



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Efficient Feature Extraction for Shape-Based Image Retrieval

Mehdi Chehel Amirani, Zahra Sadeghi Gol and Ali Asghar Beheshti Shirazi
Department of Electrical Engineering, Iran University of Science and Technology,
Narmak, Postal Code 16846-13114, Tehran, Iran

Abstract: In this study, a new method for image retrieval based on shape features is proposed that finds feature points by curvature function for each shape. The initial features are generated at these feature points. Then a supervised feature extraction method based on PCA and multilayer perceptron creates a smaller set of features from nonlinear combination of the original features. The proposed algorithm is invariant to several kinds of transformations including some articulations and modest occlusions. The retrieval performance of the approach is illustrated using the MPEG-7 shape database, which is one of the most complete shape databases currently available. Present experiments indicate that the proposed representation is well suited for object indexing and retrieval in large databases. Furthermore, the representation can be used as a starting point to obtain more compact descriptors.

Key words: Feature point, curvature function, shape descriptor, principle component analysis, neural network

INTRODUCTION

The recognition of objects in multimedia content is a challenging task and a complete solution requires the combination of multiple elementary visual and audio features via a knowledge-based inferencing process. Typically, visual features which provide complementary information, such as shape, color and texture are chosen. Shape is clearly an important cue for recognition since humans can often recognize characteristic objects solely on the basis of their shapes.

Many shape analysis techniques have been proposed over the past three decades. An extensive survey of shape matching in computer vision can be found by Loncaric (1998). Common criteria used for shape representation for reliable shape matching and retrieval include: uniqueness, invariance to translation, scale, rotation and symmetric transformations, scalability, compactness, extraction and matching efficiency and robustness to different kinds of distortions (Petraakis *et al.*, 2002). Shape distortions may be present as a result of a perspective transformation due to a change in viewing angle or may be a side effect of the segmentation or extraction process used. Kauppinen *et al.* (1995), concluded that it is not difficult to compare very similar shapes but that difficulties arise when we have to measure the degree of similarity of two shapes which are significantly different. All of these criteria can be summarized by stating that a good shape descriptor should enable the similarity

between shapes to be quantified in a way aligned with human intuition. However, when evaluating the performance of different shape analysis techniques, many researchers (Kauppinen *et al.*, 1995) have reached the conclusion that human judgments of shape similarity differ significantly between observers, thereby making the evaluation task very difficult.

Even in the case of 2-D silhouettes, there can be two notions of similarity. Objects having similar spatial distributions can have very different outline contours and vice versa. Therefore, posing a query using region-based or contour-based criteria can result in retrieving different objects (Bober, 2001).

A contour-based descriptor encapsulates the shape properties of the object's outline (silhouette). It should distinguish between shapes that have similar region-shape properties but different contour-shape properties. Such descriptors are usually very efficient in applications where high intraclass variability in the shape is expected, due to deformations in the object (rigid or nonrigid) or perspective deformations (Bober, 2001).

Since silhouettes do not have holes or internal markings, the associated boundaries are conveniently represented by a single closed curve which can be parameterized by arc length. Early work used Fourier descriptors (Kauppinen *et al.*, 1995) which are easy to implement and are based on the well-developed theory of Fourier analysis. The disadvantage of this approach is that, after the Fourier transform, local shape information is distributed to all coefficients and not localized in the

frequency domain (Loncaric, 1998). A representation that has proven to be relevant in human vision is the medial axis transform (MAT) originally proposed by Blum (1973). This approach led Sebastian *et al.* (2001) to attempt to capture the structure of the shape in the graph structure of the skeleton. In this approach, the edit distance between shock graphs is used as a metric of similarity between shapes. Sebastian *et al.* (2001) claim that their approach is robust to various transformations including articulation, deformation of parts and occlusion. The disadvantage of this description method is that matching two graphs is very computationally expensive. Given the one-dimensional (1-D) nature of silhouette curves, a natural alternative was to use dynamic programming approaches which use the edit distance between curves for matching (Petraakis *et al.*, 2002; Gorman *et al.*, 1988). These algorithms are usually fast and invariant to several kinds of transformations including some articulation and occlusion.

Given the above observations, we conclude that a representation with enhanced discriminatory capabilities could be obtained by explicitly using the feature point. To evaluate this idea we introduce a new shape description method, where for each feature contour point, we store some information about turn angle, central angle, distance between two feature points, distance between center of masse and feature point. The registration of shape for determining the starting point and so stabilizing of method against geometrical transforms such as translation, rotation and scaling is done by a PCA (principle component analysis) transform. After calculating of initial features from each feature counterpoint, a hierarchical feature extraction stage consist of PCA and a supervised neural network combines the most important of initial features to reduce their numbers and at the same time retain as much as possible of their class discriminatory. The retrieval performance of the approach is illustrated using the MPEG-7 shape database.

