling invariance. The normalized distance histogram $\mathbf{h}_l^d$ for each layer is calculated by

$$\mathbf{h}_l^d = \frac{\mathbf{H}_l^d}{N_{R_l}} = \left[h_l^1, h_l^2, \ldots . h_l^N \right]_{1*N} \tag{14}$$

where N is the bin number of distance histogram; $\mathbf{H}_l^d = [\mathbf{H}_l^1, \mathbf{H}_l^2, \ldots \mathbf{H}_l^N]_{1*N}$ is distance histogram, and N_{R_l} is the pixel number in layer l. An example of this feature is presented in Fig. 7. In this paper, we use three layers and four bins, i.e., $L = 3$ and $N = 4$.



(a)

(b)Layer 1 (c)Layer 2 (d)Layer 3

Fig. 7. An example of the 4-bins distances histogram. (a) The 3-layered shape image. (b) The histogram of the first layer (R_1). (c) The histogram of second layer (R_2). (d) The histogram of third layer (R_3).

2.3.2: Distance. Consider a shape on layer l . Contour angle $\Delta\alpha_{l,i}$ is defined as a range (in radian) that has a continuous shape; and non-contour angle $\Delta\beta_{l,j}$ is defined as a range that has no continuous shape. A three-layer example is demonstrated in Fig. 8. In the following, we define two features to describe the relationship between shape and angle. The first one is the total contour angle that is defined as

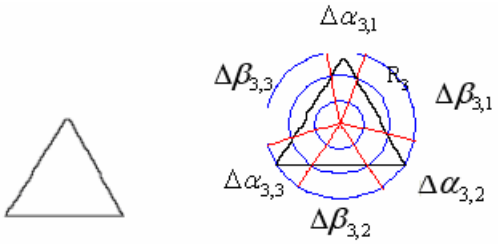
$$e_l = \frac{\sum_{i=1}^{k}\Delta\alpha_{l,i}}{\sum_{i=1}^{k}\Delta\alpha_{l,i}+\sum_{j=1}^{h}\Delta\beta_{l,j}} = \frac{\sum_{i=1}^{k}\Delta\alpha_{l,i}}{360^\circ}, \tag{15}$$

where k and h are numbers of contour and non-contour angle in layer l , respectively.

The second one is the deviation of non-contour angle, and it is defined as

$$q_l = \sqrt{\sum_{j=1}^{h}(\Delta\beta_{l,j}-\overline{\beta_l})^2}\ , \tag{16}$$

where $\overline{\beta_l}$ is the average of non-contour angles in layer l .



(a) (b)

Fig. 8. An example for extracting the angle features. (a) An extracted shape. (b) The angle differences in the third layer (R_3).

2.3.3: Pixel number. Using statistic information to capture the spatial distribution of a shape, the pixel number of the shape in each layer is computed. We denote

N_{R_l} : pixel number in l layer,

N_{Ω_p} : total pixel number of a shape.

Then, the percentage value is defined as the ratio of N_{R_l} to N_{Ω_p} in l layer, and written as

$$p_l = \frac{N_{R_l}}{N_{\Omega_p}} = \frac{N_{R_l}}{N_{R_1}+N_{R_2}+\cdots N_{R_L}}. \tag{17}$$

Obviously, this ratio is invariant to translation, rotation and scaling.

2.3.4: Shape descriptor. After the four features have been extracted, we define the shape descriptor as

$$\mathbf{F} = [\mathbf{F}_1,\mathbf{F}_2,....,\mathbf{F}_L]^T = \begin{bmatrix} h_1^{\mathrm{d}} & p_1 & e_1 & q_1 \\ \vdots & \vdots & \vdots & \vdots \\ h_L^{\mathrm{d}} & p_L & e_L & q_L \end{bmatrix} = \begin{bmatrix} h_1^1 & \cdots & \cdots & h_1^N & p_1 & e_1 & q_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ h_L^1 & \cdots & \cdots & h_L^N & p_L & e_L & q_L \end{bmatrix}_{L*(N+3)}, \tag{18}$$

where

$$\mathbf{F}_l = [h_l^d, p_l, e_l, q_l]^T. \tag{19}$$

Notation $\mathbf{F}_l$ is the shape descriptor for l layer. The features h_l^d , e_l , q_l and p_l are defined in (14), (15), (16) and (17), respectively.