FEATURE GENERATION

We assume that a size normalized contour C is represented by P contour points. Let the contour C be parameterized by arc length t : $C(t) = (x(t), y(t))$, where $t \in [0, P-1)$. The coordinate functions of C are convolved with a Gaussian kernel.

The curvature function $k(u, \delta)$ is calculated from the resulted shape as follows (Bober, 2001):

$$k(u, \delta) = \frac{x_u(u, \delta)y_{uu}(u, \delta) - x_{uu}(u, \delta)y_u(u, \delta)}{(x_u(u, \delta)^2 + y_u(u, \delta)^2)^{3/2}} \quad (1)$$

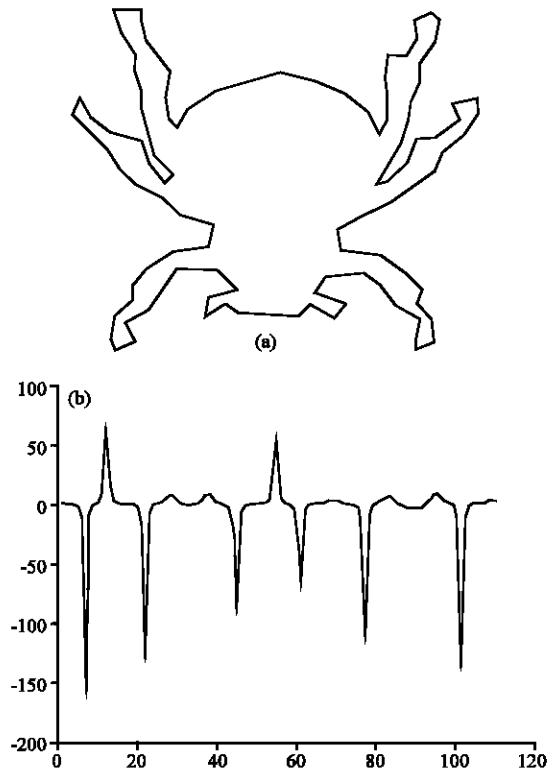


Fig. 1: Curvature function (a) original shape (b) Its curvature function

$$x_u(u, \delta) = \frac{\partial}{\partial u}(x(u) * g(u, \delta)) = x(u) * g_u(u, \delta) \quad (2)$$

$$x_{uu}(u, \delta) = \frac{\partial^2}{\partial u^2}(x(u) * g(u, \delta)) = x(u) * g_{uu}(u, \delta) \quad (3)$$

$$y_u(u, \delta) = y(u) * g_u(u, \delta) \quad (4)$$

$$y_{uu}(u, \delta) = y(u) * g_{uu}(u, \delta) \quad (5)$$

As it can be seen in (1), the curvature function depends on the first and second derivation of x and y and is a suitable explanation for shape description. Figure 1 shows a typical shape and its curvature function. The points corresponded to the extremum of the curvature function are considered as the feature points. Figure 2 shows a model of original shape and its feature points which are obtained according to extremums of the curvature function.

The obtained important points participate in several stages for extracting features. At first, principle axes are found for determining the starting point for feature generation. The computation of an object's principal axis will be showed further.

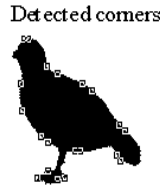


Fig. 2: Feature points of a shape image

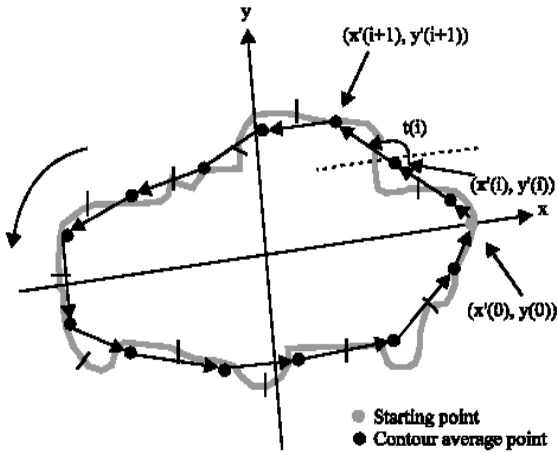


Fig. 3: Extraction of curvature turning angles

The shape-based features included turn angle, central angle, distance between two feature points, distance between center of masse and feature point are generated at selected pints. We nominate these features as F_1 , F_2 , F_3 and F_4 , respectively.

The curvature turning angles are defined as the angles made by two vectors: one that joins two contour's consecutive representative points and the other defined as the object's principal axis where the starting point in question is located (Fig. 3). This feature is calculated as follows:

$$t(i) = \tan^{-1} \left(\frac{y'(i+1) - y'(i)}{x'(i+1) - x'(i)} \right) \quad (6)$$

where, $i = 0, 1, K, P-1$ and $(x'(i), y'(i))$ is the contour's coordinates and P' is the number of selected points.

Figure 4 shows a shape with the corresponded features, wherein g_1 is center of mass, g_2 and g_3 are feature points. 1 represents central angle, 2 and 3 show distance of center and feature points and 4 shows distances between two feature points.

For calculation of these features, at first the object's centre of mass is determined:

$$m_x = \frac{1}{L} \sum_{i=0}^{L-1} x_i \quad (7)$$

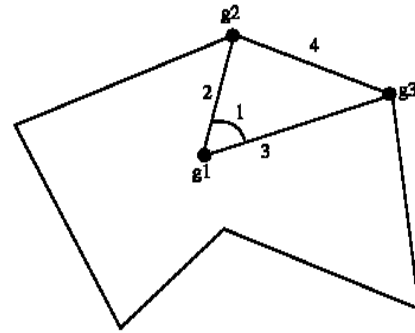


Fig. 4: Representation of features

$$m_y = \frac{1}{L} \sum_{i=0}^{L-1} y_i \quad (8)$$

where, L represents the number of contour shapels (the shapels are the pixels belonging to the object support), (x_i, y_i) and (m_x, m_y) are the i -th contour shapel coordinates and the object's centre of mass, respectively.

Distance of center and feature points are obtained as follow:

$$D_1(i) = ((x'(i) - m_x)^2 + (y'(i) - m_y)^2)^{1/2} \quad (9)$$

$$D_2(i) = ((x'(i+1) - m_x)^2 + (y'(i+1) - m_y)^2)^{1/2} \quad (10)$$

$$D_3(i) = (x'(i) - x'(i+1))^2 + (y'(i) - y'(i+1))^2)^{1/2} \quad (11)$$

The fourth feature at each selected point is the central angle:

$$CA(i) = \cos^{-1} \left(\frac{D_1(i)^2 + D_2(i)^2 - D_3(i)^2}{2D_1(i)D_2(i)} \right) \quad (12)$$

Hence the generated features at each point so called F_1 , F_2 , F_3 and F_4 are $t(i)$, $D_1(i)$, $D_3(i)$ and $CA(i)$, respectively.

FEATURE EXTRACTION

The problem of optimally selecting the statistical features is known as feature selection and feature extraction. While the feature extraction creates a smaller set of features from linear or nonlinear combinations of the original features, the feature selection chooses a subset of the original features. Feature extraction transforms a data space into a feature space that theoretically has the same dimension as the original data space. Such transformation is practically designed in a way that a reduced number of effective features represent the data set while retaining most of intrinsic information

content of the data. A reduction in the dimension of the input space is generally accompanied by losing some of information. The goal of dimensionality reduction is to preserve the relevant information as much as possible. In this section, the concepts of PCA and supervised neural networks that are used to extract the efficient features are explained.

Principle component analysis: The PCA is a well-known feature extraction technique in the multivariate analysis. It is an orthogonal transformation of the coordinate in which the data is described. A small number of principle components are usually sufficient to account for the most of structure of the data.

Let $X = \{x_n \in \mathbb{R}^d | n = 1, \dots, N\}$ represents a d-dimensional dataset. The PCA tries to find a lower dimensional subspace to describe the original dataset while preserving the information as much as possible, so a new m-dimensional dataset $Z = \{z_n \in \mathbb{R}^m | n = 1, \dots, N\}$ will be produced, where m is smaller than d. The orthogonal basis of the feature space is defined as the eigenvectors of the total class scattering matrix:

$$S_t = \frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})(x_n - \bar{x})^T \quad (13)$$

where \bar{x} is the mean vector:

$$\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n \quad (14)$$

Here we are dealing with solving the eigenproblem:

$$S_t \varphi = \varphi \Lambda \quad (15)$$

In which φ is a matrix where its columns are the eigenvectors of S_t , respectively and Λ is a diagonal matrix consisting of all the eigenvalues $\{\lambda_n | n = 1, 2, \dots, N\}$ of S_t along its principal diagonal, where it's other elements are zero.