3: EXPERIMENTAL RESULTS

To demonstrate the performance of proposed method, a set of core experiment images provided by MPEG-7 region shape database are conducted. The set A4 of MPEG-7 core experiment images are collected for test of shape descriptor invariance to perspective transform, which consists of 3101 shapes, where 330 shapes in Set A4 have been classified into 30 groups (11 similar shapes in each group). In this experiment, all of the 330 shapes from the 30 groups are used as queries to test the retrieval accuracy. As an example, some sampling images in Set A4 are shown in Fig. 9.

In order to evaluate the retrieval performance, the proposed descriptor is compared with the ART (Angular Radial Transform) [7] and the Layered Shape Descriptor (LS) [17] which partitions an image plane into several concentric circles. To make a comparison on the retrieval performance, both average retrieval rate (ARR) [18] and average normalized modified retrieval rank (ANMRR) are applied. An ideal performance will consist of ARR values equal to 1 for all values of recall. A high ARR value represents a good performance for retrieval rate, and a low ANMRR value indicates a good performance for retrieval rank.

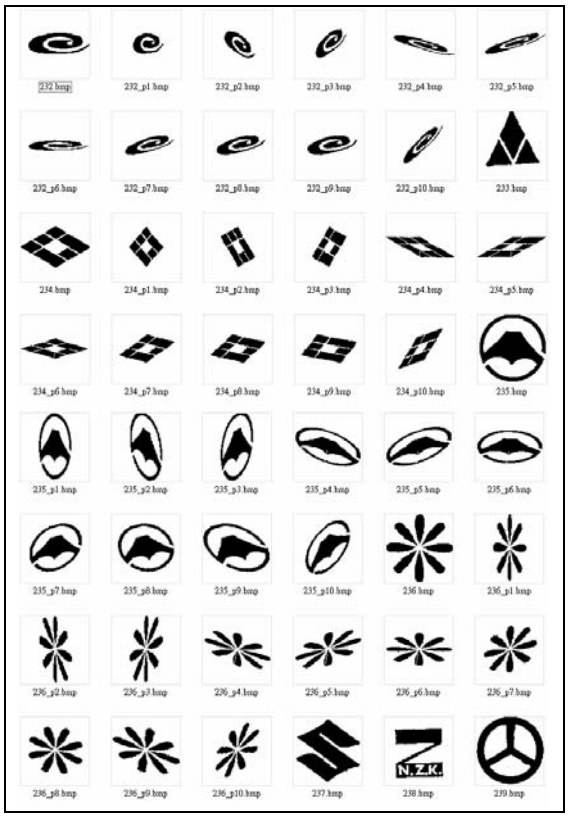


Fig. 9. Some examples in Set A4.

To test the retrieval performance for Set A4, ART, LS and proposed method have been computed. Results of retrieval performance for different descriptors are shown in Fig. 10 and Fig. 11, also listed in Table. 1. It appears from these facts that the proposed method achieves the best performance measured by ARR and ANMRR.

Evaluation / Feature	ARR	ANMRR
ART	38.70 %	0.5958
LS	30.36 %	0.6484
Proposed method	72.37 %	0.2927

Table. 1. The retrieval performance for Set A4.

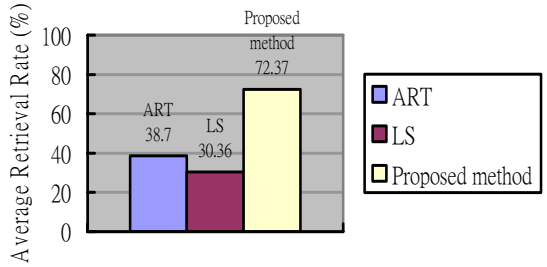


Fig. 10. The ARR measurement of the proposed, ART and LS method.