Practically, the algorithm proceeds by first computing the mean of x_n vectors and then subtracting them from this mean value. The total class scattering matrix, or the covariance matrix, is then calculated and its eigenvectors and eigenvalues are founded. The eigenvectors corresponding to the m of largest eigenvalues are retained and the input vectors x_n are projected onto the eigenvectors to give the components of the transformed vectors z_n in the m-dimensional space. This projection results in a vector containing m coefficients a_1, \dots, a_m . Hereafter the vectors are represented by a linear combination of the eigenvectors with weights a_1, \dots, a_m .

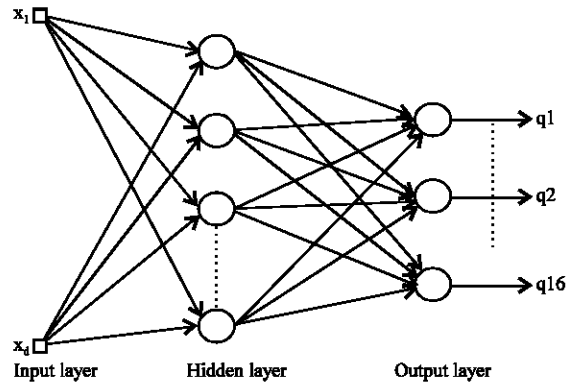


Fig. 5: Architecture of an MLP neural network

It must be pointed out that if the number of selected features be equal to the number of initial features, the aim of PCA is a conversion of features distribution for a particular purpose or computation of principle axes of a shape.

Supervised neural networks: Recently, several efforts have been done to use neural networks for feature generation and feature extraction. A possible solution for feature extraction is via multi layer perceptron (MLP). The employed network has d input and Q output nodes and a single hidden layer with m nodes and linear or nonlinear activations, where Q is the number of classes. During the training phase, the desired outputs are matched to the class number of labeled input, so the corresponded output is high and the others are low.

Such a network maps the input d-dimensional space onto the output m-dimensional subspace. This procedure is equivalent to the kernel Fisher discriminant (KFD) analysis (Muller *et al.*, 2001; Kerschen1 and Golinval, 2004). Figure 5 shows the architecture of an MLP neural network. The outputs of hidden layer are the compressed data and are called the nonlinear discriminant components.

The first step of using a neural network is the training phase in which the back-propagation algorithm is used. In implementation of such algorithm, two distinct passes of computation may be distinguished. The former is referred to as the forward pass and the latter as the backward pass. In the forward pass, the synaptic weights remain unaltered throughout the network and the function signals of the network are computed on a neuron-by-neuron basis. The Backward pass, on the other hand, starts at the output layer by passing the error signals leftward through the network, layer by layer. Weights and biases are updated iteratively until the MSE is minimized.

PROPOSED METHOD

Here, we explain the proposed method based on generated features, which defined at previously. Figure 6 shows the flowchart of the proposed method.

After determination of feature points, PCA maps their coordinates to a new space. The transform matrix is obtained by the feature points coordinates. In the new space the feature points are arranged by their angles from positive part of x-axis. Since all of the feature points participate on the transform, missing or variation a few numbers of them does not serious effect on accuracy of the feature point's arrangement.

For making the feature vector for each shape and for each kind of features that generated at previous stage, the new two dimensional plane after PCA is divided into N equal angle sectors. The vertex of these angles is the origin of the plane. The M features got from shape at each sector are localized in the feature vector regularly. M is the maximum number of features that is expected for considered database at each sector. If the number of feature points at each sector be smaller than M, the remained locations for features set to zero. Hence we have $M \times N$ features for each shape. This arrangement tends to meaning of the locations at feature vector. However, the multiplicity of features causes to fall down the speed and accuracy of the recognition. Hence, a feature extraction method is needed to make a new dataset with the smaller dimensions. The PCA is a fast procedure, but it is poor when compared to the nonlinear and supervised mapping methods. On the other hand, using a MLP network with nonlinear transfer functions has a good performance. However, the computational load and the risk of being trapped in the local minima reduce its efficiency. To solve such problems, we use a hierarchical feature extraction method that serves PCA and a neural network in order to improve the speed and accuracy simultaneously.

At the training phase of the feature extraction stage, PCA projection matrix is calculated based on training dataset. The covariance matrix of the training dataset is firstly calculated, as explained previously. Then the corresponding eigenvectors of the N_1 largest eigenvalues are selected to construct the projection matrix.

Output features from PCA are fed to a MLP neural network that its architecture was previously explained. The number of neurons at the first layer and the output layer are equal to N_1 and C, respectively. The number of neurons at the third layer is equal to N_2 (where $N_2 < N_1$). These N_2 features will be used at retrieval stage. This neural network is trained by using the back-propagation algorithm to minimize the MSE. The problem of trapping

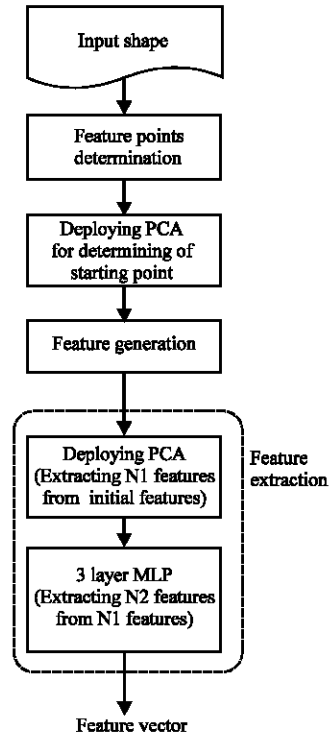


Fig. 6: The proposed algorithm

in the local minima at the training phase of neural network can be avoided by introduction of small disturbances on the weights after each epoch and evaluating and checking the results, thereby the step size or the learning rate can be controlled.

At the retrieval phase, the trained hierarchical feature extraction stage consists of the PCA and a MLP neural network provides the extracted features from the initial features got from the input shape. The similarity measure at retrieval stage is normalized Euclidean distance which is given by:

$$NED(\bar{f}^x, \bar{f}^y) = \sum_m d_m(\bar{f}^x, \bar{f}^y) \tag{16}$$

where,

$$d_m(\bar{f}^x, \bar{f}^y) = \frac{|f_m^x - f_m^y|}{\sigma(f_m)} \tag{17}$$

where $\sigma(f_m)$ is the standard deviation of the respective features over the entire database and are used to normalize the individual feature components. The numbers of ground truth images for each query image in the database are 12. The performance is measured in terms of the average retrieval rate, which is defined as

the average percentage number of patterns belonging to the same image as the query pattern in the top 12 matches.

RESULTS AND DISCUSSION

We used 16 class of shapes obtained from the MPEG-7 database that each class consist of 12 shapes. Figure 7 shows one shape from each class. We chose 3 shapes from each class randomly for training of feature extractor. To making a rich dataset for training , at first each shape from training set was resized with different factors included 1/2 , 1/4 , 2 and 4. Then each of the yield shapes was rotated with 5 random angles from 0 to 360 degree. Therefore, training dataset has 60 shapes from each class and total number of training dataset is equal to 960. The main 12 shapes from each class were used for evaluation of retrieval performance based on extracted features.

At the feature extraction stage, at first 20 features from each kind of features are selected by PCA. Since 4 kinds of features are used for processing, the total number of features for each shape after PCA is equal to 80. Then MLP selects 12 features from 80 features for retrieval.

The resultant retrieval accuracy is equal to 87.5%, while without utilizing of the neural network, it is 79.15%. This improvement is the logical consequence of nonlinear supervised combination of initial features.

For evaluating the performance of proposed method, we perform retrieval based on each kind of features after PCA. Table 1 provides a comparison of average retrieval accuracy for each kind of features include F_1 , F_2 , F_3 and F_4 as well as proposed method for database MPEG-7. It indicates that our proposed method has the best results in compare to others. Furthermore, the total performance of curvature scale space representation is 75.44% (Mokhtarian and Mackworth, 1992), which show that our method outperforms this approach, too.

The average retrieval accuracy rates based on said features and proposed method for each shape have been showed in Table 2, separately. It indicates that the extracted features via proposed method are effective at all of classes respect to other features.

It is obvious that retrieval time, at the similarity measurement stage, is depended on feature vector dimensions. From this point of view, the proposed method has a good performance due to implementation of neural network for feature extraction.

The other criterion for evaluating of a retrieval system is retrieval effectiveness (Manjunath and Ma, 1996; Kokare *et al.*, 2005) that is defined as average rate of

Table 1: Average retrieval accuracy of 16 shapes of database MPEG-7

Feature set No.1 (F_1)	Feature set No. 2 (F_2)	Feature set No. 3 (F_3)	Feature set No. 4 (F_4)	Proposed method
69.0972%	70.5295%	71.3108%	71.0503%	87.9340%

Table 2: Average accuracy rate (%) of retrieval for each shape by several feature sets