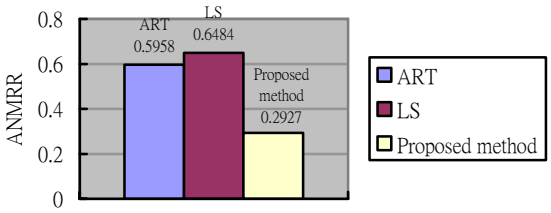


Fig. 11. The ANMRR measurement of the proposed, ART and LS method.

To test the retrieval effectiveness of asymmetrical deformation for single closed contour shapes, we generate a new dataset by using perspective transform. The new dataset consists of 770 shapes from MPEG-7 CE2, and be classified into 70 groups (11 similar shapes in each group). Fig. 12 shows some sample images of the new dataset. Fig. 13 and Fig. 14 show the results of retrieval performance of this set for different descriptors including CSS (Curvature Scale Space), ART, LS and proposed method. See Table 2 also. The average ARR value is 69.86%, and average ANMRR value is 0.34. The results demonstrate that the performance of the proposed method is better than the others.

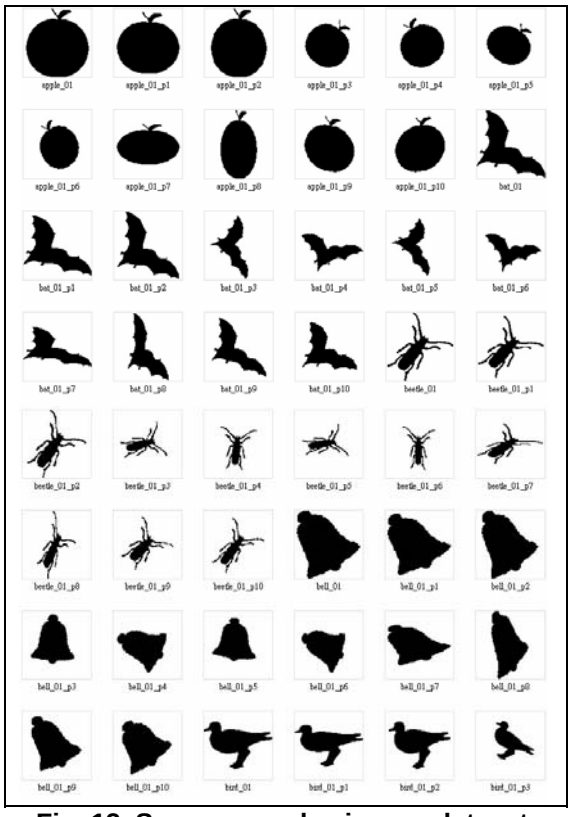


Fig. 12. Some examples in new dataset.

Feature \ Evaluation	ARR	ANMRR
CSS	61.31 %	0.4526
ART	39.84 %	0.5517
LS	64.40 %	0.3504
Proposed method	69.86 %	0.3368

Table. 2. The result of new dataset retrieval performance.

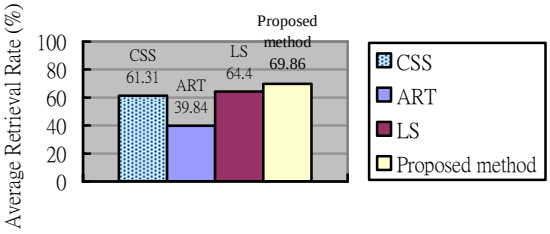


Fig. 13. The ARR measurement of the proposed method, CSS, ART and LS

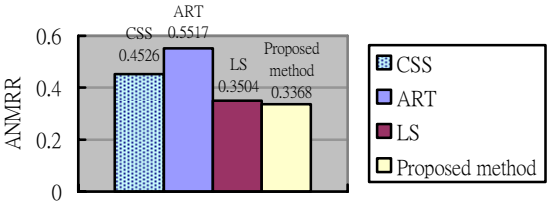


Fig. 14. The ANMRR measurement of the proposed method, CSS, ART and LS

Finally, it is worth to mention that the proposed method is a contour-based descriptor; it not only provides invariant properties but also can deal with complex shapes for shaped-image retrieval.