Type of shape	Feature set#1 (F_1)	Feature set#2 (F_2)	Feature set#3 (F_3)	Feature set#4 (F_4)	Proposed method
Class 1	47.2222	44.4444	46.5278	40.2778	95.8333
Class 2	100	100	100	100	100
Class 3	99.3056	91.6667	91.6667	91.6667	100
Class 4	53.4722	48.6111	52.0833	52.0833	81.9444
Class 5	86.8056	86.1111	88.8889	95.8333	100
Class 6	100	100	100	100	100
Class 7	59.0278	70.8333	86.8056	95.1389	95.1389
Class 8	87.5	83.3333	83.3333	83.3333	94.4444
Class 9	33.3333	48.6111	44.4444	38.8889	93.75
Class 10	100	100	100	100	100
Class 11	100	100	100	100	100
Class 12	43.0556	36.8056	30.5556	34.0278	50.6944
Class 13	39.5833	55.5556	47.9167	47.2222	93.0556
Class 14	88.1944	80.5556	84.7222	90.2778	90.2778
Class 15	40.9722	40.9722	39.5833	31.9444	47.2222
Class 16	27.0833	40.9722	44.4444	35.4167	64.5833



Fig. 7: Sample shapes from database

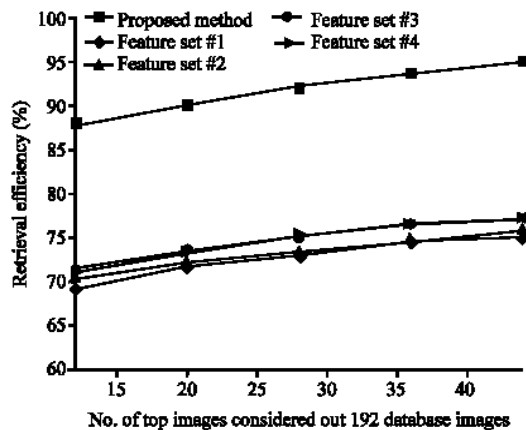


Fig. 8: Average retrieval rate of database MPEG-7 according to number of top images considered

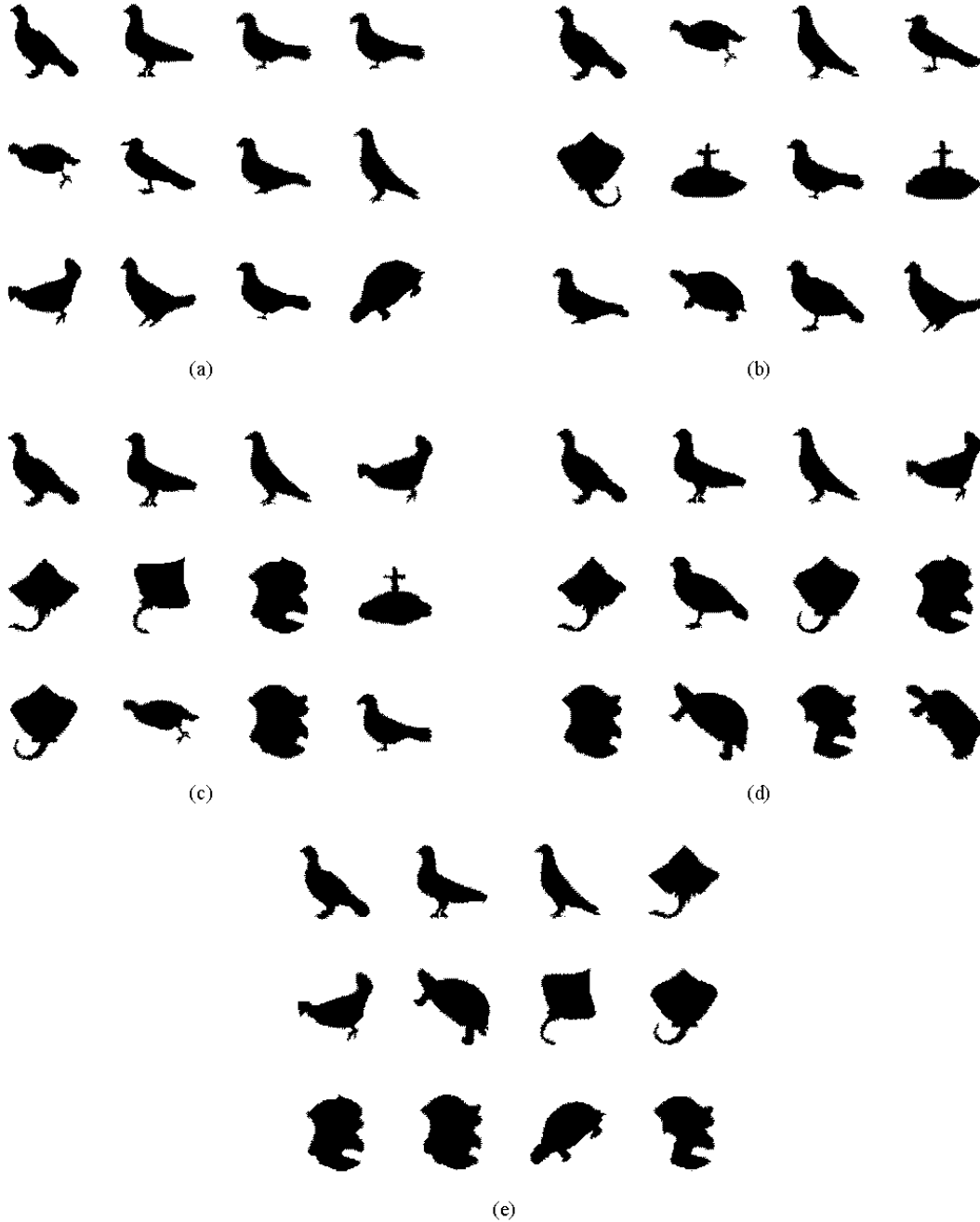


Fig. 9: Retrieved top 12 similar images for given query image using (a) proposed method, (b), (c), (d) and (e) using F_1 , F_2 , F_3 and F_4 features, respectively

retrieving relevant images versus the number of top retrieved images. Figure 8 shows a graph illustrating this comparison according to the number of top matches considered, for database MPEG-7. It is clear that the new method is superior to the others.

Two retrieval examples are shown in Fig. 9 and 10. The top left is the query image and the retrieved images

are displayed from top left to right in increasing order of distance from query image. These examples demonstrate the superior of the proposed method. Retrieval rate of the first example is 95.8333, 47.2222, 44.4444, 46.5278 and 40.2778% and for the second example is 93.7500, 33.3333, 48.6111, 44.4444 and 38.8889%, by using proposed method, F_1 , F_2 , F_3 and F_4 features, respectively.