4: CONCLUSIONS

In this paper, an efficient contour-based descriptor has been presented for asymmetrical deformation shaped-image retrieval. The proposed feature descriptor is efficient in capturing image structure; in addition, it overcomes the limitation of conventional contour-based shape descriptors. The experimental results show that the proposed method, using proper visual features, produces the best results for test of shape descriptor invariance to perspective transform.

REFERENCES

[1] P. Y. Yin and C. C. Yeh, "Content-based retrieval from trademark database," Pattern Recognition Letters, vol. 23, pp.113-126, 2002.

[2] B. Bhanu, P. Jing and Q. Shan, "Learning feature relevance and similarity metrics in image databases," In Proceedings. IEEE Workshop on Content-Based Access of Image and Video Libraries, pp. 14 –18, 1998.

[3] G. Ciocca, R. Schettini, "Content-based similarity retrieval of trademarks using relevance feedback," Patten Recognition, vol. 34, pp.1639-1655, 2001.

[4] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," Pattern Recognition, vol. 29, pp.1233–1244, 1996.

[5] Del Bimbo, P. Pala, "Visual image retrieval by elastic matching of user sketches," IEEE Trans, Pattern Anal. Mach. Intelligence, vol. 19, pp121-132, 1997.

[6] A. K. Jain and A.Vailaya, "Shape-based retrieval: a case study with trademark image databases," Pattern Recognition, vol. 31, no. 9, pp.1369-1390, 1998.

[7] A. Yamada, M. Pickering, S. Jeannin and L. C. Jens, "MPEG-7 Visual part of Experimentation Model Version 9.0-Part 5 Region-Based Shape," ISO/IEC JTC1/SC29/WG11/N3914, January 2001.

[8] R. Distasi, D. Vitulano and S. Vttulano, "A Hierarchical Representation for Content-based Image Retrieval," Journal of visual languages and computing, vol. 11, pp.369-382, 2000.

[9] H. K. Kim and J. D. Kim, "Region-base shape descriptor invariant to rotation scale and translation," Signal Processing: Image Communication, vol.11, pp.87-93, 2000.

[10] W. Y. Kim and Y. S. Kim, "A region-based shape descriptor using Zernike Moments," Signal Processing: Image Communication, pp.95–102, 2000.

[11] G. P. Babu, B.M Mehtre and M. S. Kankanhalli, "Color indexing for efficient image retrieval," Multimedia Tools Appl., vol. 1, no. 4, pp.327-348, 1995.

[12] A. Yamada, M. Pickering, S. Jeannin and L. C. Jens, "MPEG-7 Visual Part of Experimentation Model Version 9.0-Part 5 Contour -Based Shape," ISO/IEC JTC1/SC29/WG11/N3914 Pisa, January 2001.

[13] Eric Persoon and King-Sun Fu, "Shape Discrimination Using Fourier Descriptors," IEEE Trans. On Systems, Man and Cybernetics, vol. 3, pp.170-179, 1977.

[14] H. Kauppinen, T. Seppanen, and M. Pietikainen, "An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification," IEEE Trans. vol. 2, pp.201–207, 1995.

[15] G. Dudek and J.K. Tsotsos, "Shape representation and recognition from multiscale curvature," Comput. Vision Image Understanding, vol. 2, no. 68, pp. 170–189, 1997.

[16] Q.M. Tieng, W.W. Boles, "Recognition of 2D object contours using the wavelet transform zero-crossing representation," IEEE Trans. vol. 8, pp.910–916, 1997.

[17] Nai-Chung Yang, Chung-Ming Kuo, Wei-Han Chang, and Ding-Shun Huang, "An Efficient Shape-Descriptor for Shaped Image Retrieval," Proc. of 2005 IPPR Conf. on CVGIP, Taiwan, ROC.

[18] B. S. Manjunath, Jens-Rainer Ohm, Vinod V. Vasudevan, and Akio Yamada, "Color and Texture Descriptors," IEEE transaction on circuit and system for video technique, vol. 11, no. 6, 2001.