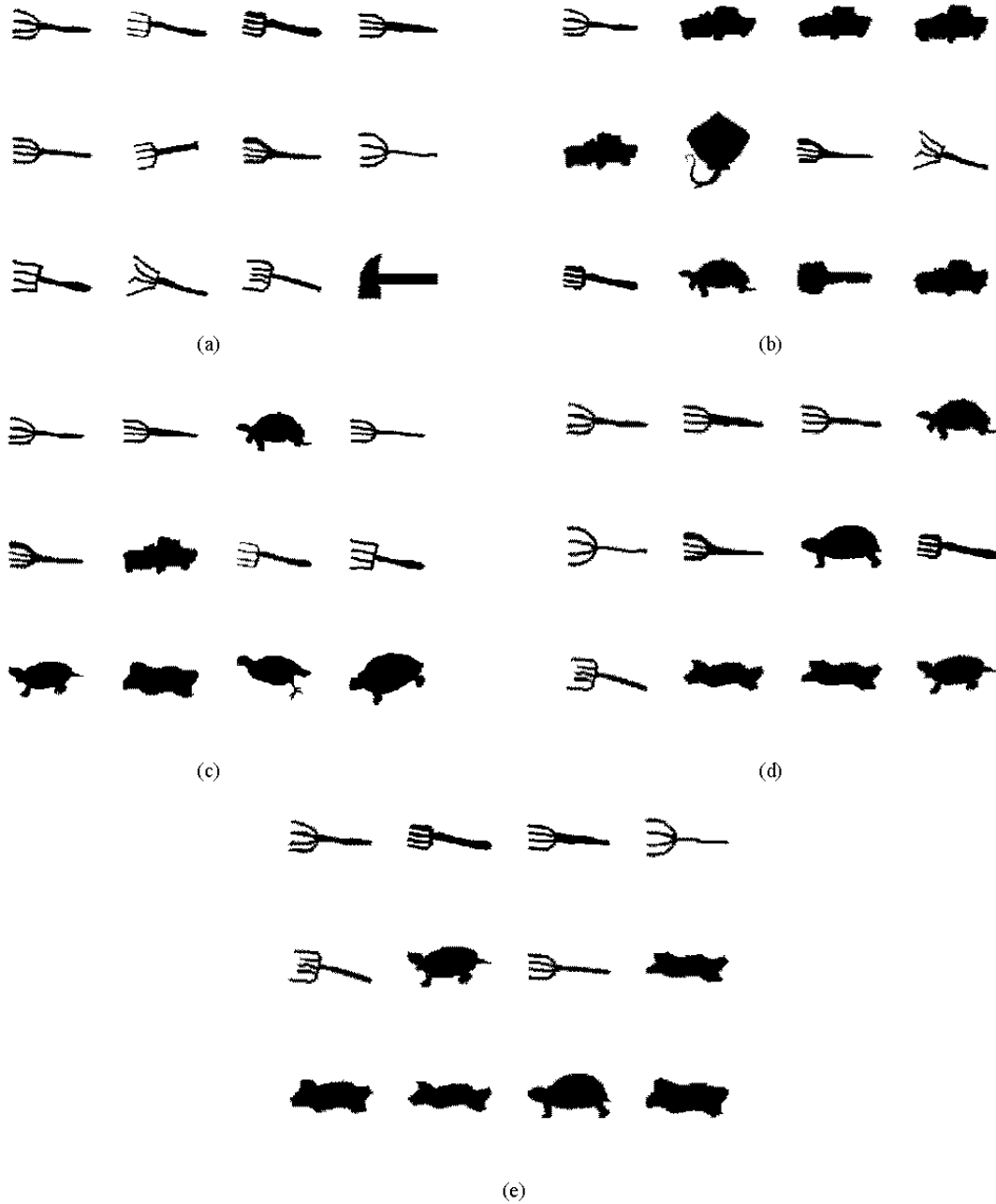


Fig. 10: Retrieved top 12 similar images for given query image using (a) proposed method, (b), (c), (d) and (e) using F_1 , F_2 , F_3 and F_4 features, respectively

CONCLUSION

We have introduced a new method for image retrieval based on shape features. Initial features are obtained at feature points, consist of turn angle, central angle, distance between two feature points, distance between

center of masse and feature point. The registration of shape for determining the starting point and so stabilizing of method against geometrical transforms such as translation, rotation and scaling is done by a PCA transform. Then, for each shape, the smaller and more efficient set of features is obtained via a hierarchical

supervised feature extractor. Retrieval results such as average retrieval accuracy, retrieval time and retrieval accuracy as a function of number of the top images showed the high performance of the proposed method in the retrieval of the MPEG-7 shape database.

ACKNOWLEDGMENT

This study is partly supported by Iran Telecommunication Research Center.

REFERENCES

- Blum, H., 1973. Biological shape and visual science (part 1). *J. Theoretic. Biol.*, 38 (2): 205-287.
- Bober, M., 2001. MPEG-7 visual shape descriptors. *IEEE Trans. Circuits Syst. Video Technol.*, 11 (6): 716-719.
- Gorman, J.W., O.R. Mitchell and F.P. Kuhl, 1988. Partial shape recognition using dynamic programming. *IEEE Trans. Pattern Anal. Mach. Intel.*, 10 (2): 257-266.
- Kauppinen, H., T. Seppanen and M. Pietikainen, 1995. An experimental comparison of autoregressive and fourier-based descriptors in 2-D shape classification. *IEEE Trans. Pattern Anal. Mach. Intel.*, 17 (2): 201-207.
- Kerschenl, G. and J.C. Golinval, 2004. Feature extraction using auto-associative neural networks. *Smart Mater. Struct.*, 13 (1): 211-219.
- Kokare, M., P.K. Biswas and B.N. Chatterji, 2005. Texture image retrieval using new rotated complex wavelet filters. *IEEE Trans. Syst. Man Cybern B Cybern.*, 35 (6): 1168-1178.
- Loncaric, S., 1998. A survey of shape analysis techniques. *Pattern Recog.*, 31 (8): 983-1001.
- Manjunath, B.S. and W.Y. Ma, 1996. Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Anal. Mach. Intel.*, 18 (8): 837-842.
- Mokhtarian, F. and A.K. Mackworth, 1992. A theory of multiscale, curvature-based shape representation for planar curves. *IEEE Trans. Pattern Anal. Mach. Intel.*, 14 (8): 789-805.
- Muller, K.R., S. Mika, G. Ratsch, K. Tsuda and B. Scholkopf, 2001. An introduction to kernel-based learning algorithms. *IEEE Trans. Neural Networks*, 12 (2): 181-201.
- Petrakis, E.G.M., A. Diplaros and E. Millos, 2002. Matching and retrieval of distorted and occluded shapes using dynamic programming. *IEEE Trans. Pattern Anal. Mach. Intel.*, 24 (11): 1501-1516.
- Sebastian, T., P. Klein and B. Kimia, 2001. Recognition of shapes by editing shock graphs. In: *Proceedings 8th International Conference Computer Vision*, Vancouver, BC, Canada